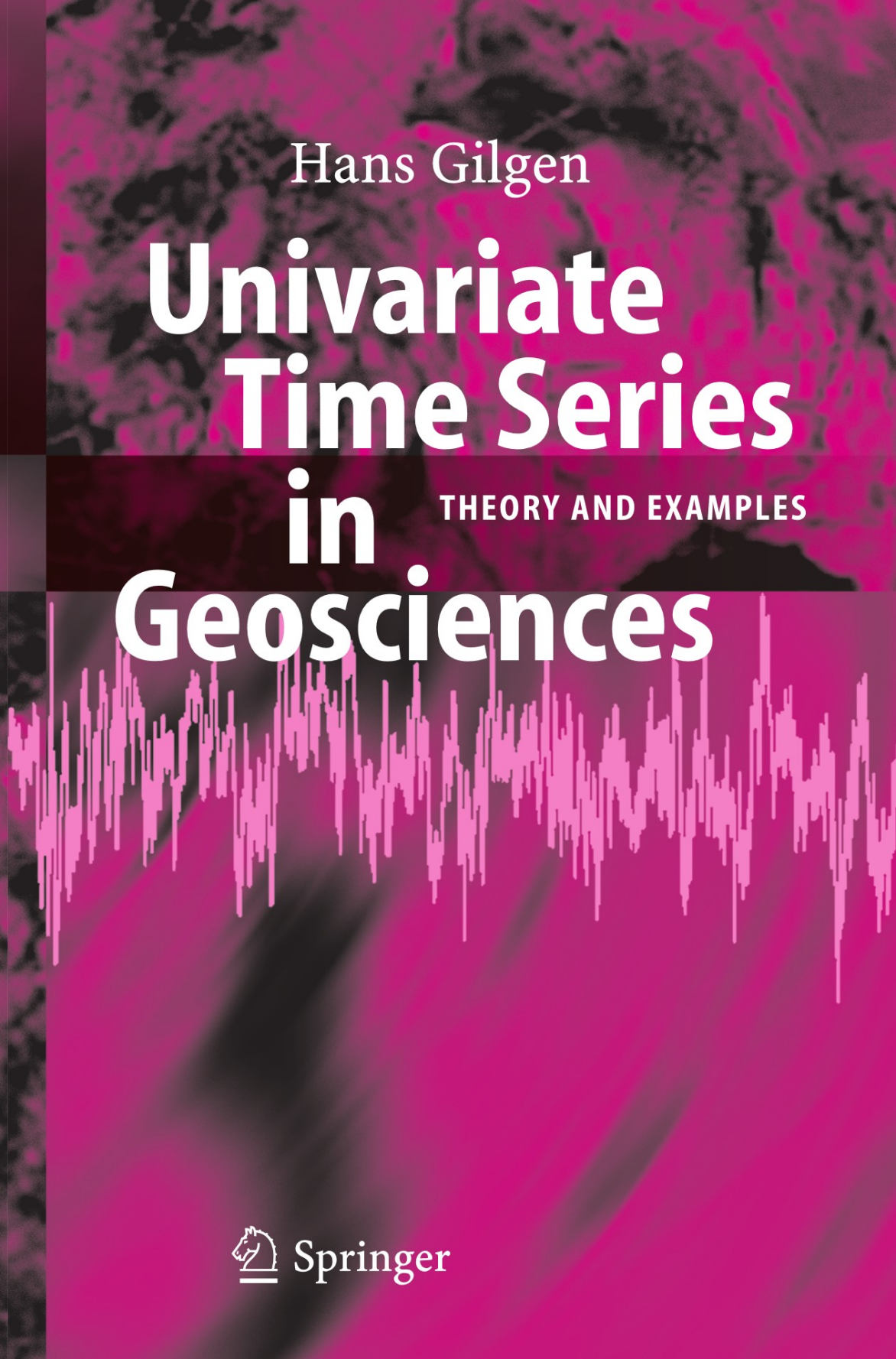

Hans Gilgen
Univariate Time Series in Geosciences
THEORY AND EXAMPLES
Springer

Hans Gilgen
Univariate Time Series in Geosciences
Theory and Examples

Hans Gilgen

Univariate Time Series in Geosciences

Theory and Examples

With 220 Figures

Springer

AUTHOR:

Dr. Hans Gilgen
Institute for Atmospheric and Climate Science
Swiss Federal Institute of Technology (ETH) Zurich
Universitätsstr. 16
8092 Zurich
Switzerland

E-MAIL: HANS.GILGEN@ENV.ETHZ.CH

ISBN 10 3-540-23810-7 **Springer Berlin Heidelberg New York**
ISBN 13 978-3-540-23810-2 **Springer Berlin Heidelberg New York**

Library of Congress Control Number: 2005933719

Springer is a part of Springer Science+Business Media
springeronline.com

Printed in The Netherlands

Cover design: E. Kirchner, Heidelberg
Production: A. Oelschläger
Typesetting: Camera-ready by the Author

Printed on acid-free paper 30/2132/AO 543210

in memory of JoJo and MaiMai
to Lüsai

Preface

La théorie des probabilités n'est au fond que le bon sens réduit au calcul.
Probability theory is, basically, nothing but common sense reduced to calculation.
Laplace, Essai Philosophique sur les Probabilités, 1814.

In Geosciences, variables depending on space and time have been measured for decades or even centuries. Temperature for example has been observed worldwide since approximately 1860 under international standards (those of the World Meteorological Organisation (WMO)). A much shorter instrumental (i.e., measured with instruments) record of the global background concentration of atmospheric carbon dioxide is available for Hawaii (Mauna Loa Observatory) only dating back to 1958, owing to difficulties inherent in the routine measurement of atmospheric carbon dioxide. Further examples of long-term records are those obtained from measurements of river discharge.

In contrast to standardised routine measurements, variables are also measured in periods and regions confined in time and space. For example, (i) ground water permeability in a gravel deposit can be approximated from grain size distributions of a few probes taken (owing to limited financial resources for exploring the aquifer), (ii) solar radiation at the top of the atmosphere has been measured by NASA in the Earth Radiation Budget Experiment for the period from November 1984 through to February 1990 using instruments mounted on satellites (since the lifetime of radiation instruments in space is limited), (iii) three-dimensional velocities of a turbulent flow in the atmospheric boundary layer can be measured during an experiment (seeing that a measurement campaign is too costly to maintain for decades), and last but not least (iv), measurements have been performed under often extremely adverse conditions on expeditions.

Many variables analysed in Geosciences depend not only on space and time but also on chance. Depending on chance means that (i) all records observed are reconcilable with a probabilistic model and (ii) no deterministic model is available that better fits the observations or is better suited for practical applications, e.g., allows for better predictions. Deterministic mod-

els have become more and more sophisticated with the increasing amount of computing power available, an example being the generations of climate models developed in the last two decades. Nevertheless, tests, diagnostics and predictions based on probabilistic models are applied with increasing frequency in Geosciences. For example, using probabilistic models (i) the North Atlantic Oscillation (NAO) index has been found to be stationary in its mean, i.e., its mean neither increases nor decreases systematically within the observational period, (ii) decadal changes in solar radiation incident at the Earth's surface have been estimated for most regions with long-term solar radiation records, (iii) Geostatistical methods for the optimal interpolation of spatial random functions, developed and applied by mining engineers for exploring and exploiting ore deposits, are now used with increasing frequency in many disciplines, e.g., in water resources management, forestry, agriculture or meteorology, and (iv) turbulent flows in the atmospheric boundary layer are described statistically in most cases.

If a variable depending on space and/or time is assumed to be in agreement with a probabilistic model then it is treated as a stochastic process or random function. Under this assumption, observations stem from a realisation of a random function and are not independent, precisely because the variable being observed depends on space and time. Consequently, standard statistical methods can only be applied under precautions since they assume independent and identically distributed observations, the assumptions made in an introduction to Statistics.

Often, geophysical observations of at least one variable are performed at a fixed location (a site or station) using a constant sampling interval, and time is recorded together with the measured values. A record thus obtained is a time series. A univariate time series is a record of observations of only one variable: a multivariate one of simultaneous observations of at least two variables. Univariate time series are analysed in this book under the assumption that they stem from discrete-time stochastic processes. The restriction to univariate series prevents this book from becoming too long.

In contrast to the other examples given, the Mauna Loa atmospheric carbon dioxide record grows exponentially, a property often found in socio-economic data. For example, the power consumed in the city of Zurich grows exponentially, as demonstrated in this book.

Subsequent to introducing time series and stochastic processes in Chaps. 1 and 2, probabilistic models for time series are estimated in the time domain. An estimation of such models is feasible on condition that the time series observed are reconcilable with suitable assumptions. Among these, stationarity plays a prominent role. In Chap. 3, a non-constant expectation function of the stochastic process under analysis is captured by means of estimating a linear model using regression methods. Chap. 4 introduces the estimation of models for the covariance function of a spatial random function using techniques developed in Geostatistics. These models are thereafter used to compute op-

timal interpolators. Optimal interpolators are a version of optimal predictors, both being estimators in the statistical sense. In Chap. 5, optimal predictors for a time series are obtained by applying models for the underlying stochastic process itself. These models are called linear processes, ARMA models or Box-Jenkins-models. Chap. 3 should be read prior to tackling Chap. 4 or Chap. 5; it is, however, not necessary to read Chap. 4 prior to accepting the challenge of Chap. 5.

Chaps. 6, 7, 8, 9 and 10 give an introduction to time series analysis in the frequency domain. In Chap. 6, Fourier analysis of deterministic functions is comprehensively (including discussions of aliasing, leakage and the width of functions in both domains, time and frequency) dealt with in order to build a solid framework supporting the application of Fourier methods to time series, which is legitimised by the Fourier representation of stationary stochastic processes introduced in Chap. 7. Under these preconditions, estimators for the spectrum of a stationary stochastic process under analysis can be calculated from a time series observed.

In Chap. 8, the periodogram is introduced. The periodogram has an asymptotic distribution that can be derived for most discrete-time stationary stochastic processes. Unfortunately, a periodogram calculated from a realisation of a process having a spectrum with a large dynamic range is severely biased and thus the periodogram is generally not a spectral estimator. Spectral estimators are calculated from tapered observations, i.e., observations modified using a data window, a technique that efficiently reduces the bias due to leakage. In Chap. 9, an estimator calculated from tapered observations is convolved with a smoothing kernel to obtain an estimator for a continuous spectrum, and Chap. 10 contains a short introduction to estimators for discrete spectra.

This book is intended to be an introduction to the analysis of time series for students of Geosciences, who are assumed to have a mathematical background of at least two semesters calculus and one semester probability theory and statistics.

I composed this book using LaTeX together with macros made available by Springer-Verlag and also prepared all figures, except for Fig. 3.10, which is credited to S. Bischof.

This book appears for the twentieth anniversary of preliminary and incomplete versions of Chaps. 2, 3, 4, 5, 6, 9 and 10 which I used in courses given to students of atmospheric and earth sciences at the Swiss Federal Institute of Technology (ETH) and the University, both in Zurich. The students learnt to estimate models for example time series (of these, the Basel temperature series has "survived" as can be seen in Table 1.1) and I learnt that clear definitions are important, obviously not a novel concept. My courses evolved together with the statistical software being available and also pertinent to teaching. The major evolutionary step was the changeover to Splus in the early 1990s. Splus proved to be the first statistical software that could

be successfully applied by the students subsequent to a short introduction. Three years ago, a small number of example programs and exercises were modified so that R could be used. The reasons for these last changes were twofold: (i) I believed (and still believe) that R will develop more dynamically than Splus and (ii) R is available under Gnu conditions.

My time series courses improved with the help of H. Künsch of the Statistics Seminar at ETH. He gave a large number of valuable suggestions and his comments on earlier versions (in German) motivated me to write in such a way that the mathematics is correct and the text remains readable for non-mathematicians. The English version was corrected by D. Scherrer who carefully read the manuscript and patiently transformed German-like constructions into English. Without the help of both, H. Künsch and D. Scherrer, this book could not have been written.

C. Hohenegger's critical comments helped to fine-tune the definitions and derivations given in this book.

I am indebted to (i) H. Jensen for entrusting me with preliminary versions of Sects. 4.3.2, 4.5.2, 4.5.3 and 4.6.3, (ii) K. Hutter for encouraging me to write this book and (iii) Atsumu Ohmura for both, his confidence in this project and funding at critical times that helped to finish this book.

I thank M. Andretta, M. Bannert, R. Barry, H. Blatter, C. Breitinger, S. Brönnimann, N. Bukowieki, P. Calanca, H. Fischer, C. Frei, M. Giansiracusa, D. Grebner, J. Gurtz, A. Hänger, H. Jensen, M. Mächler, G. Müller, Andrée and Atsumu Ohmura, A. Roesch, M. Rotach, M. Roth, W. Sawyer, P. Stamp, F. Stauffer, H. von Storch, M. Wild and M. Zappa for supplying data and/or commenting on (parts of) the manuscript. I also thank all students whose comments helped to improve the manuscript and the exercises. Any errors remaining are my responsibility and any comments are welcome.

Zurich, July, 2005

Hans Gilgen
Institute for Atmospheric and Climate Science
Swiss Federal Institute of Technology (ETH) Zurich
Universitätsstrasse 16
CH-8092 Zürich
gilgen@env.ethz.ch

Contents

1 Introduction

In this book, data with a temporal and/or spatial structure are analysed. For example, in this chapter, a widespread tool for the statistical analysis of data is used to analyse errors which occur in the measurement of a meteorological variable. This error analysis

- offers an example of data which reveal their secrets only when the time dimension is included in the analysis (in Sect. 1.4)
- introduces the statistical software R [114] (in Sect. 1.2)
- reviews the basics of statistical data analysis (in Sects. 1.3 and 1.5). This review, however, does not act as a substitute to an introduction to Statistics (e.g., [118], [143], [128], [130]).

1.1 Data in Geosciences: for Example, Surface Solar Radiation Records

Data observed in Geosciences are (i) afflicted with errors, (ii) often incomplete (i.e., values are missing because a measurement could not be performed at a given location and time), (iii) sometimes contradictory and (iv) rarely self-contradictory, namely when values not reconcilable with the body of knowledge available result from the measurement. For example, daily averages of solar radiation measured at the surface of the earth which are larger than their corresponding values at the top of the atmosphere are obviously self-contradictory due to a gross error in the measurement [59]. In addition, the connection between observations and the properties under analysis is often only approximately known. What can be inferred under these adverse circumstances from observations performed in Geosciences? Quite a lot, under adequate precautions and when the probabilistic properties of the data are taken into account, as will be demonstrated by the examples given in this book.

The observations used in this book as examples comprehend climate data (long-term records of temperature, pressure and radiation), hydrometeorological data (precipitation events), micro-meteorological data (wind speeds measured in the atmospheric boundary layer), soil parameters, seismograms, and hydrological discharge records. The example data sets are given

Table 1.1. Example data used in this book. The names of the data sets become the file names in the collection available on the data medium accompanying this book, provided that the suffix .dat is appended. In the files obtained with the suffix .des appended to the data set name, descriptions of the data are made available. Data sets with a ∘ are available in R, the statistical software introduced in Sect. 1.2, those with a ∗ can be downloaded from the internet address given in the reference(s).

	Variable(s)	Data sets	In this book
agriculture	soil parameters and crop yields	stracker	Sects. 4.3.1, 4.3.3, 4.5.1
astronomy	sun spot numbers	∗[100]	Problems 2.31, 9.21
ecology	height of tree-line	treeline	Sects. 3.6, 4.1.2, Problem 4.23
	abundance of lynx	∘lynx	Problem 9.20
seismology	seismic waves	earthquake	Sect. 9.4.1
		nucleartest	Problem 9.17
hydro-geology	hydraulic conductivity	gravel2	Problem 4.6
hydrology	discharge	inn7490	Sect. 10.2.2
meteorology/ climatology	atmospheric CO_2	∗[84]	Problems 3.23, 5.15, 9.24, Sects. 5.4.1, 5.4.2
	precipitation	ppitoct592	Sects. 4.1.1, 4.6.5
	pressure indices		
	NAO	∗[76], [69]	Problem 2.2, Sects. 2.3.1, 2.5.2, 2.8.4, 8.1, 8.2.2
	SO	∗[86]	Sects. 5.4.2, 5.4.3, 9.2.5
	solar radiation	pyrcomp	Chap. 1, Problems 2.32, 4.5
		swir696	Sect. 3.5
		swirhpb	Sect. 3.1
		swirnorde	Sects. 4.3.4, 4.5.3, 4.6.3
	temperature		
	Basel series	basel	Problem 2.7, Sects. 2.3.4, 3.2.1, 3.4, 5.4.1 and 10.1.3
	global averages	∗[75]	Problem 2.30
	turbulent flow in boundary layer	wind	Sects. 2.1, 2.3.2, 5.1.3
		turwind3	Sect. 9.4.2
economics	power consumed in Zurich	zhpower5786	Problem 3.22
		zhpower8505	

in Table 1.1. This selection is clearly biased: solar radiation data dominate in the subset of climate data, and data sets related to meteorology and/or climatology dominate in the set of all examples. This bias is ascribable to the author's work in radiation climatology.

Measuring data in Geosciences is challenging in most cases, owing to essential limitations (when the measurement disturbs the variable observed) or

limitations in the instruments, data acquisition systems and/or observational regions or periods, which can deteriorate the quality of the data. Known limitations and other (possibly unknown) shortcomings in a data set have to be accounted for when observations are used in Geosciences and consequently, introductions to the measurement of all data sets used in this book should be given. However, even abbreviated introductions are beyond the scope of this book, except for the following three examples:

1. Sect. 9.4.1 introduces the measurement of seismic waves. Seismic waves are tiny vibrations of the rock that can be recorded by seismometers. They are caused by an earthquake or an explosion, or, as background noise in a record, by natural phenomena such as ocean waves.
2. Sect. 9.4.2 introduces the measurement of velocities in a turbulent flow in the atmospheric boundary layer.
3. This section introduces the measurement of solar radiation at the surface of the earth.

Solar radiation measurements are more delicate than those of meteorological variables such as temperature, precipitation, etc., seeing that radiation instruments are less stable over longer periods (years) than thermometers, rain gauges, etc. Nevertheless, maintenance problems and problems arising from changes in (i) the technology of sensors or data acquisition systems and/or (ii) the measuring sites are shared by most long-term measurements of meteorological variables. Therefore, conclusions drawn from the discussion of solar radiation records also apply to other climate records, e.g., that changes in the observed values due to changes in the measurement have to be compensated when the data are used for climate change studies.

Solar radiation intercepted by our planet will either be absorbed or returned to space by scattering and reflection. Solar radiation is (i) scattered and reflected by clouds, dry air molecules, water vapour and aerosols in the atmosphere and (ii) reflected by the surface. The absorbed part of the solar radiation (242 Wm^{-2} (watt per square meter) on the average over the globe, obtained as 71 % (the planetary albedo is 29 % [148]) of 341 Wm^{-2} (1/4 of the solar constant [53]), all values are approximations) is the source of energy which drives the processes in the atmosphere and the oceans. Compared to this large amount of energy, the flux of heat generated in the interior of the earth (mostly from radioactive decay) through the earth's surface is negligible on the global average. Although being small, the earth's internal heat fuels convection in its mantle (the layer below its crust) and thus drives plate-tectonics which — over geological periods — allocates oceans and continents, as described in Sect. 9.4.1. Therefore, solar radiation absorbed in the atmosphere and at the surface of the earth is the primary energy source for life on our planet which is shaped by processes driven by its internal heat.

Solar radiation absorbed by our planet can be calculated from the components in the radiation budget at the top of the atmosphere: the earth absorbs and reflects solar radiation incident at the top of the atmosphere and re-

emits the exhaust heat in the form of terrestrial radiation back into space, since it is in radiative equilibrium with the sun and space: incoming and outgoing radiative fluxes balance each other on condition that no energy is stored/released by heating or cooling the oceans or as latent heat in/from the ice and snow masses. As a contrasting aside: Jupiter radiates more energy in space than it receives from the sun. The source of its internal heat is (along with radioactive decay and the tides generated by its satellites) gravitational contraction (when a gravitating object contracts, gravitational potential energy is converted into heat).

The radiative fluxes at the top of the atmophere were measured from space, using instruments mounted on satellites, in the Earth Radiation Budget Experiment (ERBE) over the period from November 1984 through to February 1990. Measuring radiative fluxes in space is a formidable task since (i) instruments have to be built which can be operated in space and, under the rough conditions prevailing there, will remain stable over relatively long (a few years) periods, since a re-calibration is too costly (ii) the instruments have to be mounted on appropriate satellites and (iii) data acquisition and archival systems have to be maintained ([8], [7]). A continuation (with improved instruments) of the ERBE records is provided by the Clouds and the Earth's Radiant Energy System (CERES) [127].

Measuring solar radiation at the surface and maintaining a network of measuring stations over decades is a task even more difficult than measuring radiative fluxes at the top of the atmosphere. This task can be accomplished on condition that

1. a consensus has been reached on the world standard maintained by the world radiation centre (WRC/PMOD) [52] which is then used to calibrate the reference instruments (secondary standards) maintained by the organisations (in the majority of cases the national weather services) that measure solar radiation
2. instruments measuring solar radiation are periodically calibrated against the secondary standards
3. measurements are guided by best practise recommendations regarding the installation of calibrated instruments, the maintenance schemes, the data acquisition systems and the data archival procedures ([57], [94])
4. recommendations for the operational measurements account for differences in climatic regions and are adapted in accordance with the development of technology without inducing systematic changes in the measured records (counter examples are given in Fig. 2.19)
5. the manned or automated stations (sites) where the instruments are mounted and their neighbourhood remain unchanged over decades
6. the measurements at a site are representative for a larger region, and
7. the data archives are maintained over a substantially long period.

Solar radiation records measured under these conditions are *homogeneous* in space and time: spatial and temporal variations in the data are not due

to changes in the observational system. The conditions enumerated above establish standard procedures for the observation of radiative fluxes and also for the subsequent data processing and archiving.

The first international standards for observing meteorological variables (e.g., temperature, precipitation, pressure) were introduced in the second half of the 19th century. Observations made with instruments prior to approximately 1860 were less standardised and are therefore often called early instrumental records. The reconstruction of a homogeneous climatological series of, e.g., monthly values, from early instrumental records is a delicate task. An example is given in Sect. 2.3.4.

Climate records obtained using standard observational procedures are homogenous and thus allow for monitoring climate change. Often, however, climate records are afflicted with inhomogeneities. Inhomogeneities stem from a variety of sources, often they are due to the deficiencies and changes enumerated in [82]. Climate data that are not homogenous can be adjusted using procedures as recommended in [46] or [109]. This difficult task is best done by the scientist responsible for the measurements since he/she has access to the station history or the meta data, i.e., the description of the measurements in the diary kept at the measuring site.

If no detailed station history is available then inhomogeneities (yet not all) can be detected using statistical methods such as those introduced in [66], [59], [88] and [140], or they can be found in relevant plots. In Fig. 2.19 for example, the amplitudes of the fluctuations in the differences in solar radiation recorded at neighbouring stations decrease abruptly, in the years 1975 and 1980, possibly due to the introduction of automated observations. Subsequent to the detection of such inhomogeneities, an adjustment remains difficult.

As a substitute for an adjustment, climate data can be flagged as doubtful due to errors and/or inhomogeneities. For example, the solar radiation data in the Global Energy Balance Archive (GEBA) [59] and also those measured at the stations of the Baseline Surface Radiation Network (BSRN) [103] undergo rigorous quality checks to assure high accuracy as well as homogeneity in the data, a prerequisite for regression analyses such as those performed in Sects. 3.1 and 3.5 to estimate decadal changes in these records ([58], [152]). Due to the quality checks applied, the data in the GEBA and BSRN are, when used in conjunction with measurements of the radiative fluxes at the top of the atmosphere, also suitable for calculating a more accurate disposition of solar radiation in the earth-atmosphere system ([3], [150]).

It remains to take a closer look at the measurement of solar radiation at the surface of the earth which stands in this book representatively for the measurement of meteorological variables. Solar radiation is measured using either a pyrheliometer or a pyranometer. A pyrheliometer is an instrument designed to measure the direct-beam solar radiation. Its sensor is oriented perpendicular to the direction of the sun and is inserted in a tube such that

it intercepts radiation only from the sun and a narrow region of the sky around the sun. Observations are made when the sky is clear. The quantity measured is called direct radiation. Pyrheliometer data are used to study the extinction of solar radiation in an atmosphere free of clouds.

Pyranometers measure the solar radiation from the sun and sky incident on a horizontal surface. They measure continuously and are exposed to all kinds of weather. Hence, the detector is shielded by a glass dome which, notwithstanding, only transmits radiation in wavelengths between 0.3 and 2.8 μm, micrometer, 1 $\mu m = 10^{-6}$ m). The glass dome has to be kept clean and dry. The detector has at least two sensing elements, one blackened such that most of the incident solar radiation is absorbed and the other coated with a white paint such that most radiation is reflected (or placed in shade to avoid solar radiation). The temperature difference between these elements is approximately proportional to the incident radiation. Annual calibrations are recommended because the responsivity of pyranometers deteriorates over time. The meteorological variable measured with a pyranometer is called global radiation or shortwave incoming radiation (SWIR).

Errors in pyranometer measurements may arise from various sources, often from instrumental deficiencies during the measurement. There are *random errors* and *systematic errors*. A random error is zero in the mean of the measurement period, a systematic error has a non-zero mean. An example of a systematic error with a time-dependent mean is the error produced by a drifting sensitivity of an instrument. Systematic errors can be detected and possibly corrected, provided that a detailed station history is available. Random errors cannot be corrected, but their statistical structure can be taken into account when the data are used, as demonstrated in Sects. 3.1.1, 3.5, 4.6.3, and 4.6.5. With a good instrument the random error of a single pyranometer measurement is 2% of the measured value.

The maintenance of the instruments, their installation and the data acquisition system is crucial to the quality of the measurements. Errors (systematic and/or random) can therefore not be excluded. Errors in the single pyranometer readings propagate to the hourly, daily, monthly, and yearly means aggregated consecutively from the original data. How much influence does maintenance have on the quality of pyranometer data? This question was investigated in a long-term pyranometer comparison project jointly performed by the Federal Office for Meteorology and Climatology (MeteoSwiss) and the Swiss Federal Institute of Technology (ETH), both in Zurich. In this project, the shortwave incoming radiation was measured by both institutions from January 1, 1989 through to December 30, 1992 at the same (Reckenholz) station, but completely independently. The instruments used in the experiment were both Moll-Gorczynsky-type thermoelectric pyranometers, made though by different manufacturers. The installation of the instruments, the maintenance schemes, and the data acquisition systems were chosen delib-

Table 1.2. A pyranometer comparison experiment was performed at Zurich-Reckenholz ($8^0 31'$E, $47^0 25'$N, 443 m a.m.s.l.) station jointly by the Federal Office for Meteorology and Climatology (MeteoSwiss) and the Swiss Federal Institute of Technology (ETH).

Pyranometer	MeteoSwiss	ETH
manufacturers	Kipp	Swissteco
calibrated	with the MeteoSwiss standard instrument	at the World Radiation Centre in Davos
installation	instrument lightly ventilated but not heated	instrument neither ventilated nor heated
data acquisition	A-net (MeteoSwiss standard)	Allgomatic
averaged over	10 minutes	5 minutes
maintenance	daily	weekly

erately not to be identical (Table 1.2) to simulate the distinct pyranometer installations of the institutions that measure and publish SWIR data [58].

From the Reckenholz measurements, hourly and daily averages were calculated consecutively. These daily averages are stored in the following format

```
89   1  1 13 12
89   1  2 18 18
...
92  12 30 18 18
```

in the file `/path/pyrcomp.dat`. This is an example of a text file containing a table of data as, for each day, it contains a line with five values. The first three represent time of the measurement as year, month and day. The fourth value is the daily mean of SWIR calculated from the MeteoSwiss measurements, and the fifth, the daily mean calculated from the ETH measurements. The unit of the radiation values is Wm^{-2}. Missing data are represented as `NA` (not available). A description of the pyranometer data is given in the file `/path/pyrcomp.des`.

This convention is used for all example data sets: the file `/path/name.dat` contains the data, the file `/path/name.des` their description. Please follow the instructions in Problem 1.1 to read the example data sets from the data medium accompanying this book.

The example data sets stored on the data medium were prepared in the years 2003 and 2004 immediately preceding the production of this book. Updated versions of the example data sets originating from the internet may be available from the sites referenced. You are encouraged to visit these sites

and possibly to reproduce the statistical analyses with the updated data sets. Then, however, you will obtain slightly different results.

Having thus transferred an example data set to your computer, you may perform a statistical analysis using R. As an example, in Sects. 1.3 and 1.4, daily averages from the pyranometer comparison experiment described above are analysed using R. R is introduced in the next section.

1.2 R

R is a programming language for the graphical representation and statistical analysis of data. Expressions in R usually contain R functions; an example is given in the following line

```
result <- function(arg1, arg2, ...)          #comment
```

In this example, `result` and `arg1, arg2, ...` are R objects. R objects can be atomic or non-atomic. Typical non-atomic objects are R vectors and matrices.

R functions and objects R are not, however, explained in detail. R and also a description of R are available from [114]; to obtain the language reference start the R help system by typing `help.start()` in R. R is available for most operating systems. Depending on the operating system, start an internet browser before invoking the R help system.

Assuming you have access to R, take advantage of the opportunity and work through this section. With the following expressions

```
rholzdayfilename <- "/path/pyrcomp.dat"    #filename
rholzdayformat <- list(year=0,month=0,day=0,smi=0,eth=0) #format
rhd <- scan(rholzdayfilename, rholzdayformat)    #read file
```

you can read all values from the file `/path/pyrcomp.dat` into `rhd`. The R object `rhd` is a collection of five parallel vectors: `rhd$year, rhd$month, rhd$day, rhd$smi, rhd$eth`.

In R, the result of operators (functions) depends on the type of operands (arguments). For example, R delivers the result 3, if you type

```
seven <- 7
four <- 4
seven - four
```

However, if the operands are vectors, R checks if they have the same length and then performs the same operation on all values. Thus, the result is a vector again. In the following example,

```
difsmieth <- rhd$smi - rhd$eth
```

the first value in the vector `rhd$eth` is subtracted from the first value in `rhd$smi` with the result stored as the first value in `difsmieth`. This operation is repeated until the operand vectors are processed to their entire length.

If an R expression contains an R function, the type of the resulting object depends on the function and its arguments. For example, when a table of data is read with `scan()` and a data format, as in the example above, the result is a collection of vectors. Is the same result delivered by the R function

`read.table()`? `read.table()` is often used to read text files containing tables of data.

The number of values in an R vector is available with the R function `length()`, e.g., as the last day of the pyranometer comparison experiment is December 30, 1992, and as 1992 is a leap year

```
length(rhd$eth)
[1] 1460
```

tells you that there are 1460 values in the vector `rhd$eth`, in accordance with the duration of the comparison experiment. The vector values are indexed, e.g., the R expressions

```
rhd$smi[1:40]
[1]  13 18 39 55 22 20 15 33 20 16 30 49 73 70 48 18 24 48 19 59
[21] 56 25 11 48 25 NA 31 72 54 29 27 70 65 35 79 79 96 NA 59 NA
rhd$eth[1:40]
[1]  12 18 35 47 21 19 15 32 20 16 29 44 57 58 45 17 24 45 18 55
[21] 52 24 11 46 20 38 30 67 44 29 26 65 62 32 75 75 90 63 53 39
```

write out the first 40 values in `rhd$smi` and `rhd$eth`. On January 1st, 1989, the first day of the experiment, the MeteoSwiss pyranometer measured 13 Wm^{-2}, 1 Wm^{-2} more than the ETH instrument. The next day both measurements were identical. On January 26th, the MeteoSwiss measurement was not available. How many days with identical values occurred in the pyranometer comparison experiment? The expressions

```
smieth <- rhd$smi[(1:length(rhd$smi))[(rhd$smi == rhd$eth)]]
ethsmi <- rhd$eth[(1:length(rhd$eth))[(rhd$eth == rhd$smi)]]
```

generate two R vectors, which contain only identical daily values from the MeteoSwiss and ETH instruments. Both vectors are identical. With the expressions

```
length(smieth)
smieth
```

the length of the vector `smieth` and its values are written out. `smieth` contains missing values, because (i) the vectors with the measurements contain missing values and (ii) the equality condition `(rhd$smi == rhd$eth)`, in the expressions above, delivers `T (true)`, even if the values compared are `NAs`, i.e., if one or both values are missing. To exclude these cases, use the R function `is.na()`, which results in `T (true)` or `F (false)` and is negated with `!`. Using

```
smi1 <- rhd$smi[(1:length(rhd$smi))[(!is.na(rhd$smi))]]
eth1 <- rhd$eth[(1:length(rhd$eth))[(!is.na(rhd$eth))]]
smi2 <- rhd$smi[(1:length(rhd$smi))
            [(!is.na(rhd$smi))&(!is.na(rhd$eth))]]
eth2 <- rhd$eth[(1:length(rhd$eth))
            [(!is.na(rhd$eth))&(!is.na(rhd$smi))]]
```

the vectors `smi1, eth1, smi2` and `eth2` are generated. The vector `smi1` contains the MeteoSwiss measurements for these days only, when the value is not missing. The length of `smi1` is 1358, i.e., MeteoSwiss measurements are

missing for 102 out of 1460 days. The figures for the ETH measurements are 28 out of 1460 days. The vector `smi2` contains the MeteoSwiss pyranometer values for days when both measurements are available. The length of this vector is 1330, i.e., during the comparison period there are 1330 days with both the MeteoSwiss and ETH measurement. Consequently, the two conditions above delivering `smi2` and the comparison `(rhd$smi == rhd$eth)` in the expressions

```
smieth1 <- rhd$smi[(1:length(rhd$smi))[(rhd$smi == rhd$eth)&
            (!is.na(rhd$smi))&(!is.na(rhd$eth))]]
ethsmi1 <- rhd$eth[(1:length(rhd$eth))[(rhd$eth == rhd$smi)&
            (!is.na(rhd$eth))&(!is.na(rhd$smi))]]
```

are needed to obtain the result, that only for 215 out of 1460 days the same shortwave incoming radiation value was obtained from the measurements.

Another strength of R is the graphical representation of data. You can choose where your graphic device will be generated, e.g., with

```
postscript(file="fig11.ps",horizontal=F,width=4,height=4)
```

the plots are written to a postscript-file. Then, you may choose how the plots are depicted. With

```
par(mfrow=c(1,3))
```

you receive, e.g., three plots side by side, as shown in Fig. 1.1. With

```
dev.off()
```

the graphic device closes (R functions `dev.xxx()` provide control over multiple graphics devices). Do not forget to close a postscript-file when the graphical description is complete.

The MeteoSwiss and ETH pyranometer daily values, made available as R vectors in this section, are analysed in the next section with `mean()` and `var()`, both R functions, under the usual statistical assumptions.

1.3 Independent Identically Distributed (Iid.) Random Variables

In this section, the basics of statistical data analysis are reviewed. As an example, the pyranometer daily values are analysed in Sects. 1.3.1 and 1.3.3 under the assumptions that they are (i) identically distributed and (ii) independent from one day to the next or previous or any other following or preceding day, i.e., that they are iid. The statistical analysis under the iid. assumptions is reviewed in Sect. 1.3.2. The pyranometer daily values, however, are not iid., as follows from a discussion in Sect. 1.3.4.

1.3.1 Univariate Analysis of the Pyranometer Data

As shown in Sect. 1.2, the data resulting from the pyranometer comparison experiment are erroneous. Do the errors afflict the histograms of the data? The R expressions

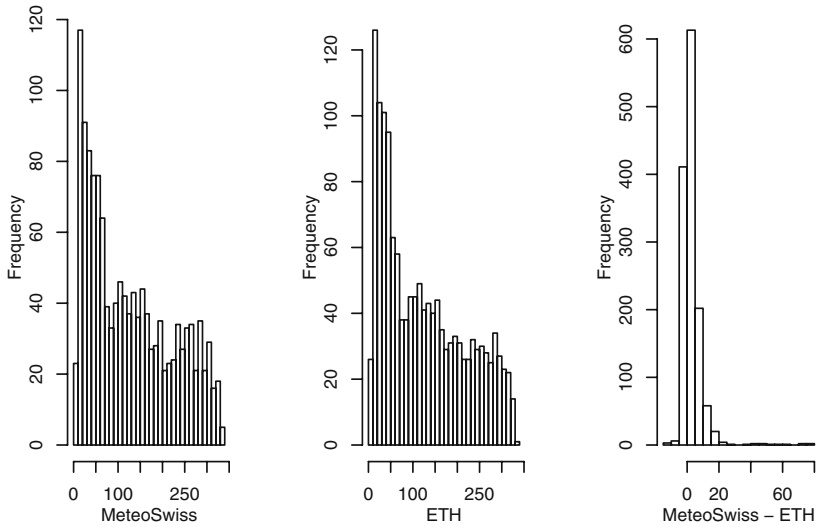


Fig. 1.1. Histograms of daily pyranometer values calculated from the measurements (in Wm^{-2}) performed by the MeteoSwiss and the ETH at Zurich-Reckenholz station (Table 1.2), and of their differences.

```
#the argument main=paste(" "," ") suppresses the title
hist(rhd$smi,breaks=34,xlab="MeteoSwiss",main=paste(" "," "))
hist(rhd$eth,breaks=34,xlab="ETH",main=paste(" "," "))
```

generate the "MeteoSwiss" and "ETH" histograms shown in Fig. 1.1, which are the histograms of the daily values of shortwave incoming radiation measured in the pyranometer comparison experiment. The histograms are very similar though the values are identical for only 1 out of 7 days. This suggests that the differences

```
dif <- rhd$smi - rhd$eth   #vector of differences
```

are small. The histogram of the pyranometer differences is obtained with

```
hist(dif,breaks=20,xlab="MeteoSwiss - ETH",main=paste(" "," "))
```

The "MeteoSwiss - ETH" histogram of these differences is also plotted in Fig. 1.1.

The "MeteoSwiss" and "ETH" histograms in Fig. 1.1, although very similar, show small deviations, e.g., in the "ETH" histogram the frequencies of small values are slightly larger than in the "MeteoSwiss" histogram. The same conclusion is drawn from the "MeteoSwiss - ETH" histogram. How much larger, one wonders, are the MeteoSwiss values in the mean?

The answer is found using the empirical moments of the MeteoSwiss and ETH measurements given in Table 1.3. These moments are calculated with the R functions `mean()` and `var()` after `NAs` have been removed from the argument vectors, e.g., with

```
dif1 <- dif[(1:length(dif))[(!is.na(dif))]]
```

Table 1.3. Empirical moments for the daily pyranometer values measured by the MeteoSwiss and ETH at Zurich-Reckenholz station.

daily value	number	mean Wm^{-2}	variance $(\mathrm{Wm}^{-2})^2$	std. dev. Wm^{-2}	covariance correlation
MeteoSwiss	1358	128.2	8924.5	94.5	8842.3 $(\mathrm{Wm}^{-2})^2$
ETH	1432	124.7	8687.3	93.7	.9974
differences	1330	3.4	47.5	6.89	

`mean(dif1)`
the mean of the daily pyranometer differences is calculated.

`mean()` calculates the empirical mean using $\hat{\mu}_X = (1/N)\sum_{i=1}^{N} x_i$, and using $\hat{\sigma}_X^2 = (1/(N-1))\sum_{i=1}^{N}(x_i - \hat{\mu}_X)^2$, `var()` calculates the empirical variance, where x_i are the observations. Both formulas are recognised referring to an introduction to Statistics. Formulas for calculating the empirical covariance and correlation are given in (1.7).

After this calculation, some questions concerning the results in Table 1.3 and Fig. 1.1 arise. These can be divided into two groups. In the first group, questions arise regarding the formulas used for the calculations as well as the accuracy of the results. In the second group the question arises, whether the error of the daily pyranometer value is random or systematic.

1. The empirical mean $\hat{\mu}_X$ is the quotient of the sum of the x_i and N, however, the empirical variance $\hat{\sigma}_X^2$ is the quotient of the sum of the squares $(x_i - \hat{\mu}_X)^2$ and $N-1$, where N is the number of observations x_i. Why then is the empirical variance calculated with the divisor $N-1$ instead of N? How exact are the results calculated by `mean()` and `var()` using these formulas? Would the results be more exact, if more than four years of data were available?
2. In Sect. 1.1, a random error was defined to be zero in the mean. Assume that the differences of the MeteoSwiss and ETH daily pyranometer values are observations of the cumulated random errors in both measurements. Are then the cumulated errors random? If the empirical mean of the differences (3.4 Wm^{-2} in Table 1.3) is not "large enough" (in the statistical sense), the cumulated errors are then random, even though the histograms and Table 1.3 suggest that they are systematic.

These questions can be answered using the definitions and propositions (1.1) to (1.6) and the results (1.15) to (1.24). As a collateral benefit, some basics from an introduction to Statistics are refreshed.

1.3.2 Unbiased and Consistent Estimators, Central Limit Theorem

The daily MeteoSwiss and ETH pyranometer values analysed above were calculated from the measurements made during the Reckenholz comparison experiment. Since error free measurements are impossible, it is assumed that the daily MeteoSwiss and ETH values are random, i.e., observations of random variables. Hence, also their differences are random. Under this basic assumption, the empirical moments in Table 1.3 are examples of estimates, as defined in (1.1).

The estimate $\hat{u}_X$ is an observed value of the estimator $\widehat{U}_X = g(X_1, \ldots, X_N)$, on condition that $\hat{u}_X = g(x_1, \ldots, x_N)$ is calculated with the formula g from the observations $x_1, \ldots, x_N$ of the random variables $X_1, \ldots, X_N$. (1.1)

In (1.1), the estimator $\widehat{U}_X$ is a *functional* of the (joint) distribution of the random variables $X_1, \ldots, X_N$, with the distribution of $g(X_1, \ldots, X_N)$ being concentrated in $\hat{u}_X$. A functional is a mapping (usually a formula), which assigns a real number to each function in a set of functions. The concept of concentration is mentioned in the remarks to the Chebyshev inequality (1.17) and discussed in an introduction to Statistics.

Often $\widehat{U}_X$ is written as $\hat{u}_X$. Once this simplification has been made, it becomes clear from the context whether $\hat{u}_X$ is an estimate or an estimator. Estimates may be found using different methods, e.g., moment, least squares and maximum likelihood estimates are known from an introduction to Statistics.

For instance, the empirical means in Table 1.3, calculated from the daily SWIR observations x_i, are estimates as defined in (1.1). In this example, the formula g is the arithmetic mean, and the estimate is calculated using $\hat{u}_X = (1/N)\sum_{i=1}^{N} x_i$ from the observed values x_i. According to definition (1.1), $\hat{u}_X$ is the observed value of the estimator $\widehat{U}_X = (1/N)\sum_{i=1}^{N} X_i$, and X_i is the theoretical daily SWIR at day i, i.e., a random variable.

With definition (1.1) alone, the questions in the remarks to Table 1.3 cannot be answered. For example, intuition suggests that more accurate means could be calculated if more than four years of data were available. This cannot be shown to be either true or false using only (1.1). To find the answers to this and the other questions in the remarks to Table 1.3, some constraining properties of the theoretical SWIR X_i at day i are therefore assumed. The usual statistical assumptions are found in (1.2).

Each observation delivers a possible value x_i of the random variable X_i. The set of the x_i is a random sample provided that
1. *an arbitrary number of observations is taken under identical conditions, and*
2. *each observation is taken independently from one another.*

(1.2)

Under these assumptions, the estimate $\hat{u}_X$ in (1.1) is calculated using g from N observations x_i, which are possible values of N identically distributed, and independent random variables X_i. The identical distributions of the X_i are required in (1.2,1), and the independence of the X_i in (1.2,2). Random variables X_i with the properties required in (1.2) are said to be independent and identically distributed (iid.). Therefore, the assumptions in (1.2) are called iid. assumptions.

Under the iid. assumptions, the estimator $\widehat{U}_X = g(X_1, \ldots, X_N)$ as defined in (1.1) is a function of N random variables X_i, which are iid. Hence, one can easily calculate its moments $\mathrm{E}\widehat{U}_X$ and $\mathrm{Var}\widehat{U}_X$, using the rules in (1.15), if (i) the estimator is linear (a weighted sum of the X_i) and (ii) the expectation $\mathrm{E}X_i$ and the variance $\mathrm{Var}X_i$ of the X_i is known. If, in addition, the distribution of X_i is known, the distribution of $\widehat{U}_X$ can then be calculated.

As an example, the moments of the mean of a random sample are calculated below. Assuming that the x_i are iid., and also assuming that μ_X is the expectation of X_i, that σ_X^2 is the variance of X_i and $\hat{\mu}_X$ is the arithmetic mean of the x_i, you obtain with (1.15,4) $\mathrm{E}\hat{\mu}_X = (1/N)\sum_{i=1}^{N}\mathrm{E}X_i = (1/N)N\mathrm{E}X_i = \mu_X$ and with (1.15,6) and (1.16,4) $\mathrm{Var}\hat{\mu}_X = (1/N^2)\sum_{i=1}^{N}\mathrm{Var}X_i = (1/N^2)N\times \mathrm{Var}X_i = (1/N)\sigma_X^2$.

The mean of a random sample is an example of an estimator without bias, which is defined in (1.3).

An estimator $\widehat{U}_X$ has no bias (is unbiased, bias-free), on condition that $\mathrm{E}\widehat{U}_X = u_X$, with u_X being the unknown theoretical value. (1.3)

If the X_i are iid. as required in (1.2), u_X in (1.3) is often an unknown parameter in the distribution of the X_i. For instance, u_X stands for μ_X or σ_X^2, assuming that X_i is normally distributed, or for λ_X, assuming that X_i is Poisson distributed. As another example, it is shown in Problem 1.4, that the empirical variance is an unbiased estimator of the theoretical variance if the X_i are iid., since $\mathrm{E}\hat{\sigma}_X^2 = \mathrm{E}\big((1/(N-1))\sum_{i=1}^{N}(x_i-\hat{\mu}_X)^2\big) = \sigma_X^2$. However, $\tilde{\sigma}_X^2 = (1/N)\sum_{i=1}^{N}(x_i-\hat{\mu}_X)^2\big)$ is not an unbiased estimate for the variance of a random sample. If an unbiased estimator is used, you can be certain that an unknown theoretical value is neither over- nor underestimated in the mean, when the estimates are calculated from many samples taken. This is a desirable property.

Another desirable property of an estimator is its consistency as defined in (1.4). If a consistent estimator $\widehat{U}_X$ is used, then the probability that the absolute difference of the estimate $\hat{u}_X$ and the true value u_X, $|\hat{u}_X - u_X|$, being less than a small ε, comes close to 1 for an arbitrarily large N. Consequently, in the mean of many random samples, the estimates become more accurate if the sample size N increases. For $N \to \infty$, $\mathrm{Var}\hat{\mu}_X \to 0$, since the estimator is defined for $N = 1, 2, 3, \ldots$.

It is assumed that $\widehat{U}_X = g(X_1, \ldots, X_N)$ *is an unbiased estimator defined for* $N = 1, 2, 3, \ldots$. *If its variance* $\mathrm{E}(\widehat{U}_X - u_X)^2 \to 0$ *for* $N \to \infty$, *then, from the Chebyshev inequality (derived in (1.17)),* $\Pr\big(|\widehat{U}_X - u_X| < \varepsilon\big) = 1 - \Pr\big(|\widehat{U}_X - u_X| \geq \varepsilon\big) \to 1$ *for* $N \to \infty$ *and for each* $\varepsilon > 0$, *is obtained.*
Then $\widehat{U}_X$ *is said to be a consistent estimator.* (1.4)

For example, the mean $\hat{\mu}_X$ of a random sample is a consistent estimator for the expectation $\mathrm{E}X = \mu_X$ of the X_i, since $\mathrm{E}\hat{\mu}_X = \mu_X$ and $\mathrm{Var}\hat{\mu}_X = \mathrm{E}(\hat{\mu}_X - \mathrm{E}\hat{\mu}_X)^2 = \mathrm{E}(\hat{\mu}_X - \mu_X)^2 = (1/N)\sigma_X^2$, $\sigma_X^2 = \mathrm{Var}X$, the variance of the X_i, as shown in the remarks to (1.2). Thus $\hat{\mu}_X$ is unbiased and $\mathrm{Var}\hat{\mu}_X \to 0$ as $N \to \infty$.

In the case of the pyranometer comparison experiment, it is concluded from (1.4) that the estimated means in Table 1.3 would be more accurate if the experiment had been performed over a longer period than the four years from 1989 to 1992. This is true under the assumption that the daily pyranometer values are in agreement with the iid. assumptions in (1.2).

This completes the answers to the first group of questions in the remarks to Table 1.3. The remaining question in the second group, i.e., whether the differences in the pyranometer daily values are due to random errors in both measurements, can only be answered if the probability distribution of the mean of the daily differences is known. Using this distribution, a confidence interval for the mean can be calculated. If then the confidence interval contains the value 0, the cumulative error is assumed to be random. Another possibility is to perform a statistical test using the distribution of the mean of the daily differences. Is it possible to assess this distribution?

Under the iid. assumptions in (1.2), it is possible to derive the probability distribution of an estimator as defined in (1.1) if the estimator, i.e., the formula g, is not too complicated and if the probability distribution of the X_i is known. This derivation is straightforward, because the distribution $F(y)$ of the function $Y = g(X_1, \ldots, X_N)$ of N random variables X_i can be calculated in (1.5)

$$\begin{aligned} F(y) &= \int\limits_{B_y} \overset{(N)}{\cdots} \int \big(f(x_1, x_2, \ldots, x_N)\big)\mathrm{d}x_1 \ldots \mathrm{d}x_N \\ &= \int\limits_{B_y} \overset{(N)}{\cdots} \int \big(f(x_1)f(x_2)\ldots f(x_N)\big)\mathrm{d}x_1 \ldots \mathrm{d}x_N \qquad (1.5) \\ B_y &= \{(x_1, \ldots, x_n) | g(x_1, \ldots, x_n) \leq y\} \end{aligned}$$

as the N-multiple integral of the N-times product of the density $f(x_i)$ of the X_i, if the X_i are independent and identically distributed with $F(x_i)$, $i = 1, \ldots, N$.

For instance, from (1.18) to (1.24) it is concluded that the mean of a random sample $\hat{\mu}_X = (1/N)\sum_{i=1}^{N} x_i$ is normally distributed with $\mathrm{E}\hat{\mu}_X = \mu_X$

and $\mathrm{Var}\hat{\mu}_X = (1/N)\sigma_X^2$, assuming that X_i is normally distributed with μ_X and σ_X^2. Thus, if the estimator is the mean of a random sample as defined in (1.2), confidence intervals are easily calculated and straightforward statistical tests are easily performed on condition that the observed random variables are normally distributed.

However, as can be seen in the histograms in Fig. 1.1, the daily MeteoSwiss and ETH pyranometer values are not normally distributed. Although the distribution of the differences of MeteoSwiss and ETH values is more symmetric, the differences are still not normally distributed. One option is to fit a distribution model to the histogram of the differences and assume that the fitted model comes close to the true probability density. Then the distribution of the sample mean is calculated from the fitted density using (1.5). When the possibly laborious calculations have been performed, confidence intervals can be calculated and statistical tests carried out.

Fortunately, this calculation is unnecessary since the mean of a random sample as defined in (1.2) is approximately normally distributed even if the distribution of the X_i is not normal. This follows from the central limit theorem as proposed in (1.6).

Let $X_1, X_2, \ldots, X_N$ be mutually independent random variables with identical distribution $F_X(x)$, and therefore, expectation $\mathrm{E}X_i = \mu_X$ *and variance* $\mathrm{Var}X_i = \sigma_X^2$, $i = 1, \ldots, N$.
Then the random variable

$$Y_N = \frac{1}{\sqrt{N}\sigma_X} \sum_{i=1}^{N} (X_i - \mu_X) \quad \textit{with distribution} \quad F_{NY}(y) \tag{1.6}$$

has asymptotically a standard normal distribution:

$$\lim_{N\to\infty} F_{NY}(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} \mathrm{e}^{-v^2/2} \mathrm{d}v.$$

According to the central limit theorem, the distribution $F_{NY}(y)$ of Y_N comes closer to the standard normal distribution, if (i) Y_N is the arithmetic mean of N independent random variables X_i which are identically distributed with $F_X(x)$, and (ii) if N increases. It is not required that $F_X(x)$ be normal. In Statistics, the key role of the iid. assumptions and of the normal distribution originates from the central limit theorem. The central limit theorem (1.6) is therefore proposed, demonstrated by examples and detailed comments and possibly proven in an introduction to Statistics.

The practitioner can conclude from the central limit theorem that the normal distribution is a useful approximation for the distribution of an arithmetic mean calculated from a random sample as defined in (1.2). This approximation holds if enough observations are available. Yet, how much is enough? If a confidence interval either for the expectation μ_Y or variance σ_Y^2 of the sample mean is calculated, the minimal N depends on the symmetry of the distribution $F_X(x)$ of the X_i. If a histogram of the observations suggests that the unknown distribution is not too far from symmetric, then no fewer than

roughly 30 observations must be made to obtain a fairly good estimator for μ_X, and no fewer than approximately 100 observations in the case of σ_X^2.

Returning to the pyranometer comparison experiment, it is assumed that the differences of the MeteoSwiss and ETH daily values are iid. as defined in (1.2). It can therefore be concluded from the central limit theorem that a normal distribution is a fairly good approximation for the distribution of the mean (3.4 Wm^{-2}, in Table 1.3) of the differences. Consequently, confidence intervals can be calculated and tests performed in Problem 1.5 with the result that the differences of the pyranometer daily values are not zero in the mean. This answers the last question in the remarks to Table 1.3.

Having performed a univariate analysis of the pyranometer daily values, we have refreshed the statistical estimation under the iid. assumptions defined in (1.2). But are the pyranometer daily values iid.? Were the conditions of the MeteoSwiss and ETH measurements identical during the four years of the comparison experiment? Are the daily MeteoSwiss and ETH values and their differences independent from day to day? We will return to these questions at the end of this section.

1.3.3 Bivariate Analysis of the Pyranometer Data

The pyranometer daily values stemming from the MeteoSwiss and ETH measurements at Zurich-Reckenholz Station can also be analysed jointly. For each day in the four year comparison period there are two values available and thus, a bivariate analysis can be performed. It is assumed that pairs (x_i, y_i) and (x_j, y_j) are independent for days $i \neq j$.

In the scatterplot on the left in Fig. 1.2, the ETH daily values are plotted against those of the MeteoSwiss for each day using the R function `plot()`

```
plot(rhd$smi,rhd$eth,type="p",xlab="MeteoSwiss",ylab="ETH",pch=".")
```

In this scatterplot, with

```
x <- c(0,340)
y <- c(0,340)
lines(x,y)
```

a straight line is drawn. All data points would be on this straight line, assuming the daily MeteoSwiss and ETH values were identical, i.e., if the measurements were not erroneous. This scatterplot shows that the ETH values are smaller (the differences amount to some Wm^{-2}) than the MeteoSwiss values in the mean. For a few days, however, differences larger than 30 Wm^{-2} are obtained. Another graph suitable for bivariate data is the two-dimensional histogram. Such a histogram of the pyranometer daily values is plotted on the right, in Fig. 1.2. In this histogram, the MeteoSwiss-ETH plane is subdivided in bi-dimensional classes, 10 Wm^{-2} wide in both dimensions. Remember that 10 Wm^{-2} is the width of the classes in the univariate histograms in Fig. 1.1. As the MeteoSwiss-ETH plane is the drawing plane of the scatterplot, on the left in Fig. 1.2, it is concluded from this scatterplot that most two-dimensional classes contain no pairs of daily values; classes on or near the diagonal of the

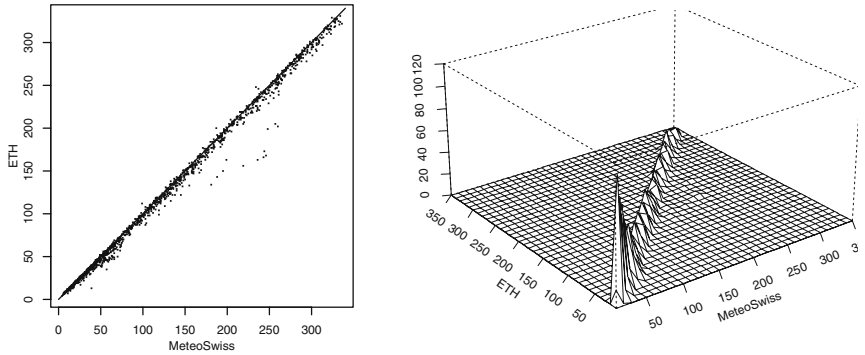


Fig. 1.2. Scatterplot (on the left) and two-dimensional histogram (on the right) of the daily pyranometer values measured in Wm^{-2} at Zurich-Reckenholz station as described in Sect. 1.1.

MeteoSwiss-ETH plane however contain many pairs. The number of pairs of daily values in a two-dimensional class is the joint frequency of the ETH and MeteoSwiss daily values. The joint frequencies are plotted on the vertical (the z-) axis, resulting in the two-dimensional histogram of the MeteoSwiss and ETH daily pyranometer values on the right in Fig. 1.2.

Using projections of the joint frequencies on the z-ETH and z-MeteoSwiss planes, one-dimensional marginal frequencies are obtained. Histograms of these marginal frequencies are plotted on the left and in the middle in Fig. 1.1. These histograms, as well as the empirical means and variances of the ETH and MeteoSwiss daily values in Table 1.3, are both very similar.

From both the one-dimensional histograms and one-dimensional moments, it is not possible to conclude on the interior of the two-dimensional scatterplot and of the two-dimensional histogram. Such a conclusion is impossible because identical one-dimensional histograms and moments can be produced by arbitrary numerous patterns of data points in the MeteoSwiss-ETH plane.

The bivariate analysis of daily pyranometer values aims at a possible statistical relationship between the ETH and MeteoSwiss values. It aims to supply some information about the interior of the two-dimensional scatterplot and of the two-dimensional histogram in Fig. 1.2. A possible solution to the problem is a reduction in the one-dimensional case by calculating the differences $d_i = x_i - y_i$ from the pairs (x_i, y_i) of the pyranometer daily values. If these differences are small, as shown on the right in Fig. 1.1, most of the data points in a scatterplot will be close to the diagonal, as shown in the scatterplot on the right in Fig. 1.2. Another possibility would be to plot a two-dimensional scatterplot and/or histogram and to calculate an estimate which describes the strength of the statistical relationship between the MeteoSwiss and ETH pyranometer daily values.

This is done by calculating the empirical covariance or correlation, as defined in (1.7).

N observations are performed, each of which delivers a pair (x_i, y_i) of possible values of the random variables (X_i, Y_i), which are assumed to be iid. $\hat{\mu}_X$ and $\hat{\mu}_Y$ are the empirical means of the x_i and the y_i, $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$ are the empirical variances. Then:

1. $\hat{c}_{XY} = 1/(N-1)\sum_{i=1}^{N}(x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y)$ *is the empirical covariance,*
2. $\hat{\varrho}_{XY} = \hat{c}_{XY}/(\hat{\sigma}_X\hat{\sigma}_Y)$ *is the empirical correlation of x_i and y_i.*

(1.7)

In definition (1.7), $\hat{c}_{XY}$ depends (as do the empirical variances $\hat{\sigma}_X^2$ and $\hat{\sigma}_Y^2$) on the unit of the x_i and y_i, but not on an additive constant. $\hat{\varrho}_{XY}$, by contrast, no longer depends on the units of the measurements. This is due to the standardisation of the empirical covariance as $\hat{c}_{XY}$ is divided by $\hat{\sigma}_X\hat{\sigma}_Y$ to obtain $\hat{\varrho}_{XY}$. $\hat{\varrho}_{XY}$ is also called the empirical correlation coefficient.

For example, under the iid. assumptions (the observations being independent from day to day and the distributions identical for every day), the empirical covariance and correlation of the daily pyranometer values at Zurich-Reckenholz station, given in Table 1.3, are calculated according to (1.7) using the R functions `var()` (or `cov()`) and `cor()`:

```
var(smi2,eth2)
cor(smi2,eth2).
```

The argument vectors `smi2` and `eth2` are generated in Sect. 1.2. They contain pairs (x_i, y_i), $i = 1, \ldots, N$, $N = 1330$, of MeteoSwiss and ETH values measured at the same day i and both available, cf. Sect. 1.2. Please note that `var()` is also used for the calculation of empirical variances in the univariate case demonstrated in Sect. 1.3.1.

If the empirical covariance is a number which describes the strength of the statistical linear relationship between x_i and y_i, then it is likely related to the differences $z_i = x_i - y_i$, (x_i, y_i) as defined in (1.7). This suggests to calculate the empirical variance of the differences. In (1.8), the differences $z_i = x_i - y_i$ are iid., if the (x_i, y_i) are iid. as assumed in (1.7).

$$(N-1)\hat{\sigma}_Z^2 = \sum_{i=1}^{N}(z_i - \hat{\mu}_Z)^2 = \sum_{i=1}^{N}\big((x_i - y_i) - (\hat{\mu}_X - \hat{\mu}_Y)\big)^2 \tag{1.8}$$

$$= \sum_{i=1}^{N}\big((x_i - \hat{\mu}_X) - (y_i - \hat{\mu}_Y)\big)^2$$

$$= (N-1)\hat{\sigma}_X^2 + (N-1)\hat{\sigma}_Y^2 - 2(N-1)\hat{c}_{XY} \tag{1.9}$$

(1.9) implies that the empirical variance of the differences $z_i = x_i - y_i$ is small provided that double the empirical covariance of the (x_i, y_i) is approximately the sum of the empirical variances of the x_i and y_i. This is illustrated by the example of the daily pyranometer values, as shown in Table 1.3.

The properties of an empirical correlation are listed in (1.10).

If $\hat{\varrho}_{XY}$ is an empirical correlation as defined in (1.7,2), then:
1. *$\hat{\varrho}_{XY}$ does not depend on the units of measurement of the x_i and y_i or an additive constant.*
2. *If $\hat{\varrho}_{XY} = \pm 1$, a perfect linear relationship between x_i and y_i exists and all pairs (x_i, y_i) are on a straight line.*
3. *If $\hat{\varrho}_{XY} = 0$, no linear relationship between x_i and y_i exists; however, another (non-linear) relationship can exist.*
4. $-1 \leq \hat{\varrho}_{XY} \leq 1$.
5. *$|\hat{\varrho}_{XY}|$ describes the strength of a possible linear relationship between the x_i and the y_i.*

(1.10)

Properties 1. and 2. in (1.10) follow, assuming $y_i = bx_i + a$, directly from definitions (1.7): straight lines with a positive slope ($b > 0$) deliver the empirical correlation 1, and straight lines with a negative slope deliver -1. Property 3. is shown by giving a counter-example, e.g. using $y_i = x_i^2$. Property 4. follows from the Cauchy-Schwarz inequality $|\boldsymbol{u}^T\boldsymbol{v}|^2 \leq (\boldsymbol{u}^T\boldsymbol{u})(\boldsymbol{v}^T\boldsymbol{v})$, $\boldsymbol{u}$ and $\boldsymbol{v}$ column vectors containing $u_i = x_i - \widehat{m}_X$ and $v_i = y_i - \widehat{m}_Y$, or $(\sum_i u_i v_i)^2 \leq (\sum_i u_i^2)(\sum_i v_i^2)$. The Cauchy-Schwarz inequality is formulated as a property of linear vector spaces in (6.4).

The empirical covariance and empirical correlation are not robust estimates: they are sensitive to outliers. Consequently, inspect scatterplots of the observations before calculating these statistics. Practical correlation analysis is, however, tricky. A non-zero empirical correlation does not necessarily mean a causal relationship between the two observed variables. In the case of the pyranometer comparison experiment, a large empirical correlation is expected for the MeteoSwiss and the ETH daily values because the measurement error is small. Nonetheless, it is not possible to obtain an empirical correlation $\hat{\varrho}_{XY} = 1$ since every pyranometer measurement is erroneous.

Under the iid. assumptions formulated in (1.2), $\hat{\varrho}_{XY}$ and $\hat{c}_{XY}$ are estimates of the joint (bivariate) second moments of two random variables X and Y.

If X and Y are random variables with the expectations μ_X and μ_Y and the variances σ_X^2 and σ_Y^2, then
1. *$c_{XY} = \mathrm{Cov}(X, Y) = \mathrm{E}((X - \mu_X)(Y - \mu_Y))$ is the covariance and*
2. *$\varrho_{XY} = c_{XY}/(\sigma_X \sigma_Y)$ is the correlation of X and Y.*

(1.11)

The correlation is often estimated. It has favourable properties when the observed variables are bivariate normally distributed. A bivariate (two-dimensional) normal distribution is often used to model the frequencies in the interior of a scatterplot. Examples are given in Table 1.4 and Fig. 1.3.

Two random variables X_1 and X_2, with the joint probability density given in (1.12), are said to be two-dimensional or bivariate normally distributed:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp(-a), \quad \text{with} \tag{1.12}$$

$$a = \frac{1}{2(1-\varrho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\varrho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right).$$

Table 1.4. Empirical moments of simulated bivariate normal distributions. Pairs (x_{1k}, x_{2k}), $k = 1, \ldots, 5000$, were simulated using (1.13) with the parameter values $\mu_1 = 100$, $\mu_2 = 50$, $\sigma_1 = 10$, $\sigma_2 = 7$, and $\varrho = 0.90, -0.5, 0.0$.

ϱ	$\hat{\varrho}$	$\hat{\mu}_{X_1}$	$\hat{\sigma}_{X_1}$	$\hat{\mu}_{X_2}$	$\hat{\sigma}_{X_1}$
0.90	0.91	99.70	10.17	49.82	7.13
−0.50	−0.51	99.99	9.98	50.01	6.98
0.00	0.00	99.86	9.89	50.14	7.03

From the special case of the standardised ($\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$) two-dimensional normal density, defined in (1.12), the one-dimensional marginal densities are derived in (1.25) to (1.30). The derived marginal densities are standard normally distributed. The very same distribution (i.e., standard normal) is assumed for the random variables X_1' and X_2' used for the definitions (1.13).

$$X_1 = \mu_1 + \sigma_1 X_1' \qquad X_2 = \mu_2 + \varrho\sigma_2 X_1' + \sqrt{1 - \varrho^2}\sigma_2 X_2' \tag{1.13}$$

In these definitions, the two random variables X_1 and X_2 are constructed as linear combinations of X_1' and X_2', both being two independent random variables with standard normal distribution. From the definitions (1.13) it is derived, in Problem 1.7, that μ_1 and μ_2 are the expectations, σ_1^2 and σ_2^2 the variances, and that ϱ is the correlation of X_1 and X_2.

Definitions (1.13) are used to simulate bivariate normal distributions. For example, in R you can simulate 5000 observations of X_1' and X_2' with

```
xs1 <- rnorm(5000,mean=0.0,sd=1.0)
xs2 <- rnorm(5000,mean=0.0,sd=1.0)
```

which then are transformed using the definitions (1.13). However, you can simulate a two-dimensional normal distribution easier with `rmvnorm()`. Using `rmvnorm()`, 5000 values of three example bivariate normal distributions are simulated. The parameters of the distributions and the empirical moments calculated from the simulations are listed in Table 1.4, with the scatterplots and marginal distributions of the simulated pairs (x_i, y_i) being plotted in Fig. 1.3. Here, different point patterns in the interior of the scatterplots generate very similar marginal frequencies.

More generally, from (1.12) and (1.13) it is concluded that two random variables with a joint density as defined in (1.12) have normal marginal distributions which do not depend on ϱ.

Assuming X and Y to be jointly normally distributed with the density given in (1.12), the empirical correlation $\hat{\varrho}_{XY}$ is an estimator for the parameter ϱ in (1.12) and (1.13). When observations (x_i, y_i) in agreement with the iid. conditions proposed in (1.2) are available, the distribution of $\hat{\varrho}_{XY}$ can be

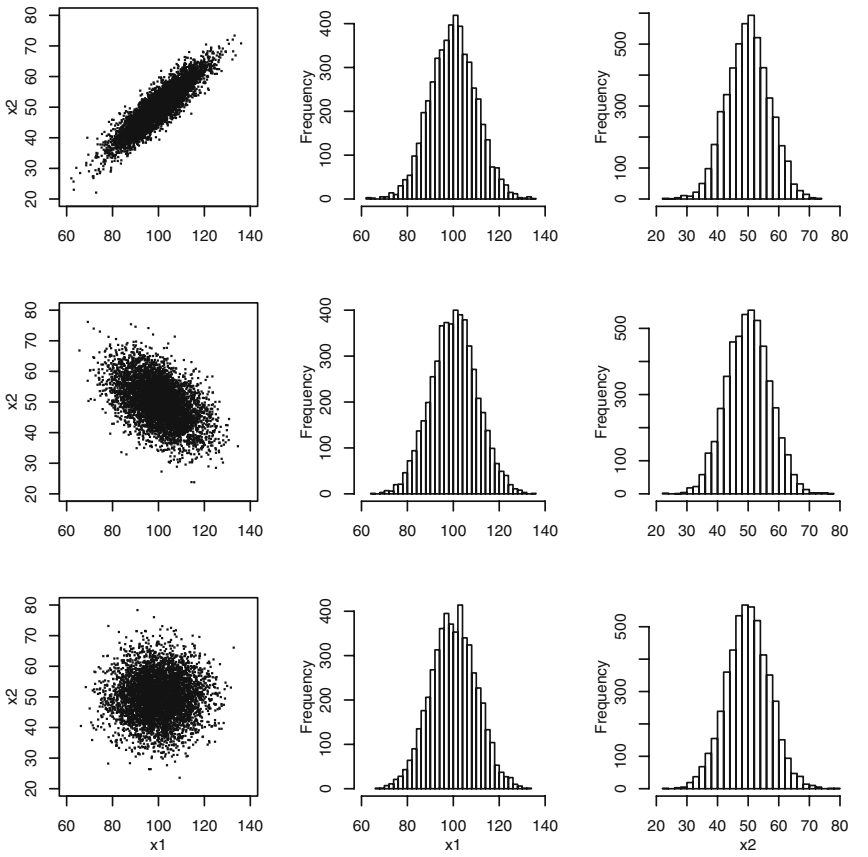


Fig. 1.3. Bivariate normal distributions: Scatterplots and marginal frequencies of the simulations using (1.13) and the values $\varrho = 0.90$, $\varrho = -0.50$ and $\varrho = 0.00$ (from top to bottom) in Table 1.4.

estimated and therefore, confidence intervals for $\hat{\varrho}_{XY}$ can further be calculated and tests performed with the procedures described in an introduction to Statistics.

1.3.4 Are the Pyranometer Daily Values Iid.?

The above uni- and bivariate analyses of data obtained from the pyranometer comparison experiment are performed under the iid. assumptions in (1.2). The succinct results show that either the MeteoSwiss or the ETH measurements, or both, are afflicted with a systematic error, and that there are, moreover, random errors in both measurements. An estimate for the systematic error is the mean of the differences in Table 1.3, and an estimate for the

cumulative random error in both measurements is the standard deviation of these differences.

However, these estimates are viable only if the pyranometer values are iid., as defined in (1.2); if not iid., they are merely preliminary approximations. The pyranometer measurements cannot be repeated under identical conditions and thus are not in agreement with condition (1.2,1).

Are the pyranometer data reconcilable with condition (1.2,2) stipulating independent measurements? Assuming that x_i is the daily value calculated from the measurements with the MeteoSwiss pyranometer and that y_i is the daily value stemming from the ETH pyranometer, (1.2,2) demands that pairs (x_i, x_j) and pairs (y_i, y_j) be independent for any days i and j over the period of the comparison experiment. The independence could have been tested under additional suitable assumptions (e.g., normality), if, for any days $i \neq j$, the correlations of the pairs (X_i, X_j) and (Y_i, Y_j) had been available. Estimates for these correlations are the empirical correlations which are calculated using (1.7) from pairs (x_{ik}, x_{jk}) or (y_{ik}, y_{jk}), $k = 1, 2, \ldots, n(i, j)$. However, $n(i, j) = 1$ for any pair of days (i, j) in the measurement period because the pyranometer measurements cannot be repeated.

Consequently, the pyranometer daily values are not in line with the condition (1.2,1) and it is not known whether they agree with the condition (1.2,2). This question is reconsidered in the remarks to Figs. 1.8 and 1.9.

1.4 Time Series

Plotting the daily pyranometer MeteoSwiss and ETH values x_i and y_i vs. the day i, i.e., vs. time, it is shown in Sect. 1.4.1 that pairs (x_i, x_j) or (y_i, y_j) are not independent for days $i \neq j$ over the period of the Reckenholz pyranometer comparison experiment. In Sect. 1.4.2, a statistical model for the error of the pyranometer daily means is proposed.

1.4.1 Time Series Plots

Each pair of daily pyranometer values has a time, which is given in the file `/path/pyrcomp.dat` by the three-tuple (year, month, day) as shown in Sect. 1.1. Data with the property that each value is associated to the time of observation are called *time series*. A time series is a sequence of observed values, as defined in (2.1). In R, the functions for the analysis of time series are made available with

```
library(ts)
```

and a special data structure is used. For example, using the vectors in the R object `rhd` generated in Sect. 1.2 from the text file with the pyranometer daily values, the expressions

```
smi <- ts(rhd$smi,start=1989,frequency=365,)
```

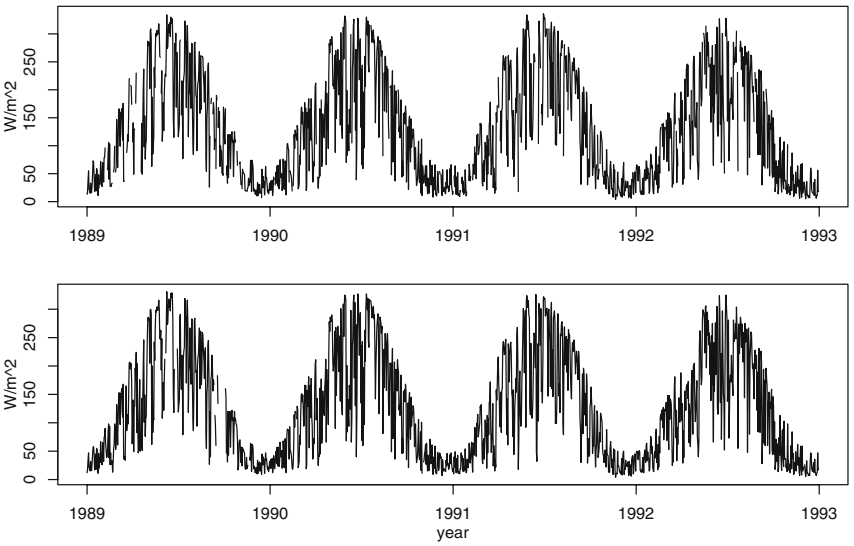


Fig. 1.4. Time series of daily pyranometer values calculated from the MeteoSwiss (above) and the ETH (below) measurements at Zurich-Reckenholz station. Time in years.

```
eth <- ts(rhd$eth,start=1989,frequency=365,)
```

produce `smi` and `eth`, which are plotted in Fig. 1.4 using

```
plot(smi,type="l",xlab="year",ylab="W/m2")
plot(eth,type="l",xlab="year",ylab="W/m2")
```

bearing in mind that `smi` and `eth` are R time series objects. In R, a time series object can be generated from a vector of values obtained by observing with a constant time step Δt. In the R function `ts()`, the argument `frequency = ...,` can be replaced by `deltat = ...,`. `deltat = ...,` is (the fraction of, cf. the R help system) the sampling period between the observations.

More technically, in the example above, the R function `ts()` associates a time domain with the argument vector `rhd$smi`, using values for the arguments `start` and `frequency` which correspond to the time when the measurements begin and to the number of observations per unit time period. Since 1992 is a leap-year, it would be advisable to generate the R time series objects

```
smif1 <- ts(rhd$smi,start=1,frequency=1,)
ethf1 <- ts(rhd$eth,start=1,frequency=1,)
```

from the pyranometer daily values. `smif1` is plotted in Fig. 1.5. In Fig. 1.5, however, the time unit is one day and the first value in the time series, which was measured on January 1st, 1989, is associated with day 1.

The systematic error in the pyranometer daily values, clearly visible in the histogram of the differences in Fig. 1.1, cannot be seen in Fig. 1.4, even though this error amounts to 3.4 Wm^{-2} in the mean (Table 1.3). From the histogram

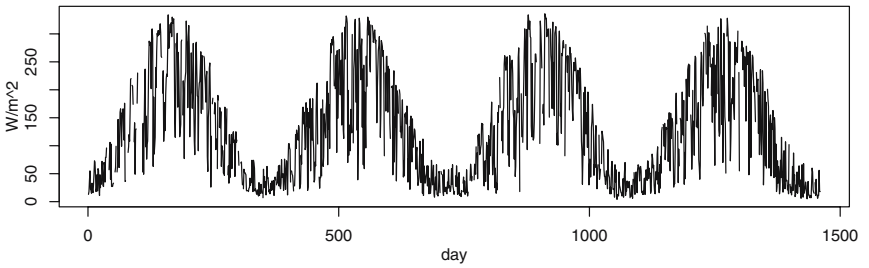


Fig. 1.5. Time series of daily pyranometer values calculated from the MeteoSwiss measurements at Zurich-Reckenholz station. Time in days, starting January 1, 1989.

of the differences in Fig. 1.1 and also from the scatterplot in Fig. 1.2, it is known that there are a few differences larger than 50 Wm^{-2}. These larger differences occurred in Spring 1992. This can been seen when the time series plots in Fig. 1.4 are compared. Therefore, from the time series plotted in Fig. 1.4, with the R expressions

```
smiapr <- window(smi,c(1992,60),c(1992,150))
ethapr <- window(eth,c(1992,60),c(1992,150))
```

time slices containing the measurements from the Spring months of 1992 are constructed. These time slices are plotted in Fig. 1.6.

The plots in Fig. 1.6 show that at the beginning of April 1992 (1992.25 on the time axis), the ETH pyranometer measured 80 Wm^{-2} less SWIR than the MeteoSwiss pyranometer did. The period with these large differences continues until May 1992 (1992.33 on the time axis). At this time and later, approximately the same SWIR values were obtained with both pyranometers.

What then happened in April 1992 to cause the unusually large differences measured in this month? Was the error larger than usual for only one or for both measurements? Which of the two measurements is more accurate in that month? By only using the available pyranometer data, these questions cannot be answered, and in the history of the MeteoSwiss and ETH measurements at Zurich-Reckenholz station, no additional information can be found that would help to find the source of these large differences.

In Fig. 1.4 you can clearly see the annual (or seasonal) cycle of SWIR at Zurich-Reckenholz station. In July, with 300 Wm^{-2}, SWIR is approximately ten times (one order of magnitude) larger than in January. Apart from the annual cycle, the time series also show large *synoptic fluctuations*, i.e., fluctuations observed within a few days. Such fluctuations are also found in the daily cloud cover at Zurich-Reckenholz station (not plotted). Clearly, the synoptic fluctuations in the SWIR time series are produced by fluctuations of the cloud cover.

When synoptic fluctuations are (at least partly) removed from the time series by smoothing, the *seasonal cycle* becomes more clearly visible. In R, use the function `filter()` to smooth a time series. However, `filter()` applies

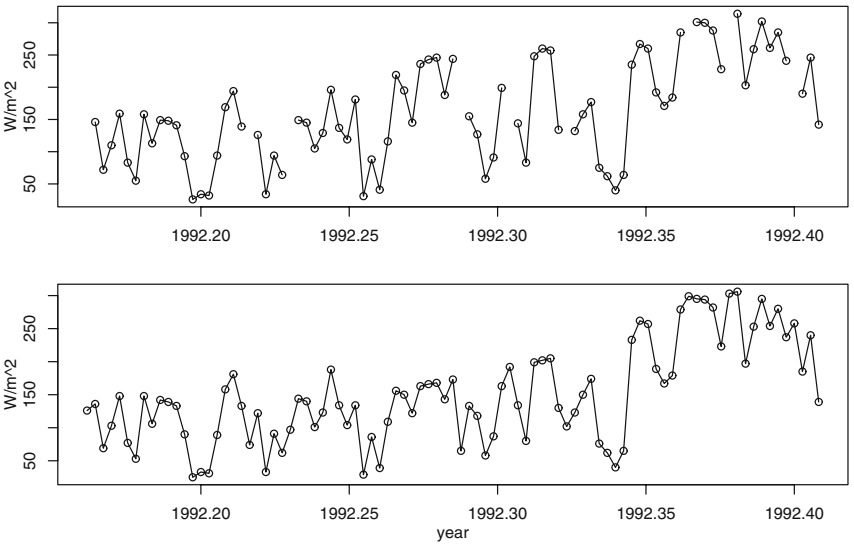


Fig. 1.6. MeteoSwiss (above) and ETH (below) pyranometer time series measured at Zurich-Reckenholz station in Spring 1992.

only if no values are missing in the time series. Since numerous pyranometer daily values are missing, as was learned in Sect. 1.2, there are noticeable gaps in the time series plots which can be seen in Fig. 1.4 and, even more clearly in Fig. 1.6.

These gaps stem from failures of the MeteoSwiss or ETH instrument and/or the corresponding data acquisition system. Generally, the ETH time series has fewer gaps than the MeteoSwiss time series. The longest period with no gaps is found in the ETH time series from January 1991 to November 1992. An R time series for this time slice is generated and smoothed with a running mean covering 11 days using

```
eth2 <- window(eth,c(1991,1),c(1992,335))
rmean <- c(1/11,...,1/11)                    #11 identical weights
eth2smo <- filter(eth2,rmean,method="convolution",sides=2)
```

and then plotted in Fig. 1.7. The R function `filter()` with argument `method="convolution"` is introduced in Sect. 2.4.

In comparison to the unsmoothed time series in Fig. 1.4, the smoothed time series in Fig. 1.7 shows more clearly that SWIR undergoes a seasonal cycle. In addition, it is clearly seen in Fig. 1.7 that periods lasting a few days with large SWIR values are followed by periods with smaller SWIR values. Large SWIR daily values are measured over periods of sunny weather with small total cloud amount; during periods with large total cloud amount, on the other hand, small SWIR values are obtained.

Fig. 1.7. ETH pyranometer daily values at Zurich-Reckenholz station from January 1991 to November 1992 smoothed with a running mean covering 11 days.

Two R time series with identical time domains can be added, subtracted, multiplied or divided. The result is a time series with the original time domain and values calculated from the original series by addition, etc. For instance, from the `smi` and `eth` time series plotted in Fig. 1.4, the time series difference `difts` is obtained with

```
difts <- smi - eth
plot(difts,type="l",xlab=" ",ylab="W/m2")
```

and then plotted in Fig. 1.8 (above). In this plot, with

```
lines(c(1989.28,1989.28),c(-15,75), lty=2)
lines(c(1989.43,1989.43),c(-15,75), lty=2)
lines(c(1992.24,1992.24),c(-15,75), lty=2)
lines(c(1992.33,1992.33),c(-15,75), lty=2)
```

two time slices are marked off with dashed vertical lines. The first starts in the middle of April and lasts until the end of May 1989, the second is in April 1992. In the first time slice, unusually small negative differences are obtained from the measurements; in the second, unusually large differences were observed, as shown in detail in Fig. 1.6. It is suspected that over both periods the error of at least one of the two measurements was larger than usual.

In a duplicate of `difts`, the values in these periods are set to `NA` and then the duplicate is plotted in Fig. 1.8 (below). This plot shows more negative differences in summer 1989 than during the summers of 1990, 1991 and 1992. In winter, the differences are generally larger than in summer. An exception is the summer of 1991, with similar differences in winter 1990/91 and 1991/92. Thus, with the exception of summer 1991, the differences are large, when small daily SWIR values were measured, and small, when larger values were observed.

This discovery gives reason to calculate the relative differences in the pyranometer measurements with

```
reldif <- difts/smi
```

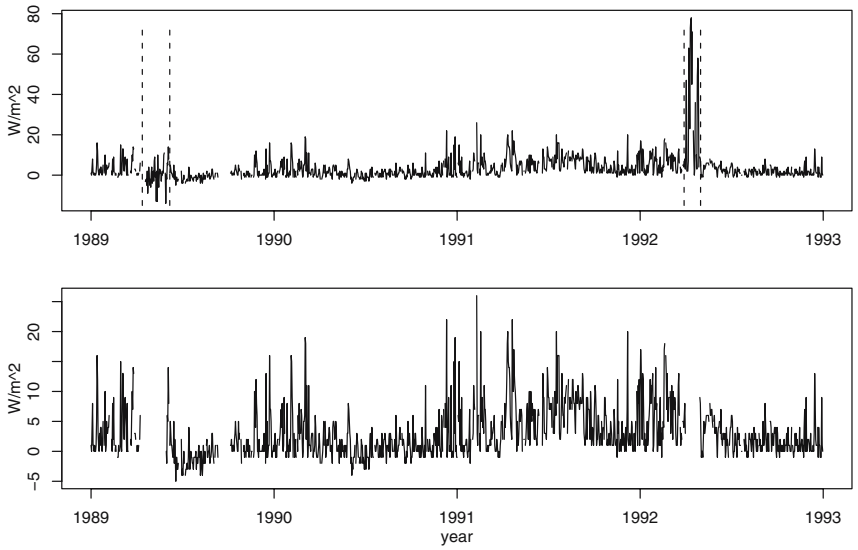


Fig. 1.8. Differences of the pyranometer daily values from the MeteoSwiss and ETH at Zurich-Reckenholz station (above), and the same differences (below), however without the values in the time slices marked off in the above plot.

The relative differences are plotted in Fig. 1.9, which shows that the relative differences are much larger in winter than in summer and thus undergo a seasonal cycle.

This annual cycle is due, on the one hand, to the slight ventilation of the MeteoSwiss pyranometer at Zurich-Reckenholz station. This device prevents dew or frost from coating the glass dome of the MeteoSwiss pyranometer provided that the temperature is not below -2^0C. On the other hand, the ETH pyranometer is not ventilated, and on cold mornings, possible dew and frost have to be evaporated, until the full radiation can penetrate the dome to reach the sensor plate. This can be seen if the original measurements (taken in intervals of 10 or 5 minutes, Table 1.2) are compared. Hence, in the mean during the season with cold mornings, the ETH measurements record less SWIR than do the MeteoSwiss measurements.

With the visual inspection of the time series plots, we have accomplished two goals. Firstly, an explanation has been found for the non-zero mean of the differences of the daily MeteoSwiss and ETH pyranometers values, given in Table 1.3. This corroborates the result from the analyses of the pyranometer daily values in Sect. 1.3. Thus it is recommended to measure SWIR with a ventilated pyranometer at least in the middle latitudes. Secondly, the question concluding Sect. 1.3, i.e., whether the daily ETH and MeteoSwiss values are independent, can now be answered.

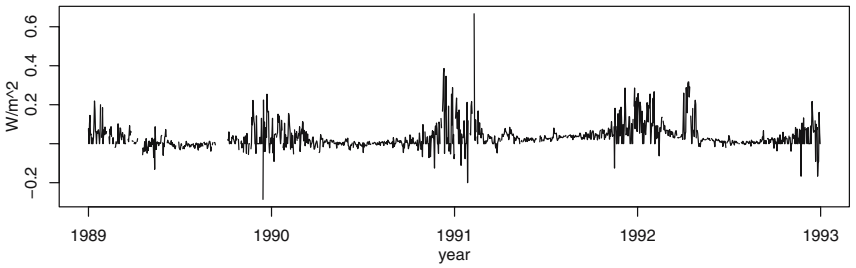


Fig. 1.9. Relative differences of the MeteoSwiss and ETH pyranometer daily values measured at Zurich-Reckenholz station.

At Zurich-Reckenholz station, a sunny day is followed by another sunny day in the majority of the cases. The cloudy days behave in the same way. Consequently, in the pyranometer time series, a large daily value of SWIR is followed by another large value and a small value is followed by another small value, in the majority of cases: the daily values of SWIR are not independent from one another. This pattern can be seen in plots where the single values can be easily counted, for example in Fig. 1.6. It is related to changes in the daily cloud cover, as SWIR is determined by the elevation of the sun (with its seasonal cycle) and, to a lesser extent, by the cloud cover (with its synoptic fluctuations). Generally, meteorological variables (SWIR, cloud cover, temperature, precipitation, etc.) observed at Zurich-Reckenholz station are not independent from day to day, because often weather conditions are stable for a few consecutive days.

The differences of the pyranometer daily values are also not independent from day to day. They mostly stem from measurements taken during the morning hours, when the unventilated ETH pyranometer is coated by dew or frost, whereas the MeteoSwiss sensor is exposed to the full radiation as it is ventilated. Since the presence or absence of dew or frost is not independent from day to day at Zurich-Reckenholz station, the differences of the pyranometer daily values are not independent.

Consequently, the MeteoSwiss and ETH pyranometer daily values observed at Zurich-Reckenholz station, and also their differences, are not independent from day to day. These values are therefore not in agreement with assumption (1.2,2). How strong is the dependence? An answer to this question is given in Problem 2.32. The daily values resulting from the Zurich-Reckenholz pyranometer comparison experiment are also not in agreement with assumption (1.2,1), seeing that the measurements cannot be repeated, as argued in the closing remarks in Sect. 1.3. As the observations are thus not iid., the empirical moments in Table 1.3 most likely do not estimate the theoretical moments in the model proposed in (1.14) without bias and consistently. Unbiased and consistent estimators are defined in (1.3) and (1.4).

1.4.2 A Model for the Error of the Pyranometer Measurements

In the pyranometer comparison experiment described in Sect. 1.1, the daily means calculated from the ETH measurements are afflicted with a systematic (non-zero mean) error. This error is due to the missing ventilation of the ETH pyranometer. In addition, both MeteoSwiss and ETH daily means are afflicted with a random error, since error-free measurements are not possible. Both error types are found in the model proposed in (1.14).

In (1.14), $X(t)$ is used for the MeteoSwiss daily means and $Y(t)$ for the ETH ones. Both are split into the true values $\mathrm{E}X(t)$ and $\mathrm{E}Y(t)$ and the random errors $e_X(t)$ and $e_Y(t)$, i.e., $X(t) = \mathrm{E}X(t) + e_X(t)$ and $Y(t) = \mathrm{E}Y(t) + e_Y(t)$. Thus, in this model, the daily value at day t, (for example $X(t)$), is the sum of the expectation of $X(t)$ at day t, $\mathrm{E}X(t)$, and the error $e_X(t)$ at day t. Thereafter, the differences $D(t) = X(t) - Y(t)$ are calculated.

$$\begin{aligned} D(t) &= X(t) - Y(t) = \big(\mathrm{E}X(t) - \mathrm{E}Y(t)\big) + \big(e_X(t) - e_Y(t)\big) \qquad (1.14)\\ \mathrm{E}D(t) &= \mathrm{E}X(t) - \mathrm{E}Y(t)\\ \mathrm{Var}D(t) &= \mathrm{Var}e_X(t) + \mathrm{Var}e_Y(t) \end{aligned}$$

Under the assumption that the difference $\mathrm{E}X(t) - \mathrm{E}Y(t)$ of the true values and the difference of the random errors $e_X(t) - e_Y(t)$ add to the difference $D(t)$ in the first line in (1.14), and because random errors are non-zero in the mean, i.e., $\mathrm{E}e_X(t) = \mathrm{E}e_Y(t) = 0$ as defined in Sect. 1.1, the expection of the differences $\mathrm{E}D(t)$ in the second line is obtained with (1.15,3). Furthermore, the variance of the differences is calculated using (1.15,8), since $e_X(t)$ and $e_Y(t)$ are independent. This independence is due to the installation of pyranometers at Zurich-Reckenholz stations, details being given in Table 1.2.

An estimate of the systematic error $\mathrm{E}X(t) - \mathrm{E}Y(t)$ is obtained by calculating an estimate for the expectation of the differences $\mathrm{E}D(t)$. Estimates of the random errors $\mathrm{Var}e_X(t)$ and $\mathrm{Var}e_Y(t)$ are obtained by calculating an estimate for $\mathrm{Var}D(t)$, if it is additionally assumed that the random errors in both measurements are equal: If $\mathrm{Var}e_X(t) = \mathrm{Var}e_Y(t)$, then $\mathrm{Var}e_X(t) = (1/2)\mathrm{Var}D(t)$.

The empirical moments of the differences in Table 1.3 are, in fact, neither unbiased nor consistent estimators for the moments $\mathrm{E}D(t)$ and $\mathrm{Var}D(t)$ in Model (1.14), as the differences are not iid.. However, these empirical moments, together with the histogram in Fig. 1.1, on the right, and the scatterplot in Fig. 1.2, motivate the diagnosis of a systematic error in either one or both measurements (in Sect. 1.3). Then, when visually inspecting the time series plots in Sect. 1.4.1, a possible cause of the systematic error is found.

Is it possible to estimate $\mathrm{E}D(t)$ and $\mathrm{Var}D(t)$ consistently and without bias, using the observed pyranometer daily differences plotted in Fig. 1.8 (above)? Do assumptions exist, which are (unlike the iid. ones) in agreement with the differences of the pyranometer daily values? If estimators can be constructed under suitable assumptions, what about their expectations, variances and distributions?

Presumably, the answers to these questions are not easy to find, and therefore, not surprisingly, prior to the answers given in Sects. 2.5.1 and 2.6, a solid theoretical framework is to be constructed in Sects. 2.1 to 2.4.

1.5 Supplements

The supplement section contains formulas for calculating the moments of linear combinations of random variables, a derivation of the Chebyshev inequality, a derivation of the distribution of a sum of independent and normally distributed random variables, a derivation of the marginal densities of a two-dimensional normal density, and some properties of the multivariate normal distribution.

1.5.1 Formulas for Calculating the Moments of Random Variables

Formulas for calculating the moments of random variables are normally found in an introduction to Statistics. Using them, you can calculate the moments of linear combinations of random variables and, as important examples, the moments of estimators, as is demonstrated in the remarks to definitions (1.1) and (1.2). Some formulas which are often used are given in (1.15) and (1.16).

Let X, Y, X_i, X_j be random variables, a, b real constants, and let E *denote an expectation,* Var *a variance,* Cov *a covariance. Then:*

$$
\begin{aligned}
&1.\ \mathrm{E}a = a\\
&2.\ \mathrm{E}(aX+b) = a\mathrm{E}X + b\\
&3.\ \mathrm{E}(X+Y) = \mathrm{E}X + \mathrm{E}Y\\
&4.\ \mathrm{E}\textstyle\sum_{i=1}^{N} a_i X_i = \sum_{i=1}^{N} a_i \mathrm{E}X_i\\
&5.\ \mathrm{Var}X = \mathrm{E}((X-\mathrm{E}X)(X-\mathrm{E}X))\\
&6.\ \mathrm{Var}(aX+b) = a^2\mathrm{Var}X\\
&7.\ \mathrm{Cov}(X,Y) = \mathrm{E}((X-\mathrm{E}X)(Y-\mathrm{E}Y))\\
&8.\ \mathrm{Var}(X\pm Y) = \mathrm{Var}X + \mathrm{Var}Y \pm 2\mathrm{Cov}(X,Y)\\
&9.\ \mathrm{Cov}(a_1X_1+b_1, a_2X_2+b_2) = a_1a_2\mathrm{Cov}(X_1,X_2)\\
&10.\ \mathrm{Var}(\textstyle\sum_{i=1}^{N} a_iX_i) = \sum_{i=1}^{N}\sum_{j=1}^{N} a_ia_j\mathrm{Cov}(X_iX_j)\\
&\qquad\qquad = \textstyle\sum_{i=1}^{N} a_i^2\mathrm{Var}X_i + 2\sum_{i<j} a_ia_j\mathrm{Cov}(X_iX_j)
\end{aligned}
\tag{1.15}
$$

If, in addition, the random variables are independent, then:

$$
\begin{aligned}
&1.\ \mathrm{E}(XY) = \mathrm{E}X\mathrm{E}Y\\
&2.\ \mathrm{E}\textstyle\prod_{i=1}^{N} X_i = \prod_{i=1}^{N} \mathrm{E}X_i\\
&3.\ \mathrm{Var}(X\pm Y) = \mathrm{Var}X + \mathrm{Var}Y\\
&4.\ \mathrm{Var}\textstyle\sum_{i=1}^{N} X_i = \sum_{i=1}^{N} \mathrm{Var}X_i
\end{aligned}
\tag{1.16}
$$

Some basic formulas, e.g., (1.15,1,2,3) or (1.16,1,2), follow in an introduction to Statistics directly from the sums or integrals in the definitions of the moments, since these operations are linear (Problem 1.2). The basic formulas are used to derive the remaining ones in (1.15) or (1.16), e.g., (1.15,4) and (1.15,7) are derived in Problem 1.3.

1.5.2 Chebyshev Inequality

Let X be a random variable with distribution $F(x)$, expectation μ_X, and variance σ_X^2, and let c and $\varepsilon > 0$ both be constant. Then calculate the probability $\Pr\big(|X - c| \geq \varepsilon\big) = \int_{|x-c|\geq\varepsilon} \mathrm{d}F(x)$ by evaluating the Stieltjes-integral (which is defined in an introduction to Statistics or in (7.19)) from $-\infty$ to $c - \varepsilon$ and from $c + \varepsilon$ to ∞. Since $|x - c| \geq \varepsilon$, $(1/\varepsilon^2)(x - c)^2 \geq 1$, and consequently

$$\int\limits_{|x-c|\geq\varepsilon} \mathrm{d}F(x) \leq \int\limits_{|x-c|\geq\varepsilon} \frac{(x-c)^2}{\varepsilon^2}\mathrm{d}F(x) \leq \frac{1}{\varepsilon^2} \int\limits_{|x-c|\geq\varepsilon} (x-c)^2\mathrm{d}F(x)$$

$$\Pr\big(|X - c| \geq \varepsilon\big) \leq \frac{1}{\varepsilon^2}\mathrm{E}\big((x-c)^2\big) = \frac{1}{\varepsilon^2}\sigma_X^2 \tag{1.17}$$

the Chebyshev inequality is obtained in (1.17). The Chebyshev inequality sets, in terms of variance or standard deviation, an upper bound on the probability of extreme deviations from a typical observation of a random variable. It thus establishes a connection between the variance and the concept of concentration, i.e., the degree to what extent the central tendency (given, e.g., by the expectation or the median) represents the distribution of a random variable. The concepts of central tendency and concentration are discussed in an introduction to Statistics.

1.5.3 Independent and Normally Distributed Random Variables

Let the random variables X_i, $i = 1, \ldots, N$, with the densities $f_i(x_i) = \big(1/(\sigma_i\sqrt{2\pi})\big)\mathrm{e}^a$, $a = (-1/2)((x_i - \mu_i)/\sigma_i)^2$ where $\mathrm{E}X_i = \mu_i$ and $\mathrm{Var}X_i = \sigma_i^2$, be independent from each other. Then calculate $Y = \sum_{i=1}^N X_i$, the sum of the X_i. You then obtain the moments $\mathrm{E}Y = \mu_Y = \sum_{i=1}^N \mu_i$ and $\mathrm{Var}Y = \sigma_Y^2 = \sum_{i=1}^N \sigma_i^2$ using the formulas (1.15,4) and (1.15,4). However, the calculation of the distribution of Y is less straightforward.

In the first step, the distribution of $Y = X_1 + X_2$ is calculated. Since the X_i are independent, (1.5) is used to obtain the density $f(y)$ of Y in (1.18) as the product of the densities $f_1(x_1)$ and $f_2(x_2)$. (1.20) follows by differentiating $F(y)$ in (1.19).

$$F(y) = \int\limits_{x_1+x_2\leq y} \int f_1(x_1)f_2(x_2)\mathrm{d}x_1\mathrm{d}x_2 = \int\limits_{x_2=-\infty}^{\infty} \int\limits_{x_1=-\infty}^{y-x_2} f_1(x_1)f_2(x_2)\mathrm{d}x_1\mathrm{d}x_2 \tag{1.18}$$

$$= \int_{-\infty}^{\infty} \Big(F_1(y-x_2) - 0\Big) f_2(x_2)\mathrm{d}x_2 = \int_{-\infty}^{\infty} F_1(y-x_2) f_2(x_2)\mathrm{d}x_2 \tag{1.19}$$

$$f(y) = \int_{-\infty}^{\infty} f_1(y-x_2) f_2(x_2)\mathrm{d}x_2 \tag{1.20}$$

$$f(y) = \frac{1}{2\pi\sigma_1\sigma_2} \int_{x_2=-\infty}^{\infty} \exp\!\left(-\frac{1}{2}\Big(\frac{(y-x_2-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\Big)\right)\mathrm{d}x_2 \tag{1.21}$$

After substituting normal densities for $f_1(x_1)$ and $f_2(x_2)$, the density of the sum of two independent and normally distributed random variables results in (1.21). If this integral can be transformed into $f(y) = \big(1/(\sigma_Y\sqrt{2\pi})\big)\mathrm{e}^a$, $a = (-1/2)((y-\mu_Y)/\sigma_Y)^2$, $\mu_Y = \mu_1 + \mu_2$, $\sigma_Y^2 = \sigma_1^2 + \sigma_2^2$, it implies that $Y = X_1 + X_2$ is normally distributed with expectation $\mu_Y = \mu_1 + \mu_2$ and variance $\sigma_Y^2 = \sigma_1^2 + \sigma_2^2$. Writing

$$u = \frac{\sigma_Y}{\sigma_1\sigma_2}\left(x_2 - \frac{\sigma_1^2\mu_2 + \sigma_2^2(y-\mu_1)}{\sigma_Y^2}\right) \qquad \text{and} \qquad v = \frac{y-\mu_Y}{\sigma_Y}$$

$$u^2 + v^2 = \frac{(y-x_2-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} \tag{1.22}$$

we obtain (1.22) with the substitutions $\mu_Y = \mu_1 + \mu_2$ and $\sigma_Y^2 = \sigma_1^2 + \sigma_2^2$. Using this result and (1.21), (1.23) is obtained, since (i) x_2 is not in the expression for v and (ii) $\mathrm{d}u/\mathrm{d}x_2 = \sigma_Y/(\sigma_1\sigma_2)$, and hence $\mathrm{d}x_2 = (\sigma_1\sigma_2)/\sigma_Y$.

$$f(y) = \frac{1}{2\pi\sigma_1\sigma_2}\mathrm{e}^{-(v^2/2)} \int_{x_2=-\infty}^{\infty} \mathrm{e}^{-(u^2/2)}\mathrm{d}x_2 \tag{1.23}$$

$$= \frac{1}{2\pi\sigma_Y}\mathrm{e}^{-(v^2/2)} \int_{u=-\infty}^{\infty} \mathrm{e}^{-(u^2/2)}\mathrm{d}u$$

$$= \frac{1}{\sqrt{2\pi}\sigma_Y}\exp\!\left(-\frac{1}{2}\big((y-\mu_Y)/\sigma_Y\big)^2\right) \tag{1.24}$$

In (1.24), the density $f(y)$ of $Y = X_1 + X_2$ follows from the result of Problem 1.12 (the error integral). Thus, the distribution of the sum of N independent and normally distributed variables has been calculated for $N = 2$ in the first step.

In the second step, we have $N = 3$ and, with a small change in notation, $Y_3 = X_1 + X_2 + X_3$ and thus $Y_3 = Y_2 + X_3$, if $Y_2 = X_1 + X_2$. Y_2 and X_3 are independent and both are normally distributed by the assumptions and the calculations above. The distribution of Y_3 can therefore be calculated the same way as the distribution of Y is calculated in (1.18) to (1.24). This

calculation delivers the result that Y_3 is normally distributed with expectation $\mu_{Y_3} = \mu_1 + \mu_2 + \mu_3$ and variance $\sigma^2_{Y_3} = \sigma^2_1 + \sigma^2_2 + \sigma^2_3$.

In the $(N-1)^{th}$ step, the distribution of $Y_N = \sum_{i=1}^{N} X_i$ is calculated on condition that, in the previous $(N-2)^{th}$ step, the distribution of $Y_{N-1} = \sum_{i=1}^{N-1} X_i$ has been calculated.

1.5.4 Bivariate (two-dimensional) Normal Density

The marginal densities of a two-dimensional normal distribution are normal. Simulated examples are given in Table 1.4 and Fig. 1.3, and, for the standardised case, this result is derived in (1.25) to (1.30). Thus, with $\mu_1 = \mu_2 = 0$ and $\sigma_1 = \sigma_2 = 1$ and the definition in (1.12), the standardised two-dimensional normal density of the random variables X_1 and X_2 in (1.25) is obtained. The two-dimensional normal density is integrated to obtain the marginal distribution of X_1 at $x_1 = a_1$ in (1.26).

$$f(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\left(-\frac{1}{2(1-\varrho^2)}(x_1^2 - 2\varrho x_1 x_2 + x_2^2)\right) \tag{1.25}$$

$$F_1(a_1) = \int_{x_1=-\infty}^{a_1} \int_{x_2=-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\varrho^2}} \exp\big(b(x_1, x_2)\big) \mathrm{d}x_2 \mathrm{d}x_1 \quad \text{with} \tag{1.26}$$

$$b(x_1, x_2) = \left(-\frac{1}{2(1-\varrho^2)}(x_1^2 - 2\varrho x_1 x_2 + x_2^2)\right)$$

Using the construction of the two-dimensional normal distribution in (1.13) as a prototype, $x_1 = u_1$ and $x_2 = u_2\sqrt{1-\varrho^2} + \varrho u_1$ are substituted in (1.25) and (1.26) to obtain $\mathrm{d}x_1 = \mathrm{d}u_1$ and $\mathrm{d}x_2 = \sqrt{1-\varrho^2}\mathrm{d}u_2$ and thus, $\mathrm{d}x_1\mathrm{d}x_2 = \mathrm{d}u_1\mathrm{d}u_2\sqrt{1-\varrho^2}$, $|\varrho| < 1$. With these substitutions, the exponent of the e-function in (1.25) and (1.26) transforms into $u_1^2/2 + u_2^2/2$ and $\varphi(u_1, u_2)$ as indicated in (1.27).

$$\varphi(u_1, u_2) = \frac{1}{2\pi\sqrt{1-\varrho^2}} \mathrm{e}^{-(u_1^2/2 + u_2^2/2)} \mathrm{d}u_1 \mathrm{d}u_2 \sqrt{1-\varrho^2} \tag{1.27}$$

$$= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(u_1^2/2)} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-(u_2^2/2)} \mathrm{d}u_1 \mathrm{d}u_2 \tag{1.28}$$

$\varphi(u_1, u_2)$ is the joint density of the random variables U_1 and U_2, $U_1 = X_1$ and $U_2 = (X_2 - \varrho x_1)/\sqrt{1-\varrho^2}$, as defined above. U_1 and U_2 are independent since their density is the product of the densities obtained in (1.28). These are standardised one-dimensional normal densities. Assuming $\varrho = 0$, you have $u_2 = x_2$, and thus, X_1 and X_2 are independent if the correlation $\varrho = 0$.

In the next step, the two-dimensional density is integrated to obtain the marginal distribution of X_1 at a_1. The inner integral in (1.29) is, however, not a proper integral.

$$F_1(a_1) = \int_{u_1=-\infty}^{a_1} \left(\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u_1^2}{2}} \right) \left(\int_{u_2=\lim_{x_2\to-\infty}\left(\frac{x_2-\varrho u_1}{\sqrt{1-\varrho^2}}\right)}^{\lim_{x_2\to\infty}\left(\frac{x_2-\varrho u_1}{\sqrt{1-\varrho^2}}\right)} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u_2^2}{2}} \mathrm{d}u_2 \right) \mathrm{d}u_1 \tag{1.29}$$

Nevertheless, since $(1/\sqrt{2\pi})\mathrm{e}^{-u_2^2/2} \geq 0$, it can be evaluated as if it were a proper one (Problem 1.13), and the marginal distribution $F_1(a_1)$ of X_1 at a_1 is obtained in (1.30). With a similar procedure, the marginal distribution $F_2(a_2)$ of X_2 at a_2 is calculated.

$$F_1(a_1) = \int_{x_1=-\infty}^{a_1} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x_1^2}{2}} \mathrm{d}x_1 \qquad F_2(a_2) = \int_{x_2=-\infty}^{a_2} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x_2^2}{2}} \mathrm{d}x_2 \tag{1.30}$$

Due to (1.30), the marginal frequencies of the simulations in Fig. 1.3 are approximately normal and do not depend on ϱ.

The two-dimensional density in (1.12) can also be written using vectors and matrices. In (1.31), $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$ is the vector of the expectations of the random variables X_1 and X_2 and $\boldsymbol{C}$ is their covariance matrix (the matrix with the variances c_{11} and c_{22} and the covariances $c_{12} = c_{21}$ of X_1 and X_2). In (1.32) the determinant and the inverse of the covariance matrix is given.

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \boldsymbol{C} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \tag{1.31}$$

$$\det|\boldsymbol{C}| = \sigma_1^2\sigma_2^2(1-\varrho^2) \qquad \boldsymbol{C}^{-1} = \frac{1}{1-\varrho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\varrho}{\sigma_1\sigma_2} \\ \frac{-\varrho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} \tag{1.32}$$

$$f(\boldsymbol{x}) = \frac{1}{2\pi\sqrt{\det|\boldsymbol{C}|}} \exp\left(-\frac{1}{2} \Big((\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \Big) \right) \tag{1.33}$$

The density in (1.33) is the two-dimensional normal density defined in (1.12). (1.33) follows immediately if the expressions in (1.31) and (1.32) are substituted in (1.12).

1.5.5 Multivariate Normal Distribution

Vectors and matrices lend themselves to defining the joint normal density of n random variables.

$$\begin{array}{rcl} \boldsymbol{x} & = & (x_1, \quad x_2, \quad \ldots, \quad x_n)^T \\ \boldsymbol{\mu} & = & (\mu_1, \quad \mu_2, \quad \ldots, \quad \mu_n)^T \end{array} \qquad \boldsymbol{C} = \begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1n} \\ c_{21} & c_{22} & \ldots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \ldots & c_{nn} \end{pmatrix}$$

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{(\det|\boldsymbol{C}|)}} \exp\left(-\frac{1}{2} \Big((\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}^{-1} (\boldsymbol{x}-\boldsymbol{\mu}) \Big) \right) \tag{1.34}$$

n random variables with a joint density, as given in (1.34), are said to be *multivariate normally distributed*. $\boldsymbol{\mu}$ is the column vector with the expectations of n random variables $(X_1, \ldots, X_n)^T = \boldsymbol{X}$, $\boldsymbol{C}$ being their invertible covariance matrix and $f(\boldsymbol{x})$ their n-dimensional density. Often, the notation $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{C})$ is used.

Multivariate normally distributed random variables are practical since linear combinations $\sum_{j=1}^{n(j)} a_j X_j$, X_j in $\boldsymbol{X}$, $n(j) \leq n$, are normally distributed as well. For a proof use the following proposition. If $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{C})$ and $\boldsymbol{A}$ is an invertible $n \times n$ matrix, then $\boldsymbol{AX} \sim N_n(\boldsymbol{A\mu}, \boldsymbol{ACA}^T)$. When a matrix with $(a_1, \ldots, a_n)$ is substituted for $\boldsymbol{A}$, the marginal distributions are obtained. For the following two special cases of linear combinations of multivariate normally distributed random variables, the densities have already been calculated.

The first case is the marginal distribution of any random variable in a set of multivariate normally distributed random variables. The marginal distribution of each random variable X_i in $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \boldsymbol{C})$ is normally distributed with expectation μ_i and the variance σ_i^2. For $n = 2$, this has been proved in (1.25) to (1.30).

In the second case, the random variables in $\boldsymbol{X}$ are assumed to be mutually independent. Under this assumption, all covariances in $\boldsymbol{C}$ become identically zero, and $\boldsymbol{C}$ contains the non-zero diagonal elements, which are the variances σ_i^2 of the X_i. The determinant and the inverse of such a covariance matrix are then easily calculated and the density in (1.35) follows.

$$\det|\boldsymbol{C}| = \sigma_1^2 \sigma_2^2 \ldots \sigma_n^2$$

$$\left(x_1 - \mu_1, \ldots, x_n - \mu_n\right) \begin{pmatrix} \sigma_{11}^{-2} & 0 & \ldots & 0 \\ 0 & \sigma_{22}^{-2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & \ldots & \sigma_{nn}^{-2} \end{pmatrix} \begin{pmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \\ \vdots \\ x_n - \mu_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} \left(\frac{(x_i - \mu_i)^2}{\sigma_i^2} \right)$$

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^{n/2} \sigma_1 \sigma_2 \ldots \sigma_n} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \left(\frac{(x_i - \mu_i)^2}{\sigma_i^2} \right) \right) \tag{1.35}$$

For $n = 2$, this result has been obtained in (1.18) to (1.24).

1.6 Problems

1.1. The data medium accompanying this book contains sub-directories `Data` and `R` in directory `UVTS`. A variety of example time series used in this book are found, together with their descriptions, in the `Data` sub-directory. Copy these files to your hard disk: `/path/` is used in this book as a placeholder for the directory where your copy of a file is located.

1.2. Are (1.15,1), (1.15,2) and (1.15,3) valid?

1.3. Are (1.15,4) and (1.15,7) valid?

1.4. Show that, under the iid.-assumptions in (1.2), the empirical variance $\hat{\sigma}_X^2 = (1/(N-1)) \sum_{i=1}^{N} (x_i - \hat{\mu}_X)^2$ is an unbiased estimate for the theoretical variance. Hint: $\mathrm{Var}\hat{\mu}_X = (1/N^2) \sum_{i=1}^{N} \mathrm{Var}X_i = (1/N^2) N \mathrm{Var}X_i = (1/N)\sigma_X^2$.

1.5. Table 1.3 contains the estimated moments of the daily pyranometer values calculated from the measurements made by the MeteoSwiss and the ETH at Zurich-Reckenholz station. a) Why does the difference of the MeteoSwiss and ETH means ($128.2 - 124.7 = 3.5$ Wm^{-2}) not equal the mean of the differences (3.4 Wm^{-2})? b) Are the MeteoSwiss and the ETH mean significantly different? Perform a statistical test using the empirical moments in Table 1.3. Then, perform this test using an appropriate R function. Hint: `help.start()` starts the R help system.

1.6. Let `dif` be the R time series of the pyranometer differences plotted above, in Fig. 1.8. From `dif`, construct R time series for the time slices: (a) the Summer months of 1990 and 1992, and (b) the Winter months of 1989/90, 1991/92 and 1992/93. Then, assuming that the observations are iid., calculate estimates for the mean and the empirical variance of the differences in each time slice.

1.7. Let X_1 and X_2 be the random variables, which are defined in (1.13) as linear combinations of the random variables X_1' and X_2'. Calculate the expectations and the variances of X_1 and X_2. Show that ϱ is the correlation of X_1 and X_2 as defined in (1.11).

1.8. Outline the region of integration for the double integral in (1.18).

1.9. In the error model (1.14), the expectation of the differences of the pyranometer daily values is the systematic error, and half of the variance of the differences is the random error.

1. Find an error model for the pyranometer daily values under the assumption that the measurements are only afflicted with a random error. Hint: assume the systematic error in (1.14) to be identically zero.
2. Assume that the differences of the pyranometer daily values are iid. and estimate the random error in the daily pyranometer values with your model. Compare your estimate with the estimate obtained, when (1.14) is being used.

1.10. The daily values for the meteorological quantities observed at Zurich-Reckenholz station are not independent, since weather conditions are likely to last for a period of a few days, as discussed in Sect. 1.4. Therefore, SWIR values for those days, which immediately follow each other, are likely to be correlated. However, estimates are unable to be calculated with (1.7), since there is only one value available for each day and the measurements cannot be repeated. Is there another possibility?

1.11. The monthly SWIR values for the stations shown in the following table are available in the files `/path/name`.

station: name, coordinates altitude in m				period of measurements from	to	filename
Toronto	79 24 W	43 40 N	116	1950	1988	`toronto.dat`
Toronto Met.Res.	79 33 W	43 48 N	149	1967	1988	`tormetres.dat`
Bracknell	0 47 W	51 23 N	73	1965	1991	`bracknell.dat`
London Wea.C.	0 7 W	51 31 N	77	1958	1991	`london.dat`

These text files contain, in each line, 12 monthly values for a year. Calculate the monthly differences for the pair (Toronto/Toronto Met.Res.) stations, and the pair (Bracknell/London Wea.C.) stations over the periods where monthly means in both time series are available. Are these differences iid. as defined in (1.2)?

1.12. Evaluate $\int_{-\infty}^{\infty} \mathrm{e}^{-bx^2}\mathrm{d}x = \sqrt{\pi/b}$. Hint: using polar coordinates and assuming that $\mathrm{d}y\mathrm{d}x = \varrho\mathrm{d}\phi \times \mathrm{d}\varrho$ is a small rectangle, provided that $\mathrm{d}\phi$ and $\mathrm{d}\varrho$ are small, evaluate $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-b(x^2+y^2)}\mathrm{d}y\mathrm{d}x = \pi/b$. The result then follows with $\mathrm{e}^{-b(x^2+y^2)} = \mathrm{e}^{-bx^2}\mathrm{e}^{-by^2}$.

1.13. Equation (1.29) is repeated below. The inner (improper) integral in this equation has to be evaluated for both finite and infinite u_1. The result is 1, if u_1 is finite, or h, $0 \le h \le 1$, if $u_1 \to -\infty$. Therefore, in the interval $-\infty \le u_1 \le b_1$, it is assumed that the inner integral evaluates to $\bar{h}(b_1)$ and, in the second, third and fourth lines, an inequality results.

$$F_1(a_1) = \int_{u_1=-\infty}^{a_1} \left(\frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u_1^2}{2}} \right) \left(\int_{u_2=\lim_{x_2\to-\infty}\left(\frac{x_2-\varrho u_1}{\sqrt{1-\varrho^2}}\right)}^{\lim_{x_2\to\infty}\left(\frac{x_2-\varrho u_1}{\sqrt{1-\varrho^2}}\right)} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u_2^2}{2}} \mathrm{d}u_2 \right) \mathrm{d}u_1$$

$$\int_{u_1=-b_1}^{a_1} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u_1^2}{2}} \mathrm{d}u_1 \le$$

$$\int_{u_1=-\infty}^{-b_1} \bar{h}(b_1) \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u_1^2}{2}} \mathrm{d}u_1 + \int_{u_1=-b_1}^{a_1} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u_1^2}{2}} \mathrm{d}u_1 = F_1(a_1)$$

$$\le \int_{u_1=-\infty}^{a_1} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{u_1^2}{2}} \mathrm{d}u_1$$

In this inequality, for $b_1 \to \infty$ the integral on the left side (in the second line) tends to the integral on the right side (in the fourth line). Is this derivation correct?

2 Stationary Stochastic Processes

Observing a geophysical variable, a time series is obtained if time is recorded together with the values being measured. As a value is a realisation of a random variable, a time series is a time slice from a realisation of a double infinite sequence of random variables, i.e., of a stochastic process. Introductions into stationary stochastic processes are given in [17], [33], and [22].

The basic definitions are given in Sects. 2.1 and 2.2. In Sect. 2.3, models for three example time series are found by trial and error. These models are examples of stochastic processes. They are linear combinations of a sequence of independent and identically distributed random variables and are treated as convolutions in Sect. 2.4. The practitioner, who has to decide whether a time series is stationary, finds some helpful hints in Sect. 2.5. In Sect. 2.6, the statistical properties of the empirical moments of a time series are derived and discussed. Sect. 2.7 introduces the optimal linear predictor. As in Chap. 1, the supplements and the problem set are found in the two last sections.

2.1 Empirical Moments of a Wind Speed Time Series

On March 11, 1987, from 22:30 until 23:00, at a height of 23.3 m (meters) above Anwandstrasse (a street in Zurich), the horizontal speed of the turbulent air flow was measured with an ultrasonic anemometer in time steps of $\Delta t = 1$ s (second) [119]. (Some remarks on the observation of a turbulent flow in the atmospheric boundary layer are found in Sect. 9.4.2.) The data obtained are made available in the text file `/path/wind.dat`, which contains 180 lines with 10 values in each line. The values are the deviations of the wind speed from a reference value and result from a Reynold's decomposition, as defined in (9.101) and (9.102), of the wind speed measurements. The unit is ms^{-1}. From this file, with the following R expressions

```
wfn <- "/path/wind.dat"
wind <- ts(scan(wfn),start=1,frequency=1,)
```

the time series `wind` is generated (R time series objects are introduced in Sect. 1.4). This time series is plotted in Fig. 2.1.

In Fig. 2.1, the wind speed tends to be close in value at adjacent time points: the time series oscillates with minor fluctuations about some local (definition (2.55) also applies to time series) mean, e.g., approximately

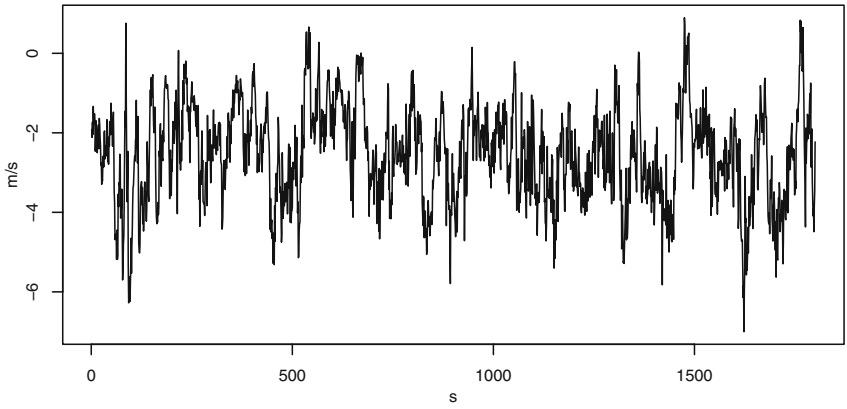


Fig. 2.1. Atmospheric turbulence. Horizontal wind speed (deviations from a reference value) in ms^{-1} (meters per second). Time in s (seconds).

$-2\ \mathrm{ms}^{-1}$ when the measurements begin at $t = 0$ s. However, at a few time points, the local mean of the time series changes abruptly, e.g., for the first time at approximately $t = 50$ s, values smaller than $-4\ \mathrm{ms}^{-1}$ are measured. After a few (usually less than approximately 60) seconds have passed, the local mean of the time series changes again. Thus, the wind speed in a turbulent flow at time t_1 is not independent from the wind speed at neighbouring time points $t_2, t_3, \ldots$. In the period with measurements, however, the wind speed appears to remain in a statistical balance, i.e., the local means fluctuate about a global (definition (2.55) also applies to time series) mean. This suggests that there was no change in the conditions that determined the turbulent flow within the observational period.

Under this assumption, the following questions are meaningful. What is the mean wind speed? What is the strength of a possible statistical linear relationship (the empirical correlation, as defined in 1.7) between wind speeds measured at adjacent time points t and $t + n\Delta t$, $n = 1, 2, 3, \ldots$, $\Delta t = 1$ s? Is there an m such that the correlation at t and $t + n\Delta t$, $n > m$, tends to zero?

In Fig. 2.1, the mean wind speed is approximately $-2.5\ \mathrm{ms}^{-1}$. This answers the first question: do graphical methods exist which would help to answer the other questions? Using the R function `lag()`

```
windb1 <- lag(wind,1)
windb2 <- lag(wind,2)
...
```

duplicates of the time series are generated, which are shifted backwards (in negative direction on the time axis) with displacements of $1, 2, 3, \ldots, \tau\ \Delta t$. The displacement τ is called *lag*. Then, from the original time series and its shifted copies with the lags $\tau = 1, \ldots, 9$, with

```
windmat <- ts.intersect(wind,windb1,windb2, ..., windb9)
```

R matrix `windmat` is constructed, taking into account the shifted time domains. The first column of `windmat` contains the original time series, the other columns being the shifted copies.

The values in an R matrix are indexed with the indices written in brackets. Remember that in Sect. 1.2, the index of an R vector is written in brackets. The `windmat` matrix for example has two indices: the first is in the sequence $1, 2, \ldots, 1791$, the time points in the original time series (-9 to account for the copy in the maximum lag 9), the second index is in the sequence $1, 2, \ldots, 10$, the number of the shifted copy (1 the original time series, 2 the copy in the lag 1, etc.). When you type `windmat[1,1]` you obtain the first value in the first column, with `windmat[1,]` the values in the first row, and with `windmat[,1]` the values in the first column are obtained. By means of this matrix, scatterplots of the time series and its lagged copies are easily produced. For example,

```
plot(windmat[,2],windmat[,1],type="p",xlab="Lag 1",ylab="Wind")
```

generates the scatterplot of the original wind time series, and its copy in the lag $\tau = 1$. Even more easily, scatterplots of a time series as well as its lagged copies are generated using `lag.plot()`. For example,

```
lag.plot(wind, lags=9, head="", layout=c(3,3))
```

produces the lag τ scatterplots, $\tau = 1, \ldots, 9$, in Fig. 2.2. Fig. 2.2 shows correlations not near zero between the original time series and its copies in the lags $\tau = 1, \ldots, 9$. These correlations decrease, as the lag increases. Hence, the properties (1.10) suggest a statistical linear relationship between wind speeds at small lags; at larger lags, however, the relationship disappears.

You can calculate numerical values, which support the conclusions drawn from the plots in Figs. 2.1 and 2.2, using the definitions (2.1).

A time series $(x_t) = (x_1, \ldots, x_N)$ *is a sequence of observed values, and* $\tau = 1, 2, \ldots, N-1$, *where* $\tau = t - u$, *is a lag, as mentioned in the remarks above. Then:*

1. $\hat{\mu}_X = (1/N) \sum_{t=1}^{N} x_t$ *is called the arithmetic mean of* (x_t),
2. $\hat{\sigma}_X^2 = (1/N) \sum_{t=1}^{N} (x_t - \hat{\mu}_X)^2 = \hat{c}_X(0)$ *the empirical variance,*
3. $\hat{c}_X(\tau) = (1/N) \sum_{t=1}^{N-\tau} (x_t - \hat{\mu}_X)(x_{t+\tau} - \hat{\mu}_X)$ *the empirical covariance function, and*
4. $\hat{\varrho}_X(\tau) = \hat{c}_X(\tau)/\hat{c}_X(0)$ *the empirical correlation function of* (x_t).

(2.1)

If the empirical moments of a time series are calculated using the definitions (2.1), realisations of the corresponding estimators are obtained. Are these estimators unbiased and consistent, as defined in (1.3) and (1.4)? The answer is given in Sect. 2.6 where the moments of these estimators are derived under the assumption that the time series is a time slice from a realisation of a stationary stochastic process, as defined in (2.8). The stationarity assumptions replace the iid. assumptions, since the values in a time series usually are not iid. and their order of succession is important. The iid. assumptions are formulated in (1.2), with estimates and estimators defined in (1.1).

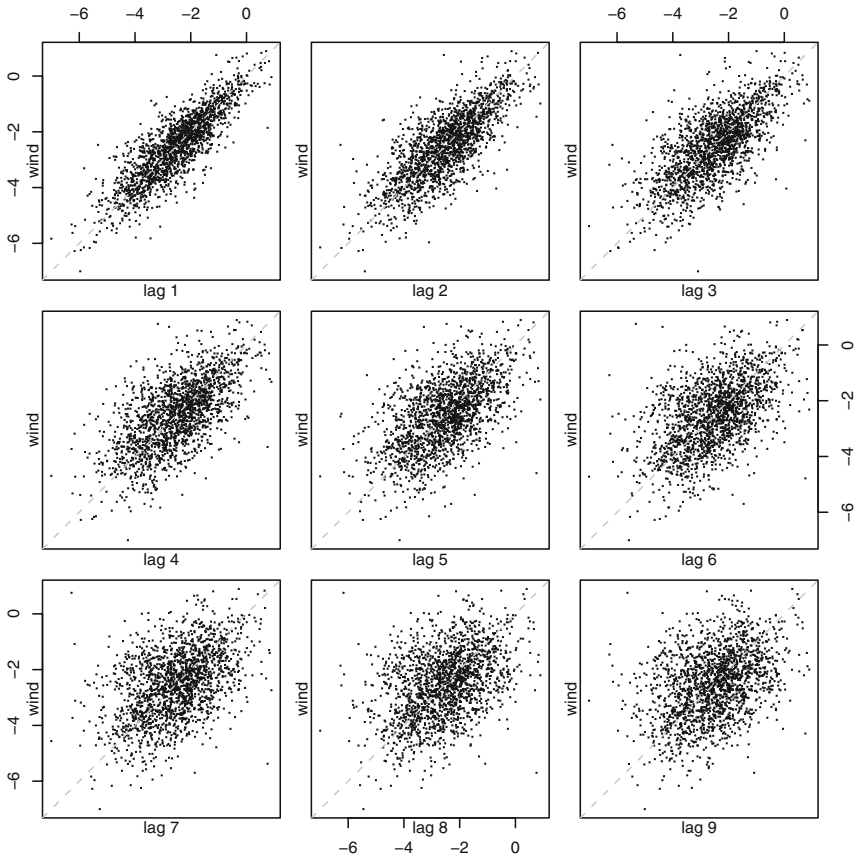


Fig. 2.2. Lag τ scatterplots (lagplots) of the wind speed time series in Fig. 2.1. The values of the original time series (in lag $\tau = 0$) are plotted on the vertical axes; on the horizontal axes, the values of the shifted time series (in lags $\tau = 1, 2, \ldots, 9$) are plotted.

The empirical second moment functions of a time series, as defined in (2.1), are not identical with the empirical variance (in the remarks to (1.2)), covariance and correlation (as defined in (1.7)) used for the calculations under the iid. assumptions with R functions `mean()`, `var()`, `cov()` and `cor()`. Unfortunately, these functions do not use the definitions in (2.1) if their arguments are R times series objects. Hence, the empirical second moments of a time series are calculated, as required in (2.1,2,3,4), using the R function `acf()`. For example, the expressions

```
lmax <- 60 #largest lag
lags <- 0:lmax    #a <- 0:5 generates the vector a
ecor <- 0:lmax    #with elements 0,1,2,3,4,5
```

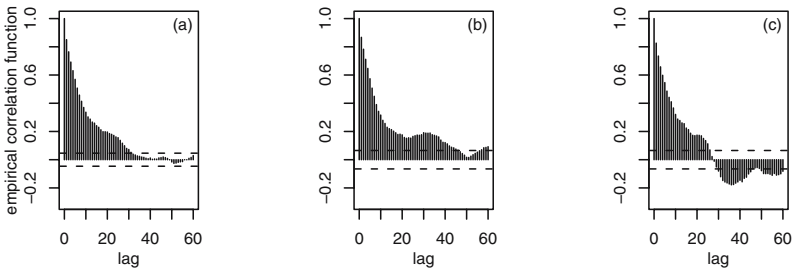


Fig. 2.3. Empirical correlation functions of the time series `wind` in Fig. 2.1, calculated from all 1800 values (a), from the first 900 values (b), and from the last 900 values (c).

```
#acf(): type="correlation" for the empirical correlations
#       type="covariance" for the empirical covariances
acfpl <- acf(wind,lag.max=lmax,type="correlation",plot=F)
for(i in 1:(lmax+1)) {ecor[i] <- acfpl$acf[(i),1,1]}
plot(c(0,lmax),c(-0.3,1),type="no",xlab="lag",
          ylab="empirical correlation function")
for(i in 1:lmax1) {lines(c(lags[i],lags[i]),c(0,ecor[i]),lty=1)}
ci <- 0.95 #confidence limit as in the remarks to (2.59)
cc <- (qnorm(0.5 + ci/2)/sqrt(acfpl$n.used))
lines(c(-2,lmax+2),c(cc,cc),lty=2)
lines(c(-2,lmax+2),c(-cc,-cc),lty=2)
```

calculate the empirical correlations of the wind speed time series using (2.1,4) and then plot the vertical lines, on the left in Fig. 2.3. The 0.95-confidence interval for no correlation, as defined in (2.59), is plotted with dashed horizontal lines. This confidence interval can also be calculated with `1.96/sqrt(acfpl$n.used)`. The R function `acf(..., plot=T)` produces the default plot of the empirical covariance or correlation function. In the default plot, the vertical axis is labelled ACF (the acronym for autocorrelation function, mentioned in the remarks to (2.6)) and the 0.95-confidence interval for no correlation, as defined in (2.59), is plotted.

The empirical correlation or covariance function is usually computed for the entire time series. Then the time series is split into two parts in the centre and the moment functions are re-calculated from the first and second half-series. For example, since there are $N = 1800$ values in the wind speed series, with R function `window()` (introduced in the remarks to Fig. 1.6) two half-series, each containing 900 values, are generated. The empirical correlations resulting from these are plotted in the middle and on the right in Fig. 2.3. In these plots, the empirical correlations are similar for lags $\tau \leq 20$. However, for $\tau > 20$, the similarity diminishes. It is recommended that you calculate and plot the empirical correlation or covariance function using different parts of a time series to be analysed, and compare the plots using the diagnostics

introduced in Sect. 2.5. This helps to decide whether the time series is stationary, as defined in (2.8). Stationarity is assumed if the empirical moments are calculated with definitions (2.1).

The empirical covariances $\hat{c}_X(\tau)$ and correlations $\hat{\varrho}_X(\tau)$ of a time series, as defined in (2.1), have properties similar to empirical covariances and correlations calculated from iid. observations, as defined in (1.7). Thus $\hat{\varrho}_X(\tau) = \hat{c}_X(\tau)/\hat{c}_X(0)$ describes the linear relationship (demonstrated in Fig. 2.2 for the wind speed example) of two measurements in lag τ. There are $N - \tau$ pairs of observations with this lag. $\hat{\mu}_X$, $\hat{\sigma}_X^2$, $\hat{c}_X(\tau)$ and $\hat{\varrho}_X(\tau)$ are not robust estimates, i.e., they are strongly influenced by extreme values. Examples are given in Fig. 3.9. Therefore, always produce a histogram of (x_t) before calculating $\hat{\mu}_X$ or $\hat{\sigma}_X^2$, and a lag τ scatterplot before calculating $\hat{c}_X(\tau)$ or $\hat{\varrho}_X(\tau)$.

The empirical covariance or correlation function can be calculated as prescribed in (2.1) provided that there are enough pairs $(x_t, x_{t-\tau})$ in the period where the original time series (x_t) and the lagged duplicate $(x_{t-\tau})$ overlap. Box and Jenkins in [16] recommend the calculation of the empirical moment functions only if a time series is not shorter than $N \approx 50$ (approximately) and the lag not larger than $\tau \approx N/4$.

Both shifting a duplicate time series by a lag τ, when the empirical covariance function is calculated, and averaging over the period of measurement when the empirical mean is calculated, only make sense on condition that the time series is stationary. The stationarity assumptions are formulated in Sect. 2.2.3. In Sect. 2.6, assuming that a time series is stationary, the expectations and variances of the estimates defined in (2.1) are calculated. However, the calculations are no longer as straightforward as when calculating the moments of the arithmetic mean of a random sample in the remarks to (1.2), because the observations in a time series are not independent.

2.2 Stationary Stochastic Processes

This section contains the basic definitions for the stochastic process and its moment functions in Sects. 2.2.1 and 2.2.2, and for the stationarity assumptions in Sect. 2.2.3

2.2.1 Stochastic Processes

Many geophysical variables depend on space and time. Thus, the measured values are often recorded together with their spatial-temporal coordinates, when a geophysical variable is observed. Examples of geophysical variables are: the meteorological variables observed worldwide at the synoptic stations, the salinity in the regions of the North Atlantic where the water freezes to sea ice, the hydraulic conductivity in an aquifer, and the percentage of humus in a soil.

The above examples demonstrate that (i) often two or more variables are observed at the same place and time and (ii) the variables are observed in a given region and/or period. If the observed variables are random and if this property, i.e., the influence of chance, is suspected of playing a role in the analysis of the measurements, it then is assumed that the observations stem from a stochastic process.

A stochastic process or a random function (X_t) *or* $X(t)$, $t \in T$, *is a sequence of random variables or a function whose values are random variables. The argument* t *of the sequence or function is called parameter.* (2.2)

The definition (2.2) allows for both multi-dimensional stochastic processes (random functions) and multi-dimensional parameters. A stochastic process (random function) is multi-dimensional, only if the random variables are multi-dimensional $\boldsymbol{X}^T(t) = \left(X_t^{(1)}, \ldots, X_t^{(p)}\right)$. Such sequences or functions are called *multivariate stochastic processes* or *multivariate random functions*.

Most geophysical stochastic processes or random functions are multivariate and their parameter is 4-dimensional, i.e., 3 in space and 1 in time. For example, an upper-air sounding usually delivers values for the pressure, temperature and humidity which are associated with longitude, latitude, altitude and time. Often only one variable is measured at a fixed location in constant time steps. Such a measurement produces a time series, as defined in (2.1).

In contrast to a random sample which delivers observations from a set of iid. random variables as stipulated in (1.2), the observations in a time series are generally not independent and their order is essential. Examples are the pyranometer measurements in Fig. 1.4 and the measurements of the windspeed in the turbulent air flow in Fig. 2.1.

In this book it is assumed that the observations of a geophysical variable are a time slice in a realisation of a stochastic process, as defined in (2.2), with integer parameter $t = \ldots, -1, 0, 1, \ldots$. Under this assumption, a value x_t has been measured for each random variable X_t, t in the observational period, and the sequence of the values (x_t) is a time series. For this *discrete-time* stochastic process, usually (X_t) or $X(t)$ is written, and for the parameter, t. t is the time, with a specified value of t being a point in time.

As an important exception, Chap. 4 treats random functions with a two-dimensional real parameter, e.g., longitude and latitude. An example is the climatological yearly means of SWIR in Germany in Fig. 4.16. These random functions are often measured at non-equidistant sites (sites which are not on a grid with uniform size of the cells) and, therefore, in (4.6), a spatial random function (regionalised variable in geostatistic terminology, cf. Chap. 4) is defined as not being a "spatial series". As another exception, *continuous-time* stochastic processes (with parameter being the time t, t real) are introduced in Sects. 7.1 and 7.2.

In the example of a stochastic process given in Fig 2.4, 6 out of all possible realisations of a stochastic process are plotted over the period $t = 1, \ldots, 400$.

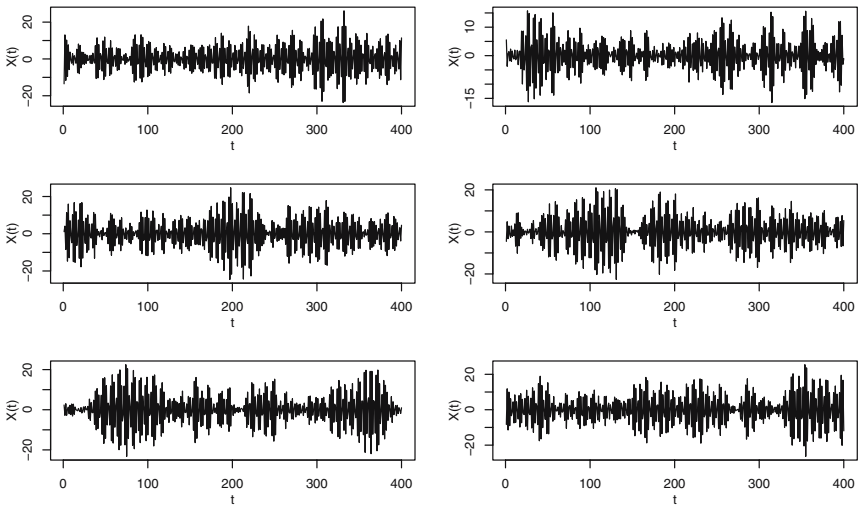


Fig. 2.4. 6 simulated realisations (nos. $100, 200, \ldots, 600$) of a stochastic process $X(t)$, plotted in the time slice $t = 1, 2, \ldots, 400$.

If all realisations had been plotted, i.e., all possible observations of the random variables $X(t)$, $t = 1, \ldots, 400$, the process would have been entirely specified for this period. However, this is not possible from a practical point of view, since there is an infinite number of realisations. Hence, a stochastic process is given by a formula or a procedure with which an arbitrary number of realisations can be simulated.

With such a procedure, 600 realisations $x(t, j)$, $t = 1, \ldots, 400$, $j = 1, \ldots, 600$, are simulated. Of these, 6 are plotted in Fig. 2.4. These simulations are used to ascertain whether the random variables $X(t)$, $t = 1, \ldots, 400$, are iid. as requested in (1.2), since any realisation is simulated independently of any other and, thus, for each random variable (X_t), $t = 1, \ldots, 400$, 600 simulated values $x(t, j)$, $j = 1, \ldots, 600$, are obtained, all of which are iid.

If, as prescribed in (1.2,1), the $X(t)$ are identically distributed, then the histograms of $x(t, j)$, $j = 1, \ldots, 600$, are similar for all t. This is true for $X(t)$, $t = 51, 52, 53, 54, 55, 301, 302, 303, 305$, as seen in Fig. 2.5, and for any other $X(t)$, $t = 1, \ldots, 400$ (not plotted). These histograms suggest that all random variables in the process have the same normal distribution with mean $\mu_X \approx 0$ and standard deviation $\sigma_X \approx 10$.

If further, as prescribed in (1.2,2), the $X(t)$ are independent of each other, then the scatterplot for any pair $(X(t), X(u))$, $t, u = 1, \ldots, 400$, $t \neq u$, is similar to the scatterplot in Fig. 1.3 at the bottom. However, assuming that pairs $(X(t), X(u))$ are bivariate normally distributed, as defined in (1.13), it is concluded from Figs. 2.6 and 2.7 that there are pairs of random variables $(X(t), X(u))$ which are not independent, since there are scatterplots similar

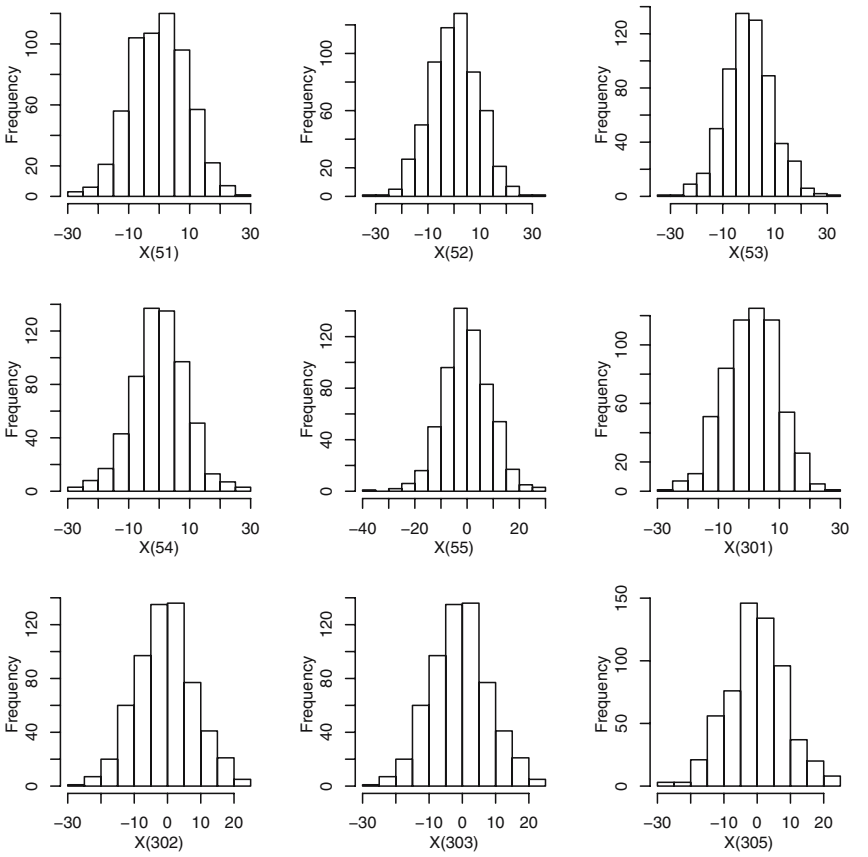


Fig. 2.5. Histograms of 600 realisations (realisations nos. $100, 200, \ldots, 600$ are plotted in Fig. 2.4) of a stochastic process (X_t), for the 9 time points $t = 51, 52, 53, 54, 55, 301, 302, 303, 305$.

to those on top and in the middle of Fig. 1.3. This shows that the $X(t)$, $t = 1, \ldots, 400$, are not iid.

In Figs. 2.6 and 2.7, there are very similar scatterplots for pairs of random variables with the same lag, e.g., only small differences can be seen when comparing the scatterplots for $(X(51), X(52))$ and for $(X(301), X(302))$. This suggests that the scatterplots for $(X(t), X(u))$, $u = t - \tau$, depend on the lag $\tau = t - u$, but not on t and u. Quite large correlations are seen in the scatterplots for small lags, with the exception of the correlation for lag $\tau = 4$, which is close to zero. In Fig. 2.7 there are also scatterplots for pairs $(X(t), X(u))$, $(t, u) = (51, 301), (51, 302), (51, 310)$, which are very similar to the scatterplot in Fig. 1.3 at the bottom. From these, assuming that pairs

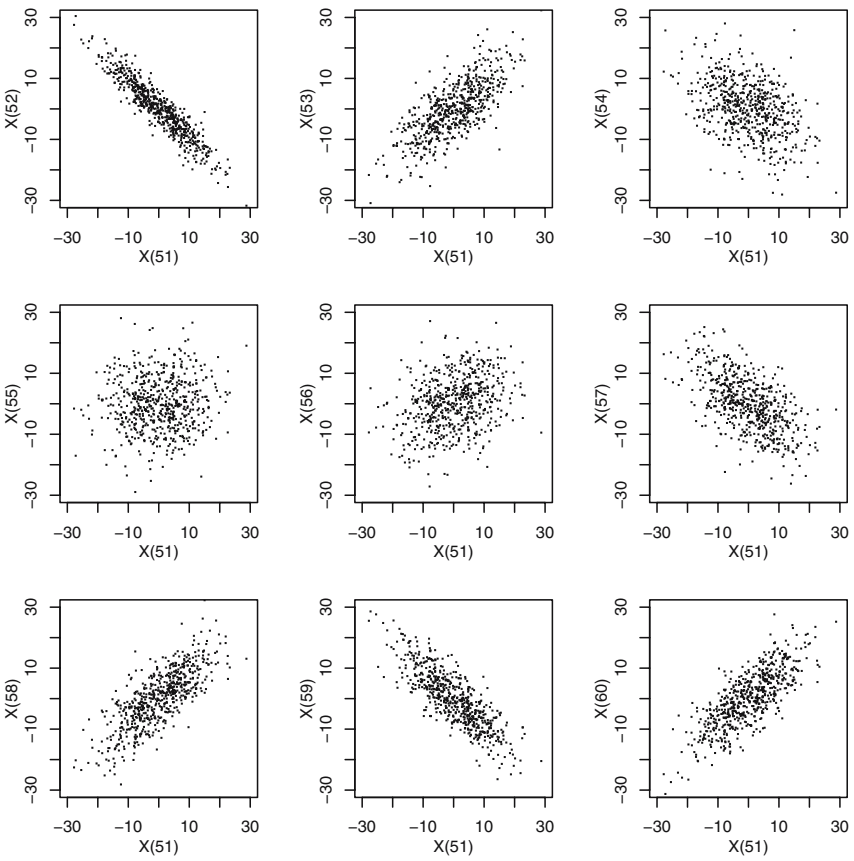


Fig. 2.6. Scatterplots $\big(X(t,j), X(u,j)\big)$, $j = 1,600$, of 600 realisations (realisations nos. $100, 200, \ldots, 600$ are plotted in Fig. 2.4) for the pairs $(t,u) = (51,52), (51,53), \ldots, (51,60)$.

$(X(t), X(u))$ are bivariate normally distributed, as defined in (1.13), it is concluded that $(X(t), X(u))$ are approximately independent if the lag is large.

Since the simulations $x(t,j)$, $j = 1, \ldots, 600$, are iid., for any $X(t)$, the mean $\mu_X(t)$ and the variance $\sigma_X^2(t)$ of $X(t)$ are estimated with $\hat{\mu}_X(t) = (1/600)\sum_{j=1}^{600} x(t,j)$ and with $\hat{\sigma}_X^2(t) = 1/(600-1)\sum_{j=1}^{600}(x(t,j) - \hat{\mu}_X(t))^2$; for any pair $(X(t), X(u))$, the covariance is estimated using the definition (1.7). These estimates are realisations of estimators, which are known to be unbiased, as defined in (1.3), and consistent, as defined in (1.4), from an introduction to Statistics.

The simulations for the example in Figs. 2.4, 2.5, 2.6 and 2.7 were done using R function `arima.sim()` and the AR[2] model given in Problem 2.22. The AR[2] model, as defined in (2.48), is an example of a *linear process*. The

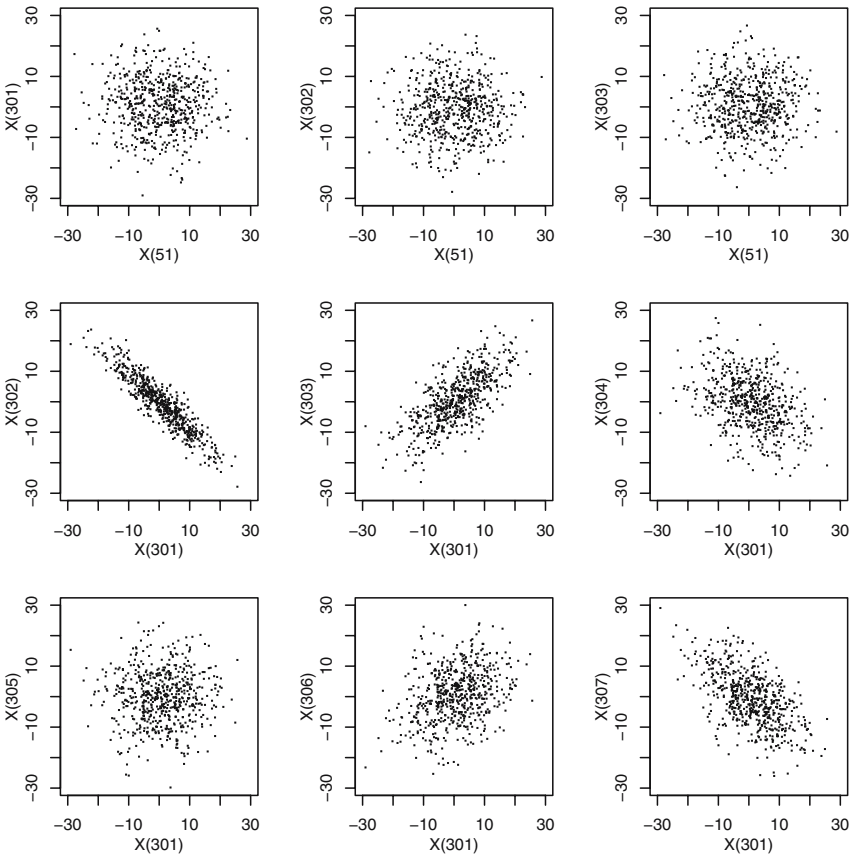


Fig. 2.7. Scatterplots $\big(X(t,j), X(u,j)\big)$, $j = 1,600$, of 600 realisations (realisations nos. $100, 200, \ldots, 600$ are plotted in Fig. 2.4) for the pairs $(t,u) = (51,301), (51,302), (51,310), (301,302), \ldots, (301,307)$.

linear process is introduced in Sect. 2.3.5 and discussed in detail in Chap. 5. The properties of a linear process and the remarks to definition (1.34) are used to derive, in Problem 2.22, that the AR[2] model used for the simulations is normal or Gaussian, as defined in (2.3).

$X(t)$ is a normal or Gaussian process, provided that all one- and multidimensional distributions of finitely many (out of possibly infinitely many) random variables $X(t), X(u), X(v), \ldots$, $t,u,v \in T$, are univariate, resp. multivariate normal distributions as defined in (1.34). (2.3)

In this definition, the joint probability distributions of all random variables in the process is given. If the number n of the random variables is finite, this

procedure is known from an introduction to Statistics: the joint n-dimensional distribution or density function is given.

The simulations for the example in Figs. 2.4, 2.5, 2.6 and 2.7 with a formula or procedure and the definition of the normal process in (2.3) demonstrate that there are two possibilities to specify a stochastic process. First, as many realisations as you wish are simulated using a formula (an AR[2] model in the example above). Second, the joint probability distributions (multivariate normal in 2.3) of all random variables in the process are given.

In Sect. 2.2.1, the stochastic process is introduced by illustrating the basic definition (2.2) with two examples. In Sect. 2.2.3 the stationarity assumptions are formulated using the moment functions of a stochastic process, which are defined in Sect. 2.2.2.

2.2.2 1st and 2nd Moment Functions

The moments of the random variables in a stochastic process are functions of the parameter, since the random variables are functions of the parameter. Hence, these are called moment functions.

Let (X_t) or $X(t)$, $t \in T$, be a stochastic process. Then:
1. $\mu_X(t) = \mathrm{E}X(t)$ *is called expectation function,*
2. $\sigma_X^2(t) = \mathrm{Var}X(t)$ *variance function,*
3. $c_X(t,u) = \mathrm{Cov}\big(X(t), X(u)\big) =$
   $\mathrm{E}\big((X(t) - \mu_X(t))(X(u) - \mu_X(u))\big)$ *covariance function, and*
4. $\varrho_X(t,u) = c_X(t,u)/\sqrt{\sigma_X^2(t)\sigma_X^2(u)}$ *correlation function.*

(2.4)

The realisations $x(t)$ of a stochastic process or random function $X(t)$ fluctuate around the expectation function $\mu_X(t)$ in a band with a height which depends on the variance function $\sigma_X^2(t)$. If $\mu_X(t)$ and $\sigma_X^2(t)$ exist, for any ordered pair of time points (t,u), $c_X(t,u)$ ($\varrho_X(t,u)$) is the covariance (the correlation) of the random variables $X(t)$ and $X(u)$. Often, $\mu_X(t)$ is called the expectation or mean and $\sigma_X^2(t)$ is called the variance, despite both, $\mu_X(t)$ and $\sigma_X^2(t)$, being functions of t. It is often useful to arrange the covariances $c_X(t,u)$ in the covariance matrix $\boldsymbol{C}_X$:

$$
\begin{array}{c}
\\
\vdots \\
X(-2) \\
X(-1) \\
X(0) \\
X(1) \\
X(2) \\
\vdots
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} \dots & X(-2) & X(-1) & X(0) & X(1) & X(2) & \dots \end{array} \\
\left(\begin{array}{ccccc}
 & & & & & & \\
 & c(-2,-2) & c(-2,-1) & c(-2,0) & c(-2,1) & c(-2,2) & \\
 & c(-1,-2) & c(-1,-1) & c(-1,0) & c(-1,1) & c(-1,2) & \\
 & c(0,-2) & c(0,-1) & c(0,0) & c(0,1) & c(0,2) & \\
 & c(1,-2) & c(1,-1) & c(1,0) & c(1,1) & c(1,2) & \\
 & c(2,-2) & c(2,-1) & c(2,0) & c(2,1) & c(2,2) & \\
 & & & & & &
\end{array}\right)
\end{array}
= \boldsymbol{C}_X .
$$

(2.5)

A stochastic process is usually described by giving its moment functions. $\mu_X(t)$ describes the expectation of the random variables in the process, $\sigma_X^2(t)$ the amplitudes of the fluctuations about $\mu_X(t)$. $c_X(t,u)$ and $\varrho_X(t,u)$ describe, because of the properties (1.10), the strength of a statistical linear relationship of the random variables $X(t)$ and $X(u)$.

If the stochastic process is multivariate (i.e., if the random variables are multi-dimensional), it is often useful to consider the covariances between the elements $X_t^{(j)}$ of the multi-dimensional random variables $\boldsymbol{X}^T(t) = \left(X_t^{(1)}, \ldots, X_t^{(p)}\right)$. These covariances are functions not only of the time, but also of the elements of the random variables. This suggests the definitions (2.6).

$$\begin{aligned} \mu_X^{(i)}(t) &= \mathrm{E}X_t^{(i)} & \left(\sigma_X^{(i)}(t)\right)^2 &= \mathrm{Var}X_t^{(i)} \\ c_X^{(jk)}(t,u) &= \mathrm{Cov}\left(X_t^{(j)}, X_u^{(k)}\right) = \mathrm{E}\Big(\left(X_t^{(j)} - \mu_X^{(j)}(t)\right)\left(X_u^{(k)} - \mu_X^{(k)}(u)\right)\Big) \\ \varrho_X^{(jk)}(t,u) &= \mathrm{Cov}\left(X_t^{(j)}, X_u^{(k)}\right) \Big/ \sqrt{\left(\sigma_X^{(j)}(t)\right)^2 \left(\sigma_X^{(k)}(u)\right)^2} \end{aligned} \tag{2.6}$$

$c_X^{(jj)}(t,u)$ is called autocovariance function, $c_X^{(jk)}(t,u),\ j \neq k$, crosscovariance function. Often, $c_X(t,u)$ ($r_X(t,u)$) is called autocovariance (autocorrelation) function even in the case of an univariate process. In this book, however, the second moment functions of an univariate process are called covariance and correlation function.

The symmetry of the covariance matrix (2.7,1) of a stochastic process follows directly from definition (2.4). However, not every symmetric matrix is a covariance matrix of a stochastic process, because a covariance matrix is required to be non-negative definite (positive semidefinite) in (2.7,2). The non-negative definiteness (positive semidefiniteness) is subsequently derived.

Let $\boldsymbol{C}_X$ be the covariance matrix of the stochastic process $X(t)$, t integer or real, with the covariances $c_X(t,u)$. Then $\boldsymbol{C}_X$ is:
1. *symmetric and* (2.7)
2. *positive semidefinite (non-negative definite). This implies that $\sum_{i=1}^n \sum_{j=1}^n b_i b_j c_X(t_i,t_j) \geq 0$ for any real b_i, b_j.*

The non-negative definiteness (positive definiteness) (2.7,2) is derived by constructing an arbitrary linear combination $Y = \sum_{i=1}^n b_i X(t_i)$ of random variables X_{t_i} in the stochastic process $X(t)$, whether t integer or real. Being a linear combination of random variables, Y is also a random variable in $L^2(\Omega, F)$ (7.10). Hence, $\mathrm{Var}Y \geq 0$, because a negative variance would be impossible. If $\mathrm{Var}Y = 0$, Y is degenerate, i.e., becomes a constant. Then, the variance of Y is calculated with (1.15,10) and $\mathrm{Var}Y = \sum_{i=1}^n \sum_{j=1}^n b_i b_j \mathrm{Cov}\left(X_{t_i}, X_{t_j}\right) = \sum_{i=1}^n \sum_{j=1}^n b_i b_j c_X(t_i,t_j)$ is obtained, where $c_X(t_i,t_j)$ is an element in the covariance matrix $\boldsymbol{C}_X$ of the stochastic process $X(t)$. Due to $0 \leq \mathrm{Var}Y$, $\sum_{i=1}^n \sum_{j=1}^n b_i b_j c_X(t_i,t_j) \geq 0$ is obtained and thus (2.7,2) is proven.

A matrix $\boldsymbol{C} = c_{ij}$, $i, j = 1, \ldots, n$, is called positive semidefinite or non-negative definite, if $\sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j c_{ij} \geq 0$ for all $(b_1, \ldots, b_n) \neq (0, \ldots, 0)$. However, if $\sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j c_{ij} > 0$, $\boldsymbol{C}$ is called positive definite. A positive definite matrix has desirable properties, for example, its inverse exists.

The properties (2.7) assert that linear combinations of random variables are also random variables with a variance that is often positive, seldom zero, but never negative. Such linear combinations are relatively easy to calculate and are therefore favoured candidates for a *prediction*. Generally, a prediction of a stochastic process is an estimator, as defined in (1.1), for a random variable in the process at a time point where no observation is available.

For example, $\widehat{X}_{t+1} = \beta_0 X_t + \beta_1 X_{t-1} + \ldots + \beta_p X_{t-p}$ is a linear prediction for X_{t+1} based on $X_t, X_{t-1}, \ldots, X_{t-p}$. Therefore, with (1.15,10) and due to the properties (2.7), the variance of the prediction error $X_{t+1} - \widehat{X}_{t+1}$ is calculated. By minimizing this variance, optimal $\beta_0, \ldots, \beta_p$ are found. Linear predictions are introduced in Sect. 2.7, important special cases of linear predictions being dealt with in Chaps. 4 and 5.

With the definitions of the stochastic process in Sect. 2.2.1 and its moment functions in Sect. 2.2.2 you are now equipped to encounter the stationarity assumptions in Sect. 2.2.3. Under the stationarity assumptions, the empirical moment functions of the wind speed measurements, in Fig. 2.1, are calculated in Sect. 2.1.

2.2.3 Stationarity Assumptions

Usually, the expectation function and the covariance function of a stochastic process, both as defined in (2.4), are not known to the practitioner. Can estimates be found for these moment functions? If a large number of independent realisations of the process have been observed, you then have many values for the random variables in the period with observations. These values are iid., as defined in (1.2), and you can calculate the estimates, known from an introduction to Statistics. This procedure is proposed in the remarks to the histograms in Fig. 2.5 and the scatterplots in Figs. 2.6 and 2.7.

However, when a geophysical variable is observed, often a time slice of only one realisation of the observed stochastic process results, and it is not possible to estimate the expectation and covariance function of the process, as (i) only one value for each random variable in the observational period is available and (ii) the measurements cannot be repeated. Obviously, the observations are not in agreement with (1.2,1). A possible way out of this apparent dead end is to assume weak stationarity, i.e., that the stochastic process, from which the observations stem, has the properties listed in (2.8).

A stochastic process (X_t) *or* $X(t)$*,* t *integer or real, as defined in (2.2), is weakly stationary, on condition that for any* t, u*:*

1. $\mu_X(t) = \mu =$ *constant and* (2.8)
2. $c_X(t, u) = c_X(t - u, 0) = c_X(\tau)$*, a function only of the lag* $\tau = t - u$*, i.e., of differences of the parameter.*

(2.8,2) requires that the covariance function be a function of the lag only, i.e., of differences of the parameter only. Hence, $c_X(t,u) = \mathrm{Cov}(X_t, X_u) = \mathrm{Cov}(X_v, X_w) = c_X(v,w)$ results on condition that X_t is weakly stationary and $u - t = w - v$ for any $t, u, v, w \in T$. For the same reason, you obtain $\sigma_X^2(t) = c_X(t,t) = c_X(0) = c_X(u,u) = \sigma_X^2(u)$, for any $t, u, \in T$. Thus, the variance function $\sigma_X^2(t)$ of a weakly stationary process is constant.

The first example of a time series stemming from a stationary process is the wind speed series in Fig. 2.1, as (i) the mean function is approximately constant in Fig. 2.1, and (ii) the empirical correlation function for lags $\tau \leq 20$ does not depend on the part of the time series used for the calculations, as seen in Fig. 2.3. Thus, the plots in Fig. 2.3 indicate that the covariances of the turbulent air flow are a function of the lag only, as required in (2.8,2).

A second example is the stochastic process with the simulated realisations in Fig. 2.4. In the simulation period, the realisations remain constant in the mean, and the histograms in Fig. 2.5 point to a constant expectation and a constant variance function. And what about the covariance function? The very similar scatterplots in the lags $\tau = 1, \ldots, 6$ in Figs. 2.6 and 2.7 indicate that covariances depend on the lag only. This indication is confirmed by the lagplots obtained from the last realisation of this process in Fig. 2.8. This simulation experiment with a stationary AR[2] model (as defined in (2.48) and the R expressions for the simulations as given in Problem 2.22) demonstrates that scatterplots for random variables in lag τ, obtained from many realisations of a stationary stochastic process, are very similar to lag plots for τ obtained from only one realisation. Please follow the hints in Problem 2.22 to demonstrate in a simulation experiment that the empirical correlations of random variables in lag τ, calculated from many simulated realisations of a stationary process, are very close in value to the empirical correlations calculated, using (2.1), from only one realisation.

As a first counter-example, the daily pyranometer values plotted in Fig. 1.4 are realisations of two stochastic processes, which are not stationary in the first moment as the summer values are much larger than the winter values in the mean.

As a second counter-example, the time series of differences of monthly SWIR values obtained for the pairs of stations in Fig. 2.19 are not stationary in the second moment, since unexpectedly, in the years 1975 and 1980, the amplitudes of the fluctuations become abruptly smaller. The abrupt decrease in amplitude points to a non-constant variance function.

However, many time series only reluctantly reveal whether they stem from a stationary process, once you look at the plots (time series plot, lag plots and plots of the empirical correlation functions) introduced in Sect. 2.1. Instructions that help to decide whether or not a time series is actually a time slice from a realisation of a stationary stochastic process are given in Sects. 2.5, 4.2 and 5.4.

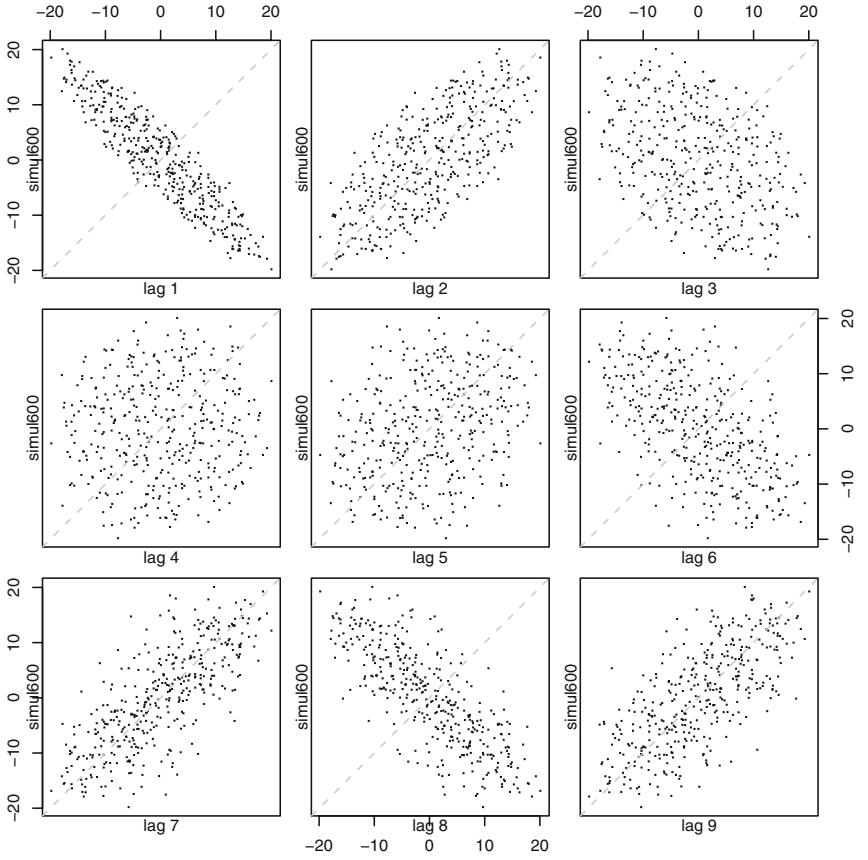


Fig. 2.8. Lagplots (cf. Fig. 2.2) obtained from realisation no. 600 produced in the simulation experiment described in the remarks to Fig. 2.4.

The covariance function of an univariate stationary stochastic process has the properties (2.9). Properties (2.9,1) and (2.9,2) directly follow from the definitions (2.4) and (2.8). Property (2.9,3) is shown using (1.15) and (1.16):

$$\begin{aligned}
\mathrm{Var}\big(X(t) \pm X(t+\tau)\big) &= \mathrm{Var}X(t) \pm 2\mathrm{Cov}\big(X(t), X(t+\tau)\big) + \mathrm{Var}X(t+\tau) \\
&= c_X(0) \pm 2c_X(\tau) + c_X(0) \\
0 &\le 2\big(c_X(0) \pm c_X(\tau)\big), \quad \text{a variance is non-negative} \\
-c_X(0) &\le c_X(\tau) \ \le \ c_X(0).
\end{aligned}$$

Property (2.9,4) is shown using the invariance of displacement, as proposed in the stationarity assumptions (2.8): $c_X(\tau) = \mathrm{Cov}\big(X(t), X(t+\tau)\big) = \mathrm{Cov}\big(X(t-t-\tau), X(0)\big) = \mathrm{Cov}\big(X(-\tau), X(0)\big) = \mathrm{Cov}\big(X(t-\tau), X(t)\big) = \mathrm{E}\big((X(t-\tau) - \mu_X(t-\tau))(X(t) - \mu_X(t))\big) = \mathrm{E}\big((X(t) - \mu_X(t))(X(t-\tau) -$

$\mu_X(t-\tau))\big) = \mathrm{Cov}\big(X(t), X(t-\tau)\big) = c_X(-\tau).$

The second moment functions of a stationary stochastic process have the following properties:

1.		$\varrho_X(\tau) = c_X(\tau)/c_X(0)$
2.	$c_X(0) = \sigma_X^2$	$\varrho_X(0) = 1$
3.	$-c_X(0) \le c_X(\tau) \le c_X(0)$	$-1 \le \varrho_X(\tau) \le 1$
4.	$c_X(-\tau) = c_X(\tau)$	$\varrho_X(-\tau) = \varrho_X(\tau)$

(2.9)

However, the cross-covariance and cross-correlation functions of a stationary multivariate process lack the properties (2.9).

If (X_t) is a stationary process with integer parameter t, the covariances in the covariance matrix $\boldsymbol{C}_X$, as defined in (2.5), depend (due to (2.9)) only on differences of the indices with every (negative-sloping) diagonal of $\boldsymbol{C}_X$ containing identical values, i.e., $\boldsymbol{C}_X$ has constant diagonals. In the main diagonal is the constant variance $c_X(0)$ of the process, in the following subdiagonals are the covariances for the lags $\tau = 1$, $\tau = 2$, etc. Equations with $\boldsymbol{C}_X$ are preferably solved with the procedure mentioned in (5.14).

Properties (2.7) pertaining to the covariance function $c_X(\tau)$ of a stationary stochastic process (X_t) are equivalent to $C_X(s) \ge 0$. $C_X(s)$ is the Fourier transform of $c_X(\tau)$ and is known as the *spectrum* of (X_t). Because it is non-negative, the spectrum $C_X(s)$ has an interpretation as mean power per frequency (a power in the physical sense is non-negative). Spectra of stationary stochastic processes are estimated in Chaps. 8 (special cases only), 9 and 10, after the necessary preparations have been made (in Chaps. 6 and 7). The estimation of a spectrum is often the starting point for the physical description of the random process that produced the realisation observed.

Summarizing Sect. 2.2.3, the first and second moment functions of a weak stationary process remain invariant under displacements in the parameter domain, because the moment functions in (2.8) are functions of differences of parameter values and not functions of the parameter itself. Beside the weak stationarity, there is also the strong stationarity. A stochastic process is strongly stationary if all stochastic properties remain invariant under displacements in the parameter domain, i.e., if the joint probability distributions of the process do not depend on the parameter, but only on differences of parameter values. Subsequently in this book, stationarity means weak stationarity, as defined in (2.8).

Using the basic definitions provided in Sect. 2.2 (stochastic process, its moment functions, the stationarity assumptions) stochastic processes with integer parameter are guessed, in Sect. 2.3, as models for the wind speed observations in Fig. 2.1 and for three more example time series. These are two indices for the North Atlantic Oscillation (NAO) and the yearly means of the Basel temperature series. The models are fitted by trial and error, the standard estimation procedure for these models being allocated to Chap. 5.

2.3 Discrete-time Stochastic Processes

In this section, example stochastic processes with integer parameter $t = \ldots, -1, 0, 1, \ldots$ are introduced. The general definition of these models (AR[p], MA[q], ARMA[p, q] models often fitted to time series with integer parameter) and their estimation is postponed to Chapter 5. These models are constructed with a white noise process, which is defined in Sect. 2.3.1. In Sects. 2.3.2 and 2.3.4, the AR[1] and MA[1] models are introduced. In Sect. 2.3.3, the random walk is defined. Both the stationary AR[1] and MA[1] models are linear processes which in turn are defined in Sect. 2.3.5.

Parallel to these models, three example time series are introduced: in Sect. 2.3.1, two indices for the North Atlantic Oscillation (NAO) and in Sect. 2.3.4, the Basel temperature series.

2.3.1 White Noise Process

The zonal circulation over the North Atlantic controls the flux of heat and moisture from the North Atlantic into Eurasia and from Eurasia into the Arctic. The intensity of the advection of sensible and latent energy has a large influence, particularly in winter, on the temperatures and precipitation over the Northern Atlantic and surrounding continents. North Atlantic zonal circulation is maintained by two quasi-permanent pressure centres: the Icelandic low and the Azores high. The difference of time averaged sea level pressure measurements at stations located in the centres of activity, i.e., the Azores and Iceland, is an index for the strength of the zonal circulation. Such an index can be constructed back to 1860 (Hurrell's index in [70], using measurements from the Azores and Iceland), or even back to 1820 (Jones' index in [77], using early (defined in Sect. 2.3.4) measurements from Gibraltar and Iceland). An example of reconstructing a climatological time series from early instrumental records is given in Sect. 2.3.4, where the Basel temperature series is introduced.

Hurrell's index is constructed from the normalised pressure differences for the months December, January, February and March. It is available from [69], and plotted on the top in Fig. 2.9. Jones provides in [76] a time series of the normalised monthly pressure differences. From this time series, the mean of the pressure differences in December, January, February and March is computed in Problem 2.2 and is plotted as Jones' index on the top in Fig. 2.10.

The NAO indices provide a coarse characterisation of the climatic processes in the North Atlantic and surrounding regions, e.g., in winters with large positive NAO indices, westerly winds are stronger than on the average over the North Atlantic and produce a more oceanic winter climate in Europe with higher temperatures than on the average. Reviews of the phenomena related to the North Atlantic climate variability are given in [90], [141], and [68].

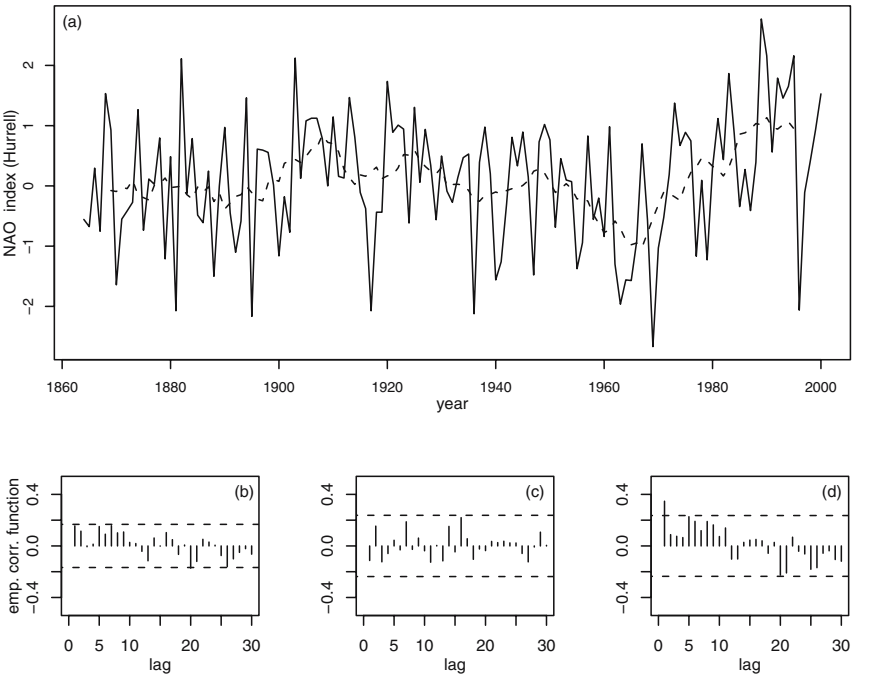


Fig. 2.9. NAO index (Hurrell) from 1864 through to 2000 (above, a) with 11-year running mean and empirical correlation functions calculated from the entire series, and the first and second half-series (below, b, c, d).

Are the NAO indices stationary? On the top in Figs. 2.9 and 2.10, both indices fluctuate around a local (cf. definition (2.55)) mean (approximated by an 11-year moving average, as defined in (2.39) until approximately 1900, whereas from 1905 through to approximately 1930 the local mean of the time series is rather high, thereafter decreasing until approximately 1970. From that time onwards, both time series increase in the local mean.

However, these decadal oscillations in the first moment functions of the NAO indices are not noticeably large since only small differences are found in the means of the indices calculated from the first and the second half of the observational periods (0.1100 and 0.0942 in the case of Hurrell's index). This diagnostic, i.e., comparing the means of a time series calculated from the first and the second half of the observational period, is recommended in Sect. 2.5.1. The absence of large differences in these means points to a constant expectation function. It is therefore assumed that both NAO indices are stationary in the first moment, i.e., the observations are in agreement with (2.8,1). In the case of Jones' index, the constant expectation function is discussed in Sects. 2.5.2 and 2.8.4.

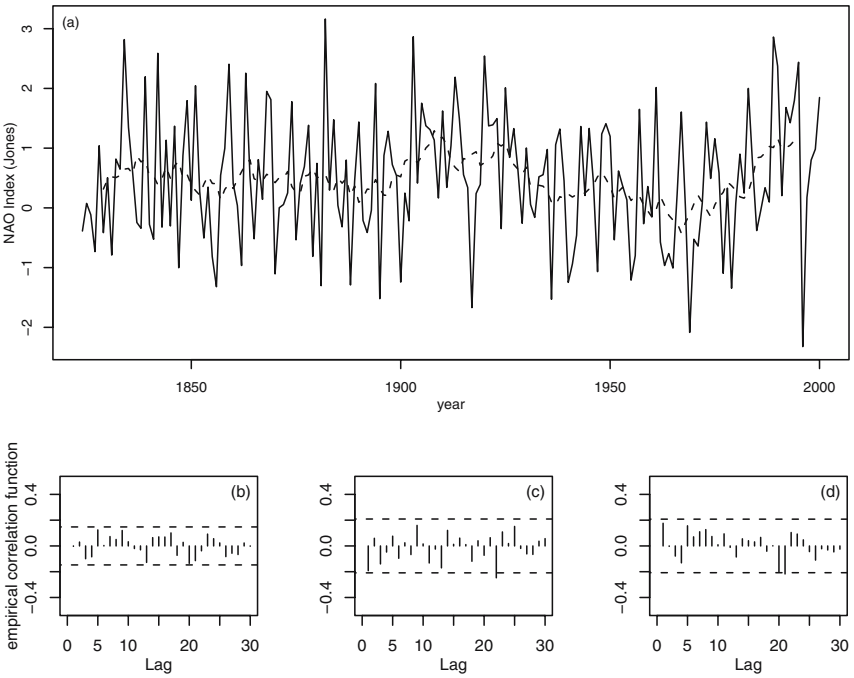


Fig. 2.10. NAO index (Jones) from 1824 through to 2000 (above, a) with 11-year running mean and empirical correlation functions calculated from the entire series, and the first and second half-series (below, b, c, d).

In both time series plots, it can also be seen that the amplitudes of the fluctuations remain approximately constant over the periods observed, and it is thus assumed that the variance functions of both indices remain constant.

The empirical correlation functions of Jones' index, on the bottom in Fig 2.10, are quite similar and accordingly do not depend on the time-slice used for the calculations. At the bottom in Fig. 2.9, however, the empirical correlation function of Hurrell's index calculated from the second half-series shows a slightly larger correlation in the first lag than those calculated from the entire series and the first half-series. If this small deviation is ignored, the six empirical correlation functions, plotted in Figs. 2.9 and 2.10, are quite similar. It is then assumed that both NAO indices are stationary in the second moment function, i.e., that the observations are in agreement with (2.8,2). Consequently, being stationary in the first and the second moment functions, both NOA indices are assumed to be stationary.

Unlike the empirical correlation function of the wind speed in the turbulent atmospheric flow in Fig. 2.3, the empirical correlation functions of the NAO indices remain (with a few exceptions) inside the 0.95-confidence interval for no correlation, as defined in (2.59). In agreement with the small cor-

relations, there are no discernible patterns in the plots of the indices (unlike those clearly recognisable in the wind speed measurements plotted Fig. 2.1).

This suggests that the NAO indices are weak white noise processes, as defined in (2.10). The definition (2.10) is the usual definition of the white noise process as being a sequence of uncorrelated random variables.

The sequence of random variables (W_t), $t = \ldots, -1, 0, 1, \ldots$, *with its moment functions* $\mu_W(t)$, $\sigma_W^2(t)$ *and* $c_W(t,u)$ *is called a weak white noise process, on condition that:*
1. $\mu_W(t) = \mu_W$ *and* $\sigma_W^2(t) = \sigma_W^2$, *i.e., do not depend on* t, *and*
2. $c_W(t,u) = \mathrm{Cov}(W_t, W_u) = 0$ *for* $t \neq u$. (2.10)

Stronger than (2.10) is the definition (2.11) of the strict white noise process which uses (joint) distribution functions of its random variables.

The sequence of random variables (W_t), $t = \ldots, -1, 0, 1, \ldots$, *with probability distributions* $F_t(w_t)$, *is called a strict white noise process, on condition that:*
1. $F_t(w_t) = F_u(w_u)$, *for* $t \neq u$, *and*
2. $F_{\underset{a\ finite\ number}{\ldots,-1,0,1,\ldots}}(\ldots, w_{-1}, w_0, w_1, \ldots) = \ldots F_{-1}(w_{-1}) F_0(w_0) F_1(w_1) \ldots$ (2.11)

Subsequently in this book, a white noise process means a weak white noise process, as defined in (2.10).

In (2.11), the joint distribution function of the random variables W_t in an arbitrary finite time slice $t_1 \leq t \leq t_2$ of the stochastic process is the product of the distribution functions of the W_t, $t_1 \leq t \leq t_2$. This is the usual definition of two or more (but a finite number) independent random variables. This definition is used, among other respects, for deriving the distribution of a function of independent random variables in (1.5).

As a direct consequence of (2.11), random variables W_t and W_u in a strict white noise process are independent for $t \neq u$.

The moment functions of a strict white noise process (W_t) are obtained easily with (1.15) and (1.16). Since the W_t are identically distributed as required in (2.11,1), $\mu_W(t) = \mathrm{E}W_t = \mu_W$, $\sigma_W^2(t) = \mathrm{Var}W_t = \sigma_W^2$ and $c_W(t,u) = \mathrm{Cov}(W_t, W_u) = \mathrm{E}\big((W_t - \mu_W)(W_u - \mu_W)\big)$, for $t \neq u$, are obtained. Since the W_t are independent as required in (2.11,2), $c_W(t,u) = \mathrm{E}(W_t - \mu_W) \times \mathrm{E}(W_u - \mu_W) = (\mathrm{E}W_t - \mu_W)(\mathrm{E}W_u - \mu_W) = 0$, for $t \neq u$, results. Thus, the independence required in (2.11,2) implies that pairs of random variables in a strict white noise process are not correlated. The reverse is however not true, since from $\mathrm{Cov}(W_t, W_u) = 0$ in (2.10,2) it is not possible to deduce that W_t and W_u are stochastically independent without postulating additional properties of the process. Hence, the set of stochastic processes as defined in (2.10) contains the set of processes satisfying (2.11) as a subset.

From both definitions, moment functions $\mu_W(t) = \mu_W$, $c_W(t,u) = \sigma_W^2$, for $t = u$ and $c_W(t,u) = 0$, for $t \neq u$, in agreement with (2.8), are obtained, and thus the white noise process is stationary.

In Problem 2.2, a white noise process (W_t) is estimated as a model for Jones' NAO index in Fig 2.10. Since $c_W(t,u) = 0$, if $t \neq u$ by definition, its expectation μ_W, its variance σ_W^2 and its distribution $F_t(w_t)$ remain to be estimated. Using this model one can calculate the probability that an NAO index larger than, e.g., 2.0, is observed.

In (7.76) it becomes self-evident why processes as defined in (2.10) or (2.11) are called "white noise".

A white noise process is easy to estimate as demonstrated in Problem 2.2; but unfortunately, it is not suitable as a model for most time series, since in a time series the observations are usually correlated. This raises the need for models that incorporate correlations or dependencies between the observations. Two relatively simple models with this property are introduced in Sects. 2.3.2 and 2.3.4.

2.3.2 First Order Autoregressive (AR[1]) Process

Fig. 2.1 suggests that the speed x_t of the turbulent atmospheric flow at time t depends on the speeds x_{t+n} at neighbouring times $t+n$, $-10 < n < 10$ approximately. The dependence of the observation x_i on the neighbour observations is confirmed by the empirical correlation functions in Fig. 2.3. Apart from this dependence, abrupt random changes are also seen in the plotted wind speed time series. In the example first order autoregressive processes given in (2.12),

Let (W_t) be a white noise process with $W_t \sim N(0,1)$. Then:

$$\begin{aligned} &1.\ (X_t - \mu_X) = \begin{cases} 0 & t = 1 \\ 0.9 \times (X_{t-1} - \mu_X) + W_t & t = 2,3,4\ldots \end{cases} \\ &2.\ (Y_t - \mu_Y) = \begin{cases} 0 & t = 1 \\ 0.5 \times (Y_{t-1} - \mu_Y) + W_t & t = 2,3,4\ldots \end{cases} \end{aligned} \tag{2.12}$$

are examples of first order autoregressive processes.

the dependence is expressed by multiplying the wind speed at time $t-1$ with a constant weight in order to obtain the wind speed at time t: $X_t = a^{(1)} \times X_{t-1}$ or $Y_t = a^{(2)} \times Y_{t-1}$, and the randomness by adding a random variable to this product: $X_t = a^{(1)} \times X_{t-1} + W_t$ or $Y_t = a^{(2)} \times Y_{t-1} + W_t$. (W_t) is a white noise process, as defined in (2.10). Both processes start in $t = 1$ with $X_1 = Y_1 = 0$, if $\mu_X = \mu_Y = 0$.

Using (2.12), e.g., $(x_t) = (0.000, 0.755, -1.726, -3.311, -2.044, \ldots)$ is obtained from the realisation $(w_t) = (0.090, 0.755, -2.406, -1.757, \ldots)$, $t = 1,2,3,4,\ldots$, of a white noise process with $W_t \sim N(0,1)$. With the same realisation, $(y_t) = (0.000, 0.755, -2.028, -2.772, -0.449, \ldots)$, is obtained, if $\mu_X = \mu_Y = 0$.

Assuming a zero mean of the first order autoregressive process (X_t) in (2.12), i.e., $\mu_X = 0$, a realisation of (X_t) is obtained with

```
wnlength <- 1000    #length of realisation
```

```
w <- rnorm(wnlength,mean=0.0,sd=1.0)  #white noise
ax <- 0.9           #in (2.12)
x <- 1:wnlength     #vector for simulated process
x[1] <- 0           #in (2.12)
for (i in 2:wnlength) {x[i] <- ax*x[i-1] + w[i] }
```

Much more quickly, however, is an autoregressive process simulated with `arima.sim()`, as demonstrated in Problem 2.22.

In the R expressions above, `rnorm(wnlength,mean=0.0,sd=1.0)` generates a realisation (w_t) of a white noise process with $W_t \sim N(0,1)$, from whence (x_t) is computed as required in (2.12). In (2.12), the mean of the white noise process is identically zero, and thus the first order autoregressive process contains three parameters. As an example, when simulating (X_t) with the R expressions above, the first parameter is $\mu_X = 0$, the second $\sigma_W^2 = 1$ and with the third, $a^{(1)} = 0.9$, x_t is obtained from w_t and the previous x_{t-1}. The parameters in (2.12) should on no account be confused with the parameter in definition (2.2), which is the argument of a stochastic process or a random function.

Simulated realisations of both example first order autoregressive processes (X_t) and (Y_t), as specified in (2.12), are plotted in Fig. 2.11. In both plots, the simulations fluctuate with small random amplitudes within small neighbourhoods; in larger neighbourhoods, the amplitudes of the fluctuations become relatively larger, even though the fluctuations remain random.

When the wind speed time series in Fig. 2.1 is compared with the simulations on the left in Fig. 2.11, it is seen that the simulation of (X_t) is more similar to the wind speed series than the simulation of (Y_t). The same ranking results from a comparison of the empirical correlation functions of the wind speed series in Fig. 2.3 with the correlation functions on the right in Fig. 2.11. It follows that $X(t)$ is a more suitable model for the wind speed series than $Y(t)$.

Nevertheless, the (X_t) model can also be improved upon, since the simulation in Fig. 2.11 (above, on the left) fluctuates between -6 and $+6$, whereas the observations in Fig. 2.1 only fluctuate between -6 and 0. An improved model is obtained in Problem 2.3 comparing the plots in Figs. 2.1 and 2.3 with the plots obtained from simulations with unchanged third parameter $a^{(1)} = 0.9$, but with values other than $\mu_X = 0$ for first (the mean of the process) and $\sigma_W^2 = 1$ for second (the variance of the white noise process) parameter of the first order autoregressive model.

When comparing the plots in Figs. 2.1 and 2.3 with those in Fig. 2.11 one tries to fit a model with two parameters (assuming $\mu_X = 0$ in (2.12)) to the measurements of the wind speeds in the turbulent atmospheric flow. This trial and error procedure is very useful to illustrate the roles being played by the model parameters. Parameter estimates as defined in (1.1), however, would be a better tool in the hand of the practitioner. Do such estimates

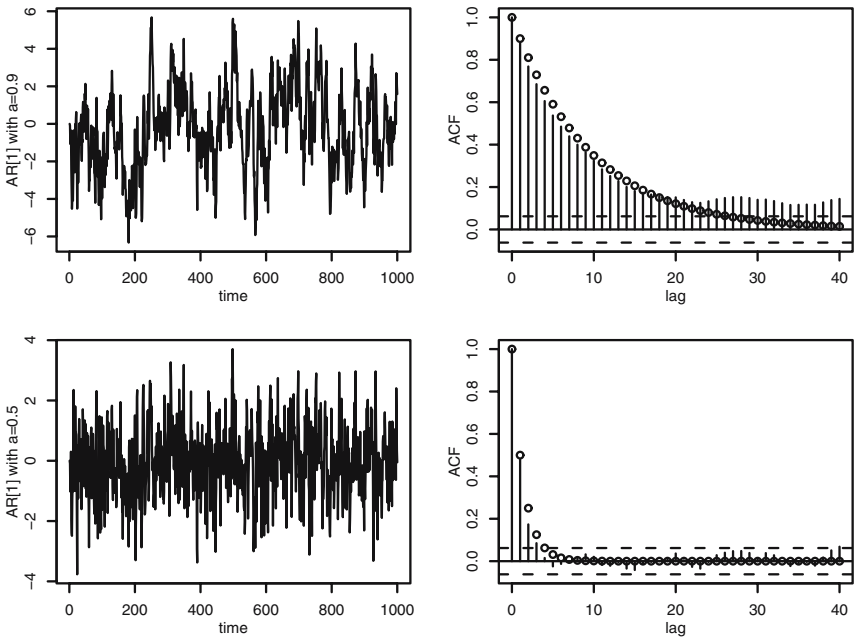


Fig. 2.11. First order autoregressive processes (X_t) and (Y_t), as specified in (2.12). On the left, simulations of (X_t) and (Y_t), on the right, with $\circ$, the theoretical correlation functions $c_X(\tau)$ and $c_Y(\tau)$ as derived in (2.16), together with the empirical correlation functions calculated from the simulations using (2.1).

exist? Answers to this and other questions related to the estimation of an autoregressive process are found in the concluding remarks of this section.

The definition of the first order autoregressive model is given in (2.13).

If (W_t) is a white noise process with $\mu_W = 0$ and σ_W^2, then $X_t = aX_{t-1} + W_t$ is called a first order autoregressive process (AR[1] process or AR[1] model) with $X_t - aX_{t-1} = W_t$ being its difference equation. A process (V_t) with an expectation function $\mu_V(t) \neq 0$ is an AR[1] model, on condition that $(X_t) = (V_t - \mu_V(t))$ is an AR[1] model. (2.13)

The difference equation of the AR[1] model in (2.13) is linear, of order 1, and its coefficients are constant. A solution is obtained with repeated substitutions and assuming an initial condition, e.g., $X_{-N} = 0$ ($X_1 = 0$ is assumed for the examples given in (2.12)).

$$\begin{aligned} X_t &= W_t + aX_{t-1} \\ &= W_t + a(W_{t-1} + aX_{t-2}) = \ldots = \sum_{u=-(N-1)}^{t} a^{t-u} W_u \end{aligned} \tag{2.14}$$

Thus in definition (2.13), the AR[1] model (X_t) is a linear combination of the present and past variables of the generating white noise process (W_t).

Is an AR[1] model stationary? The answer is found (as in the case of the white noise process in the remarks to (2.11)) when the moment functions of the process are calculated using (1.15) and (1.16). Thus, from the weighted sum of the W_t in (2.14),

$$\begin{aligned} \mathrm{E}X_t &= \mathrm{E}\sum_{u=-(N-1)}^{t} a^{t-u}W_u \quad = \mathrm{E}(W_u)\sum_{u=-(N-1)}^{t} a^{t-u} \\ &= \begin{cases} \mu_W(1-a^{t+N})/(1-a) & \text{for } a \neq 1 \\ \mu_W(t+N) & \text{for } a = 1 \end{cases} \end{aligned} \tag{2.15}$$

expression (2.15) is obtained for the expectation function, and, therefore, (X_t) is stationary in the first moment function, if $\mu_W = 0$, as required in definition (2.13).

However, allowing for $\mu_W \neq 0$ in this definition, $\mathrm{E}X_t$ would be a function of $N + t$ and thus, for $|a| < 1$, $\mathrm{E}X_t = \mu_X \approx \mu_W/(1-a)$ would result because a geometric series converges for large $N+t$. Such a process would be stationary in the first moment function only if a long time had elapsed since its initialisation.

The second moment functions of an AR[1] process are derived in (2.66) with the results (2.16) and (2.17):

$$\begin{aligned} c_X(t,\tau) &= \mathrm{Cov}(X_t, X_{t+\tau}) \\ &= \begin{cases} \sigma_W^2 a^{|\tau|}(1-a^{2(t+N)})/(1-a^2) & \text{for } |a| \neq 1 \\ \sigma_W^2(t+N) & \text{for } |a| = 1 \end{cases} \end{aligned} \tag{2.16}$$

$$\sigma_X^2 = \mathrm{Cov}(X_t, X_t) \;=\; \begin{cases} \sigma_W^2(1-a^{2(t+N)})/(1-a^2) & \text{for } |a| \neq 1 \\ \sigma_W^2(t+N) & \text{for } |a| = 1 \end{cases} \tag{2.17}$$

The covariance and the variance functions in (2.16) and (2.17) are functions of $t + N$, the time elapsed since the process was initialised. If however $|a| < 1$, then for large $N + t$, $\sigma_X^2 \approx \sigma_W^2/(1-a^2)$ and $c_X(\tau) = \mathrm{Cov}(X_t, X_{t+\tau}) \approx \sigma_W^2 a^{|\tau|}/(1-a^2)$ are obtained, and thus, the covariance function decreases with increasing lag τ.

Since stationarity means that both the first and second moment functions of a stochastic process are immune against displacements in the time domain, it is concluded that an AR[1] model with $|a| < 1$ is not stationary if the start time of the process is given. For example, due to (2.14), the AR[1] model (X_t) starts at time point $t = -N$ with $X_{-N} = 0$, and consequently, $x_{-N} = 0$ is required for all realisations (x_t). Nevertheless, an AR[1] process with $|a| < 1$ "forgets" its initial condition if "enough" time has elapsed since its initialisation: the process becomes stationary. This property is formulated letting $N \to \infty$ in the lower bound $u = -(N-1)$ of the sum in (2.14). Thus, the sum on the left side in (2.18) is obtained. Both sums in (2.18) are equal as shown in Problem 2.4.

$$\sum_{u=-\infty}^{t} a^{t-u} W_u \quad = \quad \sum_{v=0}^{\infty} a^v W_{t-v} \qquad\qquad |a| < 1 \tag{2.18}$$

A finite number of random variables sums up to a random variable. If an infinite number of random variables is being summed up, what is obtained? More specifically, does the sum constructed in (2.18) from an infinite number of weighted white noise variables exist? A plausible answer is found on having a closer look at the conditions for convergence of $\sum_{v=0}^{\infty} a^v w_{t-v}$, i.e., realisations of $\sum_{v=0}^{\infty} a^v W_{t-v}$. $\sum_{v=0}^{\infty} a^v w_{t-v}$ converges for each realisation $(w_t, w_{t-1}, \ldots,)$ of $(W_t, W_{t-1}, \ldots,)$, if $|w_t| \leq c$, $|w_{t-1}| \leq c$, etc. Each $(w_t, w_{t-1}, \ldots,)$, however, contains values $|w_t| > c$. As these values are weighted with the rapidly falling $|a|^v$, their contribution to the sum is small if the variance of the white noise process is finite, as required in (2.10). This is an intuitive approach to the convergence in (2.19) which is derived in (7.15).

The sum in (2.19) converges in mean square, as defined in (7.11) (cf. the remarks to (6.7) and (6.5)) to a limiting random variable X_t,

$$X_t \overset{\text{ms}}{=} \sum_{u=0}^{\infty} a^u W_{t-u} \qquad \text{with } |a| < 1 \tag{2.19}$$

if the moments of W_t exist, as stipulated in (2.13) and in (2.10).

Summarizing Sect. 2.3.2, the AR[1] process was introduced and the following properties were obtained with a closer investigation:

1. An AR[1] process is stationary, if $|a| < 1$ and if an infinitely long time has elapsed since its initialisation. The usual AR[1] models, however, produce simulations which become stationary after a relatively short initialisation period, as is learned from the experiments made in Problem 2.3.
2. If $|a| > 1$, the AR[1] model grows exponentially (Problem 2.3).
3. If $a = 1$, the process is called a random walk process as defined in Sect. 2.3.3.

2.3.3 Random Walk

With $a = 1$ in the difference equation of the AR[1] model as defined in (2.13) and with repeated substitutions as in (2.14) using the initialisation $X_1 = W_1$, the random walk process as defined in (2.20)

$$\begin{aligned} &\textit{If } (W_t) \textit{ is a white noise process with } \mu_W = 0 \textit{ and } \sigma_W^2\textit{, then} \\ &X_t = \textstyle\sum_{u=1}^{t} W_u\textit{, with } t \geq 1\textit{, is called a random walk process.} \end{aligned} \tag{2.20}$$

is shown to be an AR[1] model with $a = 1$. Consequently, the moment functions of a random walk process can be obtained with $a = 1$ in (2.15), (2.16) and (2.17). Since $\mu_W = 0$ is required in (2.20), a random walk process is stationary in the first moment function; in the second moment function, however,

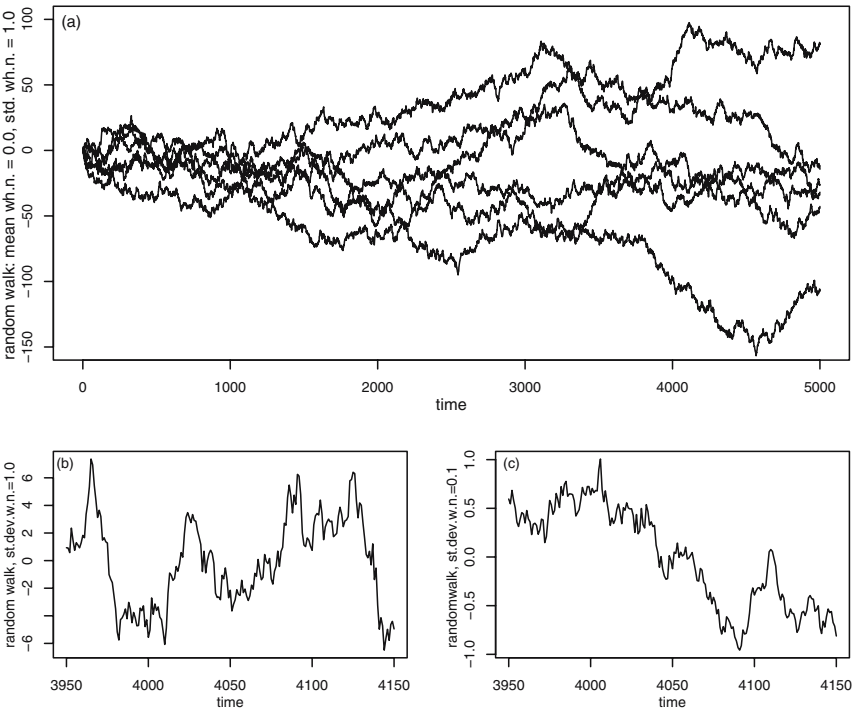


Fig. 2.12. Six simulations of a random walk as defined in (2.20) with W_u being normally distributed with $\mu_W = 0$ and $\sigma_W^2 = 1$ (above, a); and below, for $3950 \leq t \leq 4150$, simulations $x^{(1)}(t) - m^{(1)}$ with $\sigma_W^2 = 1$ (b) and $x^{(2)}(t) - m^{(2)}$ with $\sigma_W^2 = 0.1$ (c), $m^{(1)}$ and $m^{(2)}$ being the means of the simulations for the period $3950 \leq t \leq 4150$.

it is not stationary, as its variance function increases with increasing time elapsed since its initialisation. As a consequence of the increasing variance function, simulated realisations of a random walk process diverge, as shown above in Fig. 2.12. Alternatively, the moment functions of a random walk process are derived in (2.69) and (2.67) directly from its definition (2.20).

The simulated random walks plotted in Fig. 2.12 (above) show, despite the constant expectation function of the underlying process, gradual changes in their local means being typical for random walks. The amplitudes of these gradual changes depend on the variance σ_W^2 of the generating white noise process in (2.20), since, when calculating $x_t = x_{t-1} + w_t$ using definition (2.20), the probability that a w_t not close to $\mu_W = 0$ is added to x_{t-1} increases with increasing σ_W^2, and decreases with decreasing σ_W^2. The dependence of the changes in the local means on σ_W^2 is demonstrated below in Fig. 2.12.

The random walk process and the AR[1] process defined in Sects. 2.3.3 and 2.3.2 have difference equations $X_t - X_{t-1} = W_t$ and $X_t - aX_{t-1} = W_t$ with W_t being in a white noise process, as defined in Sect. 2.3.1. It is also possible to construct a stochastic process not using differences of the process itself, but differences $Y_t = W_t - bW_{t-1}$ of the generating white noise process. This process is introduced in Sect. 2.3.4.

2.3.4 First Order Moving Average Process

The need to construct the first order moving average process $Y_t = W_t - bW_{t-1}$ using differences of a white noise process (W_t) arises from a closer investigation of the yearly temperature values measured in Basel from 1755 through to 1955. This climatological time series is plotted in Fig. 2.13 (above). This time series is reconstructed in [13] from early instrumental observations, i.e., observations made prior to approximately 1860 when the first international standards for the observations of meteorological variables (e.g., temperature, precipitation, pressure) were introduced, as discussed in Sect. 1.1.

When reconstructing the Basel temperature series as described in [13], Bider and co-authors found in the archives of the Federal Office for Meteorology and Climatology records of the observations made by d'Annone from 1755 through to 1804; d'Annone's instruments, however, were missing. They circumvented this difficulty by rebuilding the instruments using the blueprints found together with the temperature records. Then, with the replicas of d'Annone's instruments installed in a reconstruction of d'Annone's historical site on the one hand, and modern instruments installed according to the climatological standard on the other, they performed a comparison experiment. With the results from this experiment, they converted the d'Annone records to meet the modern temperature standard. Thus pushing back the beginning of the Basel temperature series to 1755, they made a homogeneous series of monthly temperature values, from January 1755 through to December 1957, available in [13]. From the contiguous series of monthly values, the yearly temperature values plotted above, in Fig. 2.13, are obtained.

Are the Basel yearly temperature values a time slice from a realisation of a stationary stochastic process?

Above in Fig. 2.13, the time series fluctuates around a local (cf. definition (2.55)) mean which is here, in contrast to the NAO indices in Figs. 2.9 and 2.10, not approximated using a moving average. These decadal fluctuations of the local means are small. On the secular scale however, i.e., over periods of approximately 100 years, the Basel temperatures increase and a significant secular trend is found in Sect. 3.3. Consequently, the time series is not stationary in the expectation function.

The empirical correlation functions calculated from the entire and the second half-series and plotted below, in Fig. 2.13, show small positive correlations in the first and second lag, whereas the first half-series together with its empirical correlation function seems to stem from a white noise process.

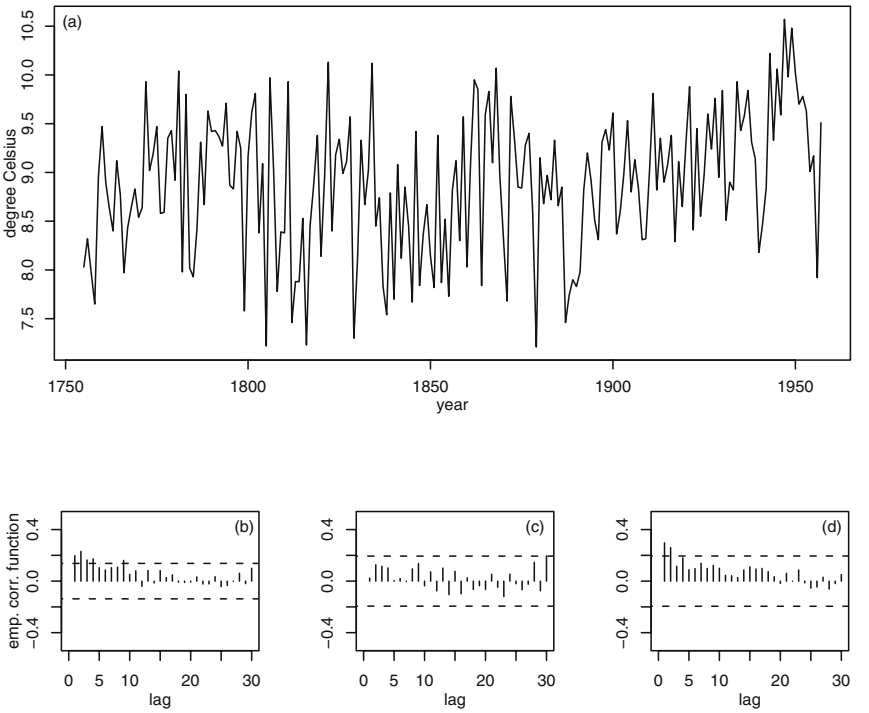


Fig. 2.13. Basel temperature series. Yearly values in ^{0}C (degree Celsius) from 1755 through to 1957 as given in [13] (above, a), with empirical correlation functions (below, b, c, d) calculated from the entire series, and the first and second half-series.

Thus, the yearly values in the Basel 200-years temperature series are not stationary in the covariance function.

As the Basel yearly temperature values (x_t), $1755 \leq t \leq 1957$, are not in agreement with the stationary assumptions (2.8), you may ask whether they are a time slice from a realisation of a random walk process, the random walk process being (except for an exponentially growing AR[1] model) the only non-stationary process introduced so far. This would be the case, reconcilable with definition (2.20), only if the temperature differences $y_t = x_t - x_{t-1}$ could be shown to stem from a white noise process. Hence, these differences are calculated, as defined in (2.39), using `diff()` (introduced in Problem 2.1) and plotted above, in Fig. 2.14.

The time series of the differences (y_t) oscillates around an approximately constant expectation function $\mu_Y = 0$ and thus is assumed to be stationary in the first moment function. The amplitudes of the oscillations are slightly larger in the period from 1800 through to 1870 than in the preceding and following periods, thus pointing to a non-constant variance function. However,

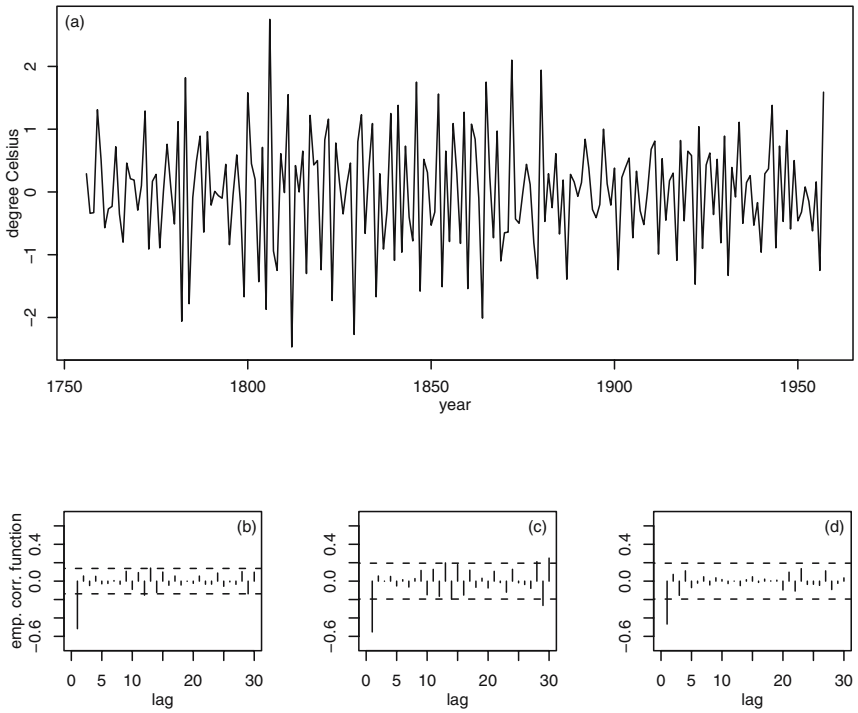


Fig. 2.14. Differences $y_t = x_t - x_{t-1}$ of the yearly values (x_t) in the Basel temperature series in ^{0}C (above, a), with empirical correlation function (below, b, c, d) calculated from the entire series, and the first and second half-series.

from the plots of the empirical correlation functions below, in Fig. 2.14, it is concluded that (y_t) is stationary in its covariance function, and, consequently, also in its variance function.

Above, in Fig. 2.14, it is seen that large differences y_t in year t are followed by small y_{t+1} in the next year $t+1$ in the average over the observational period. This clearly discernible sawtooth pattern in (y_t) is in line with an empirical correlation $\hat{\varrho}_Y(1) \approx -0.5$ in the first lag, as seen below in Fig. 2.14. Since $\hat{\varrho}_Y(1) \approx -0.5$ is far outside the 0.95-confidence interval for zero correlation, as defined in (2.59), the Basel temperature differences (y_t) above, in Fig. 2.14, obviously do not stem from a white noise process, and, consequently, the Basel yearly temperature values (x_t) above, in Fig. 2.13, are not a time slice from a realisation of a random walk process.

Inspecting the plots in Figs. 2.13 and 2.14, it is concluded that the Basel yearly temperature values stem neither from a stationary AR[1] model nor from a random walk process. Since, as a third possibility, no exponential growth is seen in Fig. 2.13 , a non-stationary AR[1] model with $|a| > 1$ can

also be excluded. Therefore, none of the models introduced so far is suitable for this time series.

The difference time series (y_t), as constructed above from the Basel yearly temperature series (x_t), is stationary, as concluded from the plots in Fig. 2.14. The stationarity suggests that a model (Y_t) for the differences (y_t) is less complex (in the sense that no trend in the expectation function has to be accounted for) than a model (X_t) for the original observations (x_t). Assuming that a model (Y_t) does exist, $X_t = \sum_{u=1}^{t} Y_u$, $t \geq 1$, can then be obtained adding the Y_t.

From the plots in Fig. 2.14 it becomes obvious that (y_t) does not stem from a white noise process due to its too large $|\hat{\varrho}_Y(1)| \approx 0.5$, as argued above. It is also apparent that (y_t) does not stem from a stationary AR[1] process, since $\hat{\varrho}_Y(\tau) \approx 0.0$ for $\tau > 1$, whereas the empirical correlation function $\hat{\varrho}_X(\tau)$ of a stationary AR[1] process (X_t) decays exponentially, in accordance with (2.16) and as demonstrated in Fig. 2.11.

The exponential decay of the correlation function $\varrho_X(\tau)$ of a stationary AR[1] model (X_t) is suspected to originate from X_t which (as shown in (2.14)) is a weighted sum of the W_t, W_{t-1}, W_{t-2}, ..., i.e., the present and past random variables in the generating white noise process. When reducing the number of the W_t in this sum to the present and only one past random variable as defined in (2.21), it is anticipated that a process with a correlation function is obtained which will be in agreement with the empirical correlation function of the differences below, in Fig. 2.14.

If (W_t) is a white noise process with $\mu_W = 0$ and σ_W^2, then $Y_t = W_t - bW_{t-1}$ is called a first order moving average process (MA[1] process or MA[1] model). (2.21)
(V_t) with an expectation function $\mu_V(t) \neq 0$ is an MA[1] model, on condition that $(Y_t) = (V_t - \mu_V(t))$ is an MA[1] model.

From the definition of the MA[1] process (2.21), using (1.15) and (1.16), the expectation function (2.22) and the covariance $c_Y(t, \tau)$

$$\begin{aligned} \mathrm{E}Y_t &= \mathrm{E}(W_t - bW_{t-1}) = \mathrm{E}W_t - b\mathrm{E}W_{t-1} = 0 \qquad (2.22)\\ c_Y(t,\tau) &= \mathrm{Cov}(Y_t, Y_{t+\tau}) = \mathrm{Cov}(W_t - bW_{t-1}, W_{t+\tau} - bW_{t+\tau-1}) \\ &= \mathrm{Cov}(W_t, W_{t+\tau}) - b\mathrm{Cov}(W_t, W_{t+\tau-1}) \\ &\qquad -b\mathrm{Cov}(W_{t-1}, W_{t+\tau}) + b^2\mathrm{Cov}(W_{t-1}, W_{t+\tau-1}) \end{aligned}$$

are obtained and the covariance and correlation functions (2.23) follow when summing up $c_Y(t, \tau)$ for the lags $\tau = 0$ and $\tau = \pm 1$. For lags $\tau < -1$ and $\tau > 1$, these functions are identically zero:

$$c_Y(\tau) = \begin{cases} (1+b^2)\sigma_W^2 \\ -b\sigma_W^2 \\ 0 \end{cases} \qquad \varrho_Y(\tau) = \begin{cases} 1 & \text{for } \tau = 0 \\ -b/(1+b^2) & \text{for } \tau = \pm 1 \\ 0 & \text{for } |\tau| > 1 \end{cases} \qquad (2.23)$$

From (2.22) and (2.23) it is concluded that a MA[1] model is stationary in the first and second moment functions for finite weights b.

Given $\varrho_Y(1)$, b can be obtained as solution of equation $b^2 + (1/\varrho_Y(1))b + 1 = 0$, and, consequently, b can be determined only by assuming an additional property (the invertibility as proposed in (5.32)) of the MA[1] process.

When an MA[1] model is fitted to an observed time series the parameters b, μ_Y and σ_W^2 are not known. In the example of the differences (y_t) of the Basel yearly temperature values plotted in Fig. 2.14, the model parameters are obtained only by trial and error: realisations are simulated with several values for each parameter and, subsequently, plots of the simulations are compared with the plot of the observed differences (y_t). The model contains three parameters, and, if simulations using five values for each of them are generated, then 125 simulations have to be compared with the observed differences. Hence, in Problem 2.8, $\mu_Y = 0$ is assumed with fewer comparisons having to be made. Using the same assumption, i.e., that the expectation function of the process is identically zero, the parameters in the AR[1] models (2.12) for the wind speed measurements in a turbulent air flow are found after a few attempts.

In Fig. 2.15, simulations of an MA[1] model with $b = 0.9$, $b = 2.0$, $b = 0.1$ as well as $\mu_Y = 0$ and $\sigma_W = 0.7$ are plotted. With $b = 0.9$, a simulation (in the second plot from the top) is obtained which is quite similar to the observations (in the first plot, from the top): the amplitudes of the fluctuations and the negative correlation of values following each other agree, and, in addition, the theoretical correlation $\varrho_Y(1) \approx -0.5$ of the model calculated with (2.23) is not too far from the empirical correlations $\hat{\varrho}(1)$ shown below, in Fig. 2.14. In contrast, the simulations with $b = 2.0$ ($b = 0.1$) fluctuate with too large (too small) amplitudes as compared to the observations.

Therefore, the MA[1] process (Y_t), as defined in (2.21), with $b = 0.9$, $\mu_Y = 0$, $\sigma_W = 0.7$, is preferred for modelling the differences of the Basel yearly temperature values. Using realisations (y_t) of (Y_t), the Basel yearly temperature values are simulated in Problems 2.8 and 2.9 by calculating the sums $x_t = x_{t-1} + y_t$.

2.3.5 Linear Process

With definitions (2.13) and (2.21), useful models for two example time series are generated from a white noise process: (i) an AR[1] model with $a = 0.9$, $\mu_X = 0$ and $\sigma_W = 1$ delivers realisations very similar to the wind speeds measured in a turbulent air flow and plotted in Fig. 2.1, and (ii) simulations using an MA[1] model with $b = 0.9$, $\mu_Y = 0$ and $\sigma_W = 0.7$ are very difficult to distinguish from the observed differences of the Basel yearly temperature values plotted above, in Fig. 2.15.

Both models are fairly parsimonious as they have only three parameters. The parsimony improves the ratio N/p (N the length of the time series, p the

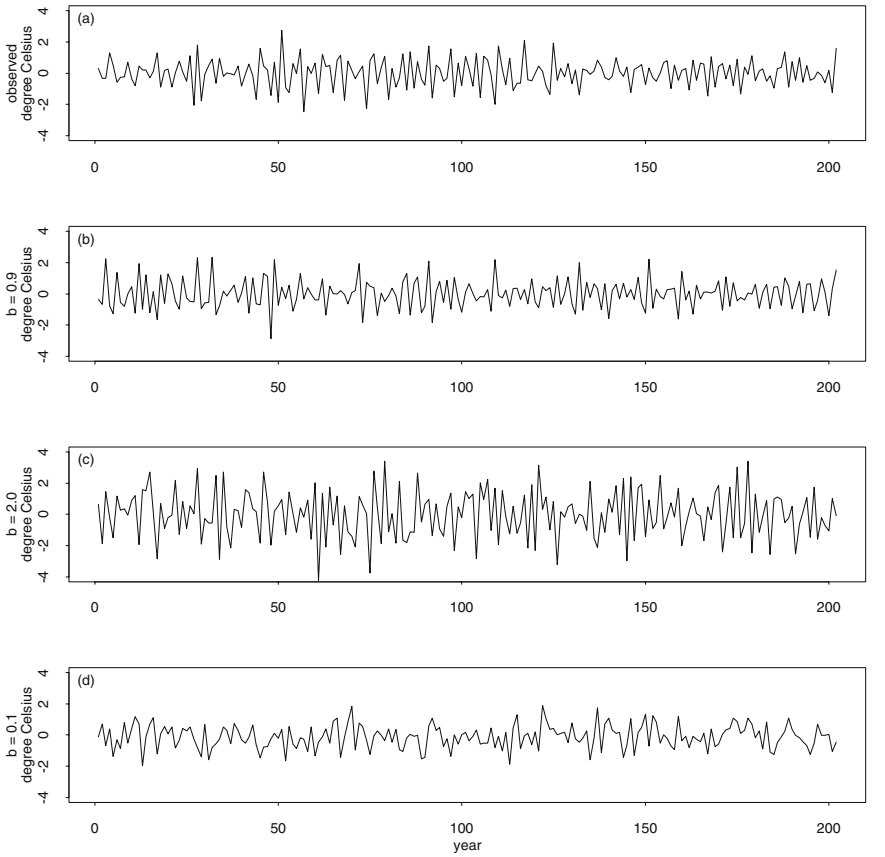


Fig. 2.15. On the top (a), differences $y_t = x_t - x_{t-1}$ of the yearly values (x_t) in the Basel temperature series in ^{0}C, as plotted in Fig. 2.14; below (b,c,d), three simulations with MA[1] models as defined in (2.21) with $\sigma_W = 0.7^0$C and $b = 0.9$, $b = 2.0$ and $b = 0.1$.

number of values to be estimated). When N/p increases often (but not always) the variance of an estimator obtained from (independent) observations decreases, and, therefore, parsimonious models are often superior to models with a large number of parameters. For example, the empirical moment functions of a time series, as defined in (2.1), are estimates for the expectation and covariance functions of the stationary stochastic process from which the observations stem, and, as proposed in (2.56) and (2.58), the variances of these estimates decrease with increasing duration of the observational period. The very same property is shared by the estimates of the parameters in the AR[1] and MA[1] models being calculated from a stationary time series, as derived in Secs. 5.2 and 5.3.4.

Both, the MA[1] and AR[1] models, are linear combinations of random variables being in a white noise process. In the case of a MA[1] model, only two non-zero elements are in the sequence $(1, -b, 0, 0, \ldots)$ of the coefficients being multiplied as weights with the random variables, and there is no doubt that the sum of weighted white noise variables, as defined in (2.21), exists. In the case of an AR[1] model, there is an infinite number of weights in the sequence $(a^0, a^1, a^2, \ldots)$ which, however, decays rapidly if $|a| < 1$. The rapid decay is used in the derivation of (7.15) to show that the infinite linear combinations in (2.18) converge, as proposed in (2.19).

In the general case of a process being generated from white noise, the weights of the white noise variables are no longer in either a finite and short sequence or in an infinite sequence decaying geometrically, and thus, the weights are required to converge absolutely, as defined in (2.24).

$$\begin{array}{l}\textit{Sequence } (x_t),\ t = \ldots, -1, 0, 1, \ldots, \textit{ converges absolutely provided} \\ \textit{that series } s_t = |x_{-t}| + \ldots + |x_0| + \ldots + |x_t| \textit{ converges for } t \to \infty.\end{array} \tag{2.24}$$

When constructed with a sequence of weights that converges absolutely, the linear process, as defined below in (2.25), exists as a linear combination of white noise variables

$$\begin{array}{l}\textit{Let } (b_t) = (b_0, b_1, \ldots) \textit{ be a sequence that converges absolutely} \\ \textit{and } (W_t) \textit{ a white noise process with } \mu_W(t) = 0 \textit{ and } \sigma_W^2(t) = \sigma_W^2. \\ \textit{If } Y_t = \sum_{u=0}^{\infty} b_u W_{t-u}, \textit{ then } (Y_t) \textit{ is called a linear process.} \\ (V_t) \textit{ with an expectation function } \mu_V(t) \neq 0 \textit{ is a linear process} \\ \textit{on condition that } (Y_t) = (V_t - \mu_V(t)) \textit{ is a linear process.}\end{array} \tag{2.25}$$

because the infinite sum in (2.25) converges in mean square to a limiting random variable Y_t. For a proof proceed as demonstrated in Sect. 7.2.1: there, the convergence in mean square of the sum in (2.19) is shown, using the absolute convergence of the geometric sequence (a^u) for $|a| < 1$

$$\begin{aligned} \mathrm{E}(Y_t) &= \mathrm{E}\left(\sum_{u=0}^{\infty} b_u W_{t-u}\right) = \mu_W \sum_{u=0}^{\infty} b_u \\ c_Y(\tau) &= \sigma_W^2 \sum_{u=0}^{\infty} b_u b_{u+\tau} = \sigma_W^2 \sum_{u=0}^{\infty} b_u b_{\tau+u} \qquad \tau \geq 0 \\ \sigma_Y^2 &= c_Y(0) = \sigma_W^2 \sum_{u=0}^{\infty} b_u^2 \end{aligned} \tag{2.26}$$

and, consequently, the linear process is stationary. From (2.26), the expectation and covariance functions of the AR[1] and MA[1] models can be derived, both of these models being linear processes (Problem 2.10).

(b_t) in definition (2.25) is a one-sided sequence of weights. One-sided sequences are most useful for the construction of linear processes, as a time point t divides the time domain into past and future. It is a property of the

natural processes governing most variables observed in Geosciences, that the past influences the present (at time t), and that the future has no influence on the present. For instance, the present wind speed x_t in the turbulent air flow, analysed in Sect. 2.1, can be modelled, in Sect. 2.3.2, with the past x_{t-1} and a present realisation of a white noise variable.

Sometimes, a linear process is constructed with a two-sided sequence $(\ldots, b_{-1}, b_0, b_1, \ldots)$. In these alternative constructions, the sums from 0 to ∞ in (2.25) and (2.26) are replaced by sums from $-\infty$ to ∞. An example of such a two-sided linear process is given in Fig. 7.16.

However, a model generated from a white noise process with a one- or two-sided sequence is usually not suitable for a stochastic process with a parameter not being the time and possibly being multi-dimensional. For instance, a location given by its geographical coordinates (λ, ϕ), does not generally divide the surface of the earth into regions with specific properties. In this case, a rule (implicit or explicit) is used to indicate which points (on the surface) are neighbours, i.e., are in a (possibly hypothetical) spatial relationship as a framework for the construction of a model [41].

This section is summarised as follows. The example time series under analysis are the wind speed measurements in a turbulent air flow plotted in Fig. 2.1, the NAO indices plotted above, in Figs. 2.9 and 2.10, and the Basel temperature yearly values plotted above, in Fig. 2.13. A white noise process, as defined in (2.10), is fitted in Problem 2.2 as a model to the long-term NAO index (Jones), whereas models constructed from a white noise process are more suitable for the other example series: by trial and error, an AR[1] model is fitted to the wind speed series, and a MA[1] model to the differences of the Basel temperature yearly values.

When the simulated realisations are compared with an example time series, both guesses, regarding the form of the model (reflecting the dependences as seen in the plots of the time series and its correlation function) and the values for the model parameters, are evaluated at the same time. When a simulation is similar to the observations it is assumed that a suitable model has been surmised, however, when no similarity can be obtained, it is not known, which guess (the model or the parameter values, or both) is the source of the dissimilarity. When assessing these models by trial and error, no estimators for the model parameters with their statistical properties must be known. The need of little knowledge is an advantage with respect to the theory introduced so far, however, using trial and error, no attempt is made to find all suitable models or even the best model.

Given an example time series, how can a linear process be found as a suitable model? Since it is obvious that there are autoregressive and moving average models with two and more parameters (AR[p] and MA[q] models, or combined models, e.g., $X_t = a_1 X_{t-1} + a_2 X_{t-2} + W_t - b_1 W_{t-1}$ is an ARMA[2,1] model), the need for a systematic approach to the selection and identification of a model suitable for the observed time series arises.

When is a model a linear process and thus stationary? Assuming that a model is found to be in agreement with the observations, are computational procedures for the estimation of its parameters available? Can statistical properties of the estimates be derived? Answers to these questions are given in Chap. 5.

2.4 Convolution I

The linear process defined in (2.25) can be written as a convolution. Prior to giving examples in Sect. 2.4.4, it is shown in Sect. 2.4.1 that a convolution is a linear and time-invariant transformation and in Sects. 2.4.2 and 2.4.3, some properties of convolution sums are derived. The properties of the convolution become more clearly visible in the frequency domain as shown in Sect. 6.8.

2.4.1 Linear and Time-invariant Transformations

The first order linear differential equation $x(t) + ax'(t) = w(t)$ is solved in Problem 2.11. The stationary solution $x(t) = \int_{-\infty}^{t} g(t-u)w(u)\mathrm{d}u$, with $g(t) = 0$, for $t < 0$, and $g(t) = (1/a)\mathrm{e}^{-(1/a)t}$, for $t \geq 0$, is obtained. Using the same procedure, the stochastic first order linear difference equation of the AR[1] model, $X_t - aX_{t-1} = W_t$, as defined in (2.13), can be solved, i.e., if the following dissimilarities are accounted for:

- the time is integer, $t = \ldots, -1, 0, 1, \ldots$, not real
- the functions are random, not deterministic
- the known function is a white noise process.

Thus, the solution of $X_t - aX_{t-1} = W_t$ is the sum of the general solution of $X_t - aX_{t-1} = 0$ and of a particular solution of $X_t - aX_{t-1} = W_t$. $X_t = Ca^t$ is the general solution of $X_t - aX_{t-1} = 0$, since $Ca^t - aCa^{t-1} = 0$ for all t and arbitrary random variables or real constants C. $X_t = \sum_{u=-\infty}^{t} a^{t-u}W_u = \sum_{u=0}^{\infty} a^u W_{t-u}$ is a particular solution of $X_t - aX_{t-1} = W_t$, since $a^{t-t}W_t + \sum_{u=-\infty}^{t-1} a^{t-u}W_u - a\sum_{u=-\infty}^{t-1} a^{t-1-u}W_u = W_t$. Thus, the solution of the AR[1] difference equation, as given in (2.27), is obtained.

$$X_t = Ca^t + \sum_{u=0}^{\infty} a^u W_{t-u} \tag{2.27}$$

The solution in (2.27) contains the infinite sum in (2.18) which converges for $|a| < 1$, as proposed in (2.19).

In the solution (2.27) of the first order linear difference equation, C depends on the initialisation of the stochastic process (X_t), e.g., with $X_0 = 0$, $C = -\sum_{u=0}^{\infty} a^u W_{-u}$ is obtained. If $|a| < 1$, then Ca^t becomes identically zero for large t, and the solution reduces to $X_t = \sum_{u=0}^{\infty} a^u W_{t-u}$, being called the stationary solution of the first order linear difference equation.

$X_t = \sum_{u=0}^{\infty} a^u W_{t-u}$, $t = \ldots, -1, 0, 1, \ldots$, is a linear process, as defined in (2.25), since the sequence (a^u) converges absolutely with $|a| < 1$.

The stationary solution of the first order linear difference equation is an example of a convolution sum. This becomes obvious when the sequence of the weights is re-expressed: $\sum_{u=0}^{\infty} a^u W_{t-u} = \sum_{u=-\infty}^{\infty} b(u) W_{t-u}$, if $b(u) = (\ldots, 0, 0, a^0, a^1, a^2, \ldots)$ with $|a| < 1$. The stationary solution of the first order linear differential equation, $x(t) = \int_{-\infty}^{t} g(t-u)w(u)\mathrm{d}u$, with $g(t) = 0$, for $t < 0$, and $g(t) = (1/a)\mathrm{e}^{-(1/a)t}$, for $t \geq 0$, is an example of a convolution integral. Convolution sums and integrals are defined in (2.28):

1. $h(t) = \big(f(t)\big) * \big(g(t)\big) = f * g(t) = \int_{-\infty}^{\infty} f(u)g(t-u)\mathrm{d}u$ *is called the convolution of* $f(t)$ *and* $g(t)$*,* t *being a real number.*
2. $c_i = (a_i) * (b_i) = a * (b_i) = \sum_{j=-\infty}^{\infty} a_j b_{i-j}$ *is called the convolution of* (a_i) *and* (b_i)*,* i *being an integer number.*

(2.28)

Definitions (2.28) are not restricted to a specific type of sequences or functions. Usually, the sequences and functions are real-valued; however, stochastic processes as defined in (2.2) quite frequently occur in convolutions, e.g., in (2.27), where a white noise process is convolved with a real-valued sequence. The existence of a convolution sum or integral is dependent on the type of sequences or functions being convolved, e.g., with $|a| < 1$, the sum in (2.27) converges in mean square, as proposed in (2.19). Convolution sums obtained from real-valued sequences converge under the conditions formulated in (2.36).

What are the properties of convolution sums or integrals? It is recommended to start an investigation with definition (2.29).

Let a function x *(here without argument) be a mapping (a rule or device) which associates a value* $x(t)$ *in the domain of values, with each value* t *in the domain of definition. Both domains are often a subset of the real numbers. Now define* $\mathcal{L}$ *as a mapping which associates with each function* x *in a set of functions a new function* $\mathcal{L}(x)$*, with* x *and* $\mathcal{L}(x)$ *being in the same set of functions. In mathematics,* $\mathcal{L}$ *is called a transformation, transform or an operator, in engineering,* $\mathcal{L}$ *is known as a filter.*

(2.29)

For example, the derivative of a function is a transformation: it associates a new function $(\mathrm{d}/\mathrm{d}t)(x)$ with each x in a set of functions. The derivative is also a linear transformation, as defined in (2.30).

Let $x_1, x_2, \ldots$ *be functions and* $\mathcal{L}(x_1), \mathcal{L}(x_2), \ldots$ *transformations of* $x_1, x_2, \ldots$*; and let* $a_1, a_2, \ldots$ *be real constants. If* $\mathcal{L}(a_1x_1 + a_2x_2 + \ldots) = a_1\mathcal{L}(x_1) + a_2\mathcal{L}(x_2) + \ldots$*, then* $\mathcal{L}$ *is called linear.*

(2.30)

Examples of linear operators (transformations) are the differentiation $y = (\mathrm{d}/\mathrm{d}t)(x)$ and the integration $x = \int_0^t y(s)\mathrm{d}s$ of a real-valued function x of the

real argument t, if the derivative and the integral exist. Both, differentiation and integration, are to be found in an introductory course in differential and integral calculus.

Another example of a transformation is the shift operator, as defined in (2.31). It usually represents a time delay by τ, i.e., a lag τ.

The operator $y = \mathcal{S}(x, \tau)$, as defined in (2.29), τ a real number, is called a shift operator provided that $y(t) = x(t - \tau)$. (2.31)

A straightforward example of a shift operator is obtained with $x(t) = \cos(t)$ and $\tau = \pi$: $y(t) = \mathcal{L}\big(x(t), \tau\big) = \cos(t - \pi) = -\cos(t)$.

The shift operator is a linear transformation, as defined in (2.30). This proposition is derived in Problem 2.12.

Using the shift operator, as defined in (2.31), time-invariant operators (transformations) are defined in (2.32).

Let $\mathcal{S}$ be a shift operator as defined in (2.31), and $\mathcal{L}$ a transformation as defined in (2.29). If $\mathcal{L}(\mathcal{S}(x, \tau) = \mathcal{S}(\mathcal{L}(x), \tau)$, then $\mathcal{L}$ is said to be time-invariant. (2.32)

If a time-invariant transformation $\mathcal{L}$ is used together with a shift operator $\mathcal{S}$, then the result of the transformations does not depend on their order of succession, i.e., applying $\mathcal{S}$ and then $\mathcal{L}$ delivers the same result as applying $\mathcal{L}$ and then $\mathcal{S}$. For example, the differentiation of a function is a time-invariant transformation; the multiplication of a function with another function is, however, not a time-invariant transformation, as derived in Problem 2.12.

Another example of a linear transformation is the convolution integral $y(t) = \int_{-\infty}^{\infty} g(t - u)x(u)\mathrm{d}u$, as defined in (2.28). Is the convolution integral not only linear but also time-invariant? If $\mathcal{L}$ is the *integral transform* in (2.33), with *kernel* $g(t, u)$ and $x(t)$ being functions such that the integral exists, t and u real, then a linear transformation, as defined in (2.30), is obtained. The integral is shifted as defined in (2.31) to obtain the second line. In the third line, $\mathcal{L}$ is required to be time-invariant, as defined in (2.32).

$$y(t) = \mathcal{L}\big(x(t)\big) = \int_a^b g(t, u)x(u)\mathrm{d}u, \quad a = -\infty \text{ and } b = \infty \tag{2.33}$$

$$y(t - \tau) = \int_{-\infty}^{\infty} g(t - \tau, u)x(u)\mathrm{d}u = \mathcal{S}\big(\mathcal{L}(x(t)), \tau\big)$$

$$\mathcal{L}\big(\mathcal{S}(x, \tau)\big)(t) = \int_{-\infty}^{\infty} g(t, u)x(u - \tau)\mathrm{d}u \qquad u = v + \tau$$

$$= \int_{-\infty}^{\infty} g(t, v + \tau)x(v)\mathrm{d}v \qquad w = t - \tau$$

$$y(w) = \int_{-\infty}^{\infty} g(w + \tau, v + \tau)x(v)\mathrm{d}v \qquad g(w - v, v - v) = f(w - v)$$

$$y(w) = \int_{-\infty}^{\infty} f(w - v)x(v)\mathrm{d}v \tag{2.34}$$

With $u = v + \tau$ substituted for the integration variable, the fourth line is obtained, and the fifth results from the substitution $w = t - \tau$ [18]. When comparing the fifth with the first line, $g(t,u) = g(t+\tau, u+\tau)$ will follow, if the transformation $\mathcal{L}$ is time-invariant. Thus, demanding time invariance, the set of all functions $g(t,u)$ in the first line is reduced to those functions with $g(t,u) = g(t+\tau, u+\tau)$. Substituting $\tau = -v$ in $g(w+\tau, v+\tau)$, $g(w-v, v-v) = g(w-v, 0) = f(w-v)$ is obtained and the integral (2.34) follows.

Accordingly, if an integral transform $\mathcal{L}$ (2.33) is required to be a time-invariant integral of a product with two functions, then $\mathcal{L}$ is a convolution integral, as defined in (2.28). The result is that, both, a convolution integral and, plausibly, a convolution sum, are linear and time-invariant transformations.

The convolution integral $x(t) = \int_{-\infty}^{t} g(t-u)w(u)\mathrm{d}u$, t, u real, with $g(t) = 0$, for $t < 0$ and $g(t) = (1/a)\mathrm{e}^{-(1/a)t}$, for $t \geq 0$, is the stationary solution of the linear differential equation $x(t) + ax'(t) = w(t)$. The convolution sum $X_t = \sum_{u=0}^{\infty} a^u W_{t-u} = \sum_{u=-\infty}^{\infty} a^u W_{t-u}$, with $a^u = \ldots, 0, 0, a^0, a^1, a^2, \ldots$, t, u an integer, is the stationary solution of the difference equation $X_t - aX_{t-1} = W_t$ in the AR[1] model. These results are derived in Problem 2.11 and in the remarks to (2.27). Therefore, and because of the convolution being time-invariant, the stationary solution of the linear differential equation $x(t)$ is displaced to $x(t-\tau)$, if the known function $w(t)$ changes its position to $w(t-\tau)$, and the stationary AR[1] model (X_t) is displaced to $(X_{t-\tau})$, if the white noise process (W_t) changes its position to $(W_{t-\tau})$.

As a third property, besides being linear and time-invariant, a convolution integral or sum are both commutative. In the case of the convolution integral, a proof is straightforward, because the arguments of the functions being integrated sum up to a constant $(= t)$. If the integral on the left side in (2.35) exists, with the substitutions $t - u = s$, $u = t - s$, $(\mathrm{d}u)/(\mathrm{d}s) = -1$, $\mathrm{d}u = -1\mathrm{d}s$, the upper bound $u = -\infty$ becomes $s = +\infty$, and the lower bound $u = +\infty$ becomes $s = -\infty$, and thus, the desired result (2.35) is obtained:

$$\begin{aligned}\int_{-\infty}^{+\infty} f(u)\, g(t-u)\mathrm{d}u &= \int_{+\infty}^{-\infty} f(t-s)g(s)(-1\mathrm{d}s) \\ &= -\int_{+\infty}^{-\infty} f(t-s)g(s)\mathrm{d}s = \int_{-\infty}^{+\infty} f(t-s)g(s)\mathrm{d}s \qquad (2.35)\end{aligned}$$

In the case of the convolution sum, however, its commutativity is derived using a procedure similar to the one applied in Problem 2.4 for the proof that the convolution sums in (2.18) are equal.

In contrast to the convolution sum $c_i = (a_i) * (b_i) = \sum_{j=-\infty}^{\infty} a_j b_{i-j} = \sum_{j=-\infty}^{\infty} b_j a_{i-j} = (b_i) * (a_i)$, the sum $c_i = (a_i) \star (b_i) = \sum_{j=-\infty}^{\infty} a_j b_{i+j}$ is not commutative. This is shown by a counter-example using two short sequences. The $\star$-operator is defined in (6.102). For example, $c_Y(\tau)/\sigma_W^2 = \sum_{u=0}^{\infty} b_u b_{\tau+u}$ in (2.26) is a $\star$-transformation.

Convolution sums of two real-valued sequences, or of a real-valued sequence and a stochastic process with integer parameter, occur quite often in the study of times series and stochastic processes. Besides being linear, time-invariant and commutative, as shown above, these convolution sums have the properties which are derived in the following Sects. 2.4.2, 2.4.3 and 2.4.4.

2.4.2 Existence and Calculation of Convolution Sums

A convolution sum, as defined in (2.28) exists if both, possibly infinite, sequences to be convolved decay "fast enough" such that their convolution is a finite real number for each t. For example, a sequence decays "fast enough" if it converges absolutely, as defined in (2.24). Since absolutely convergent sequences remain as such even if their values are re-ordered in any possible way, and since sums of re-ordered sequences remain absolutely convergent (on condition that the original sequence is absolutely convergent), proposition (2.36) is obtained.

Let the real-valued sequence $(c_i) = \sum_{j=-\infty}^{\infty} a_j b_{i-j}$ *be the convolution of the real-valued sequences* (a_i) *and* (b_i).
1. (c_i) *converges absolutely on condition that both sequences,* (a_i) *and* (b_i)*, converge absolutely.*
2. (c_i) *exists, provided that at least one of both* (a_i) *and* (b_i) *converges absolutely and the other one is bounded.*

(2.36)

The calculation of convolution sums is demonstrated with the following examples. All example convolution sums exist, as the sequences convolved contain a finite number of non-zero elements. These sequences are convolved in conjunction with zero-padding, i.e., zeros being appended at both ends. Examples with an infinite number of non-zero elements are given in Sect. 2.4.3.

A convolution sum is calculated as demonstrated in (2.37) where the convolution sums of two short example sequences, e.g., $(a_i) = (a_1, a_2, a_3, a_4)$ and $(b_i) = (b_1, b_2)$ are calculated, as defined in (2.28,2), assuming $b_i = 0$ for $i < 1$ and $i > 2$ and $a_i = 0$ for $i < 1$ and $i > 2$, i.e., zero-padded sequences.

$$\begin{aligned}
c_1 &= \ldots + a_0 b_1 + a_1 b_0 + a_2 b_{-1} + \ldots &&= 0 \\
c_2 &= \ldots + a_0 b_2 + a_1 b_1 + a_2 b_0 + \ldots &&= a_1 b_1 \\
c_3 &= \ldots + a_0 b_3 + a_1 b_2 + a_2 b_1 + a_3 b_0 + \ldots &&= a_1 b_2 + a_2 b_1 \\
c_4 &= \ldots + a_0 b_4 + a_1 b_3 + a_2 b_2 + a_3 b_1 + \ldots &&= a_2 b_2 + a_3 b_1 \\
c_5 &= \ldots + a_2 b_3 + a_3 b_2 + a_4 b_1 + \ldots &&= a_3 b_2 + a_4 b_1 \\
c_6 &= \ldots + a_2 b_4 + a_3 b_3 + a_4 b_2 + \ldots &&= a_4 b_2 \\
c_7 &= \ldots + a_2 b_5 + a_3 b_4 + a_4 b_3 + \ldots &&= 0
\end{aligned} \tag{2.37}$$

Short sequences are easily convolved using paper, pencil and a calculator by means of the paper strip device as demonstrated in (2.38). The paper strip

device is found in [18]. When calculating using this method, the values in (a_i) are written in a column on the left margin of a sheet, whereas the values in (b_i) are written on another paper strip in reversed order, from bottom to top. The order of the values in (b_i) is reversed to account for b_{i-j} in (2.28,2), j the summation variable.

If the values in (a_i) and (b_i) are written on equi-spaced lines, pairs of values (a_i, b_i) appearing on the same line are obtained when the paper strip is moved stepwise downwards. At each step, the products $a_i b_i$, a_i and b_i being on the same line, are added, and thus the convolution sum is obtained. Four example positions of (b_i) relative to (a_i) are given in (2.38), where the short sequences used for the demonstration (2.37) are convolved using the paper strip device.

$$
\begin{array}{lccccccccl}
(a_i) & i=1 & & i=2 & & i=3 & \ldots & i=6 & & (c_i) \\
\ldots & b_2 & & & & & \ldots & & & \ldots \\
0 & b_1 & & b_2 & & & \ldots & & \longrightarrow & c_1 = 0b_1 \ + \ 0b_2 = 0 \\
a_1 & & & b_1 & & b_2 & \ldots & & \longrightarrow & c_2 = a_1 b_1 \ + \ 0b_2 = a_1 b_1 \\
a_2 & & & & & b_1 & \ldots & & \longrightarrow & c_3 = a_2 b_1 + a_1 b_2 \\
a_3 & & & & & & \ldots & & & \ldots \\
a_4 & & & & & & \ldots & b_2 & & \ldots \\
0 & & & & & & \ldots & b_1 & \longrightarrow & c_6 = 0b_1 \ + \ a_4 b_2 = a_4 b_2 \\
\ldots & & & & & & \ldots & & & \ldots
\end{array}
\tag{2.38}
$$

Calculating a convolution sum as demonstrated in (2.38) is very instructive: e.g., a moving average is obtained when the sequences $(a_k) = (1.7, 1.5, 1.8, 2.0)$, $k = 1, 2, 3, 4$, and $(b_j) = (0.5, 0.5)$, $j = 1, 2$, are convolved:

$$
\begin{array}{lccccccl}
(a_k) & i=1 & i=2 & i=3 & \ldots & i=6 & & (c_i) \\
\ldots & 0.5 & & & \ldots & & & \ldots \\
0 & 0.5 & 0.5 & & \ldots & & \longrightarrow & c_1 = 0 \times 0.5 \ + \ 0 \times 0.5 = 0 \\
1.7 & & 0.5 & 0.5 & \ldots & & \longrightarrow & c_2 = 1.7 \times 0.5 \ + \ 0 \times 0.5 = 0.85 \\
1.5 & & & 0.5 & \ldots & & \longrightarrow & c_3 = 1.5 \times 0.5 \ + \ 1.7 \times 0.5 = 1.6 \\
1.8 & & & & \ldots & & & \ldots \\
2.0 & & & & \ldots & 0.5 & & \ldots \\
0 & & & & \ldots & 0.5 & \longrightarrow & c_6 = 0 \times 0.5 \ + \ 2.0 \times 0.5 = 1 \\
\ldots & & & & \ldots & & & \ldots
\end{array}
$$

Long sequences are convolved using `filter()` with argument `method = "convolution"`. For example, R expressions

```
a <- c(0,0,1.7,1.5,1.8,2.0,0,0)
b <- c(0.5,0.5)
c1 <- as.vector(filter(a,b,method="convolution",sides=1))
> c1 [1] NA, 0.0, 0.85, 1.60, 1.65, 1.90, 1.00, 0.00
```

compute the convolution sums of the example sequences convolved above with the paper strip device. In this example,

```
t <- -0:7     #indices of the convolution sum as defined in (2.28)
```

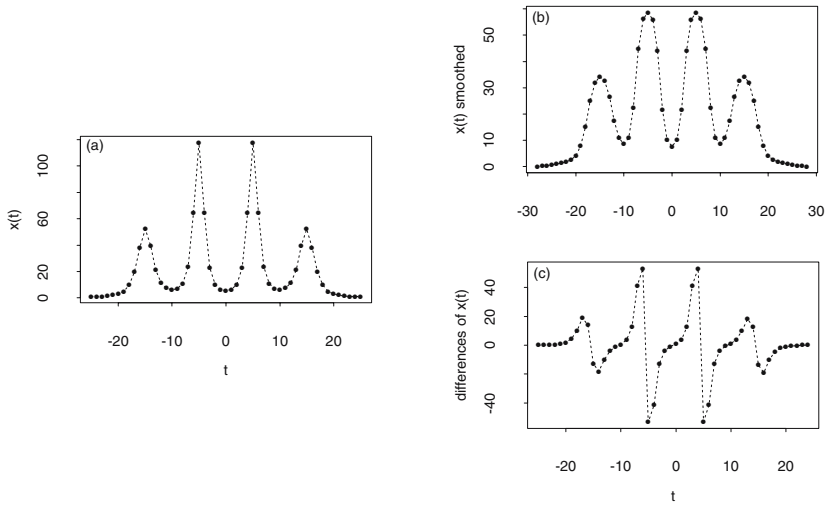


Fig. 2.16. The sequence (x_t) (on the left, a) is convolved twice: (i) with a moving average to obtain the smoothed sequence on the right (above, b), and (ii), with $(1, -1)$ to obtain the sequence of its differences on the right (below, c).

generates an R vector with indices as defined in (2.28).

R vectors containing the sequences to be convolved are usually zero-padded using `c()`. For example, the sequence `x`, containing $N = 51$ values as plotted on the left in Fig. 2.16, is zero-padded prior to being convolved to the smooth sequence on the right, above, in Fig. 2.16.

```
#x can be generated with the R expression in /path/runmeandif.R.
xpad <- rep(0,5)                #zeroes for padding
x <- c(xpad,x)                  #on the left
x <- c(x,xpad)                  #on the right
g <- c(0.2,0.2,0.2,0.2,0.2)     #moving average
xsmooth <- as.vector(filter(x,g,method="convolution",sides=2)
tsmooth <- -30:30               #indices for xsmooth
```

When using R function `filter(x,g,...,)` with argument `..., sides=2, ...`, the resulting sequence is displaced by τ, τ being half the length of R vector `g`. Both at the start of the smoothed sequence and at its end, two values are missing, because they cannot be calculated. This convolution is an example of a moving average.

Both, moving averages and differences of time series, are quite often calculated as convolution sums using definitions (2.39). Note that, when smoothing a sequence (x_t) as defined in (2.39,1), the resulting convolution $y(t)$ becomes smoother with increasing *width* $T_2 - T_1$ of (f_t), i.e., with an increasing number of non-zero values in (f_t), and thus, the form of (y_t) is strongly influenced by the width of (f_t). Additional possibilities to describe the form (wide, narrow,

etc.) of a sequence or a function are introduced in Sect. 6.6.

Let (x_t) be a time series, (f_t) a sequence, and $T_1 < T_2$.
1. *If $f_t = 0$ for $t \le T_1$ and $T_2 \le t$ as well as $\sum_{t=-\infty}^{\infty} f_t = 1$, then $(y_t) = (x_t) * (f_t)$ is called a moving average of (x_t). Often, (f_t) is chosen with identical values and with centre in $t = 0$: $f_t = 1/(2T + 1)$, for $-T \le t \le T$, $f_t = 0$, for all other t.*
2. *If $f_0 = 1$, $f_1 = -1$ and $f_t = 0$ for all other t, then $(y_t) = (x_t) * (f_t) = x_t - x_{t-1} = \Delta x_t$ is called the differences of (x_t).*

(2.39)

Often, a time series is smoothed to remove fluctuations within short periods. In the pyranometer daily means obtained in the Zurich-Reckenholz comparison experiment and plotted in Fig. 1.4, the seasonal cycle becomes more clearly visible, in Fig. 1.7, after the time series has been convolved to its 11-day moving average. As a second example, the NAO indices in Figs. 2.9 and 2.10 are smoothed to their 11-year moving average in order to see the decadal fluctuations in these time series more easily.

Thus, when a time series (x_t) is smoothed to (y_t) by convolving with a weight sequence (f_t), as defined in (2.39,1), only fluctuations with periods larger than the width $T_2 - T_1$ of (f_t) remain in (y_t). Hence, those fluctuations become (more easily) visible that were previously (at least partly) hidden by fluctuations with periods shorter than the width of (f_t).

Differences, as defined in (2.39,2), are obtained convolving with $\Delta^1 = (f_0 = 1,\ f_1 = -1)$; differences of differences, i.e., second and higher order differences are obtained convolving with $\Delta^2 = (f_0 = 1,\ f_1 = -2,\ f_2 = 1)$, $\Delta^3 = (f_0 = 1,\ f_1 = -3,\ f_2 = 3,\ f_3 = 1)$, ..., Δ^d.

$$\begin{aligned} \Delta^1(x_t) &= x_t - x_{t-1} \\ \Delta^2(x_t) &= \Delta^1(x_t) - \Delta^1(x_{t-1}) \ = \ x_t - 2x_{t-1} + x_{t-2} \\ &\dots = \\ \Delta^d(x_t) &= \Delta^{d-1}(x_t) - \Delta^{d-1}(x_{t-1}) \end{aligned} \tag{2.40}$$

For example, the sequence on the left, in Fig. 2.16, is convolved to its moving average and its differences on the right, in Fig. 2.16.

Another example for a convolution as defined in (2.39,2), is the first order differences of the yearly values in the Basel temperature series plotted above, in Figs. 2.14 and 2.15. These differences are obtained with `diff()` or with `filter()`

```
#tsbtemp contains the Basel yearly temperature values
#with diff(), differences as defined in (2.39,2), are obtained
diftsbtemp <- diff(tsbtemp,lag=1,differences=1)
#or explicitely as convolution, tsbtemp not zero-padded
diftsbtemp <- filter(tsbtemp,c(1,-1),method="convolution",sides=2)
```

from the Basel yearly temperature values being plotted in Fig. 2.13. This example demonstrates that a time series with a constant expectation function

can be obtained from a time series not being stationary in the first moment, when the first order differences are calculated. More examples of time series with stationary differences are given in Sect. 5.4.1.

The introduction of the Fourier transform in Sect. 6.3 provides a means for a more systematic approach, in Sect. 6.8.2, to the properties of differences and moving averages defined in (2.39).

2.4.3 Inverse Sequences

A convolution integral is shown to be, in (2.34) and (2.35), a linear, time-invariant and commutative transformation, and, plausibly, also a convolution sum is linear, time-invariant and commutative. Under this transform, does an identity sequence exist? If it does, what are the conditions for the existence of an inverse sequence?

1. $(e_i) = (\ldots, 0, e_0 = 1, 0, \ldots)$ *is the identity sequence, since for any sequence* (a_i)*:* $(a_i) * (e_i) = (e_i) * (a_i) = (a_i)$.
2. *If* $(a_i)^{-1} * (a_i) = (e_i)$*, then* $(a_i)^{-1}$ *is the inverse of* (a_i).
3. *There are two inverses. If* $(a_i)^{-1} = 0$ *for* $i > 0$*, then* $(a_i)^{-1}$ *is the left-inverse of* (a_i)*, if* $(a_i)^{-1} = 0$ *for* $i < 0$*, then* $(a_i)^{-1}$ *is the right-inverse of* (a_i). (2.41)
4. *Subsequently, in this book,* $(a_i)^{-1}$ *is always the right-inverse.*

When convolving the sequences in (2.42) with their left- (L.) and right-inverses (R.) using the paper strip device (2.38), the identity sequence is obtained.

$$
\begin{aligned}
(\ldots, 0, -a, \underset{\substack{\uparrow \\ i=0}}{1}, 0, 0, \ldots)^{-1} &= \begin{cases} (\ldots, a^3, a^2, a, \underset{\substack{\uparrow \\ i=0}}{1}, 0, \ldots) & \text{L.} \\ (\ldots, 0, \underset{\substack{\uparrow \\ i=0}}{0}, -1/a, -1/a^2, -1/a^3, \ldots) & \text{R.} \end{cases} \\
(\ldots, 0, \underset{\substack{\uparrow \\ i=0}}{1}, -b, 0, 0, \ldots)^{-1} &= \begin{cases} \ldots, -1/b^3, -1/b^2, -1/b, \underset{\substack{\uparrow \\ i=0}}{0}, 0, \ldots) & \text{L.} \\ (\ldots, 0, \underset{\substack{\uparrow \\ i=0}}{1}, b, b^2, b^3, \ldots) & \text{R.} \end{cases}
\end{aligned}
\tag{2.42}
$$

An inverse (i.e., a right-inverse, reconcilable with convention (2.41,4)) of a sequence can be obtained with a slightly modified paper strip device (2.38): the sequence (a_i) is written in the left column on the sheet, the identity sequence (e_i) in the right column, both in normal order from top to bottom. On the paper strip, the values in $(a_i)^{-1}$ are obtained in reverse order, when the strip is moved downwards step by step with "start" in the initial position, where the first non-zero value in the inverse is calculated as $a_0^{-1} = 1/a_0$, since $a_0 a_0^{-1} = n_0 = 1$. For instance, in Problem 2.18, the sequences on the right

side in (2.43) are obtained by inverting the sequences on the left side using the modified paper strip device.

$$\begin{aligned}(1,1)^{-1} &= (1,-1,1,-1,\ldots)\\ (1,1,1)^{-1} &= (1,-1,0,1,-1,0,\ldots)\\ (1,2,1)^{-1} &= (1,-2,3,-4,5,-6,\ldots)\\ (1,2,3,4,5,\ldots)^{-1} &= (1,-2,1) = (1,-1)*(1,-1)\\ (1,-a)^{-1} &= (1,a,a^2,a^3,\ldots)\end{aligned} \tag{2.43}$$

The inverse $(a_i)^{-1}$ of a sequence (a_i) is usually found with a polynomial division as being inverse to polynomial multiplication. When two polynomials are multiplied, the coefficients of the resulting polynomial are obtained as the convolution sum of the coefficients of the original polynomials, as shown in Problem 2.20 and proposed in (2.44).

If (a_i), $i = 0,\ldots,p$, is the sequence of the coefficients of a polynomial in x, x real, $A(x) = a_0 + a_1x^1 + \ldots + a_px^p$ and (b_i), $i = 0,\ldots,q$, the sequence of the coefficients of a polynomial $B(x) = b_0 + b_1x^1 + \ldots + b_qx^q$, then $(c_i) = (a_i) * (b_i)$ is the sequence of the coefficients of $C(x) = A(x)B(x)$. (2.44)

From (2.44), $A(x)A^{-1}(x) = 1$ is obtained, $A(x)$ being the polynomial pertaining to (a_i), $A^{-1}(x)$ the polynomial pertaining to $(a_i)^{-1}$, and 1 the polynomial pertaining to the identity sequence (e_i). Consequently, $A^{-1}(x) = 1/A(x)$, and thus the coefficients of $A^{-1}(x)$ can be obtained using polynomial division. If, for example, $(1,-a)^{-1}$ is calculated with a polynomial division

$$\begin{aligned}(a_i) &= (1,-a) & A(x) &= 1-ax\\ (a_i)^{-1} &= (1,a,a^2,a^3,\ldots) & A^{-1}(x) &= \frac{1}{1-ax} = 1+ax+(ax)^2+(ax)^3+\ldots\end{aligned}$$

the result obtained is identical with the last line in (2.43). The sequence $(1,-a)^{-1}$ converges absolutely, if $|a| < 1$, its partial sums being a geometric series. The root of $A(x) = 1 - ax$ is $x_0 = 1/a$, and thus, $|a| < 1$, only if $x_0 < -1$ or if $1 < x_0$. Consequently, if the root x_0 of $A(x)$, $A(x)$ being the polynomial pertaining to the sequence $(1,-a)$, is not in the interval $-1 \leq x \leq 1$, then the inverse sequence $(1,-a)^{-1}$ converges absolutely.

A polynomial of degree n has n complex roots, some of which may be degenerate, and which may be found with a polynomial factorisation into factors of degree 1. Hence, an inverse to a sequence is found to either converge, or not to converge, using the characteristic polynomial, as defined in (2.45).

The characteristic polynomial of a real sequence (a_k) is a polynomial in z, the elements in the sequence becoming the coefficients of the polynomial: $A(z) = \sum_k a_k z^k$; z, $A(z)$ complex. (2.45)

When using the characteristic polynomial to check whether the inverse of a sequence converges, the unit circle $|z| \leq 1$ in the complex plane replaces the real interval $-1 \leq x \leq 1$ used to show the convergence of $(1, -a)^{-1}$ above.

If the roots of the characteristic polynomial $A(z)$ of a sequence (a_i) are not in the unit circle, then the inverse $(a_i)^{-1}$ to (a_i) converges absolutely (without proof). (2.46)

The roots of a characteristic polynomial are calculated with `polyroot()`, as demonstrated in Problem 2.21.

2.4.4 Examples: AR[1] and AR[2] Models

When calculating the roots of the characteristic polynomial, as proposed in (2.46), the stationary solution of the AR[1] difference equation, as defined in (2.13), is obtained with the following calculation. $A(z) = 1 - az$ is the characteristic polynomial pertaining to the weights of the AR[1] model (X_t) with difference equation $X_t - aX_{t-1} = W_t$ or, when written as convolution, $(1, -a) * (X_t) = (W_t)$. Convolving the difference equation with $(1, -a)^{-1}$, $(X_t) = (1, -a)^{-1} * (W_t)$ is obtained. The root of $A(z) = 1 - az$ is $z = 1/a$. If $|z| > 1$, i.e., if z is outside the unit circle, $|a| < 1$, and with $(1, -a)^{-1} = (1, a, a^2, a^3, \ldots)$, as calculated in (2.43),

$$X_t = \sum_{u=0}^{\infty} a^u W_{t-u} \tag{2.47}$$

the stationary solution of the first order linear difference equation is obtained. This result is also derived in the remarks to (2.16) and, alternatively, in the remarks to (2.27). The convolution sum in (2.47) converges in mean square, as proposed in (2.19). Consequently, an AR[1] model is stationary for $|a| < 1$ and $t \to \infty$.

It is obvious that there are autoregressive models with more than one parameter, i.e., AR[p] models, as defined in (5.2). Of these models, only the second order autoregressive process is defined here, with its difference equation being another example of a convolution sum.

If (W_t) is a white noise process with $\mu_W = 0$ and σ_W^2, then $X_t = a_1 X_{t-1} + a_2 X_{t-2} + W_t$ is called an AR[2] process or AR[2] model; and $X_t - a_1 X_{t-1} - a_2 X_{t-2} = W_t$ or, as a convolution sum, $(1, -a_1, -a_2) * (X_t) = W_t$ is called its difference equation. (V_t) with expectation function $\mu_V(t) \neq 0$ is an AR[2] model, on condition that $(X_t) = (V_t - \mu_V(t))$ is an AR[2] model. (2.48)

The solution of the difference equation of the AR[2] model is given in (2.49), with z_1 and z_2 being the roots of $A(z) = 1 - a_1 z - a_2 z^2$, $A(z)$ the characteristic polynomial of the sequence $(1, -a_1, -a_2)$.

$$X_t = C_1 \left(\frac{1}{z_1}\right)^t + C_2 \left(\frac{1}{z_2}\right)^t + \sum_{u=0}^{\infty} b_u W_{t-u} \qquad (b_u) = (1, -a_1, -a_2)^{-1} \quad (2.49)$$

The AR[2] model is stationary with $(X_t) = (1, -a_1, -a_2)^{-1} * (W_t)$, if the solutions z_1 and z_2 of $A(z) = 1 - a_1 z - a_2 z^2 = 0$ are not in the unit circle, i.e., $1 < |z_1, z_2|$, and if $t \to \infty$. Under these conditions, $C_1(1/z_1)^t$ and $C_2(1/z_2)^t$ become identically zero, $(b_u) = (1, -a_1, -a_2)^{-1}$ converges absolutely and the convolution sum converges in mean square, as defined in (7.11), if the moments of the W_t exist.

From the stationarity condition $1 < |z_1, z_2|$, $A(1) = -a_2 - a_1 + 1 > 0$ and $A(-1) = -a_2 + a_1 + 1 > 0$ are obtained, and further, the equations $a_2 = -a_1 + 1$ and $a_2 = a_1 + 1$ for both sides in the triangle plotted in Fig. 2.17 are arrived at. Consequently, for an AR[2] model with coefficients (a_1, a_2) in the region below these sides, stationary solutions of the characteristic polynomial are obtained.

Substituting $a_1 = a_2 - 1$ and/or $a_1 = 1 - a_2$ in $1 < |(a_1 \pm (a_1^2 + 4a_2)^{1/2})/(-2a_2)|$, $|1/a_2| > 1$ and $|a_2| < 1$ are obtained, i.e., the equations for the third side of the triangle in Fig. 2.17. Thus, an AR[2] model $X_t = a_1 X_{t-1} + a_2 X_{t-2} + W_t$, as defined in (2.48), is stationary, if its coefficients (a_1, a_2) are within its region of convergence, as defined in (2.50),

$$a_2 < -a_1 + 1 \quad \text{and} \quad a_2 < a_1 + 1 \quad \text{and} \quad |a_2| < 1 \qquad (2.50)$$

i.e., within the triangle as plotted in Fig. 2.17.

From $a_1^2 + 4a_2 \geq 0$, it is concluded that all (a_1, a_2), being on and above the curve in the triangle in Fig. 2.17, deliver real roots of the characteristic polynomial, and that all (a_1, a_2) below the curve yield complex roots.

With the coefficients (a_1, a_2) being within the triangle, in Fig. 2.17, the stationary solution $X_t = (1, -a_1, -a_2)^{-1} * W_t = \sum_{u=0}^{\infty} b_u W_{t-u}$ of the difference equation of the AR[2]-model is obtained, as (b_u) with $b_0 = 1$ converges absolutely. From this solution, the moment functions of the stationary AR[2] model can be calculated.

$\mathrm{E}X_t = \mathrm{E}\sum_{u=0}^{\infty} b_u W_{t-u} = \mathrm{E}W_t \sum_{u=0}^{\infty} b_u = \mu_W \sum_{u=0}^{\infty} b_u$, and thus, the expectation function μ_X of the AR[2] model (X_t) is the product of the expectation $\mathrm{E}W_t = \mu_W$ of the generating white noise process and the sum of the inverse $(1, -a_1, -a_2)^{-1}$ of the sequence of the weights in the model. The very same result is obtained from the remarks to (2.15) in the case of the stationary AR[1] model. Hence, in both cases, $\mu_W = 0$ implies $\mathrm{E}X_t = 0$.

Less straightforward than the derivation of the expectation function is the derivation of the covariance function. The stationary solution $X_t = W_t + \sum_{u=1}^{\infty} b_u W_{t-u}$ of the AR[2] difference equation is written for the model variable in lag τ, $X_{t-\tau} = W_{t-\tau} + \sum_{u=1}^{\infty} b_u W_{t-\tau-u}$, from whence, with $b_0 = 1$ and $b_u = 0$ for $u < 0$, $X_{t-\tau} = W_{t-\tau} + \sum_{u=\tau+1}^{\infty} b_{u-\tau} W_{t-u}$ is obtained. This equation is multiplied with the AR[2] difference equation

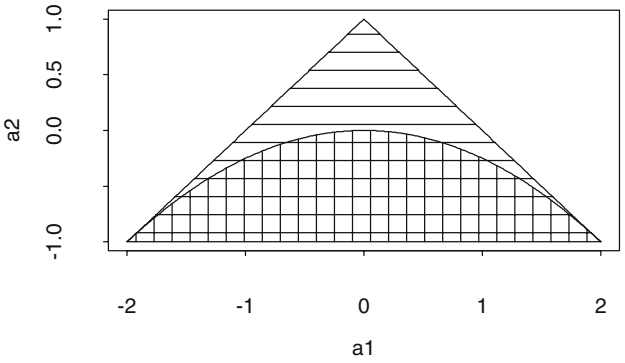


Fig. 2.17. The coefficients (a_1, a_2) of a stationary AR[2] model are within the triangle, as defined in (2.50): with real roots of the characteristic polynomial above $a_2 = -(1/4)a_1^2$ and with complex roots below.

$$(X_t - a_1 X_{t-1} - a_2 X_{t-2}) X_{t-\tau} = W_t \Big(W_{t-\tau} + \sum_{u=\tau+1}^{\infty} b_{u-\tau} W_{t-u} \Big)$$

and the expectations

$$\begin{aligned} &\mathrm{E}(X_t X_{t-\tau}) - a_1 \mathrm{E}(X_{t-1} X_{t-\tau}) \\ &\qquad - a_2 \mathrm{E}(X_{t-2} X_{t-\tau}) = \mathrm{E}(W_t W_{t-\tau}) + \sum_{u=\tau+1}^{\infty} b_{u-\tau} \mathrm{E}(W_t W_{t-u}) \end{aligned}$$

are then calculated. With $\mu_X = 0$, as obtained above from $\mu_W = 0$ in definition (2.48), the expectations of the products become covariances. The covariances in the infinite sum are identically zero because of (2.10,2), and, substituting $\tau = 0, 1, 2, \ldots$, the covariances and correlations in (2.51) are obtained:

$$\begin{aligned} c_X(0) - a_1 c_X(1) - a_2 c_X(2) &= \sigma_W^2 \\ \sigma_X^2 \big(\varrho_X(0) - a_1 \varrho_X(1) - a_2 \varrho_X(2) \big) &= \sigma_W^2 \\ c_X(1) - a_1 c_X(0) - a_2 c_X(1) &= 0 \\ \sigma_X^2 \big(\varrho_X(1) - a_1 \varrho_X(0) - a_2 \varrho_X(1) \big) &= 0 \\ c_X(2) - a_1 c_X(1) - a_2 c_X(0) &= 0 \\ \sigma_X^2 \big(\varrho_X(2) - a_1 \varrho_X(1) - a_2 \varrho_X(0) \big) &= 0 \end{aligned} \tag{2.51}$$

using the symmetry $c_X(\tau) = c_X(-\tau)$, as proposed in (2.9,4). Above, in equation $\sigma_X^2(\varrho_X(1) - a_1 \varrho_X(0) - a_2 \varrho_X(1)) = 0$ with the correlation for $\tau = 1$, $\varrho_X(0) = 1$ is substituted and, consequently, $\varrho_X(1) - a_1 - a_2 \varrho_X(1) = 0$ and $\varrho_X(1)$ are obtained. $\varrho_X(2)$ and σ_X^2 are calculated with further substitutions.

$$\begin{aligned}
\varrho_X(1) &= \frac{a_1}{1-a_2} \\
\varrho_X(2) &= \frac{a_1^2}{1-a_2} + a_2 \;=\; \frac{a_1^2 + a_2(1-a_2)}{1-a_2} \\
\sigma_X^2 &= \frac{\sigma_W^2(1-a_2)}{1-a_2-a_1^2-a_1^2a_2-a_2^2-a_2^3} \\
&= \frac{\sigma_W^2(1-a_2)}{(1+a_2)(1+a_1-a_2)(1-a_1-a_2)}
\end{aligned} \tag{2.52}$$

In (2.51), the general equation for the covariances and correlations for $\tau > 2$ is omitted. This equation is a linear difference equation $\varrho_X(\tau) - a_1\varrho_X(\tau - 1) - a_2\varrho_X(\tau - 2) = 0$ for $\tau > 0$ with initial conditions $\varrho_X(0) = 1$ and $\varrho_X(1) = a_1/(1-a_2)$. With continued substitutions, $\varrho_X(\tau)$, $\tau > 2$, follow, e.g., $\varrho_X(3) = a_1(a_1^2 + 2a_2 - a_2^2)/(1 - a_2)$. This difference equation has stationary solutions in the triangular convergence region, as defined in (2.50) and plotted in Fig. 2.17, since the characteristic polynomial $A(z) = 1-a_1z-a_2z^2$ pertains to both difference equations: the stochastic one, $X_t - a_1X_{t-1} - a_2X_{t-2} = W_t$, of the AR[2] model as defined in (2.48), and the deterministic one, $\varrho_X(\tau) - a_1\varrho_X(\tau - 1) - a_2\varrho_X(\tau - 2) = 0$, regarding the correlations.

If $a_1^2 + 4a_2 \geq 0$, then z_1 and z_2 are real and $|\varrho_X(\tau)|$ damps out exponentially, with $\varrho_X(\tau)$ possibly alternating in sign. If $a_1^2 + 4a_2 < 0$, then z_1 and z_2 are complex and $\varrho_X(\tau)$ is a function that oscillates with variable amplitude (Problem 2.21). An AR[2] model with this property is called pseudo-periodic. A pseudo-periodic AR[2] process and its duplicate in lag np, n an integer number and p the period of the pseudo-periodic behaviour, show quite large correlations for small n; when n becomes larger, however, the correlations decrease. A realisation of a pseudo-periodic AR[2] process is similar to a sinusoidal oscillation with a variable amplitude, i.e., an oscillation that always recovers, when its amplitudes have decreased, and thus never damps out. For instance, the realisations plotted in Fig. 2.4 stem from a pseudo-periodic AR[2] model.

In (2.51), the equations $\varrho_X(1) - a_1\varrho_X(0) - a_2\varrho_X(1) = 0$ and $\varrho_X(2) - a_1\varrho_X(1) - a_2\varrho_X(0) = 0$ are the linear difference equations of the AR[2] correlation function $\varrho_X(\tau)$ for the lags $\tau = 1$ and $\tau = 2$. From these equations $\varrho_X(1) = a_1 + a_2\varrho_X(1)$ and $\varrho_X(2) = a_1\varrho_X(1) + a_2$ are obtained, and, when using double-indexed coefficients in definition (2.53), i.e., a_1 and a_2 replaced by a_{12} and a_{22},

$$\begin{aligned}
\varrho_X(1) &= a_{12} + a_{22}\varrho_X(1) \\
\varrho_X(2) &= a_{12}\varrho_X(1) + a_{22}
\end{aligned} \tag{2.53}$$

the Yule-Walker equations of the AR[2] model are obtained. In these equations, the second index of the coefficients is the order of the process. The Yule-Walker equations can be solved for the model coefficients, i.e., $a_{12} = \varrho_X(1)(1 - \varrho_X(2))/(1 - \varrho_X^2(1))$ and $a_{22} = (\varrho_X^2(1) - \varrho_X(2))/(\varrho_X^2(1) - 1)$.

These solutions are known as the Yule-Walker estimates of the coefficients in an AR[2] model, after estimates $\hat{\varrho}_X(\tau)$ have been substituted for the model correlations $\varrho_X(\tau)$. The $\hat{\varrho}_X(\tau)$ are computed as empirical correlations from a time series, using definition (2.1,4).

The AR[1] model as defined in (2.13) and the AR[2] model as defined in (2.48) are special cases of the AR[p] process, as defined in (5.2). The Yule-Walker equations of an AR[p] model are derived in (5.4) and (5.5). Besides the solutions of the Yule-Walker equations (with the theoretical correlations replaced by the empirical ones), estimates for the coefficients in an AR[p] model can be obtained with the procedures introduced in Sect. 5.2.

2.5 When Is a Time Series Stationary?

The empirical moments calculated from the time series under analysis using definition (2.1) are estimates for the expectation and covariance functions of the stochastic process from which the time series stems on condition that the process is assumed to be stationary. But when is a process in agreement with the stationarity assumptions as defined in (2.8)?

For example, yearly values from the Basel temperature series as plotted in Fig. 2.13 and monthly values of the atmospheric CO_2 measured at Mauna Loa Observatory, Hawaii, as plotted in Fig. 5.9 are assumed to be non-stationary in their first moment functions. In both cases there is climatological evidence that the expectation function of the process is not constant because (i) the temperatures in Europe (and in the global mean) have increased over the last 120 years (Problem 2.30) and (ii) the atmospheric CO_2 has increased in the global mean for which the Mauna Loa measurements are representative.

Examples for processes assumed to be stationary over relatively long periods (decades) are the number of lynxes caught in the Mackenzie River District (Canada) as plotted in Fig. 9.38 and the Southern Oscillation (SO) index as plotted in Fig. 5.14. Other geophysical variables are stationary over a relatively short period. For example, the horizontal speed in a turbulent atmospheric flow was measured above Anwandstrasse in Zurich for a few days in March 1987. From the measurements recorded, the wind speed values in the time slice from 22:30 until 23:00 on March 11, 1987, as plotted in Fig. 2.1, were selected due to their stationarity, using the wind direction values which were measured and recorded together with the wind speed: the constant wind direction over the selected period points to constant conditions regarding the turbulent flow.

2.5.1 Diagnostics for the Stationarity of a Time Series

For many geophysical variables, however, it has not yet been ascertained from geophysical evidence whether or not these variables are stationary. An example for such a process is the NAO, resp. its indices plotted above in Figs. 2.9

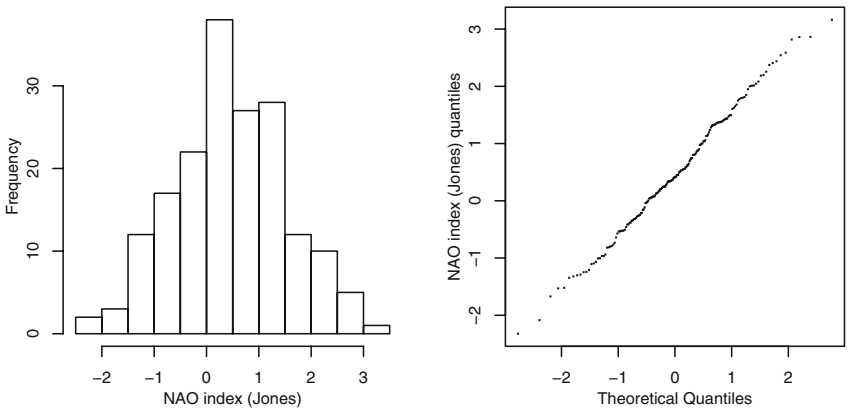


Fig. 2.18. North Atlantic Oscillation (NAO) index (Jones) from 1824 through to 2000 with mean 0.476 and empirical variance 1.161: histogram (on the left) and normal probability plot (a plot of the quantiles of the NAO index versus the quantiles of the standard normal distribution [30], on the right).

and 2.10. Hurrell in [70] argues that his NAO index has increased since approximately 1970 (and consequently, with increasing strength of the zonal circulation over the North Atlantic, winter temperatures and precipitation in the Northern Atlantic and the adjacent North Western Europe have also increased), whereas, in Problem 2.2, a white noise process (W_t) is estimated as a model for the NAO index (Jones) assuming that the observations in Fig. 2.10 (above) are a time slice from a realisation of a white noise process with expectation function μ_W and covariance function $c_W(\tau) = \sigma_W^2$ for $\tau = 0$ and $c_W(\tau) = 0$ for $\tau \neq 0$. This assumption results from (i) a small difference in the means 0.496 of the first (1824-1911) and 0.456 of the second (1912-2000) half-series of Jones' index, and (ii) a comparison of the empirical correlation functions as plotted in Fig. 2.10 (below) with the empirical correlation functions of the realisations of a white noise process in Fig. 2.22 as well as, in Problem 2.33, from the Ljung-Box statistic as defined in (2.60). This model is also assumed to be a normal process as defined in (2.3). This follows from the plots in Fig. 2.18. From both results it is concluded that Jones' NAO record is a time slice from a realisation of a normal white noise process with estimated expectation 0.476 and variance 1.161.

The NAO example demonstrates that statistical evidence concerning the stationarity of a process can be obtained even when geophysical evidence is not available. The properties of a stationary time series as enumerated in (2.54) follow from assumptions (2.8) since the moment functions of a stationary stochastic process are immune against displacements in the parameter domain. The similarity as required in (2.54,4) can be assessed by means of t-tests in the case of the first moment function, whereas similar plots of the

empirical correlation function are indicative of a stationary second moment function.

A time series (x_t), $t = 1, \ldots, N$, *is assumed to be stationary on condition that:*

1. *its mean remains approximately constant for* $t = 1, \ldots, N$,
2. *the amplitudes of its fluctuations remain approximately constant for* $t = 1, \ldots, N$, (2.54)
3. *its empirical correlation function decays rapidly (examples are given in Fig. 2.3 as well as below in Figs. 2.23 and 3.1), and*
4. *empirical moment functions calculated from a-priori defined time slices of* (x_t) *are similar.*

Properties (2.54) are to be found not only in the time series of the NAO indices (Hurrell and Jones) in Figs. 2.9 and 2.10 as shown above but also in the wind speed time series as plotted in Fig. 2.1, which, from geophysical evidence, is assumed to be stationary. In the case of the wind speeds measured in the turbulent atmospheric flow above Anwandstrasse in Zurich, it is concluded from (i) the small difference in the means (-2.378007 and -2.69577 ms^{-1}) calculated for the first and second half-series that the first moment function is constant, and from (ii) the similarity of the empirical correlation functions as plotted in Fig. 2.3 for the first 20 lags, that the second moment function depends solely on the lag. This procedure is repeated in Problem 2.25 with parts of the time series lasting 600 seconds and originating from the outset, middle and end of the wind speed time series.

Using the diagnostics as proposed in (2.54) is subject to the following restrictions:

1. Given the observational period of length N, time slices (e.g., the first and second half-series) used for calculating the empirical moment functions are defined prior to the analysis, i.e., before inspecting plots of the time series under analysis, as argued in Sect. 2.5.2. If no large fluctuations in the empirical moment functions for different time slices are obtained, it is then assumed that the time series is stationary. The result depends on the length of the time slices, i.e., on the scale applied for the comparison of the empirical moment functions.
2. If a time series is found to be stationary, it is then assumed that the process from which the observations stem is in fact stationary. However, the allowances for the fluctuations of the empirical moment functions can only be calculated if the expectation and covariance functions of the stochastic process are known. A universal test is not available since (i) the probability distribution of a test statistic is a function of the dependences in the stochastic process, and (ii) the transition from being non-stationary to stationary is continuous. An example of a stochastic process with such a transition is the AR[1] model as defined in (2.13) with the solution of its difference equation being found in (2.27).

Consequently, the assumption that a stochastic process is stationary remains a mere approximation obtained from geophysical evidence and/or from a time series found to be stationary using the diagnostics as proposed in (2.54).

These diagnostics can be enhanced by the predictions proposed in (5.70) for the control of manufacturing processes, an important application that has promoted the development of many techniques in time series analysis. Manufacturing processes are often required to be stationary to obtain constant properties of the product. A control of such a process is only possible, when the product can be described using variables for which values can be easily measured: when the time series of one or more variables under control are no longer stationary, the process is re-adjusted. A deviation from stationarity can be immediately detected by means of the procedures introduced in [17]. For example, the amount of a product formed during a chemical reaction occurring in a batch reactor as plotted in Fig. 5.3 is permitted to fluctuate with a preset tolerance (i.e., a known constant variance function) around a preset value (i.e., a known constant expectation function). When the time series is no longer stationary, the reactor has to be cleaned, or the temperature of the reactor has to be re-adjusted, etc.

In Geosciences, time series that have been found to be stationary are useful when analysing data obtained from comparison experiments. When meteorological instruments and their installations are compared, time series for the same meteorological variable but measured with different instruments in possibly different installations at the same or neighbouring stations, are subtracted from each other. From a stationary difference time series it is concluded that both instruments and their installations remained unchanged throughout the period covering the comparison experiment. A transition to non-stationarity, however, is due to a change in at least one of the measurements. Consequently, non-stationarities in a difference time series obtained from measurements of a meteorological variable at neighbour stations (with assumed identical climate) suggest that a climatological time series is not homogeneous as defined in Sect. 1.1. Procedures for the detection of inhomogeneities are given in [66], [59], [88] and [140].

For example, the error model (1.14) can be estimated provided that the differences of the pyranometer daily values, resulting from the comparison experiment described in Sect. 1.1 and plotted above, in Fig. 1.8, are assumed to be stationary (a very risky assumption indeed).

Examples of non-stationary difference time series are those calculated in Problem 1.11 and plotted in Fig. 2.19, which resulted from subtracting the time series of monthly values of SWIR measured at the neighbouring stations Toronto and Toronto Meteorological Research as well as at London Weather Centre and Bracknell. Both difference time series are non-stationary in their variance functions as the amplitudes of the fluctuations decrease abruptly in the years 1975 (Toronto) and 1980 (London) respectively. Supposedly, the abrupt decreases in the variance functions are due to the transition from

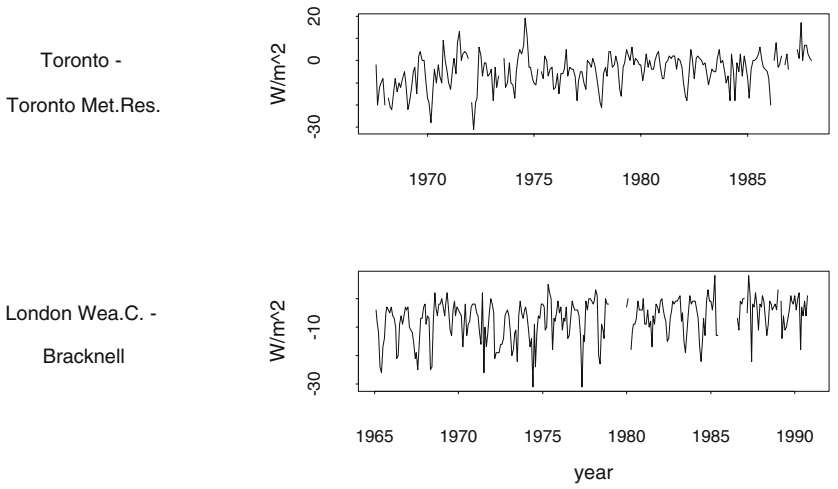


Fig. 2.19. A comparison experiment of monthly values of SWIR: differences obtained from the stations Toronto and Toronto Meteorological Research (above) as well as Bracknell and London Weather Centre (below).

manual to automated observations in Canada and the UK. The reader is invited to estimate, in Problem 2.27, the variance functions before and after the abrupt decreases.

2.5.2 Locally Stationary Time Series

Ambiguous results can occasionally be obtained from the diagnostic procedures introduced in Sect. 2.5.1 if, by violating (2.54,4), the time slices for the calculation of the empirical moment functions are defined after an inspection of the time series under analysis.

For example, R time series object `naojonests` obtained in Problem 2.2 contains observations of Jones' NAO index as plotted in Fig. 2.10. Using

```
t.test(window(naojonests,1901,1930),window(naojonests,1931,1960),
       alternative="greater")
```

the mean in the time slice from 1901 through to 1930 ($\hat{\mu}^{(1901)} = 0.92$) is found to be significantly larger (at the 0.0026 level, plotted with ∘ for the year 1930 in Fig. 2.20) than the mean in the time slice from 1931 through to 1960 ($\hat{\mu}^{(1931)} = 0.23$), and thus the conclusion is reached that Jones' index is not stationary in its first moment function. The means in these time slices have been plotted with solid horizontal lines in Fig. 2.20, the year 1930 with a broken vertical line, and the 0.05 significance level of the t-test with a broken horizontal line. In contrast, the difference in the means $\hat{\mu}^{(1831)} = 0.56$ and $\hat{\mu}^{(1861)} = 0.46$ in the time slices from 1831 through to 1860 and from 1861

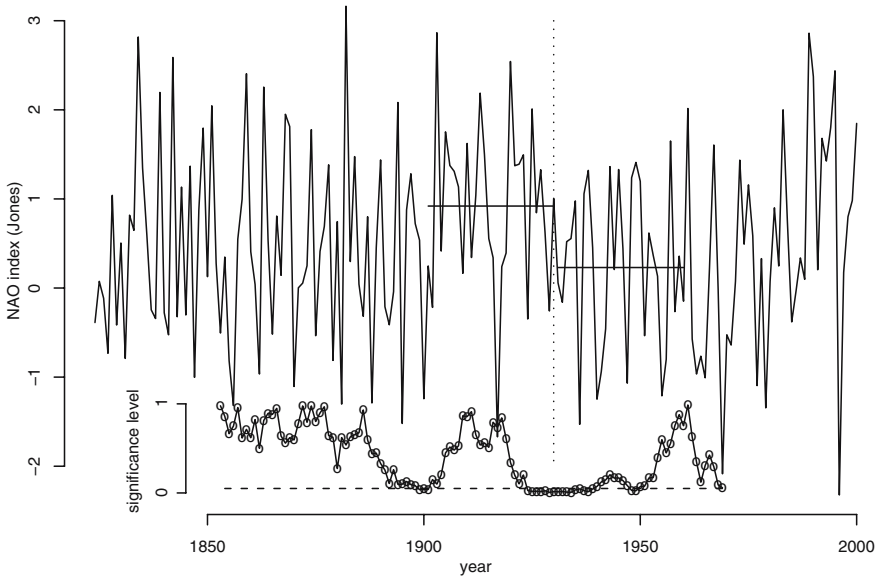


Fig. 2.20. Observations of the North Atlantic Oscillation (NAO) index (Jones) from 1824 through to 2000 with means $\hat{\mu}^{(1901)} = 0.92$ and $\hat{\mu}^{(1931)} = 0.23$ in the time slices from 1901 through to 1930 and from 1931 through to 1960 (horizontal solid lines). The significance level of two-sided t-tests comparing the means in time slices of 30 years in length before and after the years 1854 through to 1970 are plotted with $\circ$, the 0.05 level with a horizontal broken line.

through to 1890 appears to be small (significant at the 0.7133 level, plotted in Fig. 2.20 with $\circ$ for the year 1860). Thus, contradictory results are obtained from the time slices from 1831 through to 1860 and from 1901 through to 1960.

For the following discussion, let (Y_t) be Jones' NAO index and (y_t) the time series of observations y_t, $t = 1824, \ldots, 2000$ and thus $N = 177$, as plotted in Figs. 2.10 and 2.20. Means of (y_t) in two contiguous time slices, from $t - (v - 1)$ through to t and from $t + 1$ through to $t + v$, $v = 30$, for $t = 1824 + v, \ldots, 2000 - v$, are compared by means of two-sided t-tests. These t-tests produce the significance levels plotted with $\circ$ in Fig. 2.20: $n = 20$ out of $N - 2v = 117$ t-tests result in significance levels below 0.05 (as plotted with a horizontal broken line in Fig. 2.20). Obviously, significance levels below 0.05 are obtained when time slices with large differences in the means are compared and significance levels which are clearly above 0.05 result for time slices with small differences in the means. For example, (y_t) (and thus (Y_t) as well) is found to be locally stationary in $T_1' = (1831, 1832, \ldots, 1890)$ but locally non-stationary in $T_2' = (1901, 1902, \ldots, 1960)$, with both T_1' and T_2'

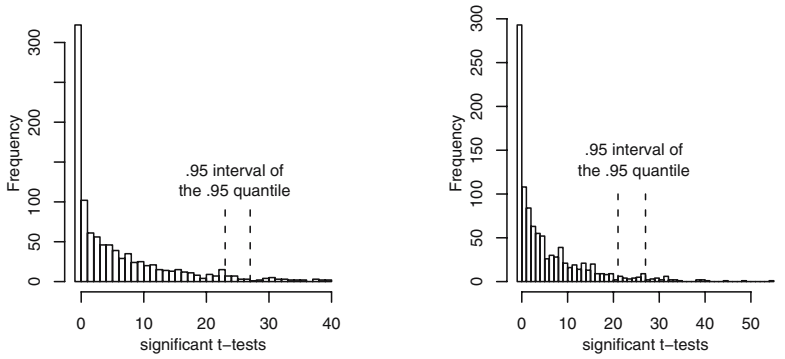


Fig. 2.21. Two histograms of the number of t-tests significant at the 0.05 level resulting when the experiment described in the remarks to Fig. 2.20 is repeated 1000 times using realisations of a standard normal white noise process.

being subsets of (i) the domain $T = (\ldots, -1, 0, 1 \ldots)$ of the parameter t of (Y_t) and (ii) the observational period.

This example for a local property of a stochastic process as defined in (2.55,2)

1. *Global properties of a stochastic process $X(t)$ as defined in (2.2) apply for all parameter values $t \in T$.*
2. *Local properties of a stochastic process $X(t)$ apply for all parameter values $t \in T' \subset T$, T' often being a time slice in the observational period.* (2.55)

demonstrates that the outcome of diagnostic procedures as proposed in (2.54) is influenced by the selection of the time slices from whence the empirical moment functions are calculated.

Clearly, the number of significant t-tests obtained in Fig. 2.20 depends on the fluctuating local mean in (y_t). Are these fluctuations reconcilable with hypothesis H_0 that (Y_t) is a normal white noise process with moments $\mu_Y = \hat{\mu}_Y = 0.476$ and $\sigma_Y^2 = \hat{\sigma}_Y^2 = 1.161$ as estimated in the remarks to Fig. 2.18? Or, formulated as a more specific question, what is the probability of obtaining, under H_0, $n = 20$ out of $N - 2v = 117$ t-tests with a significance level below 0.05 in (y_t)?

The answer could be found if the probability distribution of n could be derived from H_0 and the parameters ($N = 177$, $v = 30$, contiguous) describing the selection of the time slices for the comparison of the means in Fig. 2.20. Since an analytic solution to this problem is not easily obtained, the probability distribution of n is approximated (subsequent to a normalisation) with the histograms in Fig. 2.21, resulting with

```
lts <- 177            #length of time series
#length of time slices, 30 years are used to calculate
```

```
lmv <- 30             #climatological norms
Nexperiment <- 1000   #1000 times in a few minutes on a PC
signsample <- rep(0,Nexperiment)
for (case in (1:Nexperiment)) {     #standard normal, since the
  w <- rnorm(lts,mean=1.0,sd=1.0)   #result of a t-test does
  sign <- 0                         #not depend on linear
  for(jj in (lmv:(lts-lmv)) ) {     #transformations of the data
    tt <- t.test(w[(jj-(lmv-1)):jj],w[(jj+1):(jj+lmv)],
            alternative="two.sided")
    if(tt$p.value < 0.05) sign <- sign + 1             }
  signsample[case] <- sign                               }
hist(signsample,breaks=(-1:max(signsample)),xlab="...")
```

from two simulation experiments. The quite small number (1000) of replica in each experiment is sufficient to obtain an approximation by computing a confidence interval that includes the .95 quantile of the simulated sample with probability .95: if $c_{low} = Np - 1.96(Np(1-p))^{-1/2}$ and $c_{upp} = Np + 1.96(Np(1-p))^{-1/2}$ then, for $N = 1000$ and $p = 0.95$, the .95 quantile lies with probability 0.95 between the 936-largest and the 964-largest value, as plotted for both simulated samples in Fig. 2.21. Obviously, $n = 20$ as obtained in Fig. 2.20 is at the exterior of these confidence intervals. Hence, (Y_t) is assumed to be, in agreement with the remarks to Fig. 2.18, a normal white noise process.

The influence of selected time slices (i.e., time slices not defined prior to the analysis as required in (2.54,4)) on diagnostic procedures for the detection of instationarities in a time series can be a pitfall when climatological time series are analysed. Since it is the goal of many climatological projects to monitor climatic change, i.e., to detect a non-stationary expectation function in a homogeneous (cf. Sects. 1.1 and 2.3.4) climatological time series, climatologists are inclined towards selecting time slices with large differences in their means when applying diagnostic procedures as demonstrated in Fig. 2.20 to ascertain whether a time series is stationary in the expectation function.

When time slices with large differences in their means are preferred, false results can be obtained, as demonstrated in the following instance: it is assumed that a climatologist, seeking an instationarity in (Y_t),

1. applies diagnostic procedures with, in contradiction to (2.54,4), two selected contiguous time slices, as described in Fig. 2.20, to every realisation obtained from a standard normal white noise process (the model used above for (Y_t) since the result of a t-test is invariant under linear transformations) and
2. finds two contiguous time slices with a difference in their means, significant at the 0.05 level, when there are two such time slices in a realisation.

How often can he/she claim that the realisation under analysis is non-stationary by merely looking at the plots of the 1000 realisations produced in the simulation experiments from which the histograms in Fig. 2.21 have been

obtained? Since there are approximately 700 realisations with at least one significant t-test (there are approximately 300 realisations without a significant t-test as concluded from the histograms) he/she will detect an instationarity in approximately 7 out of 10 realisations, i.e., much more frequently than in approximately 1 out of 20 realisations, as corresponding to the significance level of the t-test. The usual t-test produces this false result since, when periods for the comparison are selected by the climatologist subsequent to inspecting plots of the time series, the experimental setup becomes undesirable for statistical inference as described in introductory textbooks.

In Geosciences we often find ourselves formulating and/or supporting scientific hypotheses after having measured the variable under study and inspected plots of the time series obtained. This setup is frequently out of line with statistical inference as presented in introductory textbooks. In this situation caution is called for when statistical tests are applied. For example, if, by violating the requirement stipulated in (2.54,4), a selection of time slices is involved in diagnostics for detecting instationarities in a time series, we then have a situation known in Statistics as *multiple comparison (test) problem* (in an analysis of variance for example, the Scheffé test allows for a-posteriori multiple comparison) combined with a *change point problem* (e.g., in [88] a two-phase regression model allows for detecting change points in climate records). In this situation, the significance of the results can be assessed in a simulation experiment, on condition that it is possible to model the selection procedure in some way ([145],[154]). The simulation of realisations of the stochastic process under analysis and assumed to be stationary is, however, only feasible if a model, e.g., a linear process as defined in (2.25), has been estimated with the methods introduced in Sect. 5.4. In R, simulations of linear processes are straightforward and a larger number of realisations can be simulated by compensating for the slow `for`-loops as demonstrated in [136] and [137].

The usually difficult modelling of a selection procedure and the estimation of a model for the stochastic process under analysis can be circumvented on condition that empirical moment functions are compared once they have been calculated from time slices defined prior to the analysis, as required in (2.54,4). A simulation experiment reflecting this requirement for the Jones' NAO time series is performed in Problem 2.36.

In this section it is demonstrated that statistical evidence concerning the stationarity of a stochastic process under analysis can be obtained using diagnostic (2.54) even when geophysical evidence is not available. The results thus arrived at can be ascertained, under favourable circumstances, by applying statistical tests and/or performing a simulation experiment. The example time series analysed is the NOA index (Jones) with the result showing that this time series is a time slice in a realisation of a white noise process. Implications of this result are discussed in Sect. 2.8.4.

2.6 Estimators for the Moment Functions of a Stationary Stochastic Process

The empirical moment functions $\hat{\mu}_X$ and $\hat{c}_X(\tau)$ of a time series (x_t), $t = 1, \ldots, N$, are defined in (2.1). Assuming that (x_t) is an observation of a stationary stochastic process (X_t), the moments of $\hat{\mu}_X$ and $\hat{c}_X(\tau)$ are calculated in Sect. 2.6.1. For most, but not all, (an example is given in Sect. 2.6.2) stationary stochastic processes, $\hat{\mu}_X$ and $\hat{c}_X(\tau)$ are unbiased and consistent estimators of the moment functions μ_X and $c_X(\tau)$.

2.6.1 Properties of the Empirical Moment Functions of a Stationary Time Series

In Sect. 1.3, the expectation and variance of the arithmetic mean $\hat{\mu}_X$ of a random sample x_i, $i = 1, 2, \ldots, N$, as defined in (1.2), are calculated and it is shown that $\hat{\mu}_X$ is an unbiased and consistent estimator of the expectation μ_X of the identically distributed random variables X_i from which the observations x_i stem. Thus, being unbiased and consistent, $\hat{\mu}_X$ has the desirable properties defined in (1.3) and (1.4). Due to the iid.-assumptions (1.2), not only the expectation and variance but also the probability distribution of $\hat{\mu}_X$ is easily calculated, if the distribution of the observed X_i is known. If this distribution is not known then the probability distribution of the arithmetic mean of a random sample is obtained with the central limit theorem as proposed in (1.6).

For the arithmetic mean of a random sample, consequently, confidence limits can be calculated and tests performed. For these desirable properties, however, we pay with the restrictive iid.-assumptions which are not reconcilable with most time series. Although not being iid., time series are often stationary as defined in (2.8). In Sect. 2.6.1 therefore, the properties of estimators of the moment functions of a stationary stochastic process are derived. These estimators are the arithmetic mean and empirical variance as well as the empirical covariance and correlation functions, calculated using (2.1) from a time series stemming from the stationary process to be estimated.

In (2.72) and (2.73), the properties (2.56) of the arithmetic mean of a time series as defined in (2.1,1) are derived.

1. *The arithmetic mean $\hat{\mu}_X$ calculated from a time slice (x_t), $t = 1, \ldots N$, of a realisation of a stationary stochastic process (X_t), $t = \ldots, -1, 0, 1, \ldots$, is an unbiased estimator for the constant expectation function μ_X of (X_t).*
2. *If the covariance function $c_X(\tau)$ of (X_t) converges absolutely as defined in (2.24) then $\hat{\mu}_X$ is a consistent estimator for μ_X as the variance of $\hat{\mu}_X$ decreases with $1/N$ when N increases.*

(2.56)

From an absolutely convergent covariance function of a stationary stochastic process it becomes plausible that the dependence of random variables with

lag τ decreases rapidly with increasing τ. Thus, with increasing length N of the time series used for the estimation, the number of independent or at least very weakly dependent observations increases and the variance of the estimator decreases.

The properties of the empirical covariance function as defined in (2.1,3) are proposed in (2.58); (2.58,1) follows from (2.57), whereas a derivation of (2.58,2) is not easy to arrive at and therefore is not given here.

Let (X_t), $t = \ldots, -1, 0, 1, \ldots$, be a stationary stochastic process with expectation function $\mathrm{E}(X_t) = \mu_X = 0$ and covariance function $\mathrm{E}(X_t X_{t+\tau}) = \mathrm{Cov}(X_t, X_{t+\tau}) = c_X(\tau)$ and let (x_t), $t = 1, \ldots, N$, be a time slice from a realisation of (X_t). $c_X(\tau)$ is estimated with the empirical covariance function $\hat{c}_X(\tau)$ as defined in (2.1,3) and thus, substituting $\mu_X = 0$ for $\hat{\mu}_X$ in (2.1,3), $\hat{c}_X(\tau) = (1/N)\sum_{t=1}^{N-\tau} X_t X_{t+\tau}$ is obtained as an estimator of $c_X(\tau)$.

The expectation of this estimator is calculated below: the result in (2.57) is obtained with the properties of the covariance matrix (2.7).

$$\begin{aligned} \mathrm{E}\hat{c}_X(\tau) &= \mathrm{E}\Big(\frac{1}{N}\sum_{t=1}^{N-\tau} X_t X_{t+\tau}\Big) = \frac{1}{N}\sum_{t=1}^{N-\tau} \mathrm{E}(X_t, X_{t+\tau}) = \frac{1}{N}\sum_{t=1}^{N-\tau} c_X(\tau) \\ &= \frac{1}{N}(N-\tau)c_X(\tau) = \begin{cases} (1-\tau/N)c_X(\tau) & \text{for } 0 \le |\tau| \le N \\ 0 & \text{for } |\tau| > N \end{cases} \end{aligned} \tag{2.57}$$

From (2.57) it is concluded that $\hat{c}_X(\tau)$ is a biased estimator of $c_X(\tau)$. However, if N is large and τ small then $\hat{c}_X(\tau)$ is approximately unbiased, τ/N being close to zero.

The variance of this estimator is obtained from the covariance of $\hat{c}_X(\tau)$, i.e., $\mathrm{Cov}\big(\hat{c}_X(\tau), \hat{c}_X(\tau+\tau')\big)$, with lag $\tau' = 0$. A derivation of this covariance is, however, less straightforward than the derivation of the variance of $\hat{\mu}_X$ in (2.73) since expectations of products of the covariance function $c_X(\tau)$ of (X_t) are involved. From the laborious calculations, it is revealed that the variance of $\hat{c}_X(\tau)$ is a function of, besides other arguments, the fourth moment function of the process (X_t), and further, that $\hat{c}_X(\tau)$ is, under the restrictions as stipulated in (2.58,2), a consistent estimator for $c_X(\tau)$.

1. *An empirical covariance function $\hat{c}_X(\tau)$ calculated using (2.1,3) from a time slice (x_t), $t = 1, \ldots N$, in a realisation of a stationary stochastic process (X_t), $t = \ldots, -1, 0, 1, \ldots$, is an approximately unbiased estimator of the covariance function $c_X(\tau)$, on condition that (i) τ is small as compared to N and (ii) $\mu_X = 0$.*
2. *$\hat{c}_X(\tau)$ is a consistent estimator for $c_X(\tau)$, provided that (i) (X_t), besides being stationary, is normal as defined (2.3), and (ii) $c_X(\tau)$ converges absolutely.*

(2.58)

(2.58,1) and (2.58,2) also apply to a stationary normal process (X_t) with $\mathrm{E}(X_t) = \mu_X \neq 0$ and an absolutely convergent $c_X(\tau)$ on condition that the non-zero expectation function is accounted for.

There are estimators for the covariance function of a stationary stochastic process with, at least in some aspects, superior statistical properties than those of $\hat{c}_X(\tau)$. For instance, $\tilde{c}_X(\tau) = 1/(N-\tau)\sum_{t=1}^{N-\tau} X_t X_{t+\tau}$ is an unbiased and consistent estimator of $c_X(\tau)$. However, $\hat{c}_X(\tau)$ as defined in (2.1,3) is preferred as being positive semidefinite. A positive semidefinite empirical covariance function is desirable in connection with the linear prediction (in Sect. 2.7 as well as in Sects. 4.4 and 5.5) and the spectral estimation (in Chaps. 8, 9 and 10). The covariance function is shown to be positive semidefinite in the remarks to (2.7), and in the case of the empirical covariance function, the positive semidefinite property is derived in Problem 2.34.

Often, the empirical covariance function $\hat{c}_X(\tau)$ is not zero for large lags even though the theoretical covariance function $c_X(\tau)$ is zero. For instance, in Fig. 9.25, the covariances of the example process are considerably overestimated for lags $\tau \geq 180$. This behaviour is due to large correlations of neighbour $\hat{c}_X(\tau)$ which fairly frequently occur. In these cases a plot of $\hat{c}_X(\tau)$ is very smooth and therefore suggests a too small variance of the estimator. This property of the empirical covariance function is demonstrated in Problem 2.3 with simulations of the AR[1] model (2.12). The pitfalls due to possibly large correlations of neighbouring $\hat{c}_X(\tau)$ can be circumvented when a long time series is observed. With increasing length of a time series stemming from a process with an absolutely convergent covariance function, in line with the remarks to (2.56), the number of approximately independent observations used for the estimation increases and, consequently, the variance of the estimator decreases.

The empirical correlation function $\hat{\varrho}_X(\tau)$ calculated, as defined in (2.1,4), from a time slice (x_t), $t = 1, \ldots N$, in a realisation of a stationary stochastic process (X_t), $t = \ldots, -1, 0, 1, \ldots$, is an estimator for the correlation function $\varrho_X(\tau) = c_X(\tau)/c_X(0)$ of (X_t). The expectation $\mathrm{E}\hat{\varrho}_X(\tau)$ and covariance $\mathrm{Cov}\big(\hat{\varrho}_X(\tau), \hat{\varrho}_X(\tau+\tau')\big)$ of the estimator are obtained by calculations that are even more laborious than those performed for the expectation and covariance of the empirical covariance function $\hat{c}_X(\tau)$. From these calculations it is concluded that the properties (2.58), with ϱ substituted for c, etc., also pertain to the empirical correlation function $\hat{\varrho}_X(\tau)$. Note that, in the case of the empirical correlation function, (2.58,2) applies not only to a stationary normal process but also to linear processes as defined in (2.25) [21].

If (W_t) is a white noise process then its correlation function $\varrho_W(\tau) = 1$ for $\tau = 0$ and $\varrho_W(\tau) = 0$ for $\tau \neq 0$ is obtained directly from definitions (2.10) or (2.11). However, the empirical correlation function $\hat{\varrho}_W(\tau)$, calculated using (2.1,4), from a time slice (w_t), $t = 1, \ldots N$, of a realisation of a white noise process (W_t) is not identically zero for lags $\tau \neq 0$ as demonstrated with the examples in Fig. 2.22. In the case of a white noise process (W_t), on the one hand, the calculation of the expectation and the variance of $\hat{\varrho}_W(\tau)$ is as laborious as the calculation of $\mathrm{E}\hat{\varrho}_X(\tau)$ and $\mathrm{Cov}\big(\hat{\varrho}_X(\tau), \hat{\varrho}_X(\tau+\tau')\big)$ in the case of a stationary process (X_t). On the other hand, however, the following

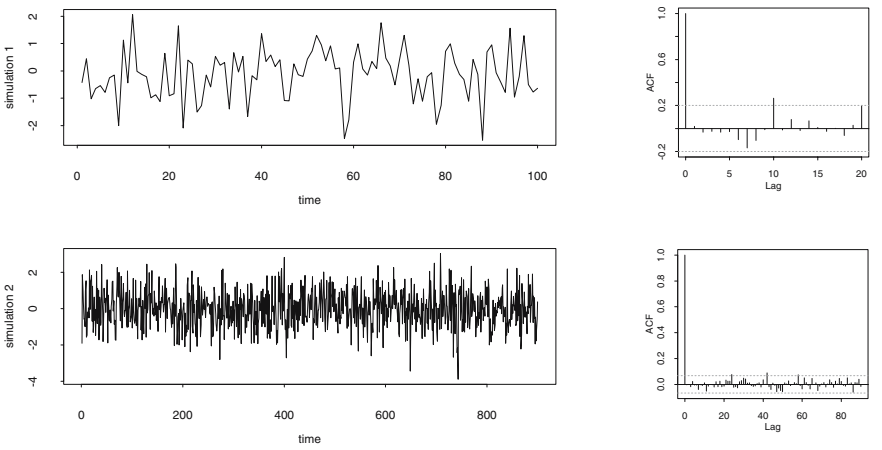


Fig. 2.22. Simulated realisations ($N_1 = 100$, $N_2 = 900$) of a white noise process with $\mu_W = 0$ and $\sigma_W^2 = 1$ and empirical correlation functions calculated from the simulations.

useful approximation (2.59) is obtained.

The empirical correlations $\hat{\varrho}_W(\tau)$ calculated using (2.1,4) from a time slice (w_t), $t = 1, \ldots N$, of a realisation of a white noise process (W_t), as defined in (2.10) or (2.11) with $\mu_W = 0$ and σ_W^2, are approximately normally distributed with $\mathrm{E}\hat{\varrho}_W(\tau) = 0$ and $\mathrm{Var}\hat{\varrho}_W(\tau) = 1/N$ for large N. (2.59)

Due to (2.59), approximately 95 out of 100 empirical correlations of the simulations in Fig. 2.22 are within $\pm 1.96 \times N^{-1/2}$ with approximately 5 out of 100 outside this confidence interval. This confidence interval is constructed as demonstrated in the remarks to Fig. 2.3 or is obtained in the default plot (default plots are shown in Fig. 2.22) of the empirical correlation function produced with R function `acf()`.

For example, in the plots of the empirical correlation functions of the NAO indices (Hurrell in Fig. 2.9) and (Jones in Fig. 2.10), one out of 30 empirical correlations are outside $\pm 1.96 \times N^{-1/2}$, and thus it is concluded that both indices stem from white noise processes.

Most time series are not a time slice from a realisation of a white noise process. Examples are the wind speed series as plotted in Fig. 2.1 with many empirical correlations outside $\pm 1.96 \times N^{-1/2}$ (as can be seen in Fig.2.3) and the differences of the yearly values in the Basel temperature series as plotted in Fig.2.14 (above) with only a few empirical correlations outside this confidence interval (as shown below, in Fig. 2.14).

These examples demonstrate that, in most cases, by using approximation (2.59), it can be excluded that a time series is a time slice from a realisation of a white noise process. Approximation (2.59) can also be used in combination with tests for white noise. For example, the K-lag *Ljung-Box statistic*

$$T_K = N(N-2)\sum_{\tau=1}^{K}\frac{\hat{\varrho}_\tau^2}{N-\tau} \tag{2.60}$$

K being much smaller than N, can be compared to an upper quantile of a chi-square distribution with K degrees of freedom, as proposed in [87] or [21]. As another possibility, a test using spectral estimation techniques as proposed in Sect. 8.2.2 can be performed. The spectrum of a white noise process is derived in (7.76).

2.6.2 Ergodic Theorems: a Cautionary Remark

(2.56) and (2.58) do not apply to all stationary stochastic processes, seeing that there are stationary processes (Y_t) with the property that $\hat{\mu}_Y$, calculated using (2.1,1) from an observed time slice (y_t), $t = 1, 2, \ldots, N$, of a realisation, is an unbiased (as defined in (1.3)) but not a consistent (as defined in (1.4)) estimator of μ_Y. In Problem 2.29 for example, $\mathrm{E}\hat{\mu}_Y(t) = 0$ and $\mathrm{Var}\hat{\mu}_Y = \sigma_V^2$ for $N \to \infty$ are obtained on condition that $Y_t = X_t + V$, $t \in T$, with (i) (X_t) being a stationary process with expectation function $\mu_X(t) = \mu_X = 0$ and absolutely convergent covariance function $c_X(\tau)$, (ii) V being a random variable with $\mu_V = 0$ and σ_V^2, and (iii) $\mathrm{Cov}(X_t, V) = 0$ for all $t \in T$. Thus, $\hat{\mu}_Y$ proves to be an unbiased but not a consistent estimator for μ_Y. Another example is the harmonic process as defined in (7.30), with its moment functions given in (7.31): it is stationary but its expectation function cannot be estimated without bias and consistently.

The conditions for a convergence (e.g., in mean square, since (2.73) implies $\mathrm{Var}\hat{\mu}_X = \mathrm{E}(\hat{\mu}_X - \mu_X)^2 \to 0$ for $N \to \infty$) of an empirical moment function (e.g., $\hat{\mu}_X$ calculated from a time slice from a realisation of a stationary stochastic process (X_t)) to the theoretical moment function of the stationary process (e.g., μ_X), are proposed in the ergodic theorems [155].

The theorist uses the convergence of an empirical moment function to derive the convergence of further estimators calculated from this moment function. For example, if $\hat{\varrho}_X(1)$ and $\hat{\varrho}_X(2)$ are substituted for $\varrho_X(1)$ and $\varrho_X(2)$, in the Yule-Walker equations (2.53), and if $\hat{\varrho}_X(1)$ and $\hat{\varrho}_X(2)$ converge to $\varrho_X(1)$ and $\varrho_X(2)$, then the estimators for the coefficients a_{12} and a_{22} of the AR[2] process converge.

The practitioner learns from the examples given above to be cautious when estimating the moment functions of a stationary stochastic process: the empirical moment functions as defined in (2.1) have the properties proposed in (2.56) or (2.58) when calculated from a realisation of most (but not of all) stationary stochastic processes (X_t).

2.7 Optimal Linear Predictions

To simplify, there are two kinds of forecasts. Examples for forecasts of the first kind are the weather forecasts produced with a model of the physical processes in the atmosphere (a general circulation model, GCM) using the observations from the synoptic stations as well as those obtained from the operational weather satellites. The equations in a GCM are solved using a computer for future time points and, from these solutions, the weather services produce the weather forecasts. Forecasts of this kind are, however, not dealt with in this book.

An example for a forecast of the second kind is the estimation of the wind speed in the turbulent atmospheric flow above the Anwandstrasse in Zurich for $t = 1801$, $t = 1802$, $t = \ldots$, using the time series (x_t), $t = 1, 2, \ldots, 1800$ plotted in Fig. 2.1, i.e., forecasts are calculated for the first, the second, etc., time point after the observational period. Another example is a forecast of the NAO index (Jones) for the year 1999 using the observations plotted in Fig. 2.10. Since this time series stems from a normal white noise process (W_t) with estimated moments $\hat{\mu}_W = 0.4713$ and $\hat{\sigma}_W^2 = 1.0858$ as shown in the remarks to Figs. 2.18 and 2.20, it is concluded from definitions (2.10) and (2.11) that the probability distribution of W_{1999} is normal with mean 0.4713 and variance 1.0858.

These examples demonstrate that a forecast of the second kind is a realisation of an estimator as defined in (1.1), calculated using an observed time slice from a realisation of a stochastic process for time points with no observations (usually for the future) under the assumption that the forecast belongs to the observed realisation. Such an estimator is called a prediction. Generally, several predictions can be calculated from the very same observations, the optimal one minimizing the prediction error.

Let the time series (x_t), $t = 1, \ldots, N$, be a time slice from a realisation of the stochastic process (X_t) with a finite variance $\sigma_X^2(t)$ for all $t \in T$ and let $X_{N+\tau}$ be a random variable pertaining to the future: $\tau > 0$. Then:

1. *If $\hat{x}_{N+\tau}$ is a realisation of estimator $\widehat{X}_{N+\tau} = \widehat{X}_{N+\tau}(X_1, \ldots X_N)$, then $\hat{x}_{N+\tau}$ is called a predicted value of $X_{N+\tau}$ and $\widehat{X}_{N+\tau}$ is called a prediction of $X_{N+\tau}$ (of (X_t) with lag τ).*
2. *An optimal prediction minimises the variance of the prediction error $\hat{e}_{N+\tau} = X_{N+\tau} - \widehat{X}_{N+\tau}$.*
   This variance is called the mean square prediction error,
   $\mathrm{PVar}\widehat{X}_{N+\tau} = \mathrm{Var}\hat{e}_{N+\tau} = \mathrm{E}\big((X_{N+\tau} - \widehat{X}_{N+\tau})^2\big)$.

(2.61)

Often used predictions $\widehat{X}_{N+\tau}$ are linear combinations of random variables X_t in the stochastic process (X_t). The variance of such a prediction is non-negative because the covariance function of the stochastic process is positive definite, as proposed in (2.7).

Let (x_t) and (X_t) be as defined in (2.61). Then
$$\begin{aligned}\widehat{X}_{N+\tau} &= a_0 X_N + a_1 X_{N-1} + \ldots + a_{N-1} X_1 \\ &= \sum_{u=0}^{N-1} a_u X_{N-u} = \boldsymbol{a}^T \boldsymbol{X}\end{aligned}$$
is called a linear prediction, with $\boldsymbol{a} = (a_0, a_1, \ldots, a_{N-1})^T$ and $\boldsymbol{X} = (X_N, X_{N-1}, \ldots, X_1)^T$ being column vectors with N values.
If $\boldsymbol{a}$ minimises the mean square prediction error
$\mathrm{PVar}\widehat{X}_{N+\tau} = \mathrm{E}\big((X_{N+\tau} - \boldsymbol{a}^T \boldsymbol{X})^2\big)$ *then $\widehat{X}_{N+\tau} = \boldsymbol{a}^T \boldsymbol{X}$*
is called an optimal linear prediction of (X_t) with lag τ. (2.62)

The predicted value $\hat{x}_{N+\tau}$ is obtained as a linear combination of the values in the observed time series (x_t), $t = 1, \ldots, N$, and the values in $\boldsymbol{a}$. The values in $\boldsymbol{a}$ are called the *coefficients* (or the *weights)* of the prediction. They are calculated as proposed in (2.63) if the process to be predicted is stationary.

If (x_t), (X_t), $X_{N+\tau}$, $\boldsymbol{a}$ and $\boldsymbol{X}$ are as defined in (2.62) and (X_t) is stationary with $\mu_X(t) = 0$ and covariance matrix $\boldsymbol{C}_X$, then:

1. $\boldsymbol{a} = \boldsymbol{C}_X^{-1}\big(\mathrm{Cov}(X_{N+\tau}, X_N), \ldots, \mathrm{Cov}(X_{N+\tau}, X_1)\big)^T = \boldsymbol{C}_X^{-1} \boldsymbol{c}_{N,\tau}$
2. $\mathrm{PVar}\widehat{X}_{N+\tau} = \mathrm{E}\big((X_{N+\tau} - \widehat{X}_{N+\tau})^2\big) = c_X(0) - \boldsymbol{c}_{N,\tau}^T \boldsymbol{a}$
3. $\mathrm{E}\widehat{X}_{N+\tau} = \mathrm{E}X_{N+\tau}$

(2.63)

(2.63) is derived under the assumption that the covariance matrix $\boldsymbol{C}_X$ of the stationary process (X_t) to be predicted is known. Under this assumption the mean square prediction error is minimised to obtain the optimality conditions. This minimisation is demonstrated in (2.74) for the first optimality condition. From the optimality conditions, proposition (2.63) is derived in (2.75), (2.76) and (2.77). For an alternative derivation, the optimality conditions are written as in (2.78) from whence (2.63) is derived in the remarks to (2.78) and in (2.80). From (2.78), the properties (2.79) and (2.81) of the optimal linear prediction are obtained, for later use in Sect. 5.5.

With a slightly modified derivation, the coefficients of an optimal linear prediction are obtained in the case of a stationary process with a constant expectation function not being identically zero, i.e., $\mu_X(t) = \mu_X \neq 0$. An unbiased optimal linear prediction is obtained with weights a_u, which not only minimise the mean square prediction error but also fulfill the universality (Sect. 4.5, (4.29) and remarks to (4.22)) condition $\sum_{u=0}^{N-1} a_u = 1$, since

$$\mathrm{E}X_{N+\tau} \;=\; \mu_X \;=\; \mathrm{E}\widehat{X}_{N+\tau} \;=\; \sum_{u=0}^{N-1} a_u \mu_X \;=\; \mu_X \sum_{u=0}^{N-1} a_u. \tag{2.64}$$

With a known covariance matrix of the process to be predicted, the column vector $\boldsymbol{a}$ containing the weights of the prediction is arrived at using (2.63,1), i.e., by multiplying the inverse of the covariance matrix $\boldsymbol{C}_X$ (as proposed in the remarks to (2.7), $\boldsymbol{C}_X^{-1}$ exists provided that $\boldsymbol{C}_X$ is positive definite) with the vector $\boldsymbol{c}_{N,\tau}$. $\boldsymbol{c}_{N,\tau}$ is the vector containing the covariances of the random variable to be predicted with the random variables at the time points with observations.

Usually, the covariance matrix $\boldsymbol{C}_X$ of the process (X_t) to be predicted is not known and therefore has to be estimated. An estimation is possible using the procedures introduced in Sect. 2.1 and discussed in Sect. 2.6. The estimates, i.e., the empirical covariances calculated as defined in (2.1,3), are substituted for the theoretical covariances, and the estimates $\hat{\boldsymbol{a}}$ for $\boldsymbol{a}$ are then calculated. The covariances between the variable (to be predicted) at $N+\tau$ and the variables at the observed time points N time steps, or further back, $\mathrm{Cov}(X_{N+\tau}, X_1), \ldots, \mathrm{Cov}(X_{N+\tau}, X_\tau)$, cannot be estimated from $x_1, \ldots, x_N$. If, however, $c_X(\tau)$ decays rapidly and N is large enough, these covariances are assumed to be identically zero. The equations thus obtained can be solved for $\hat{\boldsymbol{a}}$ in the majority of the cases, namely on condition that the empirical covariance matrix of a stationary time series is positive definite, and thus is invertible as proposed in the remarks to (2.7) (Problem 2.34). For example, in the optimal linear prediction calculated in Problem 2.35, the weights $\hat{a}_0$, $\hat{a}_1$, ... for the observations x_N, x_{N-1}, ... decrease with increasing distance from the prediction time $N+\tau$.

If a prediction is calculated from a short time series then, due to (2.63,2), the large variances of the empirical covariances propagate to the estimates $\hat{\boldsymbol{a}}$, and thus the prediction is said to be statistically not stable. If a prediction is calculated from a long time series then the solution of the equations for $\hat{\boldsymbol{a}}$ is time consuming and can become numerically unstable. Possible solutions to this problem are the use of models

1. for the expectation function (in which case, the expectation function, not the process, is estimated), in Chap. 3,
2. for the covariance function, in Chap. 4, or
3. for the process (MA and AR models as demonstrated with the examples in Sect. 2.3, generally linear processes). Predictions of linear processes are obtained in Sect. 5.5 with explicit or recursive procedures being preferable to procedures involving the solution of systems of linear equations.

2.8 Supplements

In Sect. 2.8.1, the moment functions of the AR[1] model, of the random walk and of the linear process are derived. In Sect. 2.8.2, the moments of the mean of a stationary time series are calculated. In Sect. 2.8.3, the optimal linear predictor of a stationary time series is obtained, and in Sect. 2.8.4 it is shown (at least for the white noise process) that the conditional expectation is the optimal predictor.

2.8.1 First and/or second Moment Functions of some Processes

The covariance function of an AR[1] model as proposed in (2.16) is arrived at with the initial condition $X_{-N} = 0$ from the sum in (2.14). Since $\mu_W = 0$, by definition

$$\mathrm{Cov}(X_t, X_{t+\tau}) = \mathrm{E}\big((W_t + aW_{t-1} + \ldots + a^{t+N-1}W_{-N+1}) \times (W_{t+\tau} + aW_{t+\tau-1} + \ldots + a^{t+N-1+\tau}W_{-N+1})\big)$$

and, since the covariances $\mathrm{Cov}(W_t, W_u)$ are identically zero for $t \neq u$, the second moment functions follow for $\tau \geq 0$:

$$\begin{aligned} c_X(t, t+\tau) &= \mathrm{Cov}(X_t, X_{t+\tau}) \\ &= \mathrm{Cov}(W_t, W_{t+\tau}) + a\mathrm{Cov}(W_t, W_{t+\tau-1}) + \ldots \\ &\quad + a^\tau \mathrm{Cov}(W_t, W_{t+\tau-\tau}) + \ldots \\ &\quad \ldots + a\mathrm{Cov}(W_{t-1}, W_{t+\tau}) + aa\mathrm{Cov}(W_{t-1}, W_{t+\tau-1}) + \ldots \\ &\quad + aa^{\tau+1}\mathrm{Cov}(W_{t-1}, W_{t+\tau-(\tau+1)} + \ldots \\ &= \sigma_W^2 \Big(a^\tau + a^{\tau+2} + \ldots + a^{\tau+2(t+N-1)}\Big). \end{aligned}$$

With $\mathrm{Cov}(X_t, X_{t+\tau}) = \mathrm{Cov}(X_t, X_{t-\tau})$ due to the property (2.9,4)

$$c_X(t, t+\tau) = \begin{cases} \sigma_W^2 a^{|\tau|}(1 - a^{2(t+N)})/(1 - a^2) & \text{for } |a| \neq 1 \\ \sigma_W^2 (t+N) & \text{for } |a| = 1 \end{cases} \tag{2.65}$$

$$\begin{aligned} \sigma_X^2 &= \mathrm{Cov}(X_t, X_t) \\ &= \begin{cases} \sigma_W^2 (1 - a^{2(t+N)})/(1 - a^2) & \text{for } |a| \neq 1 \\ \sigma_W^2 (t+N) & \text{for } |a| = 1 \end{cases} \end{aligned} \tag{2.66}$$

the covariance as well as the variance function of an AR[1] model (X_t) under the initial condition $X_{-N} = 0$ are obtained.

The moment functions of a random walk process $Y_t = Y_{t-1} + W_t$, $t = 1, 2, \ldots$, as defined in (2.20), are derived using initial condition $Y_0 = 0$. As $\mu_W = 0$, the expectation function $\mathrm{E}(Y_t)$ is constant, otherwise, $\mathrm{E}(Y_t)$ would be non-decreasing (for $m_W > 0$) or non-increasing (for $m_W < 0$), since

$$\mathrm{E}(Y_t) = \mathrm{E}\sum_{u=1}^{t} W_u = \sum_{u=1}^{t} \mathrm{E}W_u = \mu_W \times t. \tag{2.67}$$

As W_i and W_j are independent or uncorrelated for all pairs (i, j), a non-decreasing variance function

$$\mathrm{Var}(Y_t) = \mathrm{Var}\sum_{u=1}^{t} W_u = \sum_{u=1}^{t} \mathrm{Var}W_u = \sigma_W^2 \times t \tag{2.68}$$

is obtained (since $\sigma_W^2 > 0$), and thus the paths of the realisations in Fig. 2.12 diverge. For the covariances, with $\mathrm{Cov}(W_i, W_j) = 0$ for $i \neq j$ and $\mathrm{Cov}(W_i, W_j) = \sigma_W^2$ for $i = j$,

$$\mathrm{Cov}(Y_t, Y_u) = \mathrm{Cov}\left(\sum_{i=1}^{t} W_i, \sum_{j=1}^{u} W_j\right)$$

$$= \sum_{i=1}^{t} \sum_{j=1}^{u} \mathrm{Cov}(W_i, W_j) = \sigma_W^2 \big(\min(t,u)\big) \tag{2.69}$$

is obtained since only the terms for $i = j$ remain in the double indexed sum in (2.69). The number of these terms equals the minimum of t and u. The variances and covariances of a random walk process are summarised in the covariance matrix:

$$\begin{array}{c} \\ Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \\ \vdots \end{array} \begin{array}{c} \begin{array}{cccc} Y_1 & Y_2 & Y_3 & Y_4 \quad \ldots \end{array} \\ \begin{pmatrix} 1 & 1 & 1 & 1 & \\ 1 & 2 & 2 & 2 & \\ 1 & 2 & 3 & 3 & \\ 1 & 2 & 3 & 4 & \\ & & & & \end{pmatrix} \end{array} \sigma_W^2 = \boldsymbol{C}_Y. \tag{2.70}$$

From the above results, it is concluded that the constant moment functions of a white noise process become the proportionality factors in the moment functions of a random walk process constructed from the white noise process.

The moment functions of a linear process as defined in (2.25) can be derived from the moment functions $\mu_W = 0$ and σ_W^2 of the white noise process since the weights (b_t) converge absolutely. The first moment function is obtained easily, the second with the following calculation. If $\tau \geq 0$, then

$$\begin{aligned} c_Y(t,u) = \mathrm{Cov}(Y_t, Y_u) &= \mathrm{Cov}\Big(\sum_v b_v W_{t-v}, \sum_w b_w W_{u-w}\Big) \\ &= \sum_v \sum_w b_v b_w \mathrm{Cov}\big(W_{t-v}, W_{u-w}\big) \\ &= \sum_v \sum_w b_v b_w c_W(t-v, u-w) \end{aligned} \tag{2.71}$$

follows. From the stationarity of (W_t), $c_W(t-v, u-w) = c_W(t-v, t-\tau-w) = c_W(t-v-(t-\tau-w))$ is derived and, consequently, $c_Y(t,u) = c_Y(t, t-\tau) = c_Y(\tau) = \sum_v \sum_w b_v b_w c_W(\tau + w - v)$. In this sum remain, since $c_W = 0$ for $\tau + w - v \neq 0$, the terms with $w + \tau = v$: for $\tau = 0$, all terms with $w = v$, hence $\sum_u b_u b_u \sigma_W^2$, for $\tau = 1$ all terms with $w + 1 = v$, hence $\sum_u b_u b_{u+1} \sigma_W^2$, etc. Thus, (2.26) follows.

2.8.2 Moments of the Mean of a Stationary Time Series

Let (X_t) be a stationary stochastic process with $\mathrm{E}(X_t) = \mu_X(t) = \mu_X$ and $\mathrm{Cov}(X_t, X_{t+\tau}) = c_X(\tau)$, and let the time series (x_t), $t = 1, \ldots, N$, be a time slice from a realisation of (X_t). The arithmetic mean of the time series as defined in (2.1,1) estimates μ_X. Therefore, the estimator is $\hat{\mu}_X = (1/N) \sum_{t=1}^{N} X_t$. For the expectation of this estimator

$$\mathrm{E}\widehat{\mu}_X = \mathrm{E}\left(\frac{1}{N}\sum_{t=1}^{N} X_t\right) = \frac{1}{N}\sum_{t=1}^{N}\mathrm{E}X_t = \frac{1}{N}\sum_{t=1}^{N}\mu_X = \mu_X \tag{2.72}$$

is obtained and, thus, $\hat{\mu}_X$ is an unbiased estimator for μ_X. For the variance of this estimator

$$\begin{aligned}\mathrm{Var}\hat{\mu}_X &= \mathrm{Var}\left(\frac{1}{N}\sum_{t=1}^{N} X_t\right) = \frac{1}{N^2}\left(\sum_{t=1}^{N}\mathrm{Var}X_t \;+\; 2\sum_{t<u}\mathrm{Cov}(X_t, X_u)\right)\\ &= \frac{1}{N^2}\sum_{t=1}^{N}\sum_{u=1}^{N}\mathrm{Cov}(X_t, X_u) = \frac{1}{N^2}\sum_{t=1}^{N}\sum_{u=1}^{N} c_X(t,u)\end{aligned}$$

is calculated by summing the covariances $c_X(t,u) = c_{tu}$ in the covariance matrix

$$\boldsymbol{C}_X \;=\; \begin{pmatrix} c_{11} & c_{12} & \ddots & c_{1N} \\ c_{21} & c_{22} & \ddots & c_{2N} \\ \ddots & \ddots & \ddots & \ddots \\ c_{N1} & c_{N2} & \ddots & c_{NN} \end{pmatrix}.$$

$\boldsymbol{C}_X$ is symmetric due to (2.7,1) with, (X_t) being stationary, identical values in each diagonal as proposed in the remarks to (2.9). Hence, the covariances $c_X(t,u)$ depend on the lag $\tau = t - u$ alone. When summing the diagonals

$$\begin{aligned}\mathrm{Var}\hat{\mu}_X &= \frac{1}{N^2}\left(\sum_{t=1}^{N} c_X(0) \;+\; 2\sum_{\tau=1}^{N-1}\sum_{t=1}^{N-\tau} c_X(\tau)\right)\\ &= \frac{1}{N^2} N c_X(0) + \frac{2}{N^2}\sum_{\tau=1}^{N-1}(N-\tau)c_X(\tau)\\ &= \frac{1}{N}\sum_{\tau=-(N-1)}^{N-1}\Big(\frac{N-\tau}{N}\Big)c_X(\tau)\end{aligned}$$

is obtained. What happens to $\mathrm{Var}\hat{\mu}_X$ when the time series becomes longer, i.e., when the number of observations increases?

$$\begin{aligned}\mathrm{Var}\hat{\mu}_X &= \frac{1}{N}\sum_{\tau=-(N-1)}^{N-1}\left(\frac{N-\tau}{N}\right)c_X(\tau) \le \frac{1}{N}\sum_{\tau=-(N-1)}^{N-1}\left(\frac{N-|\tau|}{N}\right)\big|c_X(\tau)\big|\\ &\le \frac{1}{N}\sum_{\tau=-(N-1)}^{N-1}\big|c_X(\tau)\big| \le \frac{1}{N}\sum_{\tau=-\infty}^{\infty}\big|c_X(\tau)\big|\end{aligned} \tag{2.73}$$

If the covariance function $c_X(\tau)$ converges absolutely then $\sum_{\tau=-\infty}^{\infty}\big|c_X(\tau)\big| = a < \infty$, and thus $\lim_{N\to\infty}\mathrm{Var}\hat{\mu}_X = \lim_{N\to\infty}(1/N)\sum_{\tau=-\infty}^{\infty}\big|c_X(\tau)\big| = 0$. Consequently, using the Chebyshev inequality (1.17), it is concluded that $\hat{\mu}_X$ is a consistent estimator for the constant expectation function μ_X, and (2.56) follows.

2.8.3 Properties of the Optimal Linear Prediction

(2.63,1) is derived by minimizing the mean square prediction error as defined in (2.61,2) under the assumption $\mu_X = 0$:

$$\begin{aligned}\mathrm{PVar}(\widehat{X}_{N+\tau}) &= \mathrm{E}\Big(\Big(X_{N+\tau} - \sum_{u=0}^{N-1} a_u X_{N-u}\Big)^2\Big) \\ &= \mathrm{Var}X_{N+\tau} - 2\sum_{u=0}^{N-1} a_u \mathrm{Cov}(X_{N+\tau}, X_{N-u}) \\ &\quad + \sum_{u=0}^{N-1}\sum_{v=0}^{N-1} a_u a_v \mathrm{Cov}(X_{N-u}, X_{N-v}).\end{aligned}$$

Thereafter the partial derivatives $(\mathrm{d}/\mathrm{d}a_u)\big(\mathrm{PVar}\widehat{X}_{N+\tau}\big)$ are calculated for all a_u. For example, if $u = 0$ then the following expressions are obtained for each of the three terms in the above sum:

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}a_0}\ \big(\ \mathrm{Var}X_{N+\tau}\big)\ =\ 0, \\ &\frac{\mathrm{d}}{\mathrm{d}a_0}\ \Big(\ -2\big(a_0\mathrm{Cov}(X_{N+\tau}, X_N) + a_1\mathrm{Cov}(X_{N+\tau}, X_{N-1}) + \ldots \\ &\qquad\qquad +a_{N-1}\mathrm{Cov}(X_{N+\tau}, X_1)\big)\Big)\ =\ -2\mathrm{Cov}(X_{N+\tau}, X_N), \\ &\frac{\mathrm{d}}{\mathrm{d}a_0}\ \Big(\ a_0\big(a_0\mathrm{Cov}(X_N, X_N) + a_1\mathrm{Cov}(X_N, X_{N-1}) + \ldots + a_{N-1}\mathrm{Cov}(X_N, X_1)\big) \\ &\qquad +a_1\big(a_0\mathrm{Cov}(X_{N-1}, X_N) + a_1\mathrm{Cov}(X_{N-1}, X_{N-1}) + \ldots\big) \\ &\qquad +a_2\big(a_0\mathrm{Cov}(X_{N-2}, X_N) + a_1\mathrm{Cov}(X_{N-2}, X_{N-1}) + \ldots\big) \\ &\qquad \vdots \\ &\qquad +a_{N-1}\big(a_0\mathrm{Cov}(X_1, X_N) + a_1\mathrm{Cov}(X_1, X_{N-1}) + \ldots\big)\Big) \\ &\quad =\ 2a_0\mathrm{Cov}(X_N, X_N) + a_1\mathrm{Cov}(X_N, X_{N-1}) + \ldots + a_{N-1}\mathrm{Cov}(X_N, X_1) \\ &\qquad +a_1\mathrm{Cov}(X_{N-1}, X_N) \\ &\qquad +a_2\mathrm{Cov}(X_{N-2}, X_N) \\ &\qquad \vdots \\ &\qquad +a_{N-1}\mathrm{Cov}(X_1, X_N) \\ &\quad =\ 2\big(a_0\mathrm{Cov}(X_N, X_N) + a_1\mathrm{Cov}(X_N, X_{N-1}) + \ldots\big).\end{aligned}$$

Adding the three terms and setting the sum equal to zero, the following equation (in vector notation) with the first *optimality condition* of the optimal linear prediction

$$\big(\mathrm{Cov}(X_N, X_N), \ldots, \mathrm{Cov}(X_N, X_1)\big) \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{pmatrix} = \mathrm{Cov}(X_{N+\tau}, X_N) \qquad (2.74)$$

is obtained. This calculation is repeated for the partial derivatives with respect to the remaining weights $a_1, \ldots, a_{N-1}$ and the following equations are arrived at (with C written for the covariance symbol Cov):

$$\begin{pmatrix} \mathrm{C}(X_N, X_N) & \ldots & \mathrm{C}(X_N, X_1) \\ \mathrm{C}(X_{N-1}, X_N) & \ldots & \mathrm{C}(X_{N-1}, X_1) \\ \vdots & \ddots & \vdots \\ \mathrm{C}(X_1, X_N) & \ldots & \mathrm{C}(X_1, X_1) \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{N-1} \end{pmatrix} = \begin{pmatrix} \mathrm{C}(X_{N+\tau}, X_N) \\ \mathrm{C}(X_{N+\tau}, X_{N-1}) \\ \vdots \\ \mathrm{C}(X_{N+\tau}, X_1) \end{pmatrix}.$$

From these equations, the vector $(a_0, a_1, \ldots, a_{N-1})^T = \boldsymbol{a}$ containing the weights of the optimal linear prediction is obtained as the product of the inverse of the covariance matrix and the vector $\boldsymbol{c}_{N,\tau}$ containing the covariances of the random variable to be predicted with the random variables observed. If the covariance matrix $\boldsymbol{C}_X$ is positive definite as defined in the remarks to (2.7) then its inverse exists.

$$\boldsymbol{C}_X \boldsymbol{a} \;=\; \boldsymbol{c}_{N,\tau} \qquad\qquad \boldsymbol{a} \;=\; \boldsymbol{C}_X^{-1} \boldsymbol{c}_{N,\tau} \qquad (2.75)$$

Thus, (2.63,1) is derived. (2.63,2) is shown calculating the mean square prediction error from its definition (2.62)

$$\begin{aligned} \mathrm{PVar}\widehat{X}_{N+\tau} &= \mathrm{E}\big((X_{N+\tau} - \widehat{X}_{N+\tau})^2\big) \\ &= \mathrm{E}\big((X_{N+\tau} - \boldsymbol{a}^T \boldsymbol{X})^2\big) \\ &= \mathrm{E}X_{N+\tau}^2 - 2\boldsymbol{a}^T \big(\mathrm{E}(X_{N+\tau} X_N), \ldots, \mathrm{E}(X_{N+\tau} X_1)\big)^T + \mathrm{E}\big((\boldsymbol{a}^T \boldsymbol{X})^2\big) \\ &= \sigma_X^2 - 2\boldsymbol{a}^T \boldsymbol{c}_{N,\tau} + \boldsymbol{a}^T \boldsymbol{C}_X \boldsymbol{a} \end{aligned}$$

and then substituting the solution $\boldsymbol{a}$ obtained in (2.75):

$$\begin{aligned} \mathrm{PVar}(\widehat{X}_{N+\tau}) &= \sigma_X^2 - 2(\boldsymbol{C}_X^{-1} \boldsymbol{c}_{N,\tau})^T \boldsymbol{c}_{N,\tau} + \boldsymbol{C}_X^{-1} \boldsymbol{c}_{N,\tau})^T \boldsymbol{C}_X \boldsymbol{C}_X^{-1} \boldsymbol{c}_{N,\tau} \\ &= \sigma_X^2 - 2\boldsymbol{c}_{N,\tau}^T \boldsymbol{C}_X^{-1} \boldsymbol{c}_{N,\tau} + \boldsymbol{c}_{N,\tau}^T \boldsymbol{C}_X^{-1} \boldsymbol{C}_X \boldsymbol{C}_X^{-1} \boldsymbol{c}_{N,\tau} \\ &= \sigma_X^2 - \boldsymbol{c}_{N,\tau}^T \boldsymbol{C}_X^{-1} \boldsymbol{c}_{N,\tau} \\ &= c_X(0) - \boldsymbol{c}_{N,\tau}^T \boldsymbol{a}. \end{aligned} \qquad (2.76)$$

The expectation of the prediction as proposed in (2.63,3) is derived using (1.15).

$$\mathrm{E}\widehat{X}_{N+\tau} = \mathrm{E}(\boldsymbol{a}^T \boldsymbol{X}) = \sum_{u=0}^{N-1} a_u \mathrm{E}X_{N-u} = \left(\sum_{u=0}^{N-1} a_u\right) \mu_X = 0 = \mathrm{E}X_{N+\tau} \qquad (2.77)$$

Equation (2.74) with the first optimality condition is obtained deriving the mean square prediction error in (2.62) with respect to the first weight a_0. Since the calculation of the expectation is linear, it can be interchanged with the calculation of the derivatives also being linear. If we proceed in this order, i.e., by first calculating the derivatives and then the expectations, we arrive at optimality condition j as follows:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}a_j}\mathrm{E}\Big(\big(X_{N+\tau} - (a_0X_N + a_1X_{N-1} + \ldots + a_{N-1}X_1)\big)^2\Big) \\
&\quad = -2\mathrm{E}\Big(\big(X_{N+\tau} - (a_0X_N + a_1X_{N-1} + \ldots + a_{N-1}X_1)\big)X_{N-j}\Big) \\
&\quad = 0. \qquad (2.78)
\end{aligned}$$

Re-arranging

$$\begin{aligned}
&\mathrm{E}\big(X_{N+\tau}X_{N-j} - (a_0X_NX_{N-j} + a_1X_{N-1}X_{N-j} + \ldots + a_{N-1}X_1X_{N-j})\big) = 0 \\
&a_0\mathrm{E}(X_NX_{N-j)} + a_1\mathrm{E}(X_{N-1}X_{N-j}) + \ldots \\
&\qquad\qquad + a_{N-1}(X_1X_{N-j)} = \mathrm{E}(X_{N+\tau}X_{N-j})
\end{aligned}$$

and substituting $j = 0, 1, \ldots, N-1$ the equations for the optimal linear prediction proposed in (2.75) are arrived at.

The derivative in (2.78) is identically zero, when the prediction error (the mean square prediction error (2.63,2) is the variance of the prediction error) $\hat{e}_{N+\tau} = X_{N+\tau} - \widehat{X}_{N+\tau} = X_{N+\tau} - (a_0X_N + a_1X_{N-1} + \ldots + a_{N-1}X_1)$ is not correlated with X_{N-j}, and thus the *orthogonality conditions* of the optimal linear prediction are obtained in (2.79):

$$\mathrm{E}(\hat{e}_{N+\tau}X_{N-j}) = \mathrm{Cov}(\hat{e}_{N+\tau}, X_{N-j}) = 0, \qquad j = 0, 1, \ldots, N-1 \qquad (2.79)$$

The orthogonality conditions (2.79) can be obtained with linear algebra when the random variables $X_{N+\tau}, X_1, \ldots, X_N$ are assumed to be in a linear vector space (specifically in $L^2(\Omega, F)$, as defined in (7.10)). Under this assumption, the prediction $\widehat{X}_{N+\tau}$ is the projection of $X_{N+\tau}$ in the subspace of the linear combinations of $X_1, \ldots, X_N$ as shown in [21]. Using the orthogonality conditions the mean square prediction error is easily obtained in (2.80):

$$\begin{aligned}
\mathrm{PVar}\widehat{X}_{N+\tau} &= \mathrm{Var}\hat{e}_{N+\tau} = \mathrm{E}(\hat{e}_{N+\tau}\hat{e}_{N+\tau}) \\
&= \mathrm{E}\big(X_{N+\tau} - (a_0X_N + a_1X_{N-1} + \ldots + a_{N-1}X_1)\big)^2 \\
&= \mathrm{E}\Big(\hat{e}_{N+\tau}\big(X_{N+\tau} - (a_0X_N + a_1X_{N-1} + \ldots + a_{N-1}X_1)\big)\Big) \\
&= \mathrm{E}\Big(\hat{e}_{N+\tau}X_{N+\tau}\Big) \\
&= \mathrm{E}\Big(\big(X_{N+\tau} - (a_0X_N + a_1X_{N-1} + \ldots + a_{N-1}X_1)\big)X_{N+\tau}\Big) \\
&= \mathrm{E}\Big(X_{N+\tau}X_{N+\tau} - (a_0X_NX_{N+\tau} + \ldots + a_{N-1}X_1X_{N+\tau})\Big) \\
&= c_X(0) - \boldsymbol{c}_{N,\tau}^T\boldsymbol{a} \qquad (2.80)
\end{aligned}$$

with the vectors $\boldsymbol{c}_{N,\tau}$ and $\boldsymbol{a}$ as defined in (2.62) and (2.63). From $\mathrm{PVar}\widehat{X}_{N+\tau} = \mathrm{E}(\hat{e}_{N+\tau}X_{N+\tau}) = \mathrm{Cov}(\hat{e}_{N+\tau}, X_{N+\tau})$ in the above derivation of (2.80), it is concluded that the mean square prediction error is the covariance of the prediction error and random variable to be estimated.

The orthogonality conditions are the mathematical formulation of the intuitive idea that a linear prediction could be improved if $\mathrm{Cov}(\hat{e}_{N+\tau}, X_{N-j}) \neq 0$ for at least one j, by taking advantage of this very same covariance. Hence, it is also plausible that the covariance of the prediction and the prediction error are identically zero:

$$\mathrm{E}\big(\hat{e}_{N+\tau}\widehat{X}_{N+\tau}\big) = \mathrm{Cov}\big(\hat{e}_{N+\tau}, \widehat{X}_{N+\tau}\big) = 0. \tag{2.81}$$

If this were not the case, a prediction with a smaller mean square error would become possible, and thus the prediction could be improved. (2.81) is derived (e.g., in [21]) with linear algebra in the linear vector space $L^2(\Omega, F)$ as defined in (7.10).

2.8.4 The Conditional Expectation as Optimal Prediction

In the definition of the optimal prediction in (2.61) the predicted value $\hat{x}_{N+\tau}$ is assumed to belong to the same realisation as the observed values. Assuming the realisation with the observations, a prediction can be obtained if the probability of the event can be calculated that $\hat{x}_{N+\tau}$ is a realisation of the estimator $\widehat{X}_{N+\tau} = \widehat{X}_{N+\tau}(X_1, \ldots X_N)$. With the probabilities of all such events being possible, i.e., with the conditional probability density $f_{X_{N+\tau}}(X_{N+\tau}|X_1 = x_1, \ldots X_N = x_N)$, the conditional expectation $\mathrm{E}\big(X_{N+\tau}|X_1, \ldots, X_N\big)$ of $X_{N+\tau}$ can now be calculated.

Assuming a stationary process (X_t), this conditional expectation is shown to be, for example in [113], the optimal prediction of (X_t) with lag τ, its mean square prediction error as proposed in (2.82)

$$\mathrm{PVar}\widehat{X}_{N+\tau} = \mathrm{E}\Big(\big(X_{N+\tau} - \mathrm{E}\big(X_{N+\tau}|X_1, \ldots, X_N\big)\big)^2\Big) \tag{2.82}$$

being minimal. The conditional expectation $\mathrm{E}\big(X_{N+\tau}|X_1, \ldots, X_N\big)$ in (2.82) is usually, however, not a linear combination of the $X_1, \ldots, X_N$ and, thus, in general, the optimal prediction of a stationary process (the prediction with the least mean square error as defined in (2.61)) is, generally, not the optimal linear prediction as defined in (2.62).

The conditional expectation as the optimal prediction is theoretically straightforward; for practical applications, however, it is only useful for special cases as the conditional probability density is calculated from the joint probability density of the random variables $X_1, \ldots, X_N$ which is usually very difficult to obtain.

A first exception is the normal process as defined in (2.3) with its conditional expectations being optimal linear predictions. This follows from the remark to the definition of the multivariate normal distribution (1.34), namely,

that linear combinations of multivariate normally distributed random variables are normally distributed, as well.

A second exception is the white noise process with a conditional expectation that can be easily calculated. Let (W_t) be a white noise process as defined in (2.11) with expectation μ_W, variance σ_W^2 and the probability density $f_t(w_t)$ of a W_t, and let $(w_1, \ldots, w_N)$ be observations of a realisation of (W_t). The prediction of (W_t) with lag τ, $\widehat{W}_{N+\tau}$, is easily calculated as conditional expectation of the process for a future time point $N+\tau$ since the joint probability density of $W_1, \ldots, W_N$ is the product of the densities of these random variables:

$$\begin{aligned}
\widehat{W}_{N+\tau} &= \mathrm{E}(W_{N+\tau} | W_1 = w_1, \ldots, W_N = w_N) \\
&= \int_{-\infty}^{\infty} w_{N+\tau} f_{N+\tau}(w_{N+\tau} | w_1, \ldots, w_N) \mathrm{d}w_{N+\tau} \\
&= \int_{-\infty}^{\infty} w_{N+\tau} \frac{f_{1,\ldots,N,N+\tau}(w_1, \ldots, w_N, w_{N+\tau})}{f_{1,\ldots,N}(w_1, \ldots, w_N)} \mathrm{d}w_{N+\tau} \\
&= \int_{-\infty}^{\infty} w_{N+\tau} \frac{f_1(w_1) \ldots f_N(w_N) f_{N+\tau}(w_{N+\tau})}{f_1(w_1) \ldots f_N(w_N)} \mathrm{d}w_{N+\tau} \\
&= \int_{-\infty}^{\infty} w_{N+\tau} f_{N+\tau}(w_{N+\tau}) \mathrm{d}w_{N+\tau} = \mathrm{E}W_{N+\tau} = \mu_W \qquad (2.83)
\end{aligned}$$

For the mean square prediction error, $\mathrm{PVar}\widehat{W}_{N+\tau} = \mathrm{E}\big((W_{N+\tau} - \mu_W)^2\big) = \mathrm{Var}W_{N+\tau} = \sigma_W^2$ is obtained. A random variable W_t in a white noise process (W_t) cannot be estimated with a variance less than (its variance) σ_W^2. Hence, prediction (2.83) is optimal.

The mean square prediction error of $\widehat{W}_{N+\tau}$ as obtained above is identical with the mean square prediction error (2.63,2) of a linear prediction for a non-observed random variable in a white noise process, as is concluded from the following rationale. The negative-sloping diagonals in the covariance matrix of a stationary stochastic process contain identical values, as proposed in the remarks to (2.9), with the values in the subdiagonals being not identically zero in the general case. Consequently, the mean square error of an optimal linear prediction of a stationary process is less than the constant variance function of the process due to (2.63,2) or (2.76). In the case of a white noise process, however, the values in the subdiagonals of the covariance matrix are identically zero, and, consequently, the mean square error of an optimal linear prediction is the variance of the process.

For example, the observations of Jones' NAO index as plotted in Fig. 2.10 are found to be a time slice in a realisation of a white noise process with estimated expectation 0.476 and variance 1.161 in the remarks to Fig. 2.18. This result is ascertained (i) by the simulation experiment performed in Sect. 2.5.2, (ii) using the Ljung-Box statistic in Problem 2.33 and (iii) by the periodogram tests performed in Sect. 8.2.2. Using this model and (2.83), optimal predic-

tions 0.476 for the years 2001, 2002, ... are arrived at. These predictions have mean square error 1.161, as is concluded from the remarks to (2.83).

This prediction for the years following 2000 is clearly below the local (2.55,2) mean of Jones' index in the last decade of the last century, as can be seen in Figs. 2.10 and 2.20, since the mean used in the prediction is calculated from all observations available. However, a prediction using a local mean will be superior to the one obtained above using (2.83) on condition that the index continues to fluctuate about the local level attained in year 2000.

If Jones' record has the future assumed above then the outcome of the simulation experiment performed in Sect. 2.5.2 can be extrapolated:

1. an increasing number of two-sided t-tests as those in Fig. 2.20 comparing the means in time slices of 30 years duration can be performed and
2. those t-tests which are performed in addition to the 117 ones in Fig. 2.20 will result in significance levels below 0.05, as is obvious from Fig. 2.20.

Asssuming the above future therefore, the number of t-tests being significant at the 0.05 level increases with increasing length of the record. As a further consequence, it is likely that the H_0 hypothesis (the observations stem from a white noise process having a constant mean) will be discarded. If the H_0 hypothesis is discarded then Jones' index is found to be non-stationary in its mean. Despite this outcome of the experiment, it could be productive to assume that the fluctuations in the mean thus found are owed to a stationary phenomenon. Under this assumption, the model for Jones' index becomes a stationary stochastic process having a constant mean and covariances not being identically zero for lags $\tau \neq 0$, and the fluctuations in its mean can be analysed using the methods introduced in Chaps. 8, 9 and 10.

The reader is invited to repeat this simulation experiment when more data become available, using contiguous time slices of length 30 (as in Fig. 2.20) and also of lengths 25, 20 and 15.

2.9 Problems

2.1. Calculate the retreat of the Rhone glacier from the differences in the position of its tongue given for every year since 1880 in the file */path/rhone.dat.* Hint: use the R function `cumsum()`.

2.2. Assume that the NAO index (Jones), as plotted in Fig. 2.10 and made available using

```
naofile <- "path/nao12monthsjones.dat"
naoformat <- list(year=0,jan=0,feb=0,mar=0,apr=0,mai=0,jun=0,
                  jul=0,aug=0,sep=0,oct=0,nov=0,dec=0,yearave=0)
naoall <- scan(naofile,naoformat)
naoyear <- naoall$year
naojones <- rep(NA,length(naoyear))
```

```
for (i in (1:(length(naoyear)-1))) {naojones[i] <- (naoall$dec[i]+
   naoall$jan[i+1]+naoall$feb[i+1]+naoall$mar[i+1])/4 }
naojonests <- ts(naojones[3:179],start=1824,frequency=1,)
```

stems from a white noise process. Estimate the mean and the variance of this process. Is the process normal as defined in (2.3)?

2.3. Simulate an AR[1] process using definition (2.13) and the values 0.99, 0.80, 0.3 for the parameter a, the value 0.0 for the expectation μ_W, and the values 4.0, 1.0 and 0.25 for variance σ_W^2 of the generating white noise process, and then plot the empirical correlation functions obtained from the simulations. Calculate and plot the theoretical correlation functions of these AR[1] models.

2.4. Show that the infinite sums in (2.18) are equal.

2.5. Simulate an AR[1] process using definition (2.13) and the values 1.001, 1.01, 1.1 for the parameter a, the value 0.0 for the expectation μ_W, and the value 1.0 for the variance σ_W^2 of the generating white noise process.

2.6. Simulate random walk processes using (2.20). For what time points do you obtain the mimina of the simulated realisations? For what time points the maxima?

2.7. Plot the yearly values in the Basel temperature series to obtain the plot in Fig. 2.13 (above). These data are found in the text file `/path/basel.dat`. A line in `/path/basel.dat` contains 14 values, the first one is the year of the measurement followed by twelve monthly values and the last one is the yearly value. This file can be read with the following expressions

```
bsfn <- "/path/basel.dat"
z <- list(yr=0, jan=0, feb=0, mar=0, apr=0, may=0, jun=0,
jul=0, aug=0, sep=0, oct=0, nov=0, dec=0, year=0)
basel <- scan(bsfn,z)
tsbtemp <- ts(basel$year,start=1755,frequency=1,)
```

Use `diff()` to calculate the differences of time series `tsbtemp` to obtain the plot in Fig. 2.14 (above).

2.8. Simulate a MA[1] process using definition (2.21) and the values 0.9 and 0.4 for the parameter b, the value 0.0 for the expectation μ_W, and the values 3.0, 1.0, 0.5 and 0.1 for the variance σ_W^2 of the generating white noise process. Compare the simulated realisations with the differences of the yearly values in the Basel temperature series as plotted in Fig. 2.15. Plot the empirical correlation functions obtained from the simulations together with the theoretical correlation functions of these MA[1] models.

2.9. Simulate a few realisations of an MA[1] model with $b = 0.9$, $\mu_W = 0.0$ and $\sigma_W = 0.7$ and then compare the simulated realisations with the plot of the differences of Basel temperature yearly values in Fig. 2.15. Apply `cumsum()` to each simulation and then compare the results with the plot of the yearly values in the Basel temperature series as plotted in Fig. 2.13 (above).

2.10. Obtain the expectation and covariance functions of the AR[1] and MA[1] models from the moment functions of the linear process proposed in (2.26).

2.11. Given is the first order linear differential equation $x(t) + ax'(t) = w(t)$. Compare your solution with the convolution integral (2.28,1). Roughly draw the functions in this convolution integral assuming the known function is the unit step function, i.e., $w(t) = 1$ for $t \geq 0$ and $w(t) = 0$ for $t < 0$.

2.12. a) Show that the shift operator, as defined in (2.31), is a linear transformation, as defined in (2.30). b) Show that the derivation is a time-invariant transformation as defined in (2.32). c) Show that the multiplication with a function g is not a time-invariant transformation.

2.13. Calculate $(1, 5, 4, 1, 1, 1, 1, 8, 3, 1) * (0.333, 0.333, 0.333)$ using the paper strip device as demonstrated in (2.38). Repeat the calculations with `filter()` using the arguments `sides=1` and `sides=2`. Modify the paper strip device to allow for the calculation of the sum $c_i = \sum_{j=-\infty}^{\infty} a_j b_{i+j}$ in the remarks to (2.35).

2.14. Calculate $(1, 5, 4, 1, 1, 1, 1, 8, 3, 1) * (1, -1)$ using the paper strip device as demonstrated in (2.38). Repeat the calculations with `filter()` and `diff()`. Calculate the second order differences as defined in the remarks to (2.39). Why can second and higher order differences be defined recursively?

2.15. Smooth the yearly values in the Basel temperature series with a moving average over 11 years.

2.16. Smooth the wind speed observations in the turbulent atmospheric flow above Anwandstrasse as plotted in Fig. 2.1.

2.17. Apply the R function `diff()` to the yearly values in the Basel temperature series. Then apply `cumsum()` to the result obtained with `diff()`.

2.18. Check the identities (2.43) using the paper strip device as demonstrated in (2.38).

2.19. An unknown sequence (s_t) is convolved with $(a_t) = (1, 2a, 3a^2, \ldots, (t+1)a^t, \ldots)$, $t = 0, 1, 2, \ldots$, to a sequence (r_t). Show that (s_t) can be obtained from (r_t) and $(a_t)^{-1}$ using the absolute convergence of $(a_t) = (\ldots, (t+1)a^t, \ldots)$ for $|a| < 1$, since

$$\lim_{t\to\infty}\left(\frac{(t+2)a^{t+1}}{(t+1)a^t}\right) \quad = \quad |a| \lim_{t\to\infty}\left|\frac{t+2}{t+1}\right| \quad = \quad |a|$$

Hint: the solution can be found by trial and error. The examples in (2.41) and (2.43) (with minor changes) are possible inverse sequences $(a_t)^{-1}$ and can be checked with the paper strip device as demonstrated in (2.38).

2.20. Show (2.44) by multiplying the polynomials $A(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \ldots + a_p x^p$ and $B(x) = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \ldots + b_q x^q$

$$\begin{aligned} C(x) = A(x)B(x) &= a_0 b_0 + a_0 b_1 x + a_0 b_2 x^2 + a_0 b_3 x^3 + a_0 b_4 x^4 + \ldots \\ &\quad \ldots \\ &= \sum_{i=0}^{p+q} c_i x^i, \qquad \text{where} \quad c_i = \sum_{j=0}^{i} a_j b_{i-j} \end{aligned}$$

and comparing the result with (2.28).

2.21. Select weights from different regions of the a_1-a_2-plane in Fig. 2.17, e.g., `a <- c(1.0, -.5)`, `a <- c(1.0, 1.5)` and `a <- c(-0.2, 0.5)`, and obtain the roots of the characteristic polynom of the AR[2] model with R function `polyroot()`. Simulate realisations of the (stationary) stochastic processes and plot the empirical correlation functions of the realisations. Compare their theoretical with their empirical correlation functions.

2.22. The simulations for the example in Figs. 2.4, 2.5, 2.6 and 2.7 were obtained using the R function `arima.sim()` and the AR[2] model $X_t = -1.8X_{t-1} - 0.95X_{t-2} + W_t$, $W_t \sim N(0.0, 1.148412)$. The AR[2] model is defined in (2.48), the simulations being produced with

```
wnlength <- 400
wnvar <- 1.148412
w <- rnorm(wnlength,mean=0,sd=sqrt(wnvar))
wts <- ts(w,start=1,frequency=1,)
ar2 <- arima.sim(model=list(order=c(2,0,0),
                 ar=c(-1.8,-0.95)), n=wnlength, innov=wts)
```

Show that (X_t) is stationary. Show that (X_t) is a normal process.

2.23. $X_t = 2.7607X_{t-1} - 3.8106X_{t-2} - 2.6535X_{t-3} - 0.9838X_{t-4} + W_t$ with $\mu_W = 0$ and $\sigma_W^2 = 1$ is an AR[4] process. Is this process stationary?

2.24. Find the weights of a stationary AR[3] model by trial and error. How many stationary models do you obtain within five minutes?

2.25. Do the wind speed observations (x_t) in Fig. 2.1 stem from a process stationary in the expectation function? Hint: calculate the empirical means $\hat{\mu}_X$ and standard deviations $\hat{\sigma}_X$ (i) from the first and second half-series and (ii) from the first, second and third part of the series each of which contain 600 observations. Then plot $\hat{\mu}_X$ together with $\hat{\mu}_X \pm 2\hat{\sigma}_X$ for the periods from which the empirical moments are calculated.

2.26. Consider the time series of the differences in Fig. 1.8 obtained from the Zurich-Reckenholz pyranometer comparison experiment. Are there any periods in which the differences are stationary in the expectation function? Calculate the empirical first moment function using (2.1) from the observations in the selected period (periods).

2.27. The differences of the monthly values of SWIR as plotted in Fig. 2.19 are calculated in Problem 1.11. Calculate the empirical variances of the difference time series from the parts before and after 1975, resp. 1980.

2.28. Re-evaluate your answers to the questions in Problem 1.11 using the results from Problem 2.27.

2.29. Let (X_t), $t \in T$, be a stationary stochastic process with $\mu_X(t) = \mu_X = 0$ and an absolutely convergent covariance function $c_X(\tau)$, and let V be a random variable with $\mu_V = 0$ and σ_V^2, together with $\mathrm{Cov}(X_t, V) = 0$ for all $t \in T$. Show that $Y_t = X_t + V$, $t \in T$, is a stationary process, and calculate $\mathrm{E}\hat{\mu}_Y(t)$ (using (2.72)) and $\mathrm{Var}\hat{\mu}_Y$ (using(2.73)).

2.30. The Climate Research Unit, University of East Anglia, Norwich, UK, makes some climate datasets available which can be downloaded from the internet address given in [75]. For example, when downloading the time series of the global average temperatures, the text file `/path/tavegl.dat` with two lines for each year is obtained: the first line contains deviations of the global temperature averages from a reference value for the months and the year, the second line codes indicating how representative the values in the first line are.

```
1856 -0.22 ... february through to december ... -0.36
1856 14 16 15 14 14 13 13 14 16 15 15 16
1857 -0.30 ... february through to december ... -0.48
1857 18 16 18 16 17 16 16 15 17 17 15 15
1858 -0.62 ... february through to december ... -0.43
1858 17 15 16 14 16 15 16 15 15 14 14 15
...
...
```

This file is read with the R expressions

```
tempgafile <- "/path/tavegl.dat"
tempgaformat <- list(year=0, jan=0, feb=0, mar=0, apr=0, mai=0,
      jun=0, jul=0, aug=0, sep=0, oct=0, nov=0, dec=0, tjm=0,
      year1=0, jan1=0, feb1=0, mar1=0, apr1=0, mai1=0, jun1=0,
      jul1=0, aug1=0, sep1=0, oct1=0, nov1=0, dec1=0)
tempga <- scan(tempgafile,tempgaformat,multi.line=T)
```

to obtain the R object `tempga` with the vectors of the monthly and yearly values of global temperature anomalies (deviations from a reference value). The time series of the yearly values and the time series of their differences as defined in (2.39,2) are plotted in Fig. 2.23 together with the empirical correlation functions.

The empirical correlation function of the yearly global temperature values decays slowly (until lag 40) with regard to the length of the time series (150 years). For a comparison, the empirical correlation function of the wind speeds as plotted in Fig. 2.3 decays rapidly (until lag 30) with regard to the length of the time series (1800 seconds). Why do the empirical correlations of

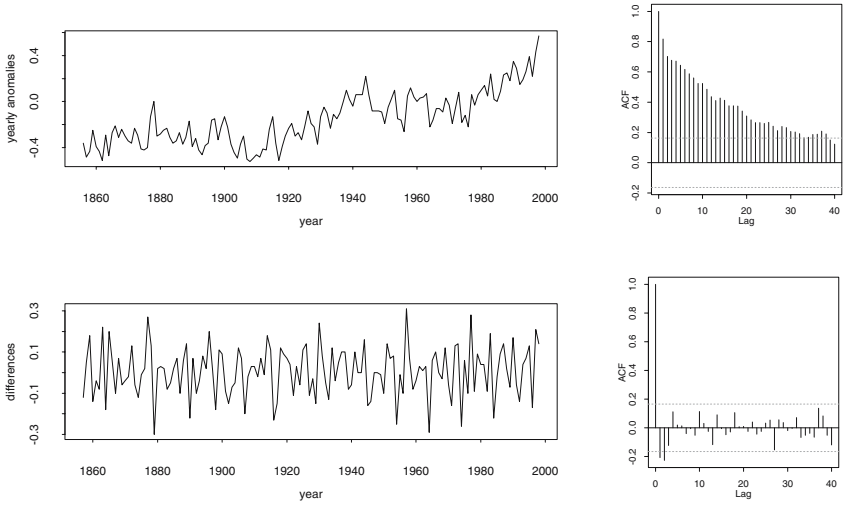


Fig. 2.23. Global average temperature since 1856: deviations from a reference value (anomalies) in ^{0}Celsius (above) and differences (below) with empirical correlation functions.

the yearly global temperature values decay so slowly? Is a stationary AR[1] model suitable for this time series?

Are the differences of the yearly global temperature anomalies as plotted below, in Fig. 2.23, a time slice from a realisation of a white noise process?

2.31. A description of the sunspot numbers can be obtained from [100]. From the address given there the sunspot numbers can be downloaded, and, after minor changes using a text editor, made available as R time series of the yearly and monthly values. Do the yearly values of the sunspot numbers stem from a stationary process?

2.32. The differences of the daily values obtained in the Zurich-Reckenholz comparison experiment described in Sect. 1.1 are plotted in Fig. 2.24 (above) for the year 1991. Are these differences a time slice from a realisation of a stationary stochastic process? The difference time series `difts` for the period of the comparison experiment is obtained using the R expressions described in Sect. 1.4; from this R time series object, and with

```
dif1991 <- window(difts,c(1991,1),c(1991,365))
```

the differences as plotted in Fig. 2.24 (above) are obtained. Unfortunately, `difts` as well as `dif1991` contain missing values, and, thus, using

```
acf(dif1991, lag.max=20, type="correlation", plot=T)
> Problem in acf(dif1991, lag.max = 20, type = "correla..:
  There are missing values in midst of time series
```

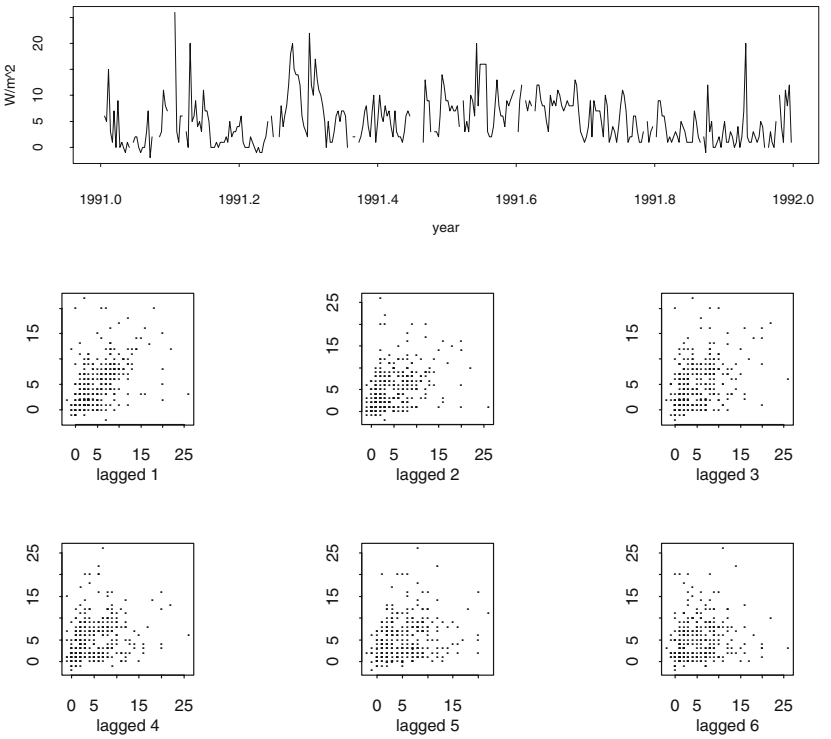


Fig. 2.24. Time series of the differences of the daily MeteoSwiss and ETH pyranometer values obtained in the Zurich-Reckenholz comparison experiment for the year 1991 in Wm^{-2} (above), lag τ scatterplots of the differences for $\tau = 1, \ldots, 6$ (below).

the empirical correlation functions cannot be successfully obtained with `acf()`. Consequently,

1. estimate these empirical correlations from the lag τ, $t = 1, \ldots, 6$, scatterplots in Fig. 2.24 (below), after having compared the lag τ scatterplots in Fig. 2.2 with the empirical correlations in Fig. 2.3
2. use lag τ scatterplots to see if these differences are a time slice from a realisation of a stationary stochastic process.

2.33. Compute the Ljung-Box statistic as defined in (2.60) for the NAO index (Jones) as plotted in Fig. 2.10 for $K = 10$ and $K = 20$ and then perform the tests.

2.34. The covariance function of a stochastic process is positive semi-definite and symmetric, these properties being required in (2.7). If the process (X_t)

is stationary and if $t = \ldots, -1, 0, 1, \ldots$, then $\sum_{t=1}^{n} \sum_{u=1}^{n} b_t b_u c_X(t,u) \geq 0$ and $c_X(\tau) = c_X(-\tau)$ are obtained. Are these properties also pertinent to the empirical covariance function $\hat{c}_X(\tau)$ calculated using (2.1,3) from a time slice (x_t), $t = 1, \ldots, N$, of a realisation (x_t), $t = \ldots, -1, 0, 1, \ldots$ of a stationary stochastic process $X(t)$?

$\sum_{t=1}^{n} \sum_{u=1}^{n} b_t b_u \hat{c}_X(t,u) \geq 0$ is obtained as shown in [95]. Let (i) $b_u = x_u$, with x_u being observations of X_u, $u = 1, \ldots, N$, and (ii) (W_t) be a white noise process with $\mu_W = 0$ and $\sigma_W^2 = 1/N$. Further, assume $\mu_X = 0$ and $\hat{\mu}_X = \mu_X = 0$ as in (2.1,3). Under these assumptions, $Y_t = \sum_{u=1}^{N} b_u W_{t-u}$ exists since a finite linear combination of random variables becomes itself a random variable. Thus, (Y_t) is a linear process as required in (2.25) and $c_Y(\tau)$ can be calculated using (2.26): $c_Y(\tau) = \sigma_W^2 \sum_{u=1}^{N-\tau} b_u b_{u+\tau} = \frac{1}{N} \sum_{t=1}^{N-\tau} x_t x_{t+\tau} = \hat{c}_X(\tau)$ is obtained, because $b_u = x_u$, $\sigma_W^2 = 1/N$ and $\hat{\mu}_X = \mu_X = 0$. Derive from these results that $\hat{c}_X(\tau)$ is positive semidefinite. Show that $\hat{c}_X(\tau) = \hat{c}_X(-\tau)$.

2.35. Calculate an optimal linear prediction for the wind speed in the turbulent atmospheric flow above the Anwandstrasse in Zurich at the time points $t = 1801$ and $t = 1802$ using (x_t), $t = 1781, 1782, \ldots, 1800$, from the time series (x_t), $t = 1, 2, \ldots, 1800$, as plotted in Fig. 2.1 and using the empirical correlation function in Fig. 2.3.

2.36. The diagnostic graphical procedures and diagnostic tests for the stationarity of a time series as described in Sect. 2.5.1 require that the time slices for the calculation and comparison of the empirical moment functions are defined prior to the analysis. In a simulation experiment reflecting this requirement for the NOA index (Jones) time series as plotted in Fig. 2.10, a-priori defined time slices, e.g., from 71 through to 100 and from 101 through to 130, of every realised sequence of values from the normal white noise process estimated in the remarks to Fig. 2.18, are compared with a t-test as follows:

```
countsign <- 0       #counter
for (case in (1:1000)) { #1000 simulations in a few seconds on a PC
  w <- rnorm(175, mean=0.4713,sd=sqrt(1.0858))
  tt <- t.test(w[71:100],w[101:130],alternative="two.sided")
  if(tt$p.value < 0.05) countsign <- countsign + 1          }.
```

Is the result in agreement with the 0.05 significance level of the t-test? Repeat the experiment for time slices from 1 through to 30 and from 101 through to 130.

3 Linear Models for the Expectation Function

Many stochastic processes are not stationary due to a systematically increasing or decreasing expectation function. A systematic increase or decrease in an expectation function is called a *trend*. A trend is estimated, in this chapter, by means of estimating a linear model for the expectation function of a stochastic process using regression methods.

In Sects. 3.1 and 3.2, both, simple and multiple linear regression methods, are introduced by estimating linear models for the trends in two example time series: (i) in the yearly values of shortwave incoming radiation measured at Hohenpeissenberg station as plotted in Fig. 3.1 and (ii) in the yearly values in the Basel temperature series as plotted in Fig. 2.13. Are the estimates reconcilable with the time series? Are the trends significant? Answers to these questions can be found when the diagnostics introduced in Sect. 3.3 are applied to the estimates.

In Sect. 3.4, seasonal linear models are introduced. A *seasonal linear model* is fitted to a time series stemming from a stochastic process with seasonal fluctuations in its expectation function, i.e., the expectation function is assumed to be periodic as defined in (6.1), with a constant and known period. It is also possible to fit a linear model to a time series showing a trend and seasonal fluctuations or a trend and fluctuations due to some other influence. Examples are given in Sects. 3.4.2 and 3.5. Linear models for the expectation function of a stochastic process with bi-dimensional parameter are estimated in Sect. 3.6. The Problems follow in Sect. 3.7.

Having estimated a linear model for the expectation function $\mu_X(t)$ of a stochastic process $X(t)$, $\mu_X(t)$ can then be estimated for time points without observations. Estimating $\mu_X(t)$ (not $X(t)$) is a means to circumvent the inversion of the empirical covariance matrix $\hat{c}_X(\tau)$ when an optimal linear prediction is calculated as required in (2.63,1). Inverting $\hat{c}_X(\tau)$ is often judged to be time-consuming and possibly delicate, in the remarks concluding Sect. 2.7. Other prediction methods are introduced in Chaps. 4 and 5.

Regression methods are comprehensively dealt with in [44]. An introduction to the estimation of linear models in Splus (and in R, since the function `lm()` is also available in R) is given in [31].

3.1 Simple Linear Regression

Simple linear regression, as defined in Sect. 3.1.2, models a *response variable* as a linear function of a *predictor variable*. When the expectation function of a stochastic process is estimated, the response variable becomes the stochastic process and the predictor variable its parameter, as demonstrated with an introductory example in Sect. 3.1.1.

3.1.1 An Introductory Example

As an introductory example, a trend in the yearly means of shortwave incoming radiation (SWIR) at Hohenpeissenberg station ($11^0 1'$ E, $47^0 48'$ N, 990 m above mean sea level) is estimated. At Hohenpeissenberg station, SWIR has been measured since 1953 using a Kipp pyranometer. The yearly values calculated from these measurements are available up to year 1993 in `/path/swirhpb.dat`. From this file, the following R expressions

```
hpbfn <- "/path/swirhpb.dat"
z <- list(year=0, swir=0)
hohenpeissenberg <- scan(hpbfn,z)
hpbyear <- hohenpeissenberg$year
hpbswir <- hohenpeissenberg$swir
hpb <- ts(hpbswir,start=1953,end=1993,frequency=1)
```

produce R vectors `hpbyear` and `hpbswir` as well as R time series `hpb`. Using

```
acf(hpb,lag.max=10, type="correlation", plot=T)
plot(hpbyear,hpbswir,type="b",xlab="year",ylab="W/m^2")
```

the time series and its empirical correlation function are plotted above, in Fig. 3.1.

The yearly means of SWIR at Hohenpeissenberg station are not stationary since the time series decreases at a constant rate over the observational period lasting four decades: i.e., it has a negative decadal trend. A decadal trend (either positive or negative) in SWIR is due to a (possibly local) climate change: the atmosphere becomes more or less transparent to solar radiation as the cloud amount, the aerosols, etc. increase and/or decrease. The empirical correlation function (plot (b) in Fig. 3.1) of this time series is calculated as required in (2.1,4) even though the process is not stationary in its expectation function. Since the stationarity assumption is violated, the estimate is presumably erroneous.

Assume that the trend in SWIR at Hohenpeissenberg station can be described, in the mean, as a straight line and that the following R expression

```
hpbswir.fit <- lm(hpbswir ~ hpbyear)
```

computes the coefficients of this straight line. Then, `hpbswir ~ hpbyear` is a linear model (here a straight line) for the expectation function of SWIR at Hohenpeissenberg as written in R: `~` separates the response variable `hpbswir` on its left from the predictor variable `hpbyear` on its right. The coefficients of the straight line, however, are not written in R. This model is the argument

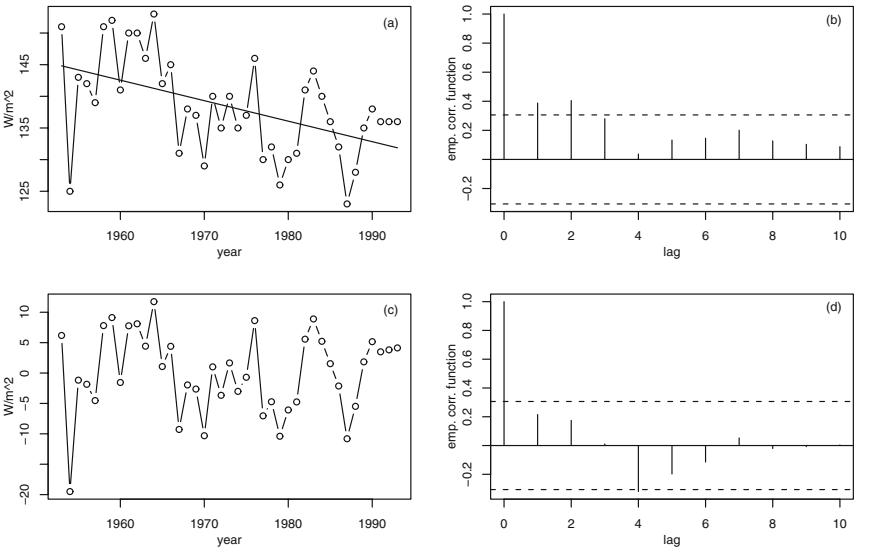


Fig. 3.1. Yearly values (in Wm^{-2}) of shortwave incoming radiation at Hohenpeissenberg station with estimated trend (above, a) with correlation function (above, b), and residuals (in Wm^{-2}) of the trend with correlation function (below, c, d).

in R function `lm()` generating `hpbswir.fit` as a result. Applying `coef()` to this resulting R object

```
coef(hpbswir.fit)
```

the coefficients are arrived at, and applying `predict()`

```
lines(hpbyear,predict(hpbswir.fit))
```

the estimated straight line is obtained (in plot (a) in Fig. 3.1). Exemplary, this straight line is a linear model for the expectation function of a stochastic process which is not stationary in its first moment function. The intercept of the linear model is 777 Wm^{-2}, i.e., the estimate for the expectation of SWIR at Hohenpeissenberg in year 0. This estimate is not feasible since the largest yearly values of SWIR at the surface of the earth measured so far amount to 280 Wm^{-2}. The estimate for the slope of the straight line is -0.3239 $\mathrm{Wm}^{-2}\mathrm{y}^{-1}$. Consequently, SWIR at Hohenpeissenberg station decreases in the mean by approximately 2.34% over 10 years in the period from 1953 through to 1993, its mean being 138 Wm^{-2}.

A negative trend of SWIR in the time slice from approximately 1950 through to approximately 1990 has been recorded at many stations in Europe [58]. Does this trend last until the year 2000? A preliminary answer is given in [152].

The yearly means of SWIR at Hohenpeissenberg station oscillate around the estimate for the expectation function: the differences between the observed values and the estimated expectations being called *empirical residu-*

als. These are the stochastic part in the linear model; in the linear model as written in R, however, they do not appear. The empirical residuals of a linear model estimated in R are calculated applying `resid()` to an object generated by `lm()`. For example, with

```
residts <- ts(resid(hpbswir.fit),start=1953,end=1993,frequency=1)
plot(hpbyear,resid(hpbswir.fit),type="b",xlab="year",ylab="W/m^2")
acf(residts, lag.max=10, type="correlation", plot=T)
```

the plots (c) and (d) below in Fig. 3.1 are obtained. Comparing plots (c) and (d) in Fig. 3.1 with those given in Fig. 2.22, it is concluded from (2.59) that the empirical residuals of the linear model for the yearly values of SWIR at Hohenpeissenberg station are a time slice from the realisation of a white noise process and are thus not correlated. Small correlations for lags 1 and 2, however, are seen in the empirical correlation function of the time series (plot (b) in Fig. 3.1). These correlations are obviously generated by the trend in SWIR at Hohenpeissenberg.

Another example for correlations generated by a trend is given in Fig. 2.23; there, however, the trend is removed by calculating the first differences of the time series, and not by calculating the residuals of a linear model. Further examples for removing a trend from a time series are given in Sect. 5.4.1.

The fluctuations of a meteorological variable from year to year, i.e., its inter-annual variability, is a property of the climate which is as important as the mean and a possible trend. An estimate for the inter-annual variability is the empirical variance of a stationary climatological time series, calculated as required in (2.1,2). If, however, a time series is not stationary in its first moment function, this estimate can be biased. For example, the empirical variance of the Hohenpeissenberg SWIR time series (60 $(\mathrm{Wm}^{-2})^2$) is larger than that of its residuals (45 $(\mathrm{Wm}^{-2})^2$), both calculated as required in (2.1,2). Both estimates of SWIR inter-annual variability include the contribution of the error in the measurements (Chap. 1).

By estimating a linear model for the expectation function of the yearly SWIR values measured at Hohenpeissenberg station, the following goals are achieved:

1. The time series is decomposed into a trend (the systematic component) and the residuals (the random component).
2. The systematic decrease of SWIR at Hohenpeissenberg station points to a possible climate change. Is the decrease significant?
3. The variance of the empirical residuals of the linear model is smaller than the variance of the time series. Is the variance of the empirical residuals a better estimate for the inter-annual variability of SWIR at Hohenpeissenberg station than the variance of the time series?

Answers to these questions can be obtained by means of simple linear regression methods as introduced in Sect. 3.1.2.

3.1.2 Simple Linear Regression

Simple linear regression models a response variable Y as a linear function of a predictor variable x. The response variable is random, the predictor variable, however, usually not. If the predictor variable is a random variable or a stochastic process then the estimates are conditional, given the realisation available. Y and x are observed jointly to obtain N pairs (y_i, x_i), $i = 1, 2, \ldots, N$. If a formula $Y = f(x)$ with unknown constants is available from prior evidence then estimates for the unknown constants are computed. If f is not available, then f is devined from a plot of the observations and constants in f are estimated. Very often, f is a straight line: this is the model introduced in (3.1). In Sect. 3.1.1, for example, SWIR at Hohenpeissenberg is the response variable, time (i.e., the parameter of the stochastic process) is the predictor variable, and f is a straight line.

For all observations $(y_1, x_1), (y_2, x_2), \ldots (y_N, x_N)$ *available, a continuous (i.e., non-discrete) random variable* Y *(the response variable) is a linear model of the predictor variable* x*:*

1. $Y_i = f(x) = b_0 + b_1 x_i + e_i$.

   The model is fitted by computing estimates $\hat{b}_0$ *and* $\hat{b}_1$ *for the model coefficients* b_0 *and* b_1 *such that the sum of the squares of the theoretical residuals,*

2. $\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} \big(Y_i - (b_0 + b_1 x_i)\big)^2$,

   becomes as small as possible (i.e., using least squares), assuming that:

3. *the* e_i *are a white noise process as defined in (2.11) with* $(\mathrm{E}e_i = 0)$ *(with expectation function identically zero) and* $\mathrm{Var}e_i = \sigma_e^2$*) (with constant finite variance), and*
4. *the* e_i *are normally distributed.*

(3.1)

In (3.1,1), $b_0 + b_1 x_i = \mu_Y(x_i)$ are the expectations of the response variable Y depending on x, and the e_i are the random (due to measurement errors and/or the natural variability of Y) deviations of the expectations from the observations. Please note that no assumptions regarding x are made, e.g., a data type (i.e., numeric or categorical) is not assumed.

In Problem 3.1, the estimators

$$\hat{b}_0 = \overline{y} - \hat{b}_1 \overline{x} \quad \text{and} \quad \hat{b}_1 = \frac{\sum_{i=1}^{N} (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{N} (x_i - \overline{x})^2} \quad \text{with} \tag{3.2}$$

$$\overline{x} = (1/N) \sum_{i=1}^{N} x_i \qquad \overline{y} = (1/N) \sum_{i=1}^{N} y_i$$

are obtained by minimizing (3.1,2). In Problem 3.2 it is shown, using only (3.1,1,2), that the estimate $\hat{b}_1$ is not biased. From the result $\mathrm{E}\hat{b}_1 = b_1$, it

is concluded that the slope of the straight line is estimated without bias provided that the linear model (3.1,1) is the true model, i.e., provided that (i) $b_0 + b_1 x_i = \mu_Y(x_i)$ and (ii) (3.1,2) is minimised using the method of least squares.

The main theme is now interrupted by a short intermezzo. The assumption that a stochastic process is stationary in its expectation function can be expressed as special case $\mathrm{E}Y_t = f(t) = \mu_Y =$ constant of a linear model: b_1 in $\mathrm{E}Y_t = b_0 + b_1 x_t$ becomes identically zero and $\sum_{t=1}^{N}(y_t - \hat{b}_0)^2$ becomes minimal for $\hat{b}_0 = (1/N)\sum_{t=1}^{N} y_t$, $\hat{b}_0$ being the mean of the time series as defined in (2.1,1).

In Problem 3.3, the variance of $\hat{b}_1$ is calculated using (3.1,3) to arrive at $\mathrm{Var}\hat{b}_1 = \sigma_e^2 / \sum_{i=1}^{N}(x_i - \overline{x})^2$. σ_e^2, the variance of the $e_i = Y_i - (b_0 + b_1 x_i)$, cannot be estimated directly since the e_i can not be observed. Therefore, the empirical residuals $\hat{e}_i = y_i - (\hat{b}_0 + \hat{b}_1 x_i)$, i.e., the difference of the observations and the estimated expectations, act as a substitute: $\hat{\sigma}_e^2 = (1/(N-2))\sum_{i=1}^{N}\hat{e}_i^2$ is the usual unbiased estimator for σ_e^2. There are $N-2$ degrees of freedom since two coefficients, b_0 and b_1, are estimated.

In Problem 3.4 it is shown, using the last assumption (3.1,2,4), that $\hat{b}_1$ is a linear combination of normally distributed random variables and is thus normally distributed with the moments as calculated above, provided that the e_i are a normal white noise process. The expectation and the variance of $\hat{b}_0$ are more difficult to obtain [44].

Statistical tests can be performed using the above results if the observations are in agreement with all assumptions in (3.1). These tests suffice to answer, e.g., the questions in the remarks concluding Sect. 3.1.1.

The question whether the systematic decrease of SWIR at Hohenpeissenberg station is significant during the observational period gives reason to formulate the null hypothesis H_0 that this stochastic process is stationary in the expectation function, i.e., that the observations neither increase nor decrease systematically in the mean and, thus, H_0: $b_1 = b_1^{(0)} = 0$ is tested versus the alternative H_1: $b_1 \neq 0$. The standardised test statistic $(\hat{b}_1 - b_1^{(0)})/(\sigma_e^2 / \sum_{i=1}^{N}(x_i - \overline{x})^2)^{1/2}$ is compared with the normal distribution if σ_e^2 is known; usually, however, σ_e^2 is not known and therefore is substituted by the empirical variance of the residuals $\hat{\sigma}_e^2$, as defined above, and the statistic

$$t = \frac{\left(\hat{b}_1 - b_1^{(0)}\right)}{\left(\hat{\sigma}_e^2 / \sum_{i=1}^{N}(x_i - \overline{x})^2\right)^{1/2}} \qquad \text{where} \tag{3.3}$$

$$\hat{\sigma}_e^2 = (1/(N-2))\sum_{i=1}^{N}\hat{e}_i^2 \quad \text{and} \quad \hat{e}_i = y_i - (\hat{b}_0 + \hat{b}_1 x_i)$$

then has a t-distribution with $N-2$ degrees of freedom. Regression software usually computes t together with its significance level. In the case of SWIR at Hohenpeissenberg station, for example, the following results are obtained

```
Call: lm(formula = hpbswir ~ hpbyear)
Residuals:
   Min     1Q Median   3Q   Max
-19.49 -4.523 1.011 5.164 11.74
Coefficients:   Value Std.Error     t value    Pr(>|t|)
(Intercept)  777.3322  177.0772      4.3898      0.0009
    hpbyear   -0.3239    0.0897     -3.6086      0.0009
Residual standard error: 6.8 on 39 degrees of freedom
Multiple R-Squared: 0.2503
F-Statistic: 13.02 on 1 and 39 degrees of freedom,
the p-value is 0.0008649
```

when `summary()` is applied to `hpbswir.fit` being generated using `lm()` in Sect. 3.1.1. `summary()` repeats the linear model given as argument to `lm()`, summarises the distribution of the residuals, and then gives the usual regression statistics.

The estimates for b_0 in line "`(Intercept)`" are not of interest in this case. The estimates for b_1 are given in line "`hpbyear`". To the estimate $\hat{b}_1 = -0.3239$ belongs a t-value of -3.6086 which is to be plotted, in Problem 3.5, in a plot of a t-distribution with 39 degrees of freedom (the difference of the number of observations and the number of estimated coefficients) available in line "`Residual standard error`". From the plot obtained, it is concluded that the trend is highly significant in accordance with the two-tailed probability value given in line "`hpbyear`".

The line "`Multiple R-Squared`" contains the R^2-statistic, for the above example, $R^2 = 0.2503$ is obtained. R^2 can be approximated by calculating $(a-b)/a$, with $a = 60$ $(\mathrm{Wm}^{-2})^2$, the empirical variance of the Hohenpeissenberg SWIR time series, and $b = 45$ $(\mathrm{Wm}^{-2})^2$, the variance of the empirical residuals of the linear model fitted in in Sect. 3.1, both calculated using (2.1,2) in Sect. 3.1. R^2 is defined in (3.4) as the proportion of "total variation" of the response variable "explained by the regression":

$$\begin{aligned} R^2 &= \frac{SS_M}{SS_Y} = \frac{SS_Y - SS_E}{SS_Y} \\ \sum_{i=1}^{N}(Y_i - \overline{Y})^2 &= \hat{b}_1^2 \sum_{i=1}^{N}(x_i - \overline{x})^2 + \sum_{i=1}^{N}\hat{e}_i^2 \\ SS_Y &= SS_M + SS_E \end{aligned} \tag{3.4}$$

In (3.4), the sum of squares of the response variable Y about its mean (or the total sum squares), SS_Y, is obtained as the sum of (i) the sum of squares due to regression (or sum of squares of the model), SS_M, and (ii) the sum of squares about regression (or the sum of squared residuals), SS_E. (3.4) is derived, in Problem 3.7, starting from SS_E which is used in (3.3) to compute the estimate for σ_e^2.

For example in Fig. 3.1, the yearly values of SWIR at Hohenpeissenberg station strongly fluctuate about the regression line, due to their inter-annual

variability, and therefore a large proportion of SS_Y remains in the residuals of the linear model, as is obvious from the results obtained when `summary(aov())` is applied to `hpbswir.fit`, the R object being generated in Sect. 3.1.1:

```
            Df Sum of Sq  Mean Sq  F Value       Pr(F)
hpbyear      1    602.07 602.0699 13.02206 0.000864918
Residuals 39   1803.15  46.2346
```

Using (3.4), $SS_M = 602.07$ (in line "`hpbyear`" above) and $SS_E = 1803.15$ (in line "`Residuals`" above) are added to arrive at $SS_Y = 2405.22$, all in $(\mathrm{Wm}^{-2})^2$. The table obtained above from `summary(aov())` is an analysis of variance table. The ratio of the mean SS_M and the mean SS_E is the F-statistic. In the example above, the calculated F-value is large since the variation due to regression ("explained by regression") is large as compared to the variation about regression ("random" variation).

$\hat{\sigma}_e^2 = 46\ (Wm^{-2})^2$, the mean SS_E (in line "`Residuals`" above), is computed as required in (3.3) and thus is an unbiased estimate for the variance of the residuals. In contrast, the estimate 45 $(Wm^{-2})^2$, calculated in Sect. 3.1.1 as required in (2.1,2), is biased (Problem 3.6).

A confidence interval for the estimate $\hat{\mu}_Y(x) = \hat{b}_0 + \hat{b}_1 x$ of $\mu_Y(x)$ can be calculated assuming (3.1,1,2,3,4). Under these assumptions, (i) $\hat{\mu}_Y(x)$ is a linear combination of normally distributed random variables, and thus is normally distributed as concluded from the remarks to definition (1.34), and (ii) the moments of this distribution are derived in Problem 3.8 using (1.15) and (1.16), $\hat{b}_0$ not being correlated with $\hat{b}_1$:

$$\mathrm{E}\hat{\mu}_Y(x) = b_0 + b_1 x = \mu_Y(x), \tag{3.5}$$

$$\mathrm{Var}\hat{\mu}_Y(x) = \sigma_e^2 \left(\frac{1}{N} + \frac{(x - \overline{x})^2}{\sum_{i=1}^{N} (x_i - \overline{x})^2} \right). \tag{3.6}$$

From (3.5) and (3.6), it is concluded that $\hat{\mu}_Y(x)$ estimates without bias and with a variance that increases when $\sum_{i=1}^{N}(x_i - \overline{x})^2$ decreases. This sum is, with $1/N$, proportional to the variance of the x_i in case of the x_i being realisations of a random variable X, and thus is related to the concentration of X as concluded from the Chebyshev inequality (1.17).

Usually σ_e^2 is not known and is therefore substituted by its estimate as required in (3.3) and thus, usually, $\hat{\mu}_Y(x)$ has a t-distribution.

In the SWIR at Hohenpeissenberg station example, the estimates for the trend are obtained, together with their standard deviations, when `predict()` is applied to `hpbswir.fit`, i.e., the R object that is generated using `lm()` in Sect. 3.1.1:

```
hpbswirexpect <- predict(hpbswir.fit, se=T)
```

From `hpbswirexpect`, the R object obtained, using the .975 quantile of a t-distribution with 39 degrees of freedom,

```
hpbswirexpect$fit + 2.02*hpbswirexpect$se.fit
hpbswirexpect$fit - 2.02*hpbswirexpect$se.fit
```

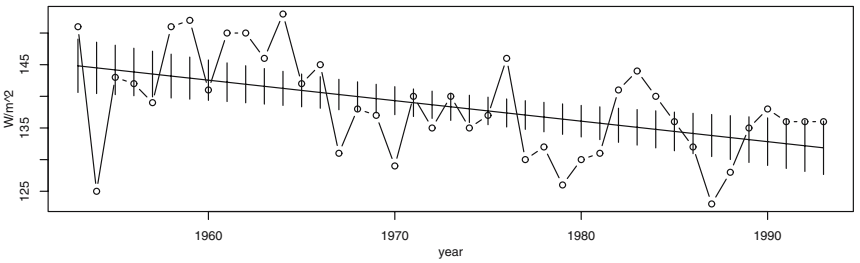


Fig. 3.2. Yearly values of shortwave incoming radiation at Hohenpeissenberg station with a .95 confidence interval for the trend.

produces intervals including the trend $\mu_Y(x)$ with probability .95, provided that the observations are in agreement with all assumptions in (3.1). The observations, the estimated trend and its confidence intervals are plotted in Fig. 3.2.

Confidence intervals as plotted in Fig. 3.2 pertain to the expectation function of SWIR at Hohenpeissenberg station and thus include, when extrapolated beyond the observational period, the expectation of SWIR with a probability of .95 in years with no observations. However, these years should not be too far from the observational period, e.g., $\hat{b}_0 = 777$ Wm^{-2} in year 0 is physically impossible.

Confidence intervals for the trend, however, do not apply to predictions. For example, it cannot be concluded from Fig. 3.2 that, in years with no observations prior to 1953 or past 1993, yearly values of SWIR at Hohenpeissenberg are with probability .95 in extrapolations of the confidence intervals. 0.95 confidence intervals for the stochastic process itself (prediction intervals), calculated as required in e.g., [44], would be much wider because only 14 out of 41 yearly SWIR values measured at this station are within the .95 confidence intervals for the trend.

Being able to predict the expectation function of a stochastic process from an estimated linear model, a first possibility to circumvent the difficulties mentioned in the remarks concluding Sect. 2.7 is arrived at provided that the observed time series is reconcilable with all assumptions in (3.1). The linear model for SWIR at Hohenpeissenberg station, for example, is in agreement with (3.1,1,2,3), because $\mu_Y = b_0 + b_1 t$, as plotted in Figs. 3.1 and 3.2, (i) obviously fits the data, (ii) is estimated using least squares, and (iii) its empirical residuals are not correlated. Do the empirical residuals also stem from a normal distribution as required in (3.1,4)? An answer to this question can be obtained using the diagnostics introduced in Sect. 3.3. As a prerequisite, the multiple linear regression is introduced in Sect. 3.2.

3.2 Multiple Regression

Multiple linear regression as introduced in Sect. 3.2 models a response variable as a linear function of at least two predictor variables. More specifically, the expectation function of a stochastic process (Y_t) is estimated as a linear function of m predictor variables $(X_{t1}), \ldots, (X_{tj}), \ldots, (X_{tm})$. If the (X_{tj}) are stochastic processes themselves then the distributions and estimates derived in this section are conditional, given the realisation of the (X_{tj}) observed. If the (X_{tj}) are not sequences of random variables then they are usually a multi-dimensional parameter of (Y_t). These two cases, however, are not discerned when a linear model is written, in this section, using capital letters for the m predictor variables.

The assumptions in Sect. 3.1.2 are adapted, in Sect. 3.2, for the case of m predictor variables. The expectations of the response variable are assumed to be, in (3.7) and (3.8), a linear function of m predictor variables. The coefficients in this linear function are estimated using least squares in (3.9) and (3.10) under the assumption (3.11), i.e., that the residuals of the model are not correlated. If, however, the residuals are correlated, then estimates can be obtained by applying (3.14), (3.15) and (3.16). Tests are available provided that the residuals are normally distributed as required in (3.17).

3.2.1 A Linear Model for the Expectation Function

The expectation function $\mu_Y(t)$ of the response variable $Y(t)$ is assumed to be a linear combination of m predictor variables $X_{t1}, \ldots, X_{tm}$ with constant coefficients $b_0, b_1, b_2, \ldots, b_m$. Non-linear terms, e.g., $\mathrm{e}^{b_j X_{tj}}$, are not allowed and constraints on the coefficients, e.g., $b_4 = b_2 b_1$, are ignored. Such a linear combination, as defined in (3.7) or, alternatively, in (3.8)

$$\mu_Y(t) = \mathrm{E}Y_t = b_0 + b_1 X_{t1} + \ldots + b_m X_{tm} \quad \text{or} \quad \mathrm{E}\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{b} \tag{3.7}$$

$$Y_t = b_0 + b_1 X_{t1} + \ldots + b_m X_{tm} + e_t \quad \text{or} \quad \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{e} \tag{3.8}$$

is called a *linear model*, the constants in the linear model are called *coefficients* or *parameters*.

Linear models are often written with matrices and, therefore, the matrices and vectors in (3.7) and (3.8) are described in detail as follows: $\boldsymbol{e}$ is the column vector with the residuals, $\boldsymbol{e}^T = (e_t)$; $\boldsymbol{b}$ is the column vector with the model coefficients, $\boldsymbol{b}^T = (b_0, \ldots, b_m)$, for which estimators $\hat{\boldsymbol{b}}$ are to be calculated; $\boldsymbol{Y}$ is the column vector of the stochastic process, $\boldsymbol{Y}^T = (Y_t)$, with the expectation function $\boldsymbol{\mu}_Y(t)$ to be estimated, and $\boldsymbol{X}$ is the matrix of the predictor variables with exclusively 1 in the first column vector, $\boldsymbol{X}_{\boldsymbol{\cdot}1} = (1, \ldots, 1)^T$, and row vectors $\boldsymbol{X}_{t\boldsymbol{\cdot}}$ containing the observations for m predictor variables for a time point t.

If linear, quadratic, etc. trends in a stochastic process $\boldsymbol{Y}$ with parameter t are supposed, then $\boldsymbol{X}$ is constructed raising t to powers $j-1$, $j = 1, \ldots, m$:

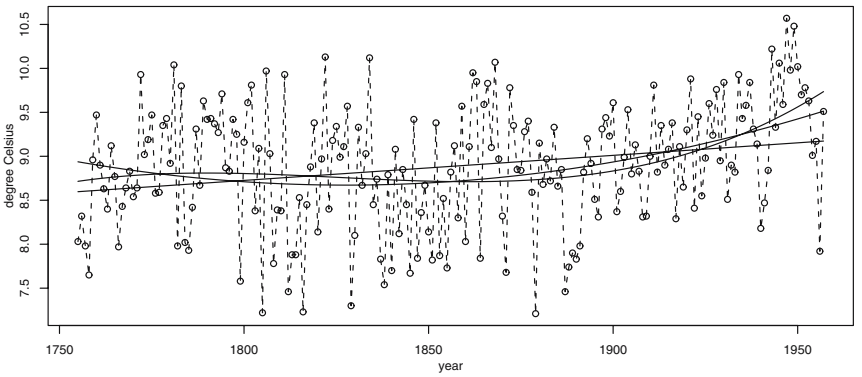


Fig. 3.3. Yearly values in the Basel temperature series as plotted in Fig. 2.13 in degree Celsius with linear models of the first (a straight line), second (a quadratic function) and third degree (a cubic function) for the expectation function.

the first column vector becomes $\boldsymbol{X}_{\cdot 1} = (t_1^0, \ldots, t_N^0)^T$, the second $\boldsymbol{X}_{\cdot 2} = (t_1^1, \ldots, t_N^1)^T$, the third $\boldsymbol{X}_{\cdot 3} = (t_1^2, \ldots, t_N^2)^T$, etc. These linear models are formulated in R as demonstrated with the following examples.

In Problem 2.7, R object `basel` is constructed from the values in the Basel temperature series. From `basel`, with

```
byeartemp <- basel$year
byeartime <- basel$yr
```

R vectors `byeartemp` and `byeartime` are obtained to be used in

```
byeartemp ~ byeartime
```

the first order linear model (a straight line) for the expectation function of the yearly values in the Basel temperature series as plotted in Fig 3.3. This is an example of a simple linear regression comparable to the SWIR at Hohenpeissenberg station example in Sect. 3.1.1 where `~` separates the response variable `byeartemp` from the predictor variable `byeartime`.

Applying `+`, one more predictor variable is added to a linear model as written in R, thus allowing for models with two or more predictor variables in the multiple regression case. In a linear model as formulated in R, symbols `+, -, :, *, ...` do not have their normal arithmetic meaning. For example, `-1` means that (i) b_0 is identically zero in the simple regression case, and (ii) the first column vector $(\boldsymbol{X}_{\cdot 1})$ in $\boldsymbol{X}$ is omitted in the multiple regression case; therefore, in both cases, the linear model is fitted through the origin. These operators, however, keep their original arithmetic meaning when used in `I(...)`. Consequently, linear models with time raised to the second, third, etc. power can be formulated in R. In

```
byeartemp ~ byeartime + I(byeartime^2)
```

for example, `byeartime` occurs twice as predictor variable, the first time raised to its first and the second time raised to its second power, and in

```
baseltemp ~ byeartime + I(byeartime^2) + I(byeartime^3)
```

`byeartime` raised to its third power is added to the linear model as third predictor variable. The estimates for these models are plotted in Fig. 3.3. The second-last and last linear models proposed above for the expectation function of the yearly values in the Basel temperature series are multiple regression models with two (t^1 and t^2) or three (t^1, t^2 and t^3) predictor variables.

Not only t raised to power p but also any other variables having been observed jointly with the stochastic process (the response variable) $\boldsymbol{Y}$ can be used as predictor variables $\boldsymbol{X}_{\cdot j}$ in a linear model as defined in (3.7) and (3.8). A special case of a predictor variable is a copy of the time series of the response variable with lag τ as introduced in Sect. 5.2.2 for the estimation of an AR[p] model.

If a formula $Y = f(X_1, \ldots, X_m)$ is known from prior evidence then estimates for the coefficients in f are computed. If, however, f is not available then f is devined from a time series plot. As a rule of thumb, estimate a linear model for the expectation function of a stochastic process only if you dare to draw a line representing the expection function in a time series plot. The selection of a model is usually not as straightforward as in the SWIR at Hohenpeissenberg station case, used as an introductory example in Sect. 3.1.1. Which of the linear models for the expectation function of the yearly values in the Basel temperature series in Fig. 3.3 is to be preferred? An answer to this question can be given, applying the diagnostics introduced in Sect. 3.3 to the estimates obtained using least squares (Sect. 3.2.2) possibly under the wishful assumptions that the residuals are iid. (Sect. 3.2.3) and normally distributed (Sect. 3.2.5).

3.2.2 Least Squares

Estimators $\hat{\boldsymbol{b}}$ for $\boldsymbol{b}$ in (3.7) and (3.8) are arrived at by minimizing the sum of the squared residuals $\boldsymbol{e}\boldsymbol{e}^T$. Using least squares, the estimators are obtained as solutions of the normal equations provided that the inverse of the symmetric matrix $\boldsymbol{X}^T\boldsymbol{X}$ exists.

$$\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{b}} = \boldsymbol{X}^T\boldsymbol{Y} \qquad\qquad \hat{\boldsymbol{b}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \tag{3.9}$$

If the expectation function $\mu_Y(t)$ is represented by the linear model (3.8) ("if the linear model is true") then the least squares estimates calculated as required in (3.9) are not biased.

$$\begin{aligned} \mathrm{E}\hat{\boldsymbol{b}} &= \mathrm{E}\big((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}\big) = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathrm{E}\boldsymbol{Y} \\ &= (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{b} = \boldsymbol{b} \end{aligned} \tag{3.10}$$

A linear model fitting the data (possibly after a transformation) can often be found and the normal equations can be solved, in most cases, to arrive at the estimates for the model coefficients.

3.2.3 Independent Observations with Constant Variance

Assumption (3.1,3) also applies to the multiple regression case: the residuals (e_t) are a white noise process with $\mu_e = 0$ and σ_e^2 or, using matrices,

$$\mathrm{E}\boldsymbol{e} = \boldsymbol{0} \qquad \text{and} \qquad \mathrm{Cov}(\boldsymbol{e}\boldsymbol{e}^T) \;=\; \boldsymbol{C}_e = \begin{pmatrix} \sigma_e^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_e^2 \end{pmatrix} = \boldsymbol{I}\sigma_e^2. \quad (3.11)$$

Assumption (3.11) is the usual statistical assumption as stipulated in (1.2,2) which implies, specifically, that (i) the variance of the residuals is constant and (ii) the residuals are non-correlated.

The residual variance can be assumed to be constant, either using knowledge obtained from prior measurements or statistically by calculating estimates $\tilde{\sigma}_e^2(t)$, for each time point t, from more than one (possibly many) independent measurements y_{tk} of Y_t, and then comparing $\tilde{\sigma}_e^2(t)$ with $\hat{\sigma}_e^2$, the regression estimate for σ_e^2. Usually, however, this comparison cannot be performed since only one realisation is available. In this case, the variance $\hat{\sigma}_e^2$ of the empirical residuals $\hat{\boldsymbol{e}}$ can be assumed to be constant by applying the diagnostics introduced in Sect. 3.3. An analysis of $\hat{\boldsymbol{e}}$ using these diagnostics should always be performed because an answer to the question can be obtained as to whether the assumptions (3.8) and (3.9) are in agreement with the observations, i.e., whether a linear model has been estimated that captures the systematic changes in the unknown expectation function $\mu_Y(t)$ of the stochastic process (Y_t).

Assuming non-correlated residuals, the covariance matrix $\mathrm{Cov}\hat{\boldsymbol{b}}$ pertaining to the estimators $\hat{\boldsymbol{b}}$ for the coefficients $\boldsymbol{b}$ in the linear model is arrived at in (3.12):

$$\mathrm{Cov}\hat{\boldsymbol{b}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\hat{\sigma}_e^2. \quad (3.12)$$

The diagonal of $\mathrm{Cov}\hat{\boldsymbol{b}}$ contains the variances of the estimators, and the subdiagonals the covariances of pairs of the estimators. If (i) the residuals are not correlated as required (3.11), (ii) the $\hat{\boldsymbol{b}}$ are least squares estimates as required in (3.9), and (iii) the linear model is the "true" model, then the $\hat{\boldsymbol{b}}$ have the smallest variance among all unbiased linear estimators.

3.2.4 Generalised least squares

If the observations are not in agreement with (3.11) (this can be obvious prior to the estimation or becomes manifest from diagnostics applied to the estimates as proposed in Sect. 3.3), then

1. the $\hat{\boldsymbol{b}}$ are unbiased estimators for $\boldsymbol{b}$, but
2. the variance of the $\hat{\boldsymbol{b}}$ has been under-estimated since no $N-p$ (p being the number of coefficients in the linear model, N the number of observations) independent observations are available as required in (3.9) for calculating a least squares estimation.

The least squares estimates can be corrected for possible dependent observations or, alternatively, dependent observations can be accounted for using generalised least squares as proposed in (3.14), both, however, only if the dependences are known.

In order to derive (3.14), the sequence of residuals (e_t) is assumed to be a stationary stochastic process with expectation function $\mu_e(t) = \mu_e = 0$ and known covariance matrix $\boldsymbol{C}_e = \boldsymbol{C}\sigma_e^2$. If $\boldsymbol{C}$ is positive definite ((2.7) requires that $\boldsymbol{C}$ be non-negative definite and symmetric) then there is a matrix $\boldsymbol{D}$ such that $\boldsymbol{D}^T\boldsymbol{D} = \boldsymbol{D}\boldsymbol{D} = \boldsymbol{C}$. Using such a $\boldsymbol{D}$, a vector $\boldsymbol{e}' = \boldsymbol{D}^{-1}\boldsymbol{e}$ is constructed with the properties (i) $\mathrm{E}\boldsymbol{e}' = \boldsymbol{0}$ since $\mathrm{E}\boldsymbol{e} = \boldsymbol{0}$, and (ii) $\mathrm{Cov}\boldsymbol{e}' = \mathrm{E}\big((\boldsymbol{e}' - \mathrm{E}\boldsymbol{e}')(\boldsymbol{e}' - \mathrm{E}\boldsymbol{e}')^T\big) = \mathrm{E}\big(\boldsymbol{e}'\boldsymbol{e}'^T\big)$. Substituting, $\mathrm{Cov}\boldsymbol{e}' = \mathrm{E}\big((\boldsymbol{D}^{-1}\boldsymbol{e})(\boldsymbol{D}^{-1}\boldsymbol{e})^T\big) = \mathrm{E}\big(\boldsymbol{D}^{-1}\boldsymbol{e}\boldsymbol{e}^T\boldsymbol{D}^{-1}\big)$ is obtained since $(\boldsymbol{D}^{-1})^T = \boldsymbol{D}^{-1}$, and $\mathrm{Cov}\boldsymbol{e}' = \boldsymbol{D}^{-1}\mathrm{E}(\boldsymbol{e}\boldsymbol{e}^T)\boldsymbol{D}^{-1} = \boldsymbol{D}^{-1}\boldsymbol{C}\sigma_e^2\boldsymbol{D}^{-1} = \boldsymbol{I}\sigma_e^2$ is arrived at. Consequently, (e'_t) is a white noise process with expectation function $\mu_{e'} = \mu_e = 0$ and variance $\sigma_{e'}^2 = \sigma_e^2$. Pre-multiplying all terms in the linear model as defined in (3.8), the model (3.13) is arrived at. The sum of the squared residuals in (3.13) is thereafter minimised to obtain, in (3.14), the *generalised least squares estimators* for $\boldsymbol{b}$,

$$\begin{aligned} \boldsymbol{D}^{-1}\boldsymbol{Y} &= \boldsymbol{D}^{-1}\boldsymbol{X}\boldsymbol{b} + \boldsymbol{D}^{-1}\boldsymbol{e} && (3.13)\\ \boldsymbol{X}^T\boldsymbol{C}^{-1}\boldsymbol{X}\boldsymbol{b} &= \boldsymbol{X}^T\boldsymbol{C}^{-1}\boldsymbol{Y} && \\ \hat{\boldsymbol{b}} &= (\boldsymbol{X}^T\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{C}^{-1}\boldsymbol{Y} && (3.14)\\ \mathrm{Cov}\hat{\boldsymbol{b}} &= (\boldsymbol{X}^T\boldsymbol{C}^{-1}\boldsymbol{X})^{-1}\hat{\sigma}_e^2 && (3.15) \end{aligned}$$

having the variances and covariances in (3.15).

The properties of the residuals $\boldsymbol{D}^{-1}\boldsymbol{e} = \boldsymbol{e}'$ in (3.13) often deviate from those of the original residuals $\boldsymbol{e}$. This has to be accounted for when the diagnostics introduced in Sect. 3.3 are applied to empirical residuals obtained from an estimation of the model (3.13), i.e., from a generalised least squares estimation.

Only when the covariances $\boldsymbol{C}\sigma_e^2$ of the residuals are known can estimates $\hat{\boldsymbol{b}}$ be calculated using (3.14). Usually, however, $\boldsymbol{b}$ as well as $\boldsymbol{C}\sigma_e^2$ are not known. How can both, the coefficients of the linear model and the covariance matrix of its residuals, be estimated? The solution to this problem is an iterative estimation.

For the first step in an iterative generalised least squares estimation, it is assumed that the residuals (e_t) are not correlated as required in (3.11) and a preliminary estimate $\hat{\boldsymbol{b}}^{(1)}$ is calculated using (3.9), i.e., ordinary least squares. To this estimate pertain preliminary empirical residuals $(\hat{e}_t^{(1)})$ as estimates for the true residuals (e_t). In the second step, the (e_t) are assumed to be a stationary stochastic process as defined in (2.8) with a covariance matrix which is estimated by calculating, using (2.1,3), the empirical $\widehat{\boldsymbol{C}}^{(1)}\sigma_e^{2\,(1)}$ from the preliminary $(\hat{e}_t^{(1)})$ obtained in the first step. $\widehat{\boldsymbol{C}}^{(1)}\sigma_e^{2\,(1)}$ is then substituted for $\boldsymbol{C}$ in (3.14) and a second, now generalised least squares estimate $\hat{\boldsymbol{b}}^{(2)}$ is

arrived at. This second $\hat{\boldsymbol{b}}^{(2)}$ is again used to calculate preliminary empirical residuals $(\hat{e}_t^{(2)})$ and the pertaining preliminary empirical covariance matrix $\widehat{\boldsymbol{C}}^{(2)}\sigma_e^{2\,(2)}$. In step k, $\widehat{\boldsymbol{C}}^{(k-1)}\sigma_e^{2\,(k-1)}$ is substituted for $\boldsymbol{C}$ in (3.14) and the $\hat{\boldsymbol{b}}^{(k)}$ obtained are used for calculating $(\hat{e}_t^{(k)})$ and $\widehat{\boldsymbol{C}}^{(k)}\sigma_e^{2\,(k)}$ The generalised least squares estimations are repeated until $\widehat{\boldsymbol{C}}^{(k-1)}\sigma_e^{2\,(k-1)} \approx \widehat{\boldsymbol{C}}^{(k)}\sigma_e^{2\,(k)}$, i.e., until the estimates for the covariance matrix become stable.

When a linear model is estimated using generalised least squares iteratively, the pitfalls in the remarks concluding Sect. 2.7 are prevalent: either possible large variances of the empirical covariances calculated from a short time series (x_t) and thus short $(\hat{e}_t)$ propagate to the estimates $\hat{\boldsymbol{b}}$ for the model coefficients, or the time needed to solve (3.14) (with the empirical covariances substituted for the theoretical ones) becomes prohibitive when a long time series is applicable. These pitfalls are circumvented by (i) modelling the residuals (e.g., using an AR[1] model with known parameter a and therefore, as concluded from (2.16), known covariance function) or (ii) estimating a model for the covariances of the residuals (an example is given in Sect. 4.3.3, Figs. 4.17 and 4.18 and in Sect. 4.5.3, Fig. 4.21).

If the observations are independent and stem from distributions which do not have identical variances then the covariance matrix $\boldsymbol{C}_e$ of the residuals is a diagonal matrix as proposed in (3.16): the covariances in the subdiagonals are thus identically zero. When $\boldsymbol{b}$ is estimated using (3.14), i.e., generalised least squares, the observations contribute with weights $\boldsymbol{C}_e^{-1}$ to the estimate and, hence, this special case of generalised least squares is called *weighted least squares* or *weighted regression*.

$$\boldsymbol{C}_e = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_N^2 \end{pmatrix} \qquad \boldsymbol{C}_e^{-1} = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_N^2 \end{pmatrix} \tag{3.16}$$

In R, a weighted least squares estimate can be calculated when the weights are given as argument to function `lm()`, e.g., `lm(..., weights=w, ...)`. Then, a generalised least squares estimation is performed using a covariance matrix with `w` as diagonal and subdiagonals which are identically zero, and thus each observation contributes with the weight given to the estimates.

Please note that weight 0 is not allowed in `lm()` to exclude an observation from the estimate. The suggested procedure is to construct, from the R vectors or matrices containing the observations, R objects without the observation to be omitted in the estimate and then to use these new R vectors and matrices in `lm()`. An example is given in Problem 3.10.

Often, however, an estimation using weighted least squares can be circumvented by taking the logarithms of the response variable. Logarithmising is the usual transformation performed to obtain a more stable variance of the response variable, and thus residuals which will then have a variance that is in (a better) agreement with (3.11). An example is given in Problem 3.22.

3.2.5 Normally distributed residuals

The probability distributions of the estimators are easily derived if the residuals are assumed to be not only iid. as required in (3.11) but also normally distributed, as stipulated in (3.17). Using these distributions, confidence intervals can be calculated and tests performed.

The residuals (e_t) *are normally distributed* $(0, \sigma_e^2)$. (3.17)

Are the assumptions proposed in Sects. 3.2.1, 3.2.2, 3.2.3 and 3.2.5 in agreement with an observed time series? This question can be answered, subsequent to the estimation of a linear model, by applying the diagnostics introduced in the next section.

3.3 Diagnostics for Linear Models

Using simple or multiple regression methods, a linear model for the expectation function $\mu_Y(t)$ of the stochastic process (Y_t) (the response variable) is estimated in accordance with at least the first and second assumptions introduced in Sect. 3.2. The first assumption (3.1,1) or (3.8) requires that the systematic changes in $\mu_Y(t)$ are captured by a linear combination of the predictor variables, i.e., a linear model. The coefficients in the linear model are estimated (the model is fitted) using least squares, as required by the second assumption stipulated in (3.1,2) or (3.9). As a rule it is further assumed (assumption no. 3) that the residuals are not correlated and that their variance is constant, as required in (3.1,3) or (3.11). On condition that the residuals are normally distributed as required in (3.1,4) or (3.17) (assumption no. 4), confidence intervals for the estimates then can be calculated.

Since at least the first and second assumptions above are required to be in line with the observations when a linear model is estimated, the estimate is always subject to the diagnostics introduced in this section. These diagnostics allow for detecting deviations from the assumptions as a prerequisite to either improve the model or to select a model from a small number of candidate estimates. For example, the diagnostics to be introduced allow for the selection, in Table 3.1, of one of the linear models estimated for yearly values in the Basel temperature series and plotted in Fig. 3.3.

3.3.1 Plots

The empirical residuals are plotted versus (i) (transformations of) the estimated model, (ii) a predictor variable, and (iii) as a sequence in the same order as the observations. These plots can give answers to the questions of whether the observations are in line with assumptions (3.8), (3.9) and (3.11). Plots of the empirical correlation function of the empirical residuals

are compared with the plots in Fig. 2.22 to ascertain whether the residuals are reconcilable with (3.11), and normal probability plots as introduced in Fig. 2.18 allow for checking assumption (3.17). It is also possible to assess the "influence" exerted by each observation on the estimate and to plot these "influences".

Plot (d) in Fig. 3.1 contains the empirical correlation function, calculated from the empirical residuals (plot (c) in Fig. 3.1) of the linear model (plot (a) in Fig. 3.1), which captures the trend in the SWIR yearly means measured at Hohenpeissenberg station. Further diagnostic plots for this linear model are obtained in Fig. 3.4 applying R functions

```
par(mfrow=c(2,2))
plot(hpbswir.fit, ask=T)
```

to R object `hpbswir.fit` estimated in Sect. 3.1.1 for the trend in the SWIR yearly means measured at Hohenpeissenberg station. Fig. 3.4 contains four plots:

1. the empirical residuals $\hat{e}(t)$ versus the estimated expectation function (the fitted values) $\hat{\mu}_Y(t)$ (above, on the left)
2. the normal probability plot of the $\hat{e}(t)$ (above, on the right)
3. the $|\hat{e}|^{1/2}(t)$ versus $\hat{\mu}_Y(t)$ (below, on the left)
4. Cook's distance plots (below, on the right).

The question of whether or not a linear model (3.8) is in agreement with the observations can be answered using plots of the empirical residuals versus the fitted model and/or versus the parameter of the time series. For example, the plot above, on the left in Fig. 3.4 and plot (c) in Fig. 3.1, both show that the linear model estimated in Sect. 3.1.1 is reconcilable with SWIR at Hohenpeissenberg station. For the yearly values in the Basel temperature series, the plots in the first and second columns in Fig. 3.5 indicate that the linear models estimated in Sect. 3.2.1 and plotted in Fig. 3.3 are all reconcilable with the observations since the residuals fluctuate with approximately constant amplitudes, i.e., they show an =-pattern or only small deviations from an =-pattern.

Fig. 3.5 contains an array of nine plots: the first row contains the plots for the first order linear model fitted to the yearly values in the Basel temperatures series, the second row those for the second order model, and the plots in the third row pertain to the third order model. In the first column, the empirical residuals of the estimates calculated in Sect. 3.2.1 for these models are plotted versus the parameter of the stochastic process (i.e., the year), in the second column versus the estimate, and the third column contains normal probability plots of the empirical residuals.

(3.11) requires that the residuals are a white noise process with constant moment functions μ_e and σ_e^2. Hence, the diagnostics introduced in Sect. 2.5 can be applied to detect non-constant moment functions $\mu_e(t) \neq \mu_e$ and/or $\sigma_e^2(t) \neq \sigma_e^2$. Examples for time series with non-constant moment functions are given in Fig. 2.19.

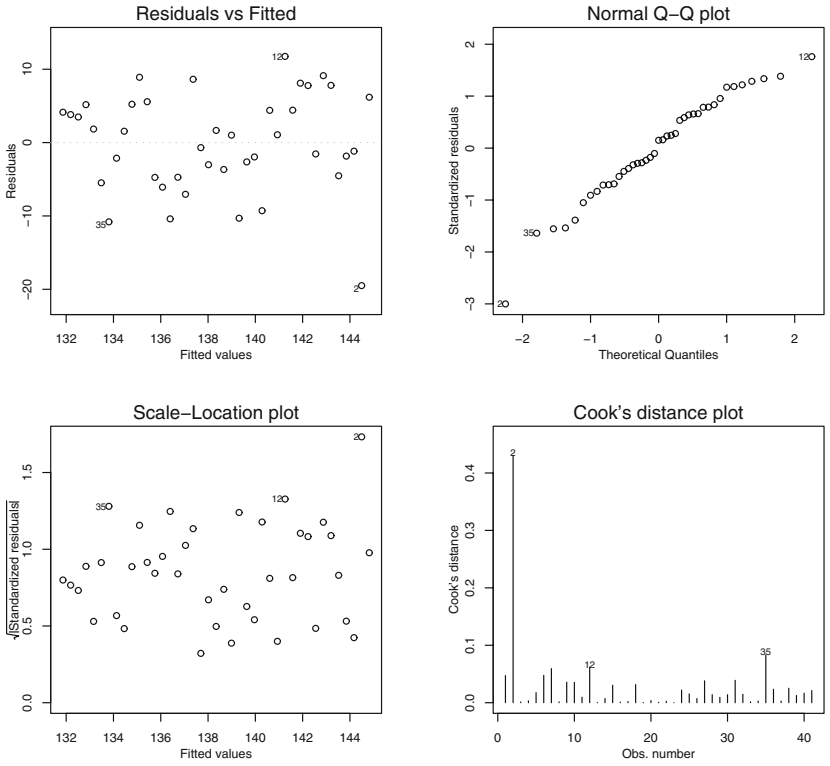


Fig. 3.4. Diagnostics plots for the trend in the shortwave incoming radiation at Hohenpeissenberg station as estimated in Figs. 3.1 and 3.2

Quite often in regression applications the variance of the observations increases with increasing mean of the response variable (Y_t) since the error introduced when (Y_t) is measured and/or the natural variability inherent in (Y_t) is proportional to its expectation function $\mu_Y(t)$. For example, the relative random error in the measurements of SWIR at Hohenpeissenberg station amounts to approximately 2% of the mean [58] and is thus much smaller than the inter-annual variability of SWIR at Hohenpeissenberg station. Consequently, the plots of the residuals in Figs. 3.1 and 3.4 show an =-pattern, i.e., the measurement error remains hidden in these plots. A larger measurement error proportional to $\mu_Y(t)$, however, would produce residuals fluctuating with increasing amplitudes, i.e., a <-pattern would be obtained. Consequently, it is assumed that the empirical residuals of a linear model for SWIR at Hohenpeissenberg station stem from a stochastic process with constant moment functions.

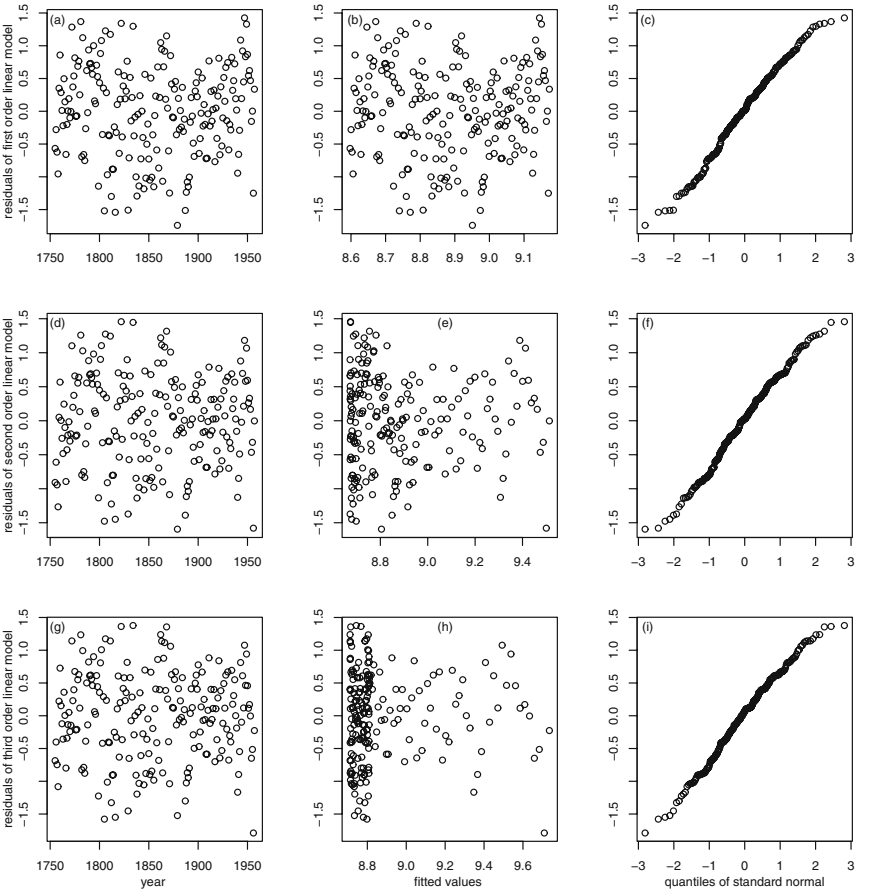


Fig. 3.5. Diagnostic plots for linear models as plotted in Fig. 3.3 for the expectation function of the yearly values in the Basel temperature series.

A <-pattern in a plot of the empirical residuals versus the fitted model (the estimated expectation function) is a strong point, generally, for a non-constant variance of the response variable. In this case, a linear model with a constant residual variance reconcilable with (3.1,3) or (3.11) is obtained when the response variable undergoes, prior to estimation, a transformation deliberately chosen to stabilise its variance. Often, the logarithms of the response variable (an example is given in Problem 3.22) are taken or a power transformation (e.g., $Y^{1/2}$, Y^{-1}, etc.) is applied [44]. In general, a variance stabilising transformation is preferred to a weighted least squares estimation using (3.16).

Empirical residuals of an estimated linear model can be plotted against any predictor variable in order to detect whether the variance in the observa-

tions depends on a predictor variable. A variance changing with a predictor variable can be stabilised by applying a power transformation to this predictor variable and then substituting the variable with its transformation or adding the transformation to the model.

Assumption (3.11) requires that the residuals (e_t) of a linear model (i) have constant moment functions $\mu_e(t)$ and $\sigma_e^2(t)$, and (ii) are independent. Plots of the empirical residuals $(\hat{e}_t)$ versus the fitted model are applied as diagnostics to see whether $\mu_e(t) = \mu_e$ and/or $\sigma_e^2(t) = \sigma_e^2$, as demonstrated above. Plots of the $(\hat{e}_t)$ versus the parameter of the time series as well as plots of the empirical correlations in the $(\hat{e}_t)$ are used to see whether the (e_t) are independent.

Note that the empirical residuals $(\hat{e}_t)$ are not independent since they have only $N - p$ degrees of freedom, N being the number of observations and p the number of parameters in the model. If, however, $(N - p)/N \approx 1$, it is then assumed that any correlations in the $(\hat{e}_t)$ stem from correlations in the (e_t) and therefore indicate observations that are not independent.

Successive random variables in a stochastic process (Y_t) are usually correlated. Correlations not induced by a trend remain in the residuals of a linear model for $\mu_Y(t)$ which, in this case, are not in agreement with (3.11). Consequently, the inspection of plots of the empirical residuals $(\hat{e}_t)$ versus the parameter of the time series and of plots of the empirical correlation function of the $(\hat{e}_t)$ are compulsory when a linear model for the expection function $\mu_Y(t)$ is being estimated. The joint application of both diagnostics is thus compulsory since stationarity is assumed when calculating the empirical correlation function of the $(\hat{e}_t)$ using (2.1,4). For example, the residuals of the linear model for the yearly values of SWIR at Hohenpeissenberg station are assumed to stem from a white noise process when plots (c) and (d) in Fig. 3.1 are compared with the plots in Fig. 2.22 using (2.59).

As an additional diagnostic, the Ljung-Box statistic (2.60) may be computed. If the response variable in a linear model is not a stochastic process then the numbers of the observations can be associated with the observed values to obtain the sequence (y_i), $i = 1, \ldots, N$, and, once the model has been estimated, the sequence of empirical residuals $(\hat{e}_i)$, $i = 1, \ldots, N$. Correlations for $(\hat{e}_i)$ can be calculated using (2.1,3,4) with i, i.e., the number of the observation, substituted for parameter t. In regression terminology, these correlations are called *serial correlations* of the empirical residuals.

As a second example, the empirical residuals of the linear models for the yearly values in the Basel temperature series are plotted versus the parameter of the time series in the first column in Fig. 3.5, and, for the first order linear model, the empirical residuals are re-plotted, together with their correlation function, in Fig. 5.7, plots (a) and (b). From these plots, it becomes obvious that small correlations remain in the empirical residuals of the first order linear model due to a non-constant first moment function of the residuals

and, therefore, it is assumed that this model does not capture the decadal fluctuations in yearly temperature values.

If substantial correlations remain in the empirical residuals of a linear model which — apart from this disadvantage — has been found to be reconcilable with the observations using the diagnostics introduced above, then: (i) the estimators $\hat{\boldsymbol{b}}$ for the model coefficients $\boldsymbol{b}$ are not biased as concluded from (3.10) but (ii) the estimates for variances pertaining to $\hat{\boldsymbol{b}}$ are too small, as concluded from Sect. 3.2.4. Underestimated variances of the estimators are usually corrected by means of calculating, under the stationarity assumption, the empirical covariance matrix of empirical residuals, and then substituting for $\boldsymbol{C}$ in (3.14). A second possibility to account for correlations in the observations is performing an iterative generalised least squares estimation as introduced in the remarks to (3.14). This iteration is routine when a spatial random function with the properties (4.27) and (4.39) is estimated using universal kriging.

A normal probability plot (as introduced in Fig. 2.18) of the empirical residuals allows to check whether the residuals are normally distributed as is required in (3.17). For example, the normal probability plots for the empirical residuals pertaining to the linear models for the yearly values in the Basel temperature series (in the column on the right in Fig. 3.5) show that the distributions of the first and second order residuals (plots c,f) are skewer than the distribution of the third order model (plot i). However, from these normal probability plots, the residuals pertaining to all models can be assumed to be approximately normal. As a consequence, the third order linear model is preferred to the second and first order ones. If, indeed, a very simple model is required with a trend remaining constant throughout the observational period (two centuries in the Basel temperature series example) then the first order linear model is preferred.

Three empirical residuals plotted in Fig. 3.4 (but not in Fig. 3.5) are given numbers. These numbers correspond to the numbers of the observations in R vectors `hpbyear` and `hpbswir` used in Sect. 3.1.1 to estimate the linear model `hpbswir.fit`, and thus are available when the plots in Fig. 3.4 are generated using `plot(hpbswir.fit,ask=T)`. Consequently, as is obvious from plots (a) and (c) in Fig. 3.1, residual no. 2 belongs to the yearly value of SWIR at Hohenpeissenberg station in year 1954, residual no. 12 pertains to the value measured in year 1964, and residual no. 35 to the 1987 value.

The observation in 1954 is approximately 20 Wm^{-2} smaller than its neighbour values and therefore is an example of an extreme value. Extreme values can be the result of gross measurement errors. The history of the measurements at Hohenpeissenberg station, however, reveals that the 1954 value was obtained by means of routine measurements performed according to the manual and therefore is an extreme value not due to a measurement error.

Extreme values have a large influence on estimators calculated using least squares: the estimates for the coefficients in the linear models (3.1,1) or (3.8)

are not robust. The influence exerted by an extreme value on a least square estimate can be assessed simply by repeating the estimate without the extreme value and then comparing the results. In Problem 3.10 for example, a larger decrease of SWIR at Hohenpeissenberg station is obtained when a linear model is estimated from a time series without the extreme 1954 value.

More sophisticated than simply repeating the estimation without the extreme values are the influence diagnostics for regressions models. For example, Cook's distances as introduced in [38] and calculated as described in [31] for the first order linear model for the Hohenpeissenberg SWIR values are plotted in Fig. 3.4. It is obvious from this plot that observation no. 2 (the 1954 SWIR value) influences the estimates to a much larger extent than do the other observations.

3.3.2 Model Selection

When two or more linear models for the expectation function of a stochastic process are estimated the question arises which of the estimates is to be preferred. An answer can be found by comparing the estimates using the statistics and diagnostics introduced in Sects. 3.1, 3.2 and 3.3.1 as demonstrated in Table 3.1 for the example of the yearly values in the Basel temperature series.

It is assumed that the expectation function of the yearly values in the Basel temperature series can be modelled using a linear function of powers of the parameter, i.e., the time in years, and, under this assumption, the estimates for the first, second and third order linear models are calculated using least squares in Sect. 3.2.1. These estimates are plotted together with the time series in Fig. 3.3.

Are the estimates reconcilable with the assumptions as stipulated in (3.8), (3.9), (3.11) and (3.17)? Possible answers should allow for excluding models which are not (or to a lesser degree than others) in agreement with the assumptions and thus for reducing the number of candidate models. Such answers can be obtained from a combination of the diagnostic plots introduced in Sect. 3.3.1 and the statistics (used as diagnostics) introduced in Sects. 3.1 and 3.2. For example, the diagnostics in Table 3.1 are applied in order to select one of the linear models for the expectation function of the yearly means in the Basel temperature series. Remarks to some of these diagnostics are given in the following paragraphs.

Evidence obtained from Geosciences should always be applied to the estimates. Often some knowledge of the observed variable is available which can then be transformed into a linear model with the innateness such that a unit and an interval of physically possible values can be associated to each coefficient. In Sect. 3.1, for example, SWIR at Hohenpeissenberg station is estimated to decrease with 2.34% over 10 years, clearly between -4% and 4% in 10 years, i.e., the trends of SWIR estimated in [58] and attributed to natural phenomena. A decrease of more than 5% in 10 years would be a very

Table 3.1. Diagnostics for the linear models for the expectation function of the yearly means in the Basel temperature series as estimated in Sect. 3.2.1 and plotted in Fig. 3.3. Geophysical evidence: (n.r.) r.: (not) reconcilable with the knowledge already available. Analysis of residuals: +, (−): normally (not normally) distributed; =: constant variance; <, (>): variance increases (decreases); ≠: variance not constant.

linear model of order	1	2	3
geophysical evidence	r.	r.	r.
number of coefficents	2	3	4
p.-values of the	0.02	0.001	0.08
t-statistics for the coefficients	0.0007	0.002	0.07
		0.001	0.07
			0.07
F-statistic	11.74	11.41	8.807
R^2	0.0552	0.102	0.1172
$\hat{\sigma}_e$	0.6945	0.6786	0.6747
analysis of residuals:			
plots in Fig. 3.5			
- normal probability plot	+	+	+
- plot vs. x_t	=	=	=
- plot vs. $\hat{\mu}_Y(t)$	=	=	=
Cook's distance	no observations with large influence		

doubtful result due to an insufficient maintenance (cf. Chap. 1) of the instruments at Hohenpeissenberg station. The knowledge available can, however, be vague: for example, most temperature time series measured at a station at the surface (of the earth) increase in the mean since global temperature increases in the mean, as shown in Fig. 2.23, or, it is known that the expectation function to be estimated increases when a predictor variable increases and that this relationship is hidden by a large measurement error, etc.

The sum of squares of the response variable about its mean can be split in two parts, as proposed in (3.4), i.e., the sum of squares due to regression and the sum of squares about regression, even when the time series is not reconcilable with all assumptions in (3.1) (or in (3.8), (3.9), (3.11) and (3.17) in the multiple regression case). Subsequently, the usual analysis of variance table and an F-value can be computed for diagnostic use.

The R^2 in (3.4) and the empirical variance of the residuals $\hat{\sigma}_e^2$ in (3.3) should be used with caution because both are influenced by the structure of the stochastic process: from a low (< 0.5) R^2, it cannot be concluded that a linear model does not fit the observed time series. In Sect. 3.1.1, for example, the first order linear model for the yearly values of SWIR at

Hohenpeissenberg station is estimated with a significant trend (-2.34% in 10 years), an $R^2 = 0.25$ and a $\hat{\sigma}_e^2 = 46\ (Wm^{-2})^2$ due to the strong inter-annual variability of SWIR at this station.

The empirical residuals should be analysed using diagnostic plots as introduced in Sect. 3.3.1 to ascertain whether (i) the linear model is reconcilable with the observations and (ii) they stem from a normal white noise process as required in (3.11) and (3.17). From an inspection of the plots in Fig. 3.5 the results discussed in Sect. 3.3.1 and summarised in Table 3.1 are obtained.

From Table 3.1 it is concluded that all linear models estimated are desirable. If statistical significance of the trend is required then the third order linear model is preferred to the first and second order one since the plots in Fig. 3.5 show that the distributions of the first and second order residuals (plots c,f) are less symmetric than the distribution of the third order model (plot i). However, the coefficients in the third order model are not significant at the .95 level, as is seen in Table 3.1. If a very simple model is required then the first order model is preferred, in which case the change in temperature is assumed to be constant over the observational period.

If no linear model fitting all observations is found then local linear models can be estimated as an alternative, i.e., models which, in agreement with definition (2.55,2), fit the observations in time slices T' of the observational period selected to obtain a locally constant mean increase or decrease of the expectation function. Such a partition of a time series can be found by smoothing, for example using a moving average as defined in (2.39). Examples of time series smoothed using a moving average are given in Fig. 1.7 as well as in Figs. 2.9 and 2.10. Local linear models for the yearly means in the Basel temperature series are estimated in Problem 3.11.

However, when estimating linear models for the expectation function in parts of a time series selected after having measured and analysed the data to some extent, the usual statistical tests cannot be performed since the time slice thus selected has not been obtained as a result of a well-defined repeatable experiment to test a scientific hypothesis and, therefore, the experimental setup is not up to the standard required for the usual statistical inference, i.e., the statistical inference as described in introductory textbooks. As a substitute, simulation experiments can be performed. In Sect. 2.5.2 for example, the selection procedure of a climatologist eager to detect a non-stationary expectation function in the NAO index (Jones), is modelled using a sequence of t-tests on the realisations of a white noise process and thus a more realistic significance of the results is obtained.

Generally, [58] and [142] recommend that caution is called for when statistical significance is claimed for a trend estimated.

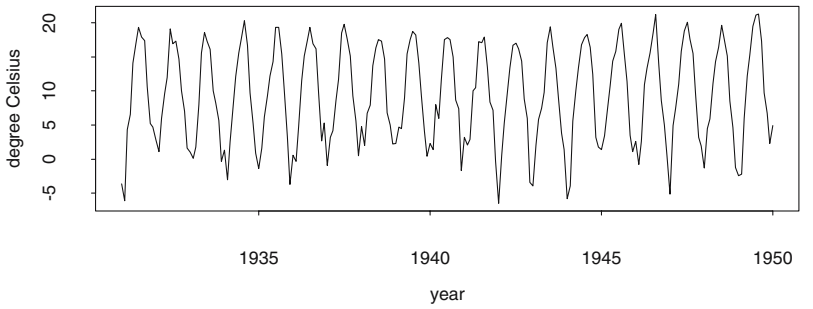


Fig. 3.6. Monthly values in the Basel temperature series from January, 1931 through to December, 1950.

3.4 Seasonal and Component Models

The homogeneous Basel temperature series made available in [13] contains yearly and monthly values in ^{0}C (degree Celsius) from 1755 through to 1957: the yearly values are plotted above, in Fig. 2.13 (together with fitted linear models, in Fig. 3.3) for the full observational period, and the monthly values plotted in Fig. 3.6 for the time slice from January 1931 through to December 1950. Fig. 3.6 shows seasonal fluctuations with a known (12 months) constant period. Seasonal fluctuations with a constant period of 12 months can be seen not only in climatological but also in many socio-economic time series because seasonal fluctuations of the atmospheric conditions have an influence on consumption (e.g., on holidays) and production (e.g., in farming or construction).

3.4.1 Seasonal Models

In a first step, it is assumed that the expectation function $\mu_X(u)$ of a stochastic process (X_u) fluctuates with a constant period and that the period is known. Then a random variable X_u can be associated to a season j and a period t and thus can be written using a two-dimensional parameter, i.e., (X_{tj}), and, vice-versa, each pair (t, j) of the two-dimensional parameter is associated with a unique time point u of the one-dimensional parameter.

In the second step, (X_{tj}), the one-dimensional process with two-dimensional parameter obtained in the first step, is assumed to be an r-dimensional process with parameter t: in Table 3.2 the stochastic process (X_u) is split into r stochastic processes, $(X_t^{(j)})$, which are called *seasonal processes*. Assumptions regarding the seasonal processes can now be made.

As a first assumption, the seasonal processes are often required to be stationary. Then the seasonal expectation functions $\mathrm{E}(X_t^{(j)}) = \mu_X^{(j)}(t) = \mu^{(j)}$

Table 3.2. Splitting up into seasonal processes

period	season 1	season 2	...	season j	...	season $r-1$	season r
$\vdots$							
t	$X_t^{(1)}$	$X_t^{(2)}$	...	$X_t^{(j)}$	...	$X_t^{(r-1)}$	$X_t^{(r)}$
$\vdots$							

and the seasonal covariance functions $\mathrm{Cov}\big(X_t^{(j)}, X_{t+\tau}^{(j)}\big) = c_X^{(j)}(\tau)$ can be estimated. With the following R expressions, for example,

```
bsfn <- "/path/basel.dat"
basel200 <- read.table(bsfn,col.names=c("yr","jan","feb","mar",
    "apr","may","jun","jul","aug","sep","oct","nov","dec","year"))
#basel200 is a dataframe in R, making the data available as R vec-
#tors and as R matrix (matrices in R are introduced in the remarks
#to Fig. 2.1): column vector basel200[,2] is vector basel200$jan)
> is.matrix(basel200)
[1] T
> length(basel200$jan)
[1] 203
#the monthly processes are assumed to be stationary: estimates for
#their constant expectation functions are obtained using mean():
mon <- 1:12
monmean <- 1:12
for (j in mon) { monmean[j] <- mean(basel200[,j+1]) }
> monmean
[1] -0.36009  1.32167  4.36896  8.85073 13.21231 16.49261
[7] 18.21724 17.55517 14.07389  8.90689  3.92068  0.70541
```

estimates are obtained for the mean monthly temperatures in Basel for the period 1755 through to 1957: this is the estimated annual cycle of the temperature in Basel for the observational period.

`basel200` is an R dataframe. A dataframe in R makes the observations available, on the one hand, in a matrix with rows containing the values of each observation and columns containing all observations of a variable, and, on the other hand, as R vectors, as demonstrated in the R expressions above. Both possibilities allow for easier calculations with multivariate data sets. A data frame is generated using R function `read.table()` from a data set available as a table in a text file. Alternatively, a data frame is generated using R function `data.frame()` from R vectors as demonstrated in the remarks to Fig. 3.7. In R, a data frame is preferably used together with function `lm()` for estimating

Table 3.3. Estimating a seasonal model for the Basel temperature series

period (year)	Jan.	Feb.	...	Nov..	Dec.	yearly
			temperature values			
t_1						
$\vdots$						
t_n						
mean	$\hat{\mu}_X^{(Jan.)}$	$\hat{\mu}_X^{(Feb.)}$	...	$\hat{\mu}_X^{(Nov.)}$	$\hat{\mu}_X^{(Dec.)}$	$\hat{\mu}_X$
seasonal fluctuation	$\hat{\mu}_X^{(\Delta Jan.)}$	$\hat{\mu}_X^{(\Delta Feb.)}$	...	$\hat{\mu}_X^{(\Delta Nov.)}$	$\hat{\mu}_X^{(\Delta Dez.)}$	

a linear model, e.g., the linear models in Figs. 3.7 and 3.11 are estimated using data frames.

The annual cycle of the temperature in Basel is estimated above, under the assumption that the seasonal (monthly) processes are stationary. Can the monthly values in the Basel temperature series be stationary albeit that the yearly values are not stationary, as shown in Fig. 3.3? An answer is found by applying the diagnostics introduced in Sect. 2.5.1 to each monthly process or by estimating linear models for the monthly expectation functions using the methods introduced in Sects. 3.1, 3.2 and 3.3. For example, linear models for the trends in the expectation functions of the January, April, July and October values in the Basel temperature series are estimated in Problem 3.12.

A second assumption requires (i) that the stochastic process (X_u) as well as the seasonal processes $(X_t^{(j)})$ be stationary with constant expectation functions $\mathrm{E}(X_u) = \mu_X(u) = \mu_X$ as well as $\mathrm{E}(X_t^{(j)}) = \mu_X^{(j)}(t) = \mu_X^{(j)}$ and (ii) that $\mu_X^{(j)}$ can be obtained as sum of μ_X and the mean seasonal fluctuations as defined in (3.18).

$$\mu_X^{(\Delta j)} = \mu_X^{(j)} - \mu_X \tag{3.18}$$

(3.18) is called a *seasonal model* and the $\mu_X^{(\Delta j)}$ are called *mean seasonal fluctuations*, or often simply *seasonal fluctuations*. Estimates for the mean seasonal fluctuations are obtained as differences of the corresponding means: $\hat{\mu}_X^{(\Delta j)} = \hat{\mu}_X^{(j)} - \hat{\mu}_X$.

$\hat{\mu}_X^{(j)}$ and $\hat{\mu}_X$ are easily obtained when the original time series (x_u) is given in a table. Each line in Table 3.3, for example, contains the period in the first column, then twelve monthly temperature values followed by their mean, i.e., the yearly value, in the last column. This scheme is often used to write seasonal data in a text file, examples being the time series of the global average temperatures available in file `/path/tavegl.dat` and analysed in Problem 2.30 and also the values in the Basel temperature series available in file `/path/basel.dat` from which R dataframe `basel200` is generated

above using `read.table()`. From this dataframe the monthly means are made available above in R vector `monmean`. Therefore, with

```
overallmean <- mean(basel200[,14]) #mean of the yearly values
monmean - overallmean
[1] -9.245418 -7.563645 -4.516354 -0.034581 4.326995  7.607290
[7] 9.331921  8.669852  5.188571  0.021576 -4.964630 -8.179901
```

the mean seasonal fluctuations as required in (3.18) are arrived at as differences between the means of the monthly values and the mean of the yearly values.

3.4.2 Component Models

If the stochastic process (X_u) under analysis is not stationary in the first moment function then a seasonal model (3.18) cannot be estimated using $\hat{\mu}_X^{(\Delta j)} = \hat{\mu}_X^{(j)} - \hat{\mu}_X$, since $\hat{\mu}_X$ cannot be calculated and the seasonal processes are likely to be not stationary. For example, the Basel temperature series is shown to stem from such a process, in Problem 3.12, and thus the estimates calculated above, in Sect. 3.4.1, can possibly be improved upon by estimating the trend and the mean seasonal fluctuations (i.e., the annual cycle) jointly, using a component model as defined in (3.19):

Let $Y(t) = \mu_Y(t) + e(t)$ be a stochastic process with a non-constant expectation function $\mu_Y(t)$ and the residuals $e(t)$. The residuals are called random component. Let further $\mu_Y(t)$ result from an operation on $\mu_Y^{(T)}(t)$ and $\mu_Y^{(S)}(t)$. The trend $\mu_Y^{(T)}(t)$ is a systematic and long-term change in $\mu_Y(t)$. The seasonal component $\mu_Y^{(S)}(t)$ is a systematic and periodic fluctuation in $\mu_Y(t)$ (periodic functions are defined in (6.1)). Then:

1. $Y(t) = \mu_Y^{(T)}(t) + \mu_Y^{(S)}(t) + e(t)$ *is called an additive and*
2. $Y(t) = \mu_Y^{(T)}(t) \times \mu_Y^{(S)}(t) \times e(t)$ *is called a multiplicative component model.*

(3.19)

A multiplicative component model becomes an additive one once the logarithms have been taken. As a side effect, a residual variance increasing with $Y(t)$ becomes more stable, such that the data are in a better agreement with (3.11) since taking the logarithms results in a variance stabilising transformation, as mentioned in the remarks to Fig. 3.4. For example, a multiplicative model is estimated in Problem 3.22.

In stochastic processes of monthly temperature values, $\mu_Y^{(S)}(t)$ is the annual cycle of the temperature and $\mu_Y^{(T)}(t)$ is a *secular trend* (secular and long-term are used interchangeably) in temperature. A very simple component model for the monthly values in the Basel temperature series is obtained, for example, assuming that the trends are identical for each month: twelve

first order linear models $b_0^{(1)} + b_1 \times t, \ldots, b_0^{(j)} + b_1 \times t, \ldots, b_0^{(12)} + b_1 \times t$ with identical coefficient b_1 are to be estimated.

The yearly values in the Basel temperature series are plotted for the full observational period in Fig. 3.3, whereas the monthly values are plotted in the time slice from January 1931 through to December 1950 in Fig. 3.6 to make the annual cycle clearly visible. The data are made available in R as dataframe `basel200` using the R expressions in the remarks to Table 3.2. From `basel200`, in a first step, R vectors `temp` for the one-dimensional process and `idx` for its one-dimensional parameter are obtained using the following R expressions:

```
temp <- 1:(length(basel200$jan)*12)
idx <- 1:(length(basel200$jan)*12)
#month 0 is in the mean of the observational period: June 1856
lagidx <- (length(basel200$jan)*12)/2
k <- 0    #fill in new vectors from data frame
for(i in 1:length(basel200$jan)) {
  for (j in mon) {    #generated in the remarks to Table 3.2
    k <- k + 1
    idx[k] <- k - lagidx
    temp[k] <- basel200[i,j+1]
  }
}
```

In a second step, a categorical R vector (i.e., a vector with values from a set of possible non-numeric values) is generated with

```
lab <- as.factor(rep(c("JAN","FEB","MAR","APR","MAY","JUN","JUL",
           "AUG","SEP","OCT","NOV","DEC"), length(basel200$jan) ) )
```

such that the monthly temperature values in `temp` can be associated with the month using `lab`. Such a vector is called a *factor* with values from a set of possible *levels* when used as a predictor variable in a linear model. For example, R expressions

```
plot(c(-1200,1200),c(-12,22),type="no",xlab="month",
           ylab="degree Celsius")
lines(idx[lab=="JAN"],temp[lab=="JAN"])
points(idx[lab=="JAN"],temp[lab=="JAN"],pch="1")
...
```

produce the seasonal time series for the months January, April and July plotted in Fig. 3.7. In the third step, dataframe `bm200` is obtained with

```
bm200 <- data.frame(idx,lab,temp)
```

as a collection of R vectors `idx`, `lab` and `temp`.

Having completed these preparations, an additive component model for the trend and annual cycle in the Basel temperature series is estimated in R, using the following linear model

```
baselcomp.fit <- lm(temp ~ lab -1 + idx, data=bm200)
baselcomp.erw <- predict(baselcomp.fit, se=T)
```

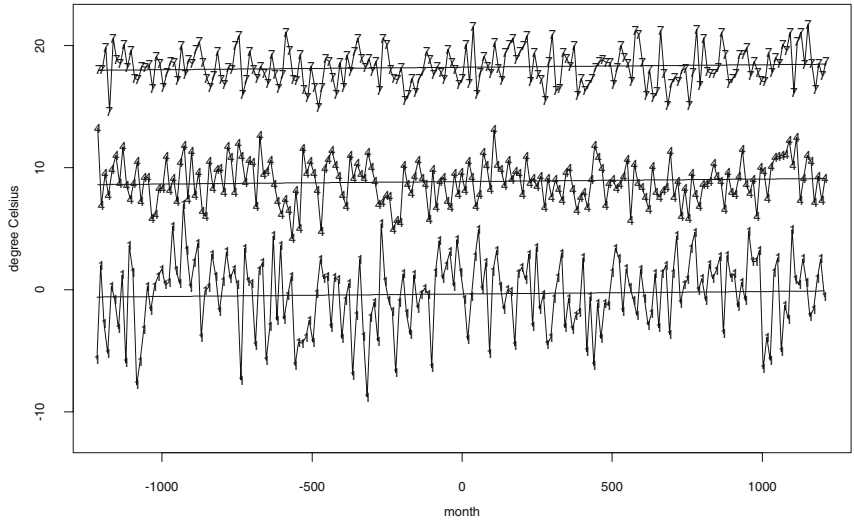


Fig. 3.7. Monthly values in the Basel temperature series: parallel first order linear models for the January, April and July values. Time in months with origin in June, 1856.

and with

```
lines(bm200$idx[bm200$lab=="JAN"],
        baselcomp.erw$fit[bm200$lab=="JAN"])
...
```

the estimates for the January, April and July expectation functions are plotted in Fig. 3.7. As required in (3.19,1), all models for the monthly expectation functions increase at an identical rate (parallel regression).

Vector `temp` in `bm200` is the response variable, and vectors `idx` as well as `lab` are predictor variables. `lab` is a categorical vector. Categorical predictor variables are not excluded in definitions (3.1) and (3.8) where only the response variable is required to be numeric. In the R expression above, the model is written `temp ~ lab -1 + idx` where `-1` means that the first column vector $\boldsymbol{X}_{\cdot 1} = (1, \ldots, 1)^T$, $\boldsymbol{X}$ being the matrix of the predictor variables in (3.8), is omitted: R calculates for each month a straight line with slope $\hat{b}_1$ and intercepts $\hat{b}_0^{(j)}$, $j = 1, \ldots, 12$. $\hat{b}_1$ is the estimate for the trend in an additive component model, as defined in (3.19,2) for the Basel temperature series, and the $\hat{b}_0^{(1)}, \ldots, \hat{b}_0^{(12)}$ are estimates of the seasonal component, i.e., annual cycle. The estimates obtained

```
> summary(baselcomp.fit)
Call: lm(formula = baseltemp   baselmlab - 1 + baselmindex)
Residuals: Min     1Q  Median    3Q    Max
         -10.87 -1.109 0.08246 1.221 10.96
```

```
Coefficients:  Value Std. Error  t value Pr(>|t|)
labAPR        8.8511     0.1348  65.6500   0.0000
labAUG       17.5548     0.1348 130.2057   0.0000
labDEC        0.7042     0.1348   5.2231   0.0000
labFEB        1.3225     0.1348   9.8091   0.0000
labJAN       -0.3591     0.1348  -2.6633   0.0078
labJUL       18.2170     0.1348 135.1179   0.0000
labJUN       16.4926     0.1348 122.3276   0.0000
labMAR        4.3696     0.1348  32.4097   0.0000
labMAY       13.2125     0.1348  97.9988   0.0000
labNOV        3.9197     0.1348  29.0726   0.0000
labOCT        8.9061     0.1348  66.0573   0.0000
labSEP       14.0733     0.1348 104.3831   0.0000
idx         0.0002     0.0001   3.6944   0.0002
Residual standard error: 1.921 on 2423 degrees of freedom
Multiple R-Squared: 0.9711
F-statistic: 6261 on 13 and 2423 degrees of freedom,
              the p-value is 0
```

are close in value to (i) the seasonal model (i.e., the annual cycle) for the Basel temperature series, as estimated in the remarks to Table 3.2, and (ii) the trends of the yearly and monthly values in the Basel temperature series as calculated in Problem 3.19. Please compare the estimates of the component model obtained above with those obtained in Problem 3.21.

The systematic components in a model as defined in (3.19) are required to capture Geophysical phenomena (e.g., a trend and an annual cycle), and the random component is required to be reconcilable with (3.11) and (3.17). Therefore, diagnostics as introduced in Sects. 3.3.1 and 3.3.2 should be carefully applied whenever a component model is estimated, as demonstrated in Sect. 3.5 where linear models for the regional trend (first systematic component) and the station effects (second systematic component) of SWIR measured at more than one station in Germany are estimated.

3.5 Trends and Station Effects of Shortwave Incoming Radiation

Measurements of meteorological variables are often analysed in a grid, e.g., when observations made at the surface (of the earth) are compared with results obtained from a general circulation model or from weather and/or other satellites. Examples for gridded temperature data are given in Problem 2.30, and, in Sects. 4.1.1 and 4.6.5, methods are introduced that allow to obtain areal precipitation means from precipitation data measured at stations. As a third example, [58] makes available means and trends of shortwave incoming radiation at the earth's surface (SWIR) in the cells of a 2.5^0 grid. In this

Table 3.4. Climatologies (means and empirical variances) for the yearly values of SWIR measured in the period from 1950 through to 1993 at the stations in grid box $\big((10^0\text{E}, 50^0\text{N}), (15^0\text{E}, 52.5^0\text{N})\big)$.

name and coordinates of station (degrees and minutes, m a.m.s.l.)				observational period		mean Wm^{-2}	empirical variance $(\text{Wm}^{-2})^2$	in Fig. 3.8	in *s*
Praha	14 26 E	50 4 N	262	1985	1991	112		1	A
Coburg	10 57 E	50 16 N	357	1981	1992	119	42	2	B
Fichtelberg	12 57 E	50 26 N	1214	1964	1993	109	67	3	C
Zinnwald	13 45 E	50 44 N	877	1991	1993	133		4	D
Chemnitz	12 52 E	50 48 N	418	1991	1993	137		5	E
Weimar	11 19 E	50 59 N	257	1991	1993	117		6	F
Wahnsdorf	13 41 E	51 7 N	271	1964	1993	114	35	7	G
Halle/Saale	11 57 E	51 31 N	96	1991	1993	124		8	H
Harzgerode	11 8 E	51 39 N	404	1991	1993	124		9	I
Braunlage	10 37 E	51 43 N	615	1957	1993	111	66	0	J
Lindenberg	14 7 E	52 13 N	???	1991	1993	118		a	K
Braunschweig	10 27 E	52 18 N	81	1958	1993	113	37	b	L
Potsdam	13 6 E	52 23 N	33	1937	1993	117	38	c	M

section, the method applied in [58] to estimate climatologies for the trends and station effects in grid cells is demonstrated by estimating a linear model for trend and station effects in grid cell $\big((10^0\text{E}, 50^0\text{N}), (15^0\text{E}, 52.5^0\text{N})\big)$.

The stations with SWIR measurements appearing in this grid cell are given in Table 3.4, and the time series of yearly values of SWIR measured at these stations are plotted in Fig. 3.8 for the period from 1950 through to 1993. The climatological mean of SWIR in this grid cell amounts to 114 Wm^{-2}, and is calculated as the mean of all yearly values measured at the stations in Table 3.4. An estimate for a climatological trend of SWIR in the above-mentioned grid cell is calculated in this section using a linear model.

The monthly and yearly values of SWIR measured at the stations given in Table 3.4 are available in text file `/path/swir696.dat`, with the SWIR values in lines identified using the number of the grid cell, the station number, the station identification within the cell, and the year of measurement. With R expressions

```
format <- list(gridnum=0,stnum=0,stcellid="",year=0,jan=0,feb=0,
               mar=0,apr=0,may=0,jun=0,jul=0,aug=0,sep=0,oct=0,
               nov=0,dec=0,yearlyswir=0)
filename <- "/path/swir696.dat"
t1 <- scan(filename,format)
y <- t1$yearlyswir
```

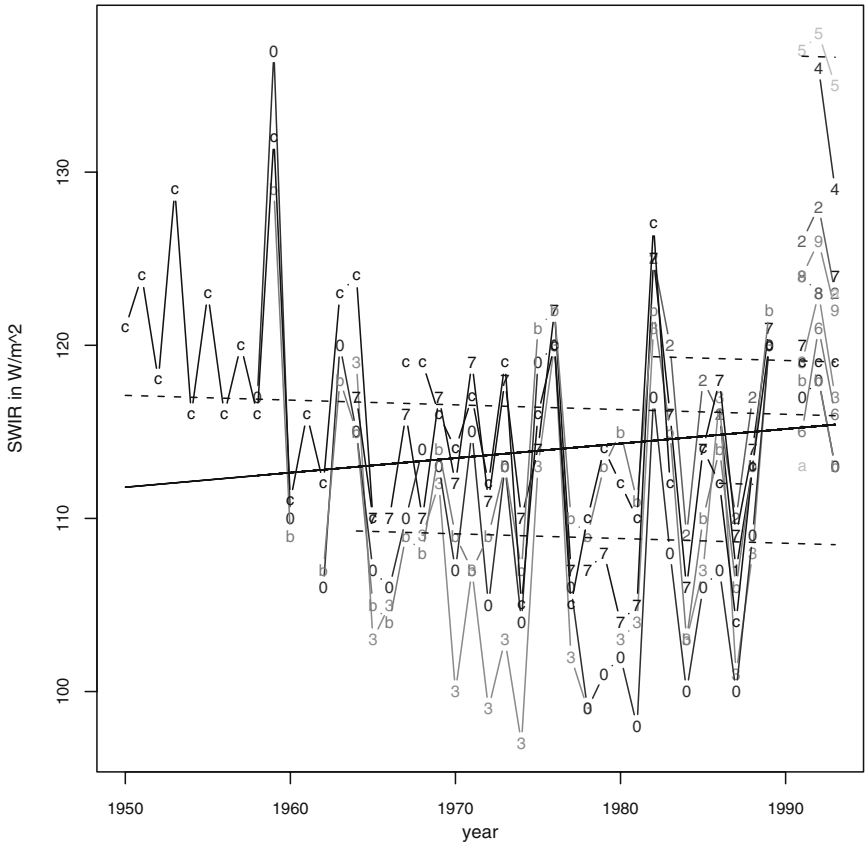


Fig. 3.8. Yearly values of SWIR measured at the stations given in Table 3.4 together with estimates for linear models as required in (3.20) (an increase of SWIR) as well as in (3.21) (a decrease of SWIR and station effects, plotted only for the stations Praha (1), Coburg (2), Fichtelberg (3), Chemnitz (5) and Potsdam (c)).

```
t <- t1$year
s <- t1$stcellid
```

the yearly SWIR values are made available together with their identification in three R vectors `y`, `t` and `s`. `y` contains the yearly values, `t` the year of the measurement and `s` the station identification as given in Table 3.4: A for Praha, B for Coburg, etc. `s` is a factor as introduced in Sect. 3.4.2 with the levels A, B, ..., M. `s` is not ordered since no natural order of the stations in a grid cell is available. The yearly SWIR values in this grid cell are plotted as times series in Fig. 3.8 with

```
xm <- matrix(NA,50,30) #less than 50 years at a station
ym <- matrix(NA,50,30) #less than 30 stations in a grid cell
```

```
for(i in 1:length(unique(s))) {
  xm[1:length(t[s==unique(s)[i]]),i] <- t[s==unique(s)[i]]
  ym[1:length(y[s==unique(s)[i]]),i] <- y[s==unique(s)[i]] }
matplot(xm,ym,type="bbbbbbbbbbbbbbbbbbbbbbbbbbbbbb",
        xlab=" ",cex=0.8,ylab=" ",lty=1)
mtext(side=1,line=2,cex=0.8,"year")
mtext(side=2,line=3,cex=0.8,"W/m^2")
```

From `y`, `t` and `s`, missing values are removed with

```
yy <- y[(1:length(y))[!is.na(y)]]
tt <- t[(1:length(y))[!is.na(y)]]
ss <- s[(1:length(y))[!is.na(y)]]
```

and the mean 114 Wm^{-2}, as well as the empirical variance 65.9 $(\mathrm{Wm}^{-2})^2$ of the yearly means of SWIR in this grid cell, are obtained with `mean()` and `var()`.

Using a linear model, the expectation functions of SWIR at the stations in Table 3.4 in grid cell $\left((10^0\mathrm{E}, 50^0\mathrm{N}), (15^0\mathrm{E}, 52.5^0\mathrm{N})\right)$ are easily estimated after the response variable (the yearly values of SWIR) and the predictor variables (the time in years, and the station identification), as obtained above, are organised in R dataframe `gridbox696`:

```
gridbox696 <- data.frame(ss,tt,yy).
```

The means and empirical variances of SWIR as given in Table 3.4 are calculated using

```
mean(yy[(1:length(yy))[ss=="A"]])
...
var(yy[(1:length(yy))[ss=="M"]])
```

however, the empirical variances are calculated only for those stations where more than 10 yearly values of SWIR are available.

The variance 65.9 $\left(\mathrm{Wm}^{-2}\right)^2$ of SWIR in grid cell $\left((10^0\mathrm{E}, 50^0\mathrm{N}), (15^0\mathrm{E}, 52.5^0\mathrm{N})\right)$, i.e., the variance calculated above from the yearly values of SWIR measured at the stations in Table 3.4, is, as seen in Fig. 3.8, due to the following sources of variability:

1. a decadal trend as SWIR decreases in the observational period, i.e., from 1950 through to 1993
2. non-identical local radiation climates at the stations generating deviations of the local means from the grid cell mean of SWIR, these deviations being the *station effects*
3. fluctuations in SWIR from year to year, i.e., the inter-annual variability as introduced in Sect. 3.1.1, and
4. an error in the pyranometer (introduced in Sect. 1.1) measurement of SWIR.

How can estimates for the decadal trend, station effects, inter-annual variability and the error of measurement be obtained?

Neglecting the station effects, a decadal trend is estimated using the linear model as defined in (3.20)

Table 3.5. Means m in Wm^{-2} and variances s^2 in $(\mathrm{Wm}^{-2})^2$ of the difference time series of SWIR calculated for pairs in the subset (Potsdam, Braunlage, Braunschweig, Fichtelberg) of the set of stations given in Table 3.4 and result of the diagnostics applied to see whether these difference time series are stationary in their first moment function (yes/no in columns stationary).

Station	Potsdam			Braunlage			Braunschweig		
–	m	s^2	stationary	m	s^2	stationary	m	s^2	stationary
Station									
Braunlage	4.4	19.6	yes						
Braunschweig	2.2	19.5	yes	-2.3	26.4	yes			
Fichtelberg	5.5	28.6	no	0.7	24.2	yes	3.3	26.2	yes

$$\mathrm{E}Y_i = a_0 + a_1 t_i \qquad i = 1, \ldots, N \tag{3.20}$$

which requires that the expectation of the yearly SWIR increases or decreases with constant rate in grid cell $\big((10^0\mathrm{E},\ 50^0\mathrm{N}),\ (15^0\mathrm{E},\ 52.5^0\mathrm{N})\big)$ from 1950 through to 1993 with

```
onlytrend.fit <- lm(yy ~ tt, data=gridbox696)   #model (3.20)
```

to obtain

```
                Value Std.Err.  t value Pr(>|t|)
(Intercept) -52.1942 104.1452  -0.5012   0.6168
         tt   0.0841   0.0527   1.5968   0.1120
```

a positive trend of 0.0841 Wm^{-2}/year, and the results given in Table 3.6. With

```
onlytrend.exp <- predict(onlytrend.fit,se=T)
lines(gridbox696$tt,onlytrend.exp$fit,lty=1)
```

this model is plotted as a solid line in Fig. 3.8. Since the residuals of this model contain fluctuations due to station effects, inter-annual variability and an error of measurement, the model needs to be improved.

A linear model for both the trend and station effects can be estimated under the assumption that the station effects remain constant in the observational period. This assumption is in approximate agreement with the observations for the pairs of stations with the longer time series of yearly SWIR values in Table 3.4, i.e., Potsdam, Braunlage, Braunschweig and Fichtelberg stations, as is concluded from Table 3.5. Thus, a linear model for the trend and station effects as defined in (3.21)

$$\mathrm{E}Y_i = a_0 + a_1 t_i + \begin{cases} a_{2A} & \text{for } s_i = \mathrm{A} \\ \ldots & \\ a_{2M} & \text{for } s_i = \mathrm{M} \end{cases} \qquad i = 1, \ldots, N \tag{3.21}$$

is assumed to be reconcilable with the yearly values of SWIR measured at the stations in grid cell $\big((10^0\text{E}, 50^0\text{N}), (15^0\text{E}, 52.5^0\text{N})\big)$ from 1950 through to 1993.

A linear model with a factor (a categorical predictor variable as introduced in Sect. 3.4.2) generally contains more parameters than can be estimated. In the linear model as defined in (3.21), for example, a_0 can be substituted with $a_0 + c$, c fixed, provided that a_{2A}, ..., a_{2M} are substituted with $a_{2A} + c$, ..., $a_{2M} + c$. Consequently, the model is not determined because it has too many coefficients due to its definition (and not because of an insufficient amount of data being available as compared with the number of coefficients to be estimated): the linear model is said to be functionally over-parametrised. This is a consequence of (3.22)

$$\begin{aligned} &\mathrm{E}Y_i = a_0 + a_1 t_i + x_{iA} + \ldots + x_{iM} \qquad i = 1, \ldots, N \\ &x_{iA} = \begin{cases} 1 & \text{if } ss_i = \text{“A”} \\ 0 & \text{else} \end{cases} \quad \ldots \quad x_{iM} = \begin{cases} 1 & \text{if } ss_i = \text{“M”} \\ 0 & \text{else} \end{cases} \end{aligned} \tag{3.22}$$

where the factor s in (3.21) is replaced by 13 indicator variables x_{iA}, ..., x_{iM} with $x_{iA} + \ldots + x_{iM} = (1, \ldots, 1)^T$, the first column vector $\boldsymbol{X}_{\cdot 1}$ in the matrix of the predictor variables in (3.7). This vector, however, pertains to coefficient a_0 (i.e., the intercept) and, consequently, the sum of all indicator variables for the levels of the factor estimates the intercept, too.

In R, prior to estimating a linear model with a k-level factor as predictor variable, a functional over-parametrisation is resolved by generating $k - 1$ indicator variables which are functionally independent of their sum [31]. Such a parametrisation is called a set of *contrasts*. By default, Helmert contrasts are used for a non-ordered factor. More suitable for the linear model (3.21) are the following contrasts

```
contrasts(C(ss,sum))
  [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
A    1    0    0    0    0    0    0    0    0     0     0     0
B    0    1    0    0    0    0    0    0    0     0     0     0
C    0    0    1    0    0    0    0    0    0     0     0     0
D    0    0    0    1    0    0    0    0    0     0     0     0
E    0    0    0    0    1    0    0    0    0     0     0     0
F    0    0    0    0    0    1    0    0    0     0     0     0
G    0    0    0    0    0    0    1    0    0     0     0     0
H    0    0    0    0    0    0    0    1    0     0     0     0
I    0    0    0    0    0    0    0    0    1     0     0     0
J    0    0    0    0    0    0    0    0    0     1     0     0
K    0    0    0    0    0    0    0    0    0     0     1     0
L    0    0    0    0    0    0    0    0    0     0     0     1
M   -1   -1   -1   -1   -1   -1   -1   -1   -1    -1    -1    -1
```

with which the station effects for all stations are easily obtained, as shown below. Using these contrasts,

```
trendandsteff.fit <- lm(yy ~ tt+C(ss,sum),data=gridbox696)
```

calculates the following estimates for model (3.21)

```
                 Value  Std. Error t value Pr(>|t|)
  (Intercept) 173.0659    101.9348  1.6978   0.0913
           tt  -0.0273      0.0514 -0.5315   0.5957
  C(ss, sum)1  -6.8335      3.6902 -1.8518   0.0657
  C(ss, sum)2   0.3692      2.1620  0.1708   0.8646
  C(ss, sum)3 -10.1861      1.5555 -6.5483   0.0000
  C(ss, sum)4  13.8166      4.4884  3.0783   0.0024
  C(ss, sum)5  17.9696      3.7064  4.8482   0.0000
  C(ss, sum)6  -1.3637      3.7064 -0.3679   0.7134
  C(ss, sum)7  -5.3756      1.5298 -3.5140   0.0006
  C(ss, sum)8   4.7893      4.4843  1.0680   0.2870
  C(ss, sum)9   5.3030      3.7064  1.4308   0.1543
 C(ss, sum)10  -8.5886      1.4805 -5.8010   0.0000
 C(ss, sum)11  -0.6970      4.4863 -0.1554   0.8767
 C(ss, sum)12  -6.4662      1.4834 -4.3591   0.0000
```

and with

```
trendandsteff.exp <- predict(trendandsteff.fit,se=T)
lines(tt[ss == "A"],trendandsteff.exp$fit[ss == "A"],lty=2)
...
```

the estimated linear models for the expectation functions at the stations Praha, Coburg, Fichtelberg, Chemnitz and Potsdam are plotted in Fig. 3.8.

$\hat{a}_{2,1}$, ..., $\hat{a}_{2,12}$, calculated above, are estimates for $a_{2,1}$, ..., $a_{2,12}$, the station effects in the parametrisation `C(ss,sum)` for the factor `ss` in `yy ~ tt+C(ss,sum)`, i.e., the formulation of model (3.21) in R. Therefore, $\hat{a}_{2,1}$, ..., $\hat{a}_{2,12}$ are estimates for the vertical displacements of the trends for stations A, ..., L in Table 3.4 with respect to the overall trend of the grid cell. At Braunschweig station (L in Table 3.4), for example, the trend is -6.4662 Wm^{-2} lower than the overall trend for the grid cell with intercept 173.0659 Wm^{-2}, and thus $173.0659 - 6.4662 = 166.5997$ Wm^{-2} is obtained for the intercept of the trend at Braunschweig station. The estimate for the station effect of the last station (Potsdam, M in Table 3.4) is obtained using $\sum_{k=1}^{12}(-1)\hat{a}_{2,k}$ in accordance with the above parametrisation.

The usual (cf. Fig. 3.4) diagnostic plots (a), (b), (c) and (d) in Fig. 3.9 for these estimates are arrived at with R expression

```
plot(trendandsteff.fit)
```

whereas plots (e) and (f) in Fig. 3.9 are generated by plotting the empirical residuals of the linear model against the predictor variables. Plot (g) in Fig. 3.9 results from the application of `acf()` to a R time series generated from the empirical residuals in the order of the observations. Plot (g) is called, in regression terminology, a serial correlation plot, as introduced in Sect. 3.3.1. Plot (h) in Fig. 3.9 is obtained with `qqnorm()` from the empirical

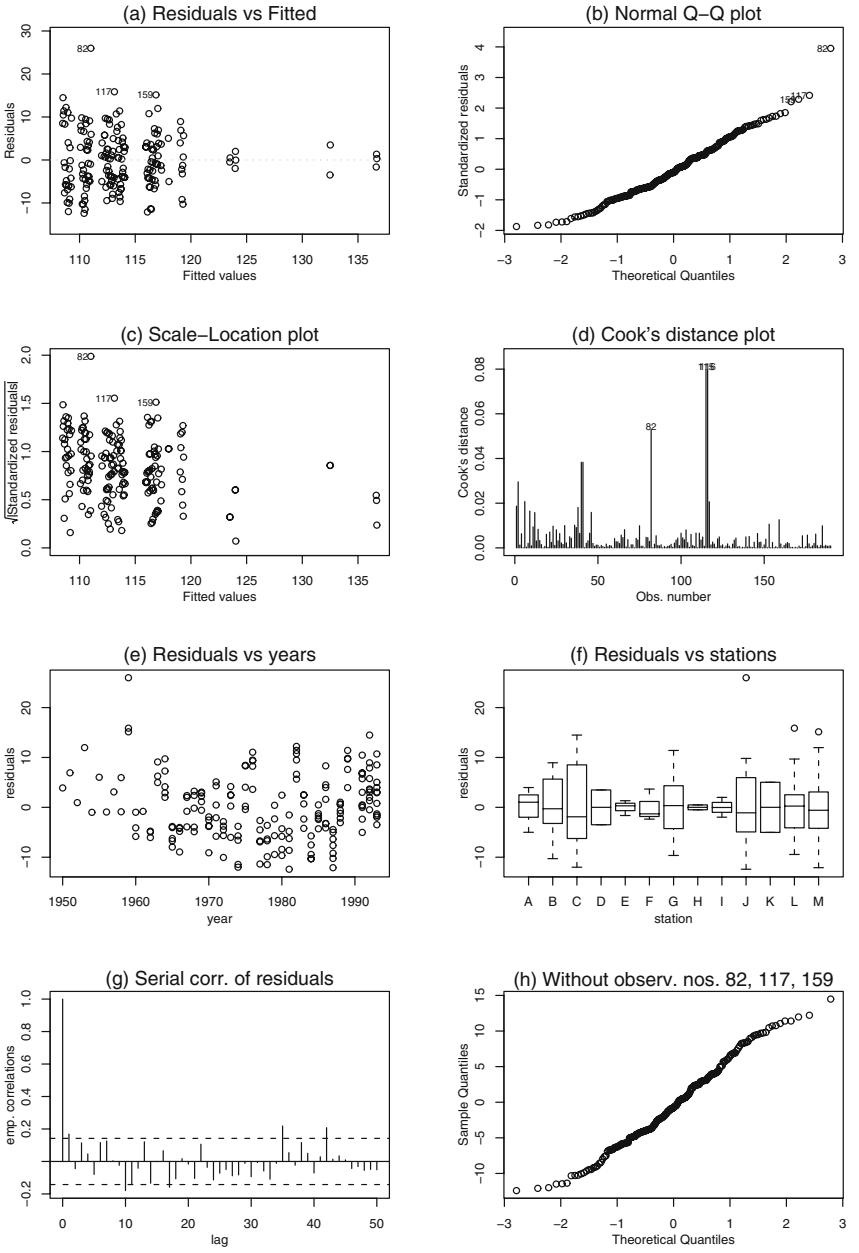


Fig. 3.9. Diagnostic plots for the trend and station effects of SWIR in grid cell $\big((10^0\text{E}, 50^0\text{N}), (15^0\text{E}, 52.5^0\text{N})\big)$ estimated using model (3.21). Plots (b) and (h) are normal probability plots, (b) using all residuals, (h) without the residuals pertaining to observations nos. 82, 117, 159 with large Cook's distances in (d).

residuals after those pertaining to observations nos. 82, 117 and 159 have been removed.

Observations nos. 82, 117 and 159 with very high yearly values (132, 129 and 137 Wm^{-2}) measured in year 1959 at stations Potsdam, Braunschweig and Braunlage (Fig. 3.8) have a large influence on the estimates (Cook's distances, plot (d) in Fig. 3.9). Since, in 1959, extraordinarily high yearly values of SWIR were measured at all stations in neighbour grid cells with measurements available for that year, it is concluded that these observations are correct due to a year with very abundant SWIR in this region.

The unusually high 1959 values also produce, in the empirical residuals of model (3.21) for the yearly values of SWIR measured at the stations in Table 3.4, serial correlations of approximately 0.2, shown in plot (g) in Fig. 3.9, for lags $35 = 117 - 82$ and $42 = 159 - 117$, i.e., the differences in numbers of those observations with the large 1959 values (observations with nos. 82 and 117 as well as with nos. 117 and 159). These correlations are due to (2.1,4) being non-robust, as mentioned in the remarks concluding Sect. 2.1. If these empirical correlations due to 1959 values and the non-robust estimates are neglected when plot (g) in Fig. 3.9 is compared with the templates in Fig. 2.22, it is then concluded from (2.59) that the empirical residuals in the order of the observations stem from a white noise process as required in (3.11) despite the $>$-pattern seen in plots (a) and (c) in Fig. 3.9.

This $>$-pattern is generated by the empirical residuals pertaining to stations Halle/Saale and Harzgerode, both with means of 124 Wm^{-2}, as well as for Zinnwald and Chemnitz with means of 133 and 137 Wm^{-2}. At these stations, however, the time series of the yearly values of SWIR are of insufficient length to claim that their inter-annual variability of SWIR is smaller than at the other stations.

The very large values measured in 1959 at stations Potsdam, Braunschweig and Braunlage can also be seen in plots (e) and (f) in Fig. 3.9 (there, however, the numbers of these observations are omitted). The $\smile$-pattern in plot (e) suggests that a slightly better fit could be obtained with a second order (i.e., a quadratic) linear model for the trend in this grid cell. The differences in the variances at the stations seen in plot (f) (boxplots are default when a numeric variable is plotted vs. a factor [31]) are assumed to be due to the differences in the number of observations available at the stations in Table 3.4.

From the normal probability plot (b) in Fig. 3.9 it is obvious that the tail of the empirical distribution of the residuals is too long to be normal, due to the residuals pertaining to the 1959 observations. When these residuals are omitted, the empirical distribution becomes approximately normal as is seen in plot (h).

Further results for model (3.21) are given in Table 3.6. Table 3.6 shows that the total sum of squares $SS_Y = 65.9 \times 189 = 12455$ $(\mathrm{Wm}^{-2})^2$ (defined in (3.4), the sum of squares pertaining to the grid cell) is split into

Table 3.6. Estimating the linear models as defined in (3.20) and (3.21) for SWIR measured at the stations in Table 3.4. DF: degrees of freedom, SS: sum of squares, MSS: mean square, i.e., SS/DF

model	N	mean in grid cell Wm^{-2}	std.dev. in grid cell Wm^{-2}	std.dev. of residuals Wm^{-2}	Trend Wm^{-2} / year	t	p-value
(3.20)	190	114	8.1	8.1	0.084	1.6	0.11
(3.21)	190	114	8.1	6.7	-0.027	-0.5	0.59

	FG	QS $(\mathrm{Wm}^{-2})^2$	MS $(\mathrm{Wm}^{-2})^2$	F	p-value
(3.20)	1	166	166	2.5	0.11
	188	12279	65		
(3.21)	1	166	166	3.7	0.05
	12	4306	359	7.8	0.00
	176	7973	45		

1. $SS_T = 166$ $(\mathrm{Wm}^{-2})^2$, the sum of squares of the trend, and $SS_E = 12279$ $(\mathrm{Wm}^{-2})^2$, the residual sum of squares when model (3.20) is estimated, and
2. $SS_T = 166$ $(\mathrm{Wm}^{-2})^2$, and $SS_S = 4306$ $(\mathrm{Wm}^{-2})^2$, the sum of squares due to station effects and $SS_E = 7973$ $(\mathrm{Wm}^{-2})^2$ when model (3.21) is estimated.

From the F-statistic given in Table 3.6 it is also concluded, under the — admittedly risky, as is obvious from plots (b) and (h) in Fig. 3.9 — assumption that the residuals are normally distributed, that the negative trend obtained for (3.21) is significant at the $\alpha = 0.05$ test level. The positive trend estimated for (3.20) is, however, not significant even for a "10% climatologist", i.e., a person prepared to run the risk of rejecting the hypothesis of no trend wrongly once in every 10 tests and not once in every 20 tests as usual.

If a linear model for both decadal trend and station effects can be estimated, then the residuals of this model will contain, as the residuals of the linear model for the yearly values of SWIR at Hohenpeissenberg station estimated in Sect. 3.1, the fluctuations generated by the inter-annual variability and the measurement error. The measurement error amounts to 2% of the mean as shown in [58], i.e., 2.3 Wm^{-2} for the mean 114 Wm^{-2} in the grid cell, and is assumed not to depend on the inter-annual variability. Under this assumption, the inter-annual variability is calculated, using (1.16,3), from the variance of the residuals of model (3.21) in Table 3.6:

$45 - 2.3 \times 2.3 = 39.6$ $(\mathrm{Wm}^{-2})^2$ is obtained for the variance of the inter-annual variability of SWIR in grid cell $\big((10^0\mathrm{E}, 50^0\mathrm{N}), (15^0\mathrm{E}, 52.5^0\mathrm{N})\big)$.

The trend, station effects and inter-annual variability estimated using (3.21) for the yearly values of SWIR measured at the stations in grid cell $\big((10^0\mathrm{E}, 50^0\mathrm{N}), (15^0\mathrm{E}, 52.5^0\mathrm{N})\big)$ (Table 3.4) are all reconcilable with the Geophysical knowledge available.

The station effects estimated above result from (i) the differences in altitude of the stations and (ii) the spatial trend of SWIR. This spatial trend can be seen in Fig. 4.16 where the SWIR means have been plotted for those stations where at least five yearly values of SWIR are available. The sources of this trend are the top-of-the-atmosphere solar radiation increasing in a southward direction and the total cloud amount decreasing with increasing distance from the ocean. This spatial trend is also supposed in the remarks to Figs. 4.17 and 4.18 and accounted for when the SWIR climatologies as plotted in Fig. 4.16 are interpolated, in Sect. 4.5.3, using universal kriging.

For the observational period from 1950 through to 1990 negative trends have been observed in Europe, except for regions which have become de-industrialised in those years [58]. Attempts to attribute this negative trend to changes in the energy balance at the surface are made in [151]. A reversal of these trends, i.e., from negative to positive, is found in [152].

The inter-annual variability of SWIR estimated for this grid cell is in line with the values estimated for the neighbour grid cells.

3.6 Trend Surfaces for the Tree-line in the Valais Alps

The expectation function $\mathrm{E}Z(x, y) = \mu_Z(x, y)$ of a random function $Z(x, y)$ with two-dimensional parameter (x, y), x and y being coordinates, is called a trend surface. In this section, linear models for trend surfaces as defined in (3.23) are estimated for the alpine tree-line in the Valais alps. The alpine tree-line is the highest altitude at which trees can grow on mountains, higher up, the vegetation period being of too short a duration to sustain trees.

The tree-lines in the Valais alps as plotted in Fig. 3.10, were observed in the field by the author in summer 1973, taking into account the absence or presence of trees and/or larger areas with typical plant associations of small shrubs, e.g., rhododendron ferrugineum or rhododendron hirsutum, species of alpine roses. The coordinates pertaining to observations (x for the West-East direction, the horizontal axis in Fig. 3.10, y for the South-North direction, the vertical axis in Fig. 3.10, z for the altitude) were obtained from excellent maps produced in scale 1:25000 by the Swiss Federal Office of Topography (SWISSTOPO) and, reading a mechanical altimeter (as electronic equipment was not available at that stage). The error in the z coordinate, originating from the limited accuracy of the altimeter, contributes to the non-systematic variability in the tree-line observations brought about by micro-scale variations in the conditions under which trees can grow at the alpine tree-line

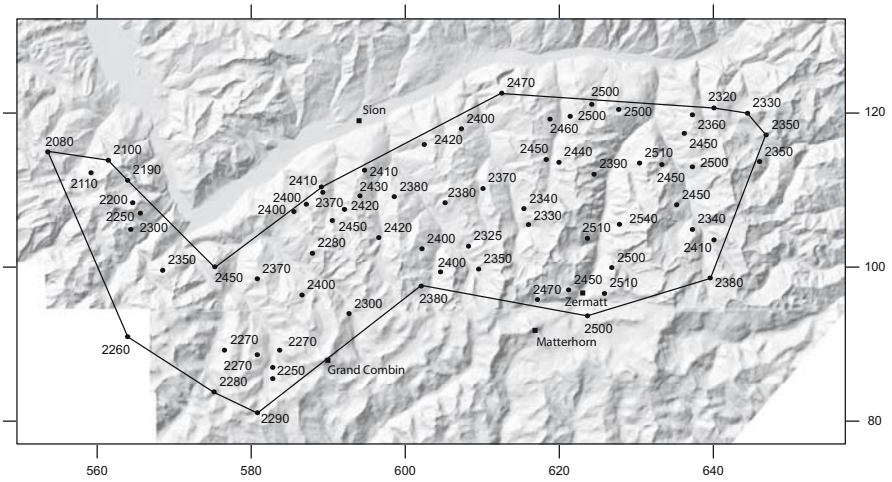


Fig. 3.10. 70 observations (in meter a.m.s.l.) of the tree-line in the Valais alps: the area under study is defined by the closed polygon. Coordinates of the Swiss Federal Office of Topography (SWISSTOPO) are in km. No digital terrain model was available for the SW and SE corners of the map.

[5]; the errors in the x and y coordinates are accounted for, bearing in mind that the estimates calculated in this section are conditional to the observed realisation (cf. the remarks introducing Sect. 3.2).

The observations of the tree-line fluctuate about a local (local properties are defined in (2.55,2)) mean, being low (approximately 2200 m a.m.s.l) in the western part of the area under study, then increasing in easterly direction until its maximum (approximately 2500 m) is arrived at in the Visp valley (including Zermatt), and then decreasing, since lower values (approximately 2350 m a.m.s.l) were observed in the most easterly part of the area under study. Obviously, the observations of the tree-line are not stationary in their mean, and, therefore, a linear model is assumed to be reconcilable with the tree-line observations as plotted in Fig. 3.10.

It is assumed that the observed tree-lines $z(x_i, y_i)$, $i = 1, \ldots, 70$, are a realisation of a random function $Z(x, y)$ as defined in (2.2) with two-dimensional parameter $(x, y) \in U$, U the area under study. Both the observations and the area under study are plotted in Fig. 3.10. The expectation function of the tree-line, $\mathrm{E}Z(x, y) = \mu_Z(x, y)$, is assumed to be a function with continuous derivatives with respect to x and y, i.e., $\mu_Z(x, y)$ is assumed to be a smooth surface. The fluctuations of the $z(x_i, y_i)$ about $\mu_Z(x, y)$ are due to (i) small-scale variations in the terrain, producing micro-climates more or less favourable to the growth of trees and (ii) the error in the readings of the altimeter.

Often, polynomials in x and y are used as linear models for $\mu_Z(x, y)$ because their form is known (at least in the case of low order polynomials) and because they are invariant under translation and rotation when complete, with respect to the order as defined in (3.23):

$$\begin{array}{lll} \text{zero order:} & \mathrm{E}Z(x,y) & = a_0 \text{ (constant)} \\ \text{first order:} & \mathrm{E}Z(x,y) & = b_0 + b_1 x + b_2 y \\ \text{second order:} & \mathrm{E}Z(x,y) & = c_0 + c_1 x + c_2 y + c_3 x^2 + c_4 xy + c_5 y^2 \\ \ldots & \ldots & \end{array} \tag{3.23}$$

The zero order model is a horizontal plane, the first order one a plane inclined in an arbitrary direction, and the second order model has an extremum, i.e., the trend surface becomes more flexible with increasing order.

The observed tree-lines as plotted in Fig. 3.10 are stored in text file `/path/treeline.dat` with three coordinates x, y and z as introduced above

```
646.000 120.700 2330
....
```

in each line. These observations are made available, with R expressions

```
treelinefn <- "/path/treeline.dat"
treelineformat <- list(x=0, y=0, z=0)
treeline <- scan(treelinefn, treelineformat)
x <- treeline$x - 600.000 #origin approximalety in the
y <- treeline$y - 100.000 #centre of the area under study
z <- treeline$z
tl.frame <- data.frame(x,y,z)
```

in a dataframe `tl.frame` (dataframes are introduced in Sect. 3.4.1) after the origin of the two-dimensional parameter (x, y) has been moved into the centre of the area under study. Using `tl.frame`, together with:

```
tl1.fit <- lm(z ~ x+y, tl.frame)
tl2.fit <- lm(z ~ x+y+I(x^2)+I(x*y)+I(y^2),tl.frame)
...
```

trend surfaces as defined in (3.23) are estimated as linear models for the expectation function $\mu_Z(x, y)$ of the tree-line in the Valais alps, and the results and diagnostics in Table 3.7 are obtained. When trend surfaces are estimated, in addition to the usual diagnostic plots introduced in Sect. 3.3.1, the empirical residuals are plotted against the two-dimensional parameter: e.g., in Fig. 3.11, no systematic pattern is seen in the empirical residuals pertaining to the trend surface of order five for the tree-line.

The plots in Figs. 3.11 and 3.12 are obtained with R function `contour()`. `contour()` requires a grid with the values to be represented. For the plots in Figs. 3.11 and 3.12, this grid is obtained as the cross product of R vectors

```
xnull <- c(-52, -48, ..., -4, 0, 4, ..., 48, 52)
ynull <- c(-24, -20, ..., -4, 0, 4, ..., 20, 24)
```

calculated with R expressions

```
xynewdim <- length(xnull)*length(ynull)
x <- rep(0,xynewdim)
```

Table 3.7. Assesment of the linear models as defined in (3.23) for the height of the tree-line in the Valais alps (slightly modified using the example in Table 3.1).

order		1	2	3	4	5	6
number of coefficients		3	6	10	15	21	28
F statistic		26.1	24.2	19.6	15.6	16.9	12.3
R^2 statistic		0.44	0.65	0.75	0.80	.87	.89
$\hat{\sigma}_e$		76	61	54	50	42	43
$\hat{e}_i$ vs.	x_i	$\neq$	$\neq$	$=$	$=$	$=$	$=$
	y_i	$\neq$	$=$	$=$	$=$	$=$	$=$
	(x_i, y_i)	$\neq$	$\neq$	$\neq$	$=$	$=$	$=$
	$\widehat{\mu}_Z(x_i, y_i)$	$\neq$	$=$	$\neq$	$=$	$=$	$=$
normal probability plot		$-$	$+$	$+$	$-$	$+$	$-$
number of extreme values		3	4	3	4	3	3

```
y <- rep(0,xynewdim)
k <- 0
for(i in 1:length(xnull)) {
  for(j in 1:length(ynull)) {
    k <- k + 1
    x[k] <- xnull[i]
    y[k] <- ynull[j]
  }
}
grid.frame <- data.frame(x,y)
```

`grid.frame` contains coordinates `x` and `y` which are used as predictor variables together with the models estimated above, i.e., at the points with coordinates in `grid.frame`, within and outside the area under study, trend surfaces as defined in (3.23) are estimated with

```
tl1pr <- predict(tl1.fit, grid.frame)
tl2pr <- predict(tl1.fit, grid.frame)
...
```

using the linear models `tl1.fit, tl2.fit, ...` obtained above. Estimates calculated with `predict()` are numerically stable, an asset when calculating higher order trend surfaces.

As an aside, the numerical stability of the estimates calculated with R functions `lm()` and `predict()` is improved by re-defining the origin of the two-dimensional parameter (x, y), i.e., by subtracting 600.000 from the x-coordinates and 200.000 from the y-coordinates, when preparing, above, data frame `tl.frame` for the estimation of the linear models.

The size of the grid is chosen to be 4 km in length, because, within this distance, only small changes are supposed to occur in the mean of the ob-

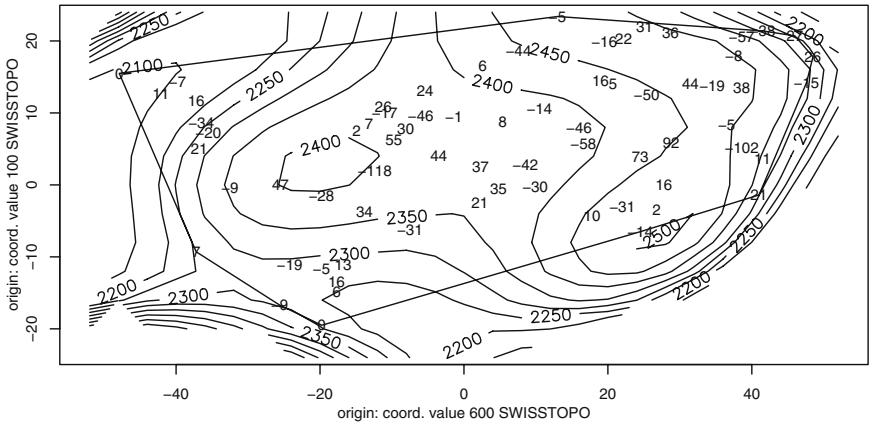


Fig. 3.11. Fifth order trend surface as defined in (3.23) for the tree-line in the Valais alps together with its residuals as estimated from the observations in Fig. 3.10. Coordinates in km, coordinate origins in SWISSTOPO coordinates, altitudes of contours in m, area under study as convex hull of the points with observations.

served tree-lines as plotted in Fig. 3.10. The changes in the mean tree-line are bounded within the area under study; outside these borders, however, larger changes can occur in the estimates and then a too low or too high (i.e., not reconcilable with the growth conditions) expectation function of the tree-line can be estimated. Therefore, from the estimates obtained above in `tl1pr, tl2pr, ...`, only those with $2000 < z < 2600$ m are selected using

```
#build matrix with predicted values for contour()
#include only points with values between 2000 and 2600 m
zpredmat <- matrix(NA,(length(xnull)),(length(ynull)))
k <- 0
for(i in 1:length(xnull)) {
  for(j in 1:length(ynull)) {
    k <- k + 1
    #for the fifth order trend surface
    if((tl5pr[k]<2600)&&(tl5pr[k]>2000)) zpredmat[i,j]<-tl5pr[k]
  }
}
```

to be represented in Figs. 3.11 and 3.12.

When plotting spatial data which depend on equi-scaled coordinates x and y, i.e., coordinates with equal scales, then the scales on both axes should also be equal in the plot. Equi-scaled plots can be obtained in R using `eqscplot()`, i.e., a scatterplot version with scales chosen to be equal on both axes, to set up the coordinate system

```
library(MASS)     #makes eqscplot() available
eqscplot(c(-52,52),c(-24,24),type="no",ratio=1,tol=0.04,
```

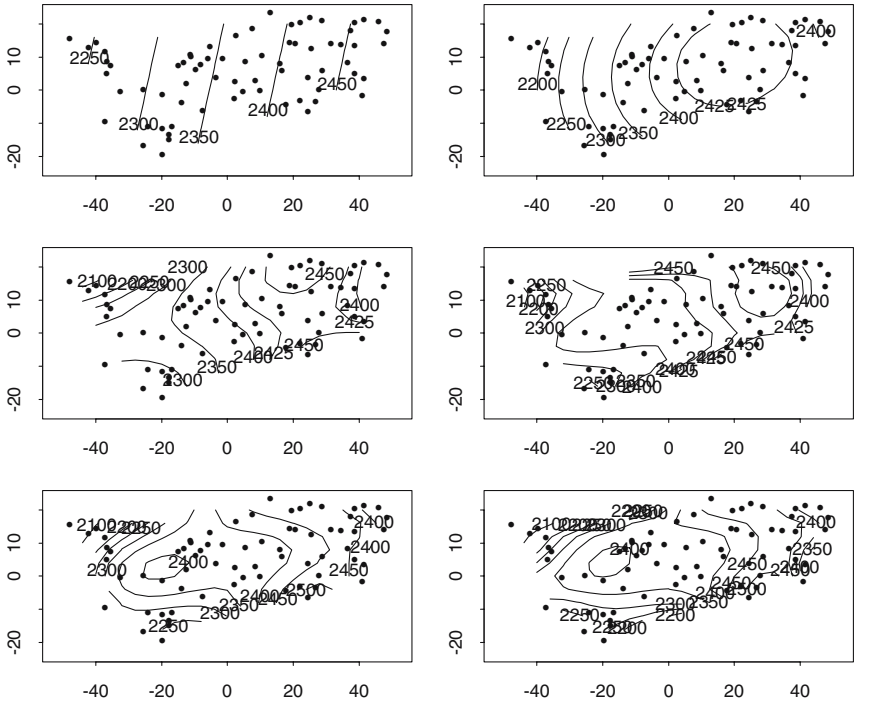


Fig. 3.12. As in Fig. 3.11, for trend surfaces of order 1 through to order 6. An additional contour (2425 m) is plotted for the trend surfaces of the second, third and fourth order.

```
            xlab="origin: coord. value 600 SWISSTOPO",
            ylab="origin: coord. value 100 SWISSTOPO")
```

for R-function `contour()`, which is thereafter used with

```
contour(xnull,ynull,zpredmat,levels=c(2000,2100,2200,2250,2300,
     2350,2400,2450,2500,2600),add=TRUE)
#plot coordinates of observations and their convex hull as a
#substitute for the area under study as defined in Fig. 3.10
xmat <- cbind(tl.frame$x,tl.frame$y)
#points(xmat)     plots points with observations as shown in
#Fig. 3.12, do not use when the observed z-values or the
#residuals are plotted as below and shown in Fig. 3.11
#text(tl.frame$x,tl.frame$y,label=round(predict(tl5.fit)))
text(tl.frame$x,tl.frame$y,label=round(resid(tl5.fit)))
hpts <- chull(xmat) #convex hull
hpts <- c(hpts, hpts[1])
lines(xmat[hpts, ])
```

to obtain the contour plots in Figs. 3.11 and 3.12.

The estimated trend surfaces of orders $1, 2, \ldots, 6$ for the tree-line in the Valais alps are plotted in Fig. 3.12. The first order trend surface does not capture the systematically decreasing tree-line in the most easterly part of the area under study, as is concluded from a comparison of Figs. 3.10 and 3.12 and the diagnostics summarised in Table 3.7. Systematic patterns indicating a model not reconcilable with the observations are also seen in the diagnostic plots of the empirical residuals pertaining to the estimated second and third order trend surfaces.

With an increasing number of coefficients in the model, the estimates come closer to the observations. Consequently, the R^2 statistic increases when $\hat{\sigma}_e$, the empirical standard deviation of the residuals, decreases until it reaches its minimum of 42 m for the fifth order trend surface. However, when the sixth order model is fitted, the minor decrease (as compared to the fifth order model) in the sum of the squared residuals does not compensate for the increase in p, i.e., the number of coefficients estimated (the empirical variance of the residuals having $N - p$ degrees of freedom): both, R^2 and $\hat{\sigma}_e$, increase, as is seen in Table 3.7. Although the fourth, fifth and sixth order trend surfaces are all reconcilable with the observations, as is seen in the diagnostic plots of their empirical residuals, the fifth order one is preferred because its empirical residuals have minimal variance and stem from a normal distribution.

Therefore, the fifth order trend surface for the tree-line in the Valais alps as plotted in Fig. 3.11 is found to be in agreement with all assumptions in Sect. 3.2 and thus is preferred to the other models in Table 3.7. The fifth order trend surface is also reconcilable with the knowledge available from plant ecology. For example, Aulitzky and Turner found in [5] that the alpine tree-line is determined by the duration of the vegetation period. The vegetation period (the period without snow cover) at constant altitude, e.g., 2200 m, is of longer duration in the interior of the alps than it is on the exterior slopes, as can be seen from the maps available in the season with snow cover for the Swiss alps from [131].

3.7 Problems

3.1. Derive the estimators in (3.2).

3.2. Apply (1.15) to show that the estimate $\hat{b}_1$ in (3.2) is not biased.

3.3. Calculate the variance of $\hat{b}_1$ in (3.2).

3.4. When is $\hat{b}_1$ normally distributed? A hint is given in the remarks to the definition of the multivariate normal distribution in (1.34).

3.5. Plot the density of the standardised normal distribution including the quantiles -4.0 and 4.0. Then add, to this plot, the densities of the t-distributions with 4 and 39 degrees of freedom. Further, plot the t-value -3.6086, i.e., the t-statistic as defined in (3.3) pertaining to the linear model estimated in Sect. 3.1.2 for trend in SWIR at Hohenpeissenberg station. What is the probability of obtaining a $t \leq -3.6086$ for this trend?

3.6. `lm()` calculates, as required in (3.3), the empirical variance for the estimated residuals of a linear model. If, in the remarks to Fig. 3.1, the empirical residuals are assumed to be a time series, then their variance is calculated as required in (2.1,2). Which estimate do you prefer?

3.7. Prove (3.4).

3.8. Calculate the moments of the estimated expectation function $\widehat{\mu}_Y(x)$ in (3.5) and (3.6).

3.9. Plot confidence intervals for the trend of SWIR at Hohenpeissenberg station as shown in Fig. 3.2. (Often, alternative graphical representations are used for these confidence intervals.)

3.10. Estimate the trend of SWIR at Hohenpeissenberg station without using the 1959 observation.

3.11. In Problem 2.7, as well as in the remarks to Table 3.2, the values in the Basel temperature series are made available as objects in R and, in Fig. 3.3, linear models for the expectation function of the yearly values in the Basel temperature series are shown. Split the series of the yearly values into parts suitable for estimating local first order linear models and then estimate the local linear models. Can the usual significance tests be applied? Possible answers can be obtained from arguments similar to those used in Sect. 2.5.2.

3.12. Plot the values for January, April, July and October available in the Basel temperature series and estimate a first order linear model for each of these months. Which monthly series are stationary in the first moment function?

3.13. The yearly values for SWIR at Hamburg station are available in a file with 17 values in each line: the first, second and third values are identifications, the fourth is the year of measurement, the remaining ones are the monthly and yearly values of SWIR in Wm^{-2}. These values are made available in R vectors using

```
fn <- "/path/hamburg.dat"
format <- list(gridnum=0,stnum=0,stcellid="",year=0,jan=0,feb=0,
               mar=0,apr=0,may=0,jun=0,jul=0,aug=0,sep=0,oct=0,
               nov=0,dez=0,yearlyswir=0)
hamburg <- scan(fn,format)
```

Plot the time series of the yearly values of SWIR at Hamburg station and its empirical correlation function.

3.14. The plot generated in Problem 3.13 shows that the yearly values of SWIR at Hamburg station are not stationary in the first moment function. Estimate the trend suggested by the plot.

3.15. Are the differences in the monthly values of SWIR calculated in Problem 1.11 stationary in their first moment functions? The differences for the Toronto and Toronto Met. Res. stations as well as for the London Weather Centre and Bracknell stations are plotted in Fig. 2.19. The following R expressions generate from `diftor`, the R time series with the Toronto differences obtained in Problem 1.11,

```
diftor1 <- as.array(diftor)
time1 <- time(diftor)
plot(time1,diftor1,type="l",xlab="year ",ylab="W/m^2")
```

R vectors for the differences and for the time, from which missing values are removed with

```
diftor2 <- diftor1[(1:length(diftor1))[!is.na(diftor1)]]
time2 <- time1[(1:length(time1))[!is.na(diftor1)]]
```

to obtain R vectors that can be used as the response and predictor variable in `lm()`. Is the estimate for the trend significant?

From the remarks to Fig. 2.19 it is concluded that the time series of the Toronto differences is not reconcilable with all assumptions in (3.1). Which of the assumptions is not met? What follows for the trend estimated above?

3.16. Estimate a linear model for the trend in the global mean temperatures obtained in Problem 2.30.

3.17. The mean annual cycle of the temperature in Basel for the period from 1755 through to 1957 is estimated in the remarks to Table 3.2 under the assumption that the monthly processes are stationary. Plot the mean annual cycle, i.e., plot the means of the monthly time series vs. the months. Then plot the empirical variances of the monthly series vs. the months. Do the empirical variances have an annual cycle?

3.18. Subtract the means calculated in Problem 3.17 from the values in the monthly series to obtain a seasonally de-meaned time series of the monthly values in the Basel temperature series.

3.19. Estimate linear models for the trend in the monthly and yearly values in the Basel temperature series and compare your results with the trend estimated in the remarks to Table 3.7 using an additive component model.

3.20. Estimate a linear model for the trend in the seasonally de-meaned time series obtained in Problem 3.18.

3.21. Estimate an additive component model for trend and annual cycle in the Basel temperature series using the R expressions in the remarks to Table 3.7, but with the origin of R vector `idx` at the start of the observational period, i.e., month 0 is December, 1754.

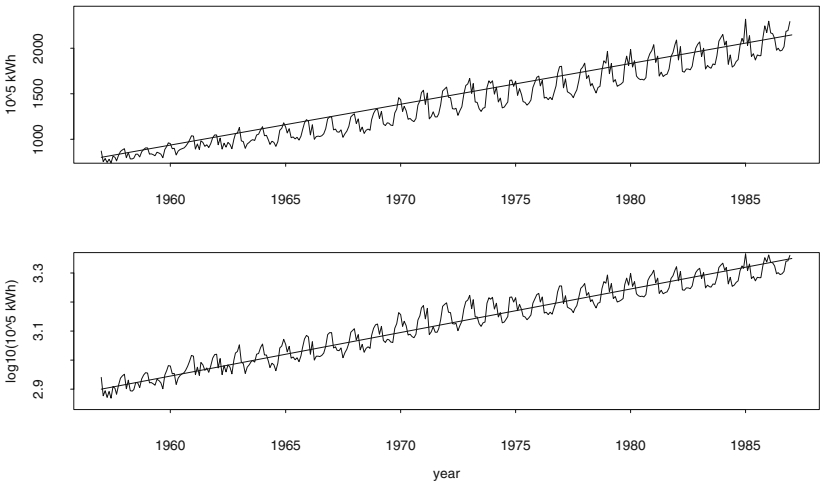


Fig. 3.13. Above, monthly values of power consumed in the city of Zurich in 10^5 kWh; below, the logarithms of the above values. The trends are drawn as straight lines using trial and error.

3.22. File `/path/zhpower5786.dat` contains the Zurich monthly consumption of power in 10^5 kWh for the period 1957 through to 1986. The first value in the file pertains to January 1957. From this file, using R expressions

```
zhpower <- ts(scan("/pfad/zhpower5786.dat"),
                          start=1957,frequency=12,)
plot(c(1957,1987),c(800,2400), type="n",xlab="",ylab="10^5 kWh")
lines(zhpower, type="l")
lines(c(1957,1987),c(800,2150), type="l") #the fourth attempt
```

the monthly values of power consumed in the city of Zurich are made available as R time series `zhpower` and then plotted above, in Fig. 3.13. Next, the logarithms are taken and plotted below in Fig. 3.13 with the straight line plotted in the second attempt. Estimate a component model as defined in (3.19) for the logarithms of the Zurich power consumption.

File `/path/zhpower8505.dat` contains the Zurich quarterly consumption of power in 10^3 kWh for the period 1985 through to 2004. Construct a homogeneous time series for the power consumption in Zurich since 1957. Estimate a component model for the logarithms of this time series.

The data in these files are a courtesy of ewz, Zurich Municipal Electric Utility, Tramstrasse 35, P.O. Box, 8050 Zurich.

3.23. The atmospheric carbon dioxide monthly values measured at Mauna Loa Observatory (Hawaii) can be downloaded from the internet address given in [84]. Estimate a component model.

4 Interpolation

An interpolation aims at approximating (the estimation of) a (random) function $Z(x)$, x being a multi-dimensional argument (parameter if the function is random, as defined in (2.2)), for argument (parameter) values without observations or, more generally, at the approximation (estimation) of linear functionals (mentioned in the remarks to (1.1)) of $Z(x)$, e.g., of integrals of $Z(x)$ over subsets in the argument (parameter) domain. Often, the argument (parameter) is two-dimensional. For example, the tree-line observations plotted in Fig. 3.10 are a realisation of a random function with a two-dimensional parameter.

A sharp distinction between interpolation and extrapolation is not advantageous when applied to functions with a multi-dimensional argument because mostly a multi-dimensional area with observations cannot be defined unambiguously as has been demonstrated for the two-dimensional case by the tree-line observations in the Valais alps: the area as defined in Fig. 3.10 is not identical with the convex hull of the points with observations in the (x, y)-plane, obtained in the remarks to Fig. 3.11. Hence, extrapolation is only used to emphasise that the function under study, namely $Z(x)$, x multi-dimensional, is approximated or estimated for x being a larger distance away from the observations $z(x_i)$, $i = 1, 2, \ldots, N$, (compared with the average distance between the x_i), i.e., for x not surrounded by x_i.

There are probabilistic and deterministic interpolation methods. Deterministic interpolation methods can be used when the function $Z(x)$ to be interpolated is quite smooth and the observations are afflicted with only a small measurement error. For example, a freehand drawing of contours on a map when interpolating a set of observations $z(x_i)$, $i = 1, \ldots, N$, is a deterministic interpolation method. Freehand contouring is often called subjective interpolation, whereas objective interpolation is used when a mathematically defined interpolation function is applied, regardless of whether the method is probabilistic or deterministic. Sect. 4.1 introduces two groups of usual deterministic interpolation methods.

If a random function $Z(x)$ is interpolated with a probabilistic method then the properties of $Z(x)$ as described by its moment functions $\mathrm{E}Z(x)$ and $\mathrm{Cov}\big(Z(x_1), Z(x_2)\big)$ are explicitly taken into account, as is shown in the remaining sections in this chapter.

In Sect. 4.2 the stationary and intrinsically stationary spatial random functions are defined. The second moment function of an (intrinsically) stationary random function can be estimated by calculating the empirical variogram, as demonstrated in Sect. 4.3. Using the empirical moments as defined in Sect. 4.3 and the estimators as defined in Sect. 4.4, interpolation methods for spatial random functions are thereafter introduced in Sect. 4.5 where the equations for the interpolation of a stationary, an intrinsically stationary and a spatial random function with a non-constant expectation function are given. Sect. 4.6 contains the supplements, including (i) the interpolation of spatial random functions from observations afflicted with an error, (ii) the possibility to define an interpolation neighbourhood, and (iii) the estimation of spatial averages.

The problems are posed in Sect. 4.7.

In the second half of the last century, methods for the optimal interpolation of spatial random functions were developed and applied by mining engineers for exploring ore deposits: *Geostatistics* emerged as a discipline with its own technique for the analysis of ore deposit data and with its own terminology. Comprehensive textbooks on Geostatistics are [41] and [35].

Optimal interpolation techniques as developed in Geostatistics have, when compared with deterministic interpolation methods, many advantages which are enumerated in the remarks concluding Sect. 4.5. Hence, optimal interpolation is the best choice when ore deposits are explored/exploited: the expensive measurements (often obtained from drilling cores) are used exhaustively (i.e., maximal use of the measurements is made) and the estimated interpolation error allows for assessing the risks of production when investing in the mining of deposits. Apart from mining, optimal interpolation is used with increasing frequency in many disciplines, e.g., in water resources management, forestry or agriculture.

Meteorological variables have been optimally interpolated since about 1960 [54]. Today, optimal (but also deterministic) interpolation procedures are an integral part of the *data assimilation*. Data assimilation is the process of combining observations (as obtained at the surface, from upper-air soundings, and from satellites) with the dynamics (as implemented in a model, usually a GCM, in the introduction to Sect. 2.7) of the atmosphere in order to obtain a comprehensive and dynamically consistent data set as an estimate of the state of the atmosphere at the time of observations [56]. The assimilated data set is (i) comprehensive, i.e., estimates are available at all model grid points, not only at those near the observations, and (ii) dynamically consistent, i.e., a sequence of assimilated data sets for time points $t_1, t_2, \ldots$ is reconcilable with the dynamic principles relevant to the atmosphere. Data sets obtained from assimilation are then used to calculate weather forecasts with GCMs. Clearly, data assimilation methods can also be applied to other dynamic systems (e.g., the ocean, a glacier or the Martian atmosphere) to obtain a "movie" of their evolution in time and space.

4.1 Deterministic Interpolation

In (4.1), a variable $z(x)$ is interpolated using a deterministic method. Some terms in (4.1) are repeated in (4.6) where a random function is interpolated by way of a probabilistic method.

1. *$z(x)$ is a deterministic function of the argument x, with x being a one- two- or n-tuple of coordinates $x^*, y^*, \ldots$, e.g., $x = (x_1^*, x_2^*, x_3^*)$ or $x = (x^*, y^*)$. $z(x)$ has been measured at points $x_1, x_2, \ldots, x_N$, and thus x_i, $i = 1, \ldots, N$, are called points with measurements. The x_i are in a bounded and known multi-dimensional domain D in which $z(x)$ is assumed to exist. Extending the two-dimensional terminology, D is often called area under study or area with observations.*
2. *$z(x_1), z(x_2), \ldots, z(x_N)$ or $z_1, z_2, \ldots, z_N$ are called data, observations, measured values or values.*
3. *The point x_0 or x is an interpolation point provided that (i) it is not a point with measurements and (ii) $z(x)$ is not known.*
4. *The interpolation of $z(x)$ at x_0 or x is a function $\hat{z}\big(x_0, x_1, \ldots x_N, z(x_1), \ldots, z(x_N)\big)$ which approximates $z(x_0)$ as good as possible; the criteria used for the approximation, however, depend on the interpolation method.*
5. *Linear functionals of $z(x)$ are calculated by applying further operations on $z(x)$, e.g., by calculating integrals.*
6. *If $\hat{z}(x_i) = z(x_i)$ for $i = 1, \ldots, N$, then the interpolation is called exact.*

(4.1)

Distance weighted methods interpolate using an interpolation function as defined in (4.1,4) with arguments calculated from distances between x and x_i, two examples being given in Sect. 4.1.1. Other interpolation methods use x_i, $i = 1, 2, \ldots, N$, to obtain a tessellation of D (i.e., in the case of two-dimensional coordinates, a collection of figures, e.g., triangles, that fill D excluding gaps and overlaps). Thereafter, the $z(x_i)$ are interpolated in the figures of the tessellation obtained, an example being given in Sect. 4.1.2.

4.1.1 Distance Weighted Interpolation

A very straightforward interpolation is a linear combination of the observations using weights which are a function of the distance $r_i = r(x_0, x_i)$ between the x_i and x_0 as defined in (4.2).

$$\hat{z}(x_0) \;=\; \sum_{i=1}^{N} a_i z(x_i) \qquad\qquad a_i \;=\; \frac{g\big(r(x_0, x_i)\big)}{\sum_{j=1}^{N} g\big(r(x_0, x_j)\big)} \tag{4.2}$$

In (4.2), the distance $r_i = r(x_0, x_i)$ is often chosen to be $|x_0 - x_i|$, the absolute value of the difference of x_i and x_0, with $g(r)$ being a function of the distance. From (4.2) it becomes obvious that an interpolation weight a_i depends only

Table 4.1. Distance weighted interpolation of climatological mean SWIR. SWIR is measured at the stations and simulated using a climate model at the grid points.

station coordinates	observed Wm^{-2}	interpolated Wm^{-2}	gridpoint long.	gridpoint lat.	model Wm^{-2}	distance km	weight
Leopoldville/	179.8	188.6	14.625	−5.046	196.8	104.1	0.2096
Kinshasa-Binza			15.750	−5.046	192.8	95.4	0.2293
15.25^0 E			15.750	−3.925	182.1	72.7	0.3013
4.347^0 S			14.625	−3.925	185.9	83.7	0.2598
St. Petersburg/	108.6	87.6	29.250	58.878	90.4	134.8	0.0345
Leningrad			30.375	58.878	90.7	121.5	0.0382
30.3^0 E			30.375	59.999	87.3	5.5	0.8479
59.967^0 N			29.25	59.999	87.7	58.5	0.0793

on $r(x_0, x_i)$, and not on $r(x_i, x_k)$, i.e., the distances between the points with observations. $\sum_{i=1}^{N} a_i = 1$ is required in (4.2) to obtain $\hat{z}(x) = c$ for $z(x) = \mathrm{c}$, c a constant.

In a first example, distance weighted interpolation allows for calculating, in [149], the differences of observed (obtained from pyranometer measurements, cf. Sect. 1.1) and simulated (from a climate model) values of shortwave incoming radiation (SWIR) at 720 stations worldwide. Since (i) climatological means of SWIR do not change abruptly, as seen in Table 3.4 and Fig. 4.16, and (ii) the error in the measurements is not large (2% of the mean for yearly SWIR values, from which the climatologies are calculated [58]), SWIR can be interpolated with the distance weighted method in the grid cells of a T106 grid with grid size 1.125 degree latitude and longitude (i.e., a high-resolution GCM grid, as shown in Table 4.1.

In this distance weighted interpolation, the interpolation neighbourhood is defined by the grid of the climate model since each station with SWIR measurements is in a grid cell: a station (located in x as defined in (4.1,3)) is surrounded by four grid points (the x_i as defined in (4.1,1)) and the interpolation distances are smaller than the diagonal of a grid cell. The inverse of the four distances calculated on a sphere with radius 6371 km is applied as interpolation function (4.1,4). For example, Table 4.1 shows that, for Leopoldville/Kinshasa-Binza station, the value interpolated from the simulated values is not too far from the observed one, whereas for St. Petersburg/Leningrad station, the observed value is much larger than the interpolated one, provided that the interpolation error is approximately the concentration (in the remarks to (1.17)) of the four simulated values about their mean.

Usually, and therefore in opposition to the above SWIR example, the x_i are irregularly scattered in D, the x are grid points and, in addition to having the properties as required in (4.2), an interpolation $\hat{z}(x, \ldots)$ as defined in (4.1,4) should meet the following demands:

1. *When not calculating spatial averages, $\hat{z}(x \ldots)$ is exact, as defined in (4.1,6) or, at least, approximates the $z(x_i)$ close to x.*
2. *The interpolation neighbourhood depends on x_i: often, $z(x_i)$ at a larger distance from x can be neglected without losing accuracy in order to reduce the time needed for the calculations.*
3. *$\hat{z}(x, \ldots)$ should depend not only on the distances between x and x_i but also on the directions of x_i as seen from x: $\hat{z}(x, \ldots)$ should not depend on $x - x_i$ solely but also on $x_i - x_k$.*

(4.3)

For example, the SYMAP algorithm [123] is reconcilable with the requirements stipulated in (4.1) and (4.3): its interpolation function depends on the absolute values and the directions of the $x - x_i$ as well as on the $x_i - x_k$ taking into account a possibly varying density of the x_i in D. Using a SYMAP version with spherical (introduced in the remarks to Table 4.1) distances, some world-wide precipitation climatologies are obtained [121]. Another SYMAP version is in use to produce daily precipitation maps for the Alps and, from these maps, a precipitation climatology for the period from 1971 through to 1990 and region $\big((2^0\text{E}, 42^0\text{N}), (19^0\text{E}, 50.0^0\text{N})\big)$ was calculated [50].

The SYMAP algorithm as described in [123] and modified as described in [50] is used to produce an areal precipitation map (shown in Fig. 4.1) from observations (rain gauge measurements) for October, 5, 1992 and region $\big((2^0\text{E}, 42^0\text{N}), (19^0\text{E}, 50.0^0\text{N})\big)$. The interpolation points are on a grid with distances of 0.3^0 in easterly and 0.22^0 in northerly directions, resulting in a grid size of approximately 25 km as seen in Fig. 4.3.

The stations with rain gauge measurements used for the interpolation as mapped in Fig. 4.1 are plotted in Fig. 4.2. From Fig. 4.2 it is obvious that the spatial density of the rain gauge stations decreases rapidly in Italy and thus cannot be assumed to be constant in the region under study. The varying density of the x_i is accounted for by using the scaled interpolation distance r_s as defined in (4.4)

$$r_s = \sqrt{\Big(\frac{x_0^* - x_i^*}{s\Delta x^*}\Big)^2 + \Big(\frac{y_0^* - y_i^*}{s\Delta y^*}\Big)^2} \qquad s = 1, 2, 3, 4 \tag{4.4}$$

with (x_0^*, y_0^*) being the coordinates of the interpolation point x_0 on the 25 km grid, (x_i^*, y_i^*) the coordinates of a x_i, $i = 1, 2, \ldots, N$, $\Delta x^* = 0.3$ and $\Delta y^* = 0.22$ the size of the grid in easterly (x^*) and northerly (y^*) directions. The algorithm tries $s = 1$, $s = 2$, $s = 3$ and as a last possibility $s = 4$, until at least three x_i are obtained within a distance of less or equal $r_s = 1$ from x_0. If $s \leq 4$ does not exist then an interpolation is not possible. For each x_0, $r_s = 1$

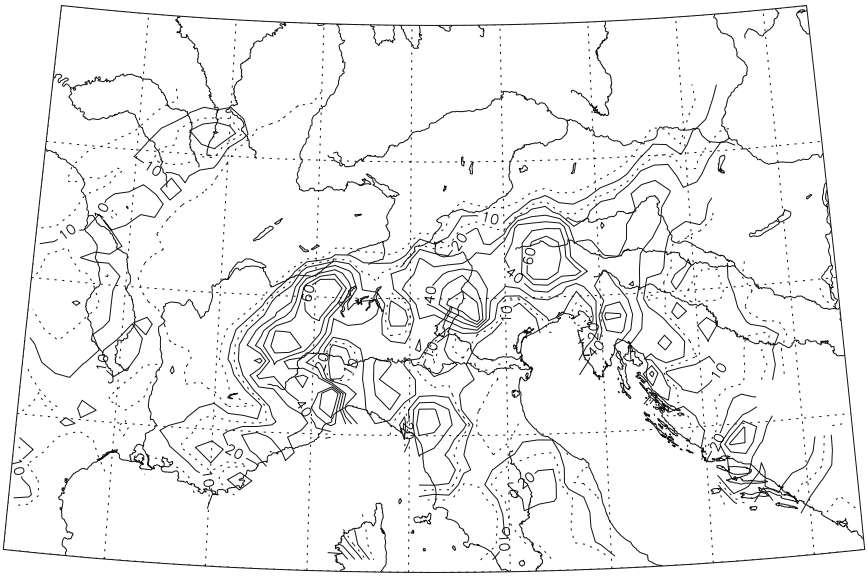


Fig. 4.1. Areal precipitation in mm for October, 5, 1992 and region $\big((2^0\text{E}, 42^0\text{N}), (19^0\text{E}, 50.0^0\text{N})\big)$ interpolated as described in [50].

describes an ellipse with half-axes $s\Delta x^*$ and $s\Delta y^*$ being the interpolation neighbourhood of the SYMAP algorithm as described in [123] and modified in [50].

The original SYMAP algorithm described in [123] determines the interpolation neighbourhood, i.e., s in (4.4), by trial and error until at least four but not more than ten observations are found in its interior. Then, in this interpolation neighbourhood, a rapidly decreasing interpolation function $g(r) = 1/r^2$ is applied by (i) neglecting observations outside and (ii) giving equal weights to observations very close to the interpolation location. This results in a "quite exact" (cf. (4.1,4)) interpolation when the interpolation location is close to locations with observations. From the interpolated values, areal averages are then obtained by smoothing, i.e., by calculating averages from the interpolated values in defined areas.

Applying a slowly decreasing interpolation function (for example, the function (4.5) as described in [50])

$$g(r_s) = \begin{cases} \big(1 + \cos(\pi r_s)\big)/2 & \text{for } r_s \le 1 \\ 0 & \text{for } r_s > 1 \end{cases} \qquad r_s \text{ as defined in (4.4)} \tag{4.5}$$

averages are obtained directly for areas dependent on $s\Delta x^*$ and $s\Delta y^*$ in (4.4), i.e., on the interpolation neighbourhood, which, obviously, depends on the size of the interpolation grid and the density of the observations.

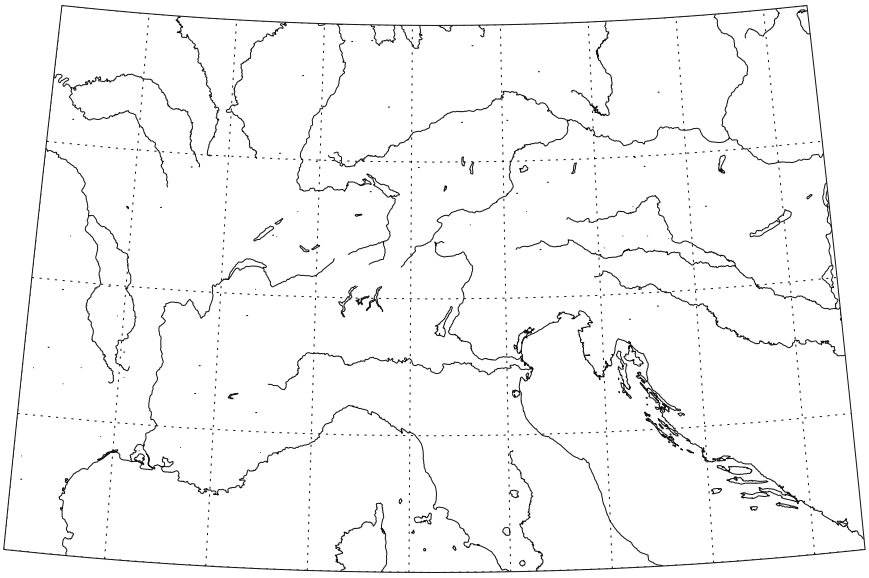


Fig. 4.2. 6623 locations with observations (rain gauge stations) used for the interpolation of the areal precipitation shown in Fig. 4.1.

The directions of the x_i as seen from x is accounted for, in the usual SYMAP algorithm, by calculating $1 - \cos(\theta_{ik})$ from the angle θ_{ik} between $x - x_i$ and $x - x_k$ and then multiplying the weight $1 - \cos(\theta_{ik})$ with $g(r_s)$ obtained for pairs of stations using (4.5). Weighting with $1-\cos(\theta_{ik})$ becomes plausible when three cases are distinguished: (i) if, seen from x, the angle between x_i and x_k is small then $1 - \cos(\theta_{ik}) \approx 0$, (ii) if x_i and x_k are in approximately opposite directions then $1 - \cos(\theta_{ik}) \approx 2$, and (iii) if the angle of x_i and x_k is approximately 90^0 then $1 - \cos(\theta_{ik}) \approx 1$. These weights are also applied in the SYMAP version described in [50].

(4.4) and (4.5) in the SYMAP version described in [50] act together when the areal precipitation for October 5, 1992 and region $\big((2^0\text{E}, 42^0\text{N}), (19^0\text{E}, 50.0^0\text{N})\big)$ is interpolated with the result as mapped in Fig. 4.1: as shown in detail in Fig. 4.3, the interpolation produces (i) using (4.4), an areal precipitation even in regions with a small density of rain gauge stations and (ii) due to (4.5), an areal precipitation smooth enough, even in regions with a high density of stations. Hence, the areal precipitation is of constant smoothness in the area under study, despite the changing density of observations. If these precipitation data were interpolated without using a variable interpolation neighbourhood then the resulting areal precipitation would appear to be smooth in regions with a low density of stations and rough in regions with a high density of stations.

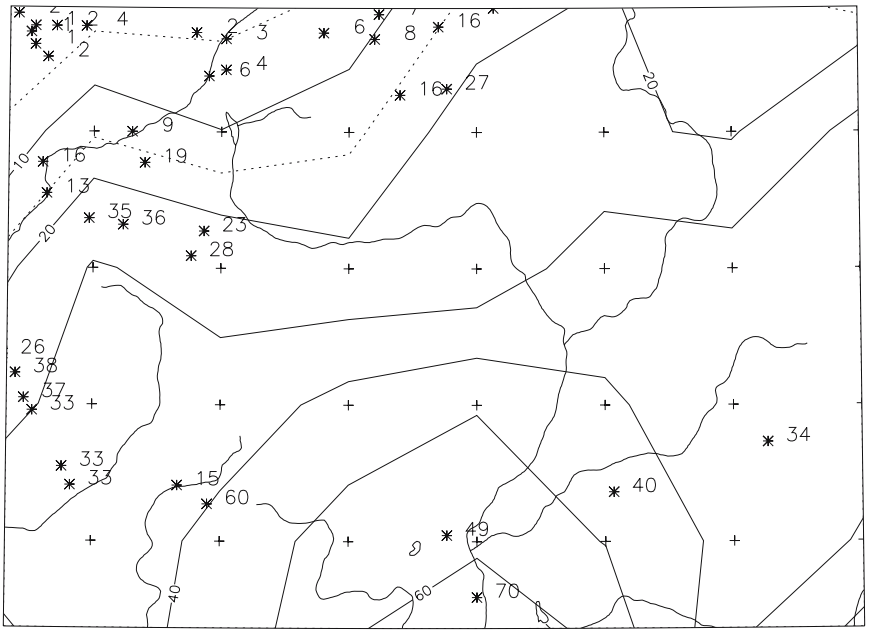


Fig. 4.3. Zooming into Figs. 4.1 and 4.2: region $\left((10^0\text{E}, 46^0\text{N}), (12^0\text{E}, 47.0^0\text{N})\right)$ with lower density of rain gauge stations in Italy than in Switzerland and Austria. $*$ locations with observations, $+$ interpolation location on a grid with sizes 0.3^0 (in west-east direction) and 0.22^0 (in south-north direction).

The areal precipitation interpolated using the modified SYMAP algorithm in Fig. 4.1 is similar to the estimate as mapped in Fig. 4.34 obtained when using block kriging, the standard probabilistic method for estimating areal means. The probabilistic method, however, also estimates the variance of the interpolation as mapped in Fig. 4.35.

4.1.2 Tessellation of the Area under Study

As an alternative to using an interpolation function which depends on the distances between interpolation points x and points with observations x_i, $i = 1, 2, \ldots, N$, deterministic interpolation methods are available that partition, using all x_i, the domain D into subsets that exhaust D and do not intersect, i.e. in the two-dimensional case, into triangles, rectangles or other polygons that fill D (excluding gaps and overlaps). The set of polygons thus obtained is called a *tessellation* of D. These polygons are then used for the interpolation: linearly on the sides, using low order polynoms in the coordinates in the interior. For example, a polygon whose interior consists of all points in D closer to x_k than to any $x_i \neq x_k$, x_i and x_k being points with

observations, is called a Voronoi polygon [6] and the resulting tessellation a Dirichlet tessellation. As a second example, a set of triangles that tessellate D is used in Akima's interpolation.

Akima, in [2], triangulates D being in the x^*-y^*-plane such that x_i, $i = 1, 2, \ldots, N$, become vertices of the triangles. The first side in the first triangle is defined by the pair (x_i, x_k) most close to each other. Then the centre c of this side is determined. From the remaining x_i, the one closest to c becomes the third vertex of the first triangle. When the first triangle is complete, further triangles are added: one for each of the remaining x_i ordered with increasing distance from c. The sides of the triangle added are obtained by drawing lines from x_i to the vertices of those triangles that are "visible", i.e., the sides of the triangle added do not intersect with any side of the triangles already existing. This order asserts that each x_i added is outside the convex hull of the x_k already triangulated, i.e., having become vertices of one or more triangles.

Then, in the interior of the triangles, Akima's algorithm calculates low-order polynomials in x^* and y^* from the partial derivatives $(\mathrm{d}/\mathrm{d}x^*)z$, $(\mathrm{d}^2/\mathrm{d}x^{*2})z$, $(\mathrm{d}/\mathrm{d}y^*)z$, $(\mathrm{d}^2/\mathrm{d}y^{*2})z$ and $(\mathrm{d}/\mathrm{d}y^*)\big((\mathrm{d}/\mathrm{d}x^*)z\big)$ such that the function defined by the polynomials becomes smooth on the sides of the triangles. The partial derivatives at an x_k are calculated using the observed z_i and the $z(x_i)$ at the n nearest x_i. Usually, $3 \leq n \leq 10$ is chosen.

For example, when the observations of the tree-line in the Valais alps as plotted in Fig. 3.10 are interpolated using Akima's algorithm with $n = 5$:

```
#function interp(), an implementation of Akima's interpolation
#algorithm, is available in Splus [11], an ancestor of R.
#tl.frame      observations of the tree-line (Sect. 3.6)
#xnull,ynull   grid with interpolation locations: interp()
#              generates the objects for contour() (Sect. 3.6)
#n             number of observations used for the
#              calculation of the partial derivatives:
#              the interpolation becomes smoother with
#              increasing n
w <- interp(tl.frame$x,tl.frame$y,tl.frame$z,
       xnull,ynull,n=5,extrap=F)
contour(w, levels=c(1800,1900,2000,2100,2200,2250,2300,2350,2400,
       2450,2500),xlab="0:600 SWISSTOPO",ylab="0:100 SWISSTOPO")
points(tl.frame)
text(tl.frame,label=z)
```

the result in Fig. 4.4 is obtained. Above, $n = 5$ was selected by trial and error to make Akima's interpolation similar to the fifth order trend surface for the tree-line in the Valais alps as plotted in Fig. 3.11. A smoother interpolation is, however, obtained with $n > 5$.

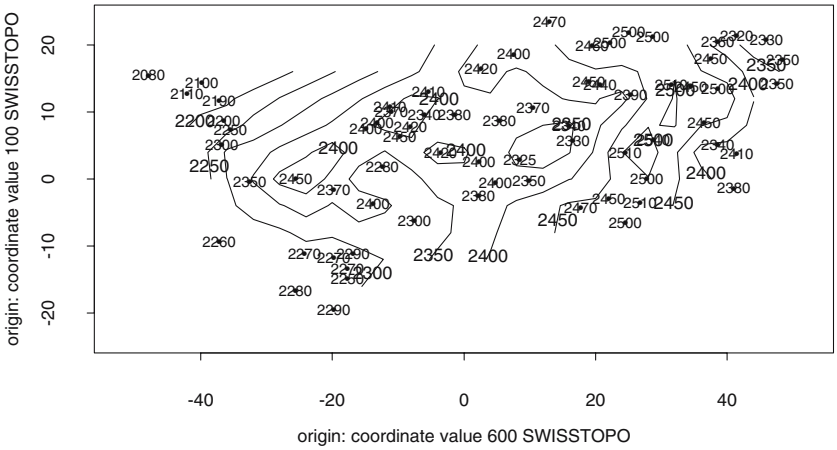


Fig. 4.4. Tree-line in the Valais alps interpolated using Akima's algorithm from the observations in Fig. 3.10. Coordinates (approximately equi-scaled) and contours as in Fig. 3.11.

4.1.3 Discussion

Often, when not high accuracy is called for, a distance weighted interpolation method is used which is simple to implement and fast. Most distance weighted methods can be easily tailored to suit specific applications: distances, interpolation neighbourhoods and interpolation functions are chosen to arrive at the desired result. An example is the version of the SYMAP algorithm introduced in Sect. 4.1.1. A tessellation of D (the domain in which the variable under analysis $z(x)$ is assumed to exist) can be useful when D is to be partitioned into subsets influenced by the x_i using, e.g., Voronoi polygons.

Deterministic interpolation methods, however, have the following disadvantages:

1. The interpolation neighbourhood is defined by the method, e.g., using (4.4) when interpolating with the SYMAP version as described in [50] or by $n = 5$ when using Akima's method for the interpolation of the tree-line in the Valais alps as demonstrated in Fig. 4.4.
2. The interpolation function (used in distance weighted methods) or the partition of D (used in tessellation methods) depend on the algorithm selected and not solely on the data.
3. The interpolation error is not explicitly part of the algorithm and thus has to be assessed somehow from the known variability of $z(x)$: the method is not optimal when applied to a random function, i.e., does not predict with the least mean square prediction error (cf. definition (2.61,2)).

4.2 The Spatial Random Function and its Variogram

It is assumed that the observations $z(x_i)$ to be interpolated stem from a realisation of a spatial random function $Z(x)$. The moments of a spatial random function can be estimated (Sect. 4.3) and then used for an interpolation of $Z(x)$ with probabilistic methods (Sect. 4.5).

1. *A random function $Z(x)$ with a (generally) four-dimensional parameter being the coordinates in space and time is called a spatial random function or regionalised variable. Often, the parameter is two-dimensional.*
2. *x, y, ... are points (locations) in the parameter space, and x^*, y^*, z^*, x_1^*, x_2^*, ... are coordinates.*
3. *A realisation $z(x)$ of $Z(x)$ is observed at points x_i, $i = 1, \ldots, N$, to obtain observations $z_i = z(x_i)$.*
4. *$Z(x)$ is assumed to exist in a known parameter domain D.*
5. *x_0 or x is an interpolation point as defined in (4.1).*

(4.6)

D is usually given by its geographical boundaries in longitude, latitude and altitude and by specifying a period if time is a dimension of x.

In D, $Z(x)$ is often assumed to be continuous. This assumption applies to many meteorological (temperature, humidity, etc.) and geometric (altitude, thickness of a deposit, etc.) variables for physical reasons. An example of a variable being non-continuous is the shortwave downward radiation absorbed at the surface: it jumps at a coast because the albedos (the fractions of radiation reflected from the ground) of water and a solid surface differ. In D, $Z(x)$ is usually non-isotropic, i.e., the properties of $Z(x)$ are not independent of direction. Temperature, for example, changes more rapidly with increasing altitude than with increasing horizontal distance.

The moment functions of a spatial random function $Z(x)$ are functions of the parameter x, as can be seen in (4.7).

Let $Z(x)$ be a spatial random function as defined in (4.6), Then:

1. $\mathrm{E}Z(x) = \mu_Z(x) = \mu(x)$ *is called expectation function or drift,*
2. $\mathrm{Cov}\big(Z(x_i), Z(x_k)\big) = c_Z(x_i, x_k) = c(x_i, x_k) = c_{ik}$
   or $\mathrm{Cov}\big(Z(x_i), Z(x_0)\big) = c_Z(x_i, x_0) = c(x_i, x_0) = c_{i0}$
   or $\mathrm{Cov}\big(Z(x_i), Z(x)\big) = c_Z(x_i, x) = c(x_i, x) = c_{i0}$
   or $\mathrm{Cov}\big(Z(x), Z(y)\big) = c_Z(x, y) = c(x, y)$ *covariance function,*
3. $\mathrm{Var}Z(x) = \sigma_Z^2(x) = \sigma^2(x)$ *variance function and*
4. $\mathrm{Var}\big(Z(x) - Z(y)\big) = 2\gamma_Z(x, y) = 2\gamma(x, y)$ *variogram and $\gamma_Z(x, y)$ or $\gamma(x, y)$ semivariogram of $Z(x)$. From (1.15,8) $2\gamma_Z(x, y) = \sigma_Z^2(x) + \sigma_Z^2(y) - 2c_Z(x, y)$ is obtained.*

(4.7)

(4.7) contains, when compared with (2.4), additional definitions often used in Geostatistics: (i) "drift" is used to avoid "trend" for an expectation function $\mathrm{E}Z(x)$ that is possibly not constant, and (ii) (half of) the variance of differences of $Z(x)$ calculated for differences in the parameter x, i.e., displacements, is called (semi-) variogram.

If a large number of independent realisations of the spatial random function $Z(x)$ have been observed then many values $z(x_i)$ are available. Since the $z(x_i)$ are iid., as defined in (1.2), the estimates known from an introduction to Statistics can be calculated for the moment functions of $Z(x)$ defined in (4.7). Usually, however, observations $z(x_i)$ stemming from only one realisation are available and, consequently, estimates can be calculated only under additional assumptions. Usually, the spatial random function $Z(x)$ under analysis is assumed to be stationary in (4.8), or intrinsic in (4.9).

The moment functions of a stationary spatial random function are invariant under translations as defined in (4.8):

A spatial random function $Z(x)$ as defined in (4.6) with moment functions as defined in (4.7) is stationary on condition that:
1. $\mathrm{E}Z(x) = \mu_Z(x) = \mu$, *i.e., its expectation function is constant, and*
2. $\mathrm{Cov}\big(Z(y), Z(x)\big) = \mathrm{Cov}\big(Z(x+h), Z(x)\big) = c_Z(h) = c_Z(y-x)$, *its covariance function depends, for x and y in D, on the translation $h = y - x$ and not on x and/or y.*

(4.8)

The translations h in (4.8) are multi-dimensional as the parameter x is multi-dimensional and, therefore, translations are possible in more than one direction whereas, in definition (2.8), the parameter t of the stochastic process $X(t)$ is one-dimensional and translations τ are only possible in time.

Many random functions are found to be not stationary: using the diagnostics introduced in Sect. 2.5, for example, a random walk process (X_t) as defined in (2.20) with $\mu_W = 0$ is not stationary since its covariance function $c_X(t,u) = \min(t,u) \neq c_X(t-u)$ as obtained in (2.69) is not a function of the translation $t - u$. Consequently, the moment functions of these processes cannot be estimated under the stationarity assumptions as required in (2.8) or (4.8).

Are estimates possible under other, less restrictive, assumptions? Less restrictive than the stationarity is the intrinsic hypothesis as defined in (4.9)

A stationary random function $Z(x)$, as defined in (4.6), with constant expectation function $\mu_Z(x) = \mu_X$ satisfies the intrinsic hypothesis, (or, equivalently, is intrinsically stationary) on condition that its differences $Z(x+h) - Z(x)$ are stationary:
1. $\mathrm{E}\big((Z(y) - Z(x)\big) = \mathrm{E}\big((Z(x+h) - Z(x)\big) = 0$ *and*
2. $\mathrm{Var}\big(Z(y) - Z(x)\big) = \mathrm{Var}\big(Z(x+h) - Z(x)\big) = \mathrm{E}\big(Z(x+h) - Z(x)\big)^2$
$= 2\gamma_Z(h) = 2\gamma_Z(y-x)$ *for all x and $x+h$ in D.*

(4.9)

which requires differences $Z(x+h) - Z(x)$ of the spatial random function to be stationary. Comparing definitions (4.8) and (4.9) it is immediately concluded that a stationary random function satisfies the intrinsic hypothesis; the inverse, however, is not true.

For example, the random walk process $X(t) = X(t-1) + W(t)$, $W(t)$ being a normally distributed white noise process with $\mu_W = 0$ and $\sigma_W^2 = 1$,

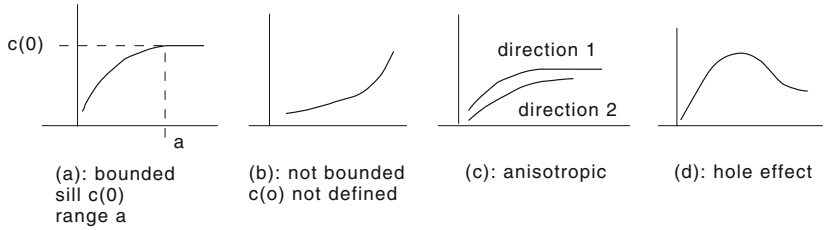


Fig. 4.5. Variogram types according to its behaviour for $|h|$ not close to 0. $|h|$ is plotted on the horizontal axis, $2\gamma(|h|)$ on the vertical axis.

is intrinsically stationary, since (i) its drift $\mu_X(t) = 0$ is constant, as obtained in (2.67), and (ii) the variance of differences $X(t) - X(u)$ depends on $h = u - t$ solely: $\mathrm{Var}\big(X(t+h) - X(t)\big) = \mathrm{Var}\big(\sum_{u=1}^{t+h} W(u) - \sum_{u=1}^{t} W(u)\big) = \mathrm{Var} \sum_{u=t+1}^{t+h} W(u) = \sum_{u=1}^{h} \sigma_W^2 = h = 2\gamma_X(h)$, as is derived from its covariance function $c_X(t,u) = \min(t,u)$ in (2.69). Thus, as the variogram of a random walk process increases linearly and is not bounded (plot (b) in Fig. 4.5), both properties are in agreement with the divergence of those realisations of a random walk process demonstrated in Fig. 2.12.

In contrast, the variogram of a stationary random function $Z(x)$ is bounded, as derived in Problem 4.4: $\gamma_Z(h) = c_Z(0) - c_Z(h)$, $c_Z(0) = \sigma_Z^2(x)$ being the constant variance function of $Z(x)$. Hence, the variogram $\gamma_Z(h)$ of a stationary random function $Z(x)$ increases (plot (a) in Fig. 4.5) and comes closer to its *sill* $c(0)$ with increasing h, provided that the covariance function $c_Z(h)$ decreases with increasing h: $\gamma_Z(h) = c_Z(0)$ for $h \geq a$, a being the *range* of the variogram.

Two additional variogram types can be distinguished by following the behaviour of $2\gamma(h)$ for translations h not close to zero. A variogram with a peak (plot (d) in Fig. 4.5) corresponds to a possibly existing (when $Z(x)$ is stationary) covariance function with a *hole effect*, pointing to a periodic second moment function. An *anisotropic variogram* (plot (c) in Fig. 4.5) depends on the translation $h = y - x$, i.e., the direction and the distance $|h| = |y - x|$, whereas an *isotropic variogram* depends solely on the distance: $2\gamma(h) = 2\gamma(|h|)$. A geometric anisotropy can be captured using the generalised distance as defined in (4.10) for spatial random functions with two-dimensional parameter.

Let x_i^* and y_i^* be the coordinates of $x \in D$ and $y \in D$,
g_{ik} be constant and let $h_i^* = x_i^* - y_i^*$. Then
$|h^{(g)}| = \big(g_{11}h_1^{*2} + g_{12}h_1^*h_2^* + g_{22}h_2^{*2} + \ldots\big)^{1/2}$ (4.10)
is called generalised distance of x and y.

Points at identical generalised distance away from a point $x_0 \in D$ are on an ellipse described by $\gamma(x - x_0) = a$ or $\gamma(x - x_0) = a$, a constant (plot (b) in Fig. 4.6).

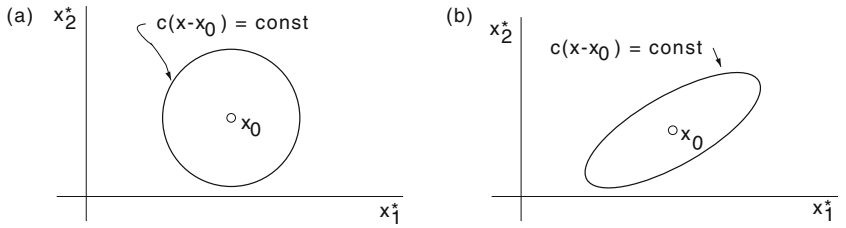


Fig. 4.6. For an isotropic spatial random function, the covariance of points on the circle and its centre x_0 is constant (plot a); for a geometrically anisotropic one, the covariance of points on the ellipse and x_0 is constant, the points on the ellipse being at identical generalised distance (4.10) away from x_0 (plot b).

In addition to terms such as sill, range, etc. used in Geostatistics to describe a variogram for translations not close to zero, the terms smooth, continuous and nugget effect are used to describe a variogram for very small translations, i.e., for h close to zero: variograms with $2\gamma_Z(h) = 0$ are either smooth in $|h| = 0$ (plot a in Fig. 4.7) or continuous (plot b in Fig. 4.7), whereas variograms increasing over a very short distance from 0 to a finite number and thereafter showing a moderate increase, as shown in plots (c) and (f) in Fig. 4.7, are said to have a *nugget effect*. The estimate for a variogram having a nugget effect, however, jumps in $|h| = 0$, as shown in plot (e) in Fig. 4.7 (variograms are estimated by calculating the empirical variogram as introduced in Sect. 4.3).

The term nugget effect was coined by Geostatisticians to describe variograms for $Z(x)$, $Z(x)$ being the amount of gold per unit volume, in deposits where gold is found in the form of nuggets. In such a deposit, $Z(x)$ jumps when moving into the interior of a nugget and, therefore, its variogram $2\gamma_Z(h) = 2\gamma_Z(y-x) = \mathrm{Var}\big(Z(y) - Z(x)\big)$ changes very rapidly for very small differences of x and y. In contrast, the empirical variogram, being calculated from measurements of $Z(x)$ in volumes larger than the nuggets (Sect. 4.6.2), jumps in $|h| = 0$. In general, an empirical variogram with a jump in $|h| = 0$ which is not due to a measurement error (Sect. 4.6.3) is said to have a nugget effect, i.e., a spatial random function with a large natural variability for very small translations has a variogram with a nugget effect.

A white noise process has a variogram with a pure nugget effect as shown in plot (d) in Fig. 4.7. The white noise process (W_t) with integer parameter t is defined in (2.10) or (2.11). If, however, t is real, then a stochastic continuous (defined in (7.16,1))) white noise process does not exist: in the remarks to (7.22) it is shown that $\mathrm{Cov}\big(Z(x_i), Z(x_k)\big) = 0$ for $x_i \neq x_k$, with $|x_i - x_k|$ very small and close to zero, is not possible on condition that $Z(x)$ is a stochastic continuous random function with real parameter x.

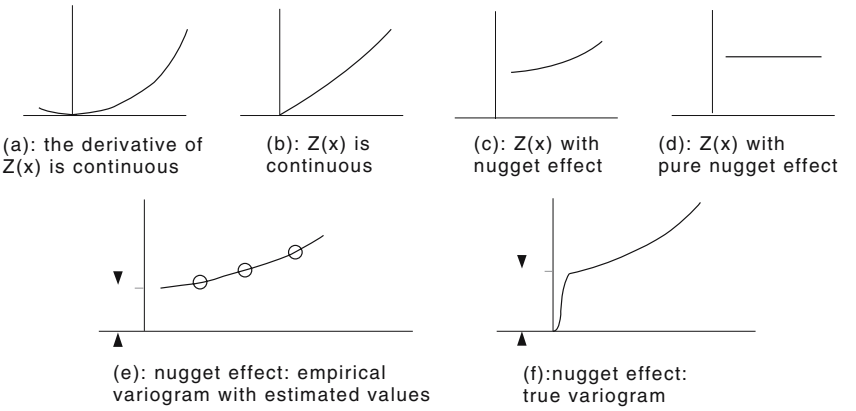


Fig. 4.7. Variogram types according to its behaviour for $|h|$ not close to 0. $|h|$ is plotted on the horizontal axis, $2\gamma(|h|)$ on the vertical axis. Below, nugget effect in the (empirical) variogram for h in higher resolution.

4.3 Estimating the Variogram

The variogram $2\gamma_Z(h)$ of a spatial random function $Z(x)$ is estimated with the empirical variances of differences $z(x_i) - z(x_k)$ provided that many independent observations are available for all x_i. In Geosciences, however, often only one realisation of $Z(x)$ has been observed and, therefore, only one $z(x_i)$ is available for all x_i. In this case, it is assumed that $Z(x)$ is intrinsically stationary and the empirical variogram $2\hat{\gamma}_Z(h)$ is calculated as an estimate for $2\gamma_Z(h)$ by partitioning the set of all differences $x_i - x_k$ using direction and distance.

The empirical variogram $2\hat{\gamma}_Z(h)$ *is an estimator for the variogram* $2\gamma_Z(h)$ *of a spatial random function* $Z(x)$*, satisfying the intrinsic hypothesis as defined in (4.9), provided that*

$$2\hat{\gamma}_Z(h) = \big(1/n(h)\big) \sum_{x_i - x_k \,\in\, \text{class}(h)} \big(Z(x_i) - Z(x_k)\big)^2 \tag{4.11}$$

is calculated for those $n(h)$ *pairs* (x_i, x_k) *with differences* $x_i - x_k \approx \pm h$ *in the anisotropic case and* $x_i - x_k \approx |h|$ *in the isotropic case.*

Differences $x_i - x_k \approx \pm h$ as required in (4.11) are obtained by sorting the $x_i - x_k$ into classes of distance and direction as shown in Fig. 4.8; thereafter, values $2\hat{\gamma}_Z(h)$ are calculated for each distance class in the isotropic case and for each combined distance and direction class in the anisotropic case.

In the isotropic case, an empirical variogram can be calculated without using direction and distance classes as demonstrated in (4.40).

Definition (4.11) is, even in combination with Fig. 4.8, a skeleton, the body of the variogram estimation only becoming visible when empirical variograms are calculated for the examples given in Sects. 4.3.1, 4.3.3 and 4.3.4.

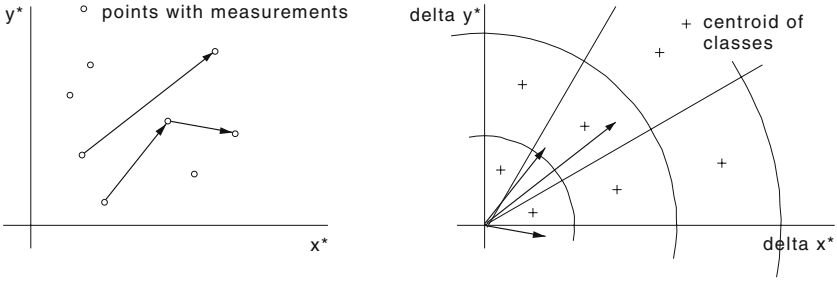


Fig. 4.8. Sorting the $x_i - x_k$ into classes of distance and direction shown for two-dimensional x: three differences in the (x^*, y^*)-plane (on the left), the same differences in the $(\Delta x^*, \Delta y^*)$-plane being partitioned into classes of distance and direction as required in (4.11) (on the right).

4.3.1 Empirical Variograms I: the Isotropic Case

Cereals used for bread making and forage crops are cultivated under crop rotation on plot "Strassenacker" of the Inforama Rütti located in Zollikofen (near Berne, Switzerland). On this plot, the soil type undergoes variations and, consequently, also soil parameters show spatial variations. For example, potassium, in water soluble and exchangeable form (K, written as oxide K_2O in this chapter) which can be taken up by the roots via diffusion, has values varying between 1.6 and 7.1 mg in 100 g of soil, i.e, a range from a medium to a very abundant K_2O content. As K_2O is important for many physiological processes (including photosynthesis), crop growth may not be constant over the plot and, consequently, also biomass and grain yield may show spatial variations. If the spatial variability of soil properties and yield within a plot are known from measurements and interpolation then soil and crop management can be carefully chosen to arrive at a sustained yield.

In plot "Strassenacker", soil parameters and crop yields depend on past and present natural (peri-glacial sedimentation at the end of the last ice-age and subsequent climate and vegetation history) and antropogenic processes (farming methods practised since the plot became arable land during Celtic or Roman times) with intricate inter-dependencies and are, therefore, assumed to be spatial random functions. These random functions can be predicted using the methods introduced in Sect. 4.5 provided that their covariance functions or their variograms are known. As estimates, empirical variograms are calculated using (4.11) on condition that observations are available for a sufficiently large number of points in the "Strassenacker" plot.

In year 1995 (the crop was oats), the yield (biomass and grain) and some soil parameters were measured in plot "Strassenacker" [63]. The grain yield and biomass were obtained by mowing a square with side 1 m about each point x_i with measurements, threshing and then weighing immediately (to obtain the fresh weight) as well as after some time in the compartment drier

(to obtain the dry weight). Consequently, the yield values are averages covering 1 m^2. Values for the soil parameters were obtained from probes with 5 cm diameter in the first 20 cm of soil, and, consequently, soil parameter values are averages over approximately 400 cm^3 of soil. The resulting data are available in text file `/pfad/stracker.dat`, one line for each measurement with comma-separated values, in the first line the abbreviated names (short names in the following paragraphs) of the variables measured:

```
X,Y,pH,OrgMat,Ntot,CNratio,grain,biomass,potass,phosph
0,0,6.9,4.1,0.29,8.2,587.5,1147.3,7.1,0.73
8,0,6.2,2.7,0.29,5.4,450.4,909.3,3.9,0.56
...
```

This file is read with

```
stracker <- read.table("/path/stracker.dat", sep=",", header=T)
```

to obtain R dataframe `stracker` with vectors

```
#X         coordinate, distances of observations are 15 m, and
#          8 m and 12 m at margins of plot
#Y         coordinate, distances of observations are 15.66 m
#pH        pH-value
#OrgMat    organic matter (in %)
#Ntot      total nitrogen (in %)
#CNratio   ratio of C (carbon) and N (nitrogen)
#grain     grain yield (g dry matter per m2)
#biomass   biomass (plants without roots, g dry matter per m2)
#potass    potassium in water soluble and exchangeable
#          form K2O (mg per 100 g of soil)
#phosph    phosphorus in water soluble and exchangeable
#          form P2O5, (mg per 100 g of soil)
```

from which, for example, histograms of the K_2O measurements as plotted above, in Fig. 4.9, are obtained with

```
hist(stracker$po,xlab="potassium")   #...$po for ...$potass
hist(log(stracker$po),xlab="log(potassium)")
```

and the plot below, in Fig. 4.9, of the logarithms of the K_2O measurements is obtained using

```
plot(c(0,160), c(0,160), type = "no", xlab = "m", ylab="m")
title("log(potassium)")
text(stracker$X, stracker$Y,label=round(log(stracker$po),2))
```

with R function `log()` calculating the logarithm.

Do the K_2O observations $z(x_i)$ as plotted below, in Fig. 4.9, stem from a realisation of a stationary spatial random function $Z(x)$? No systematic increase or decrease is seen in the $z(x_i)$ and, therefore, the expectation function $\mu_Z(x)$ is assumed to be constant over the "Strassenacker" plot, as required in (4.8,1) or (4.9,1). Is $Z(x)$ stationary in its covariance function as required in (4.8,2)? An answer to this question is given by calculating the empirical variogram $2\hat{\gamma}_Z(h)$: if $2\hat{\gamma}_Z(h)$ is bounded, i.e., increases until it reaches a sill

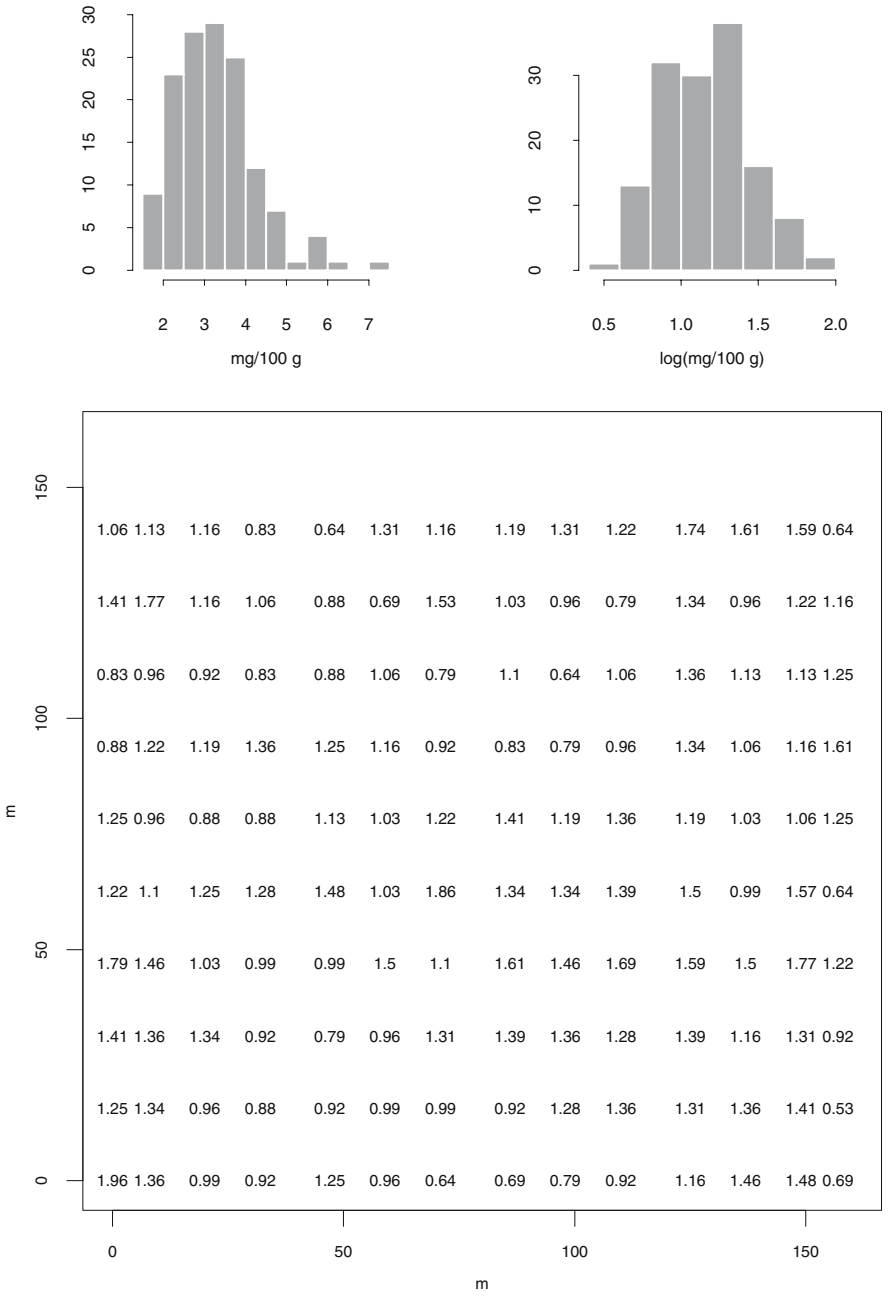


Fig. 4.9. Observations of potassium, in water soluble and exchangeable form K_2O (log(mg in 100 g of soil)) in plot "Strassenacker' (below) and its histograms prior and subsequent to taking the logarithm (above).

as shown in Fig. 4.5 (a), then $Z(x)$ is assumed to be stationary in the second moment function.

In R, functions for spatial analysis are made available by invoking `library(spatial)`. In 1999, when the examples in this chapter were analysed, the functions in the spatial library [136] available at that time for Splus were found to be inferior in performance to those available in Gstat as described in [107] and [105]. Since 2003, Gstat is also available as R package ([106] and link "R spatial projects" available under [114]), the author, however, prefers the Gstat stand-alone version [105] because of its interfaces with Geographical Information Systems (GIS) as argued in the remarks concluding Sect. 4.5.1.

Gstat in its stand-alone version reads the observations from text files with all values pertaining to one observation in one line, separated by spaces, provided that these files contain GeoEAS header lines as described in [105]: the first header line contains a description of the data, the second the number of the variables, the third and following header lines contain the descriptions of the variables, one line for each variable. For example, the following GeoEAS file `/path/stracker.eas`

```
stracker
11
number of observation
X
...
short names as in the first line in file /path/stracker.dat above
...
phosph
1 0 0.00 6.9 4.1 0.29 8.2 587.5 1147.3 7.1 0.73
...
```

with the "Strassenacker" observations is obtained by writing R dataframe `stracker` generated above on file `/path/stracker.eas` and thereafter adding the header lines using a text editor. The numbers in the first column in `/path/stracker.eas` are generated when the observations are read into R dataframe `stracker`, as the lines have no numbers in file `/path/stracker.dat`.

Having prepared the GeoEAS table file `/path/stracker.eas`, please invoke the Gstat user interface for the variogram estimation by typing

```
gstat -i
```

and thereafter select one of the following possibilities from the menu:

1. `enter/modify data` to read a GeoEAS file with one or more variables, as Gstat is designed for the estimation of multivariate spatial random functions
2. `choose variable` to select one or two variables
3. `calculate what` to select the moment to be estimated
4. `cutoff, width` to choose the maximum distance and width of the distance classes to be used for computing the empirical variogram as defined in (4.11) and Fig. 4.8

5. `direction` to choose the direction classes to be used for calculating the empirical variogram
6. `variogram model` to select the model to be fitted to the empirical variogram (examples are given in Fig. 4.10)
7. `fit method` to select the method for the fitting of the model selected
8. `show plot` to calculate and plot the results.

Select `enter/modify data` and then answer the following questions

```
Choose identifier or enter new one [] : Potassium
definition of 'Potassium' (variable 1):
Enter data file name [] : /path/stracker.eas
Enter x coordinate column number [0] : 2
Enter y coordinate column number [0] : 3
Enter variable column number [0] : 10
Log transform data [n] : Yes
Average duplicate observations [n] : Yes
/path/stracker.eas (GeoEAS file)
attribute: log(potass) [x:] X: [ 0, 157]
n:                  140 [y:] Y: [ 0, 141]
sample mean:   1.16752 sample std.: 0.284795
Press any key to continue...
```

to initialise Gstat for calculating the empirical variogram of the K_2O observations in the "Strassenacker" soil:

1. The identifier `Potassium` is used for the K_2O observations (observation of potassium in water soluble and exchangeable form) in the "Strassenacker" soil by specifying the columns with the observations of the variable under analysis in the GeoEAS file `/path/stracker.eas`.
2. The logarithm of the observations is taken since the histogram on the right in Fig. 4.9 is more symmetric than the histogram on the left.
3. The mean of observations with identical coordinates is calculated to prepare for a possible interpolation which applies the kriging systems as derived in Sect. 4.5.

Gstat then calculates the mean and the standard deviations given above and is ready to compute the empirical variogram of the variable under analysis as required in (4.11) and Fig. 4.8.

It is supposed that K_2O is globally (as defined in a version of (2.55) for spatial random functions) stationary and, therefore, empirical variograms are meant to cover a maximal distance of 140 m, i.e., the length of the shorter side of the "Strassenacker" plot becomes the maximal distance in the $(\Delta x^*, \Delta y^*)$-plane on the right in Fig. 4.8. This maximal distance is called `cutoff` in Gstat. The $(\Delta x^*, \Delta y^*)$-plane is then partitioned into classes 45^0 wide for directions (`direction` in gstat) 0^0 (positive y^* axis), 45^0, 90^0 (positive x^* axis) and 135^0, as well as distance classes 10 m wide (`width` in Gstat). Then, Gstat computes the empirical variograms in Fig. 4.10, plots (a), (b), (c) and (d).

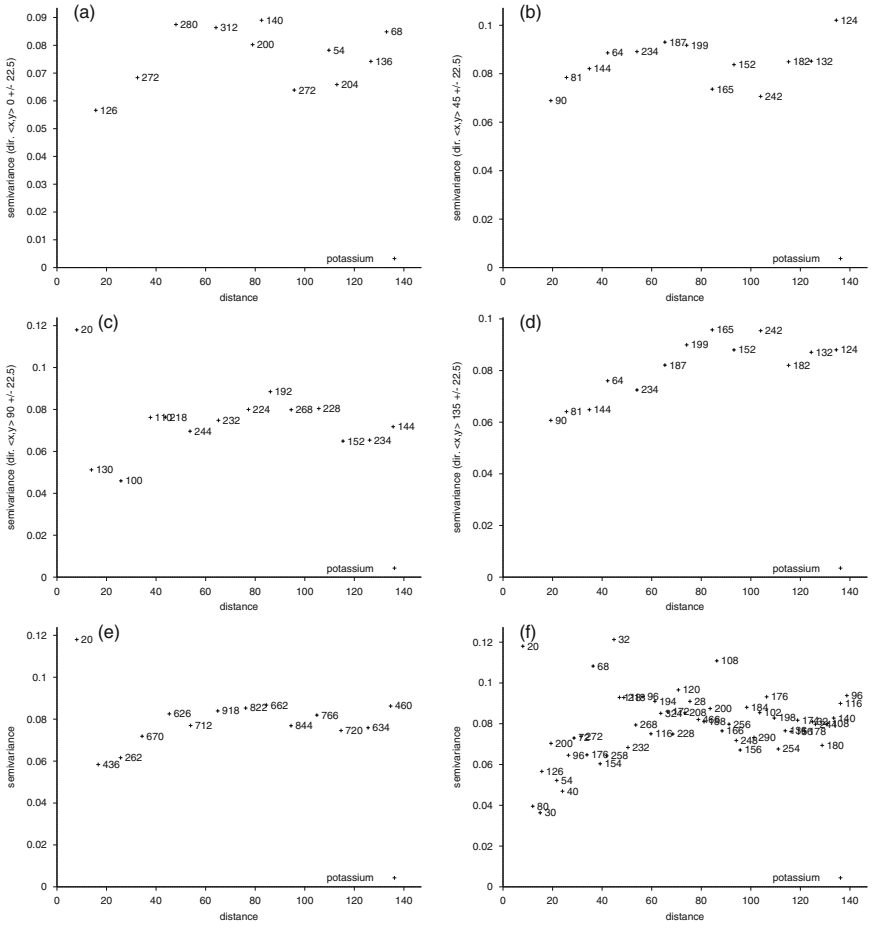


Fig. 4.10. Empirical semivariograms $\hat{\gamma}(h)$ in $(\log(\text{mg in 100 g soil}))^2$ of K_2O in the "Strassenacker" soil are plotted with + and $n(h)$, $n(h)$ being the number of the differences $h = x_i - x_k$, x_i and x_k being points with observations, in a distance and/or direction class. (a), (b), (c) and (d): calculated for $|h| \leq 140$ m with 10 m wide distance classes for the directions 0^0, 45^0, 90^0 and 135^0 clockwise from the direction of positive y; (e) and (f): for all directions with widths 10 and 2.5 m.

There are 20 differences $h = x_i - x_k$ (x_i and x_k being points with observations) with $|h| = 8$ m, which are calculated from pairs (x_i, x_k) at the left and right margins of "Strassenacker" plot as can be seen in Fig. 4.9. Hence, the estimate for the smallest distance class for direction 90^0 (plot (c) in Fig. 4.10) (i) is not representative for the "Strassenacker" plot, and (ii) has a larger variance than those pertaining to the remaining distance and direction classes being calculated from larger numbers (between 100 and 300) of pairs (x_i, x_k).

The points with observations x_i in the "Strassenacker" plot are on a grid with identical cells (except those points close to the left and right margins where the cells become smaller in x^*-direction). An allocation of the x_i in a grid is an advantage when interpolating and a disadvantage when calculating the empirical variogram as no estimates are obtained for $|h|$ smaller than the size of the grid. This disadvantage could have been avoided when in 1995, in addition to probing at 140 grid points, observations were made at approximately 30 points irregularly scattered over the "Strassenacker" plot.

The empirical semivariograms for directions 0^0, 45^0, 90^0 and 135^0 (plots (a), (b), (c) and (d) in Fig. 4.10) increase, from approximately 0.06 (log(mg in 100 g of soil))^2 (this unit is omitted in the following paragraphs) for $|h| \approx 20\ m$, until they arrive at their sills between 0.08 and 0.10 for ranges $40 \leq |h| \leq 60$ m. Because these sills and ranges are approximately identical for all four directions, the K_2O variogram is assumed to be isotropic from a comparison with plot (c) in Fig. 4.5 and, therefore, empirical semivariograms are calculated by partitioning the $(\Delta x^*, \Delta y^*)$-plane on the right in Fig. 4.8 solely into distance classes, being 10 and 2.5 m wide (plots (e) and (f) in Fig. 4.10). These statistics are often called *isotropic empirical semivariograms* or *empirical semivariograms for all directions*.

When calculating an isotropic empirical variogram, the number of the differences $|h|$ in a distance class decreases with decreasing width and, therefore, the empirical variance of the estimates $\hat{\gamma}(h)$ calculated as required in (4.11) increases when the width of the classes decreases, as fewer observations are available for each estimate. For example, when estimating the isotropic K_2O variogram, the estimates calculated for 10 m classes are closer to an assumed theoretical smooth model than those for classes being 2.5 m wide, as is obvious from plots (e) and (f) in Fig. 4.10.

Using distance classes being 2.5 m wide, estimates $\hat{\gamma}(h) \approx 0.04$ are calculated for $10 \leq |h| < 20$ m. Thus, more evidence is obtained to substantiate that the K_2O variogram increases for increasing distances $|h|$; in this general pattern, however, $\hat{\gamma}(h) \approx 0.12$ for $7.5 \leq |h| < 10$ m (calculated solely from observations at the left and right margins of the "Strassenacker" plot) does not fit. Neglecting this estimate, the K_2O semivariogram is assumed to have a sill of 0.08 and a range of 50 m and, consequently, to be bounded, reconcilable with a stationary spatial random function as shown in plot (a) in Fig. 4.5.

Summarizing these results, K_2O in the "Strassenacker" soil is assumed to be isotropic and stationary in its covariance function.

The observations of the "Strassenacker" soil parameters are averages covering approximately 400 cm^3 soil resulting from probes with 5 cm diameter and 20 cm depth. This diameter is small when compared to the side of the "Strassenacker" plot and thus the soil values are assumed to be point observations as required, in Sect. 4.6.2, when an empirical variogram is calculated.

4.3.2 Variogram Models

An analytical function (an analytical variogram or variogram model) is (preferably, as shown in the remarks concluding Sect. 4.4) fitted to an empirical variogram prior to interpolating a spatial random function with the kriging systems derived in Sect. 4.5. Frequently used variogram models are introduced in Fig. 4.11.

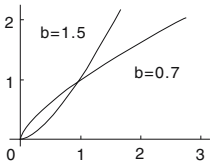
The variogram model to be fitted to an empirical variogram is selected and its parameters can be "estimated" by trial and error when advantage is taken of the following hints:

1. The slope in $|h| = 0$ is obtained by drawing a straight line through the first values of the empirical variogram. For example, a model for the isotropic and bounded K_2O semivariogram is not smooth in $|h| = 0$ as can be seen in plots (e) and (f) in Fig. 4.10: a spherical or an exponential model is fitted and not a Gaussian model.
2. The intersection of the straight line (obtained above) with the vertical axis is an approximation of the nugget effect, provided that the observations are not afflicted with a measurement error. For example, the K_2O semivariogram has a nugget effect of approximately 0.015 (log(mg in 100 g of soil))^2 (this unit is omitted in the following paragraphs) when the estimates 0.12 for $7.5 \leq |h| < 10$ m are neglected because they stem solely from observations at the left and right margins of the "Strassenacker" plot.
3. The sill and also range of the model (the range is identical with parameter a in the case of both spherical and cubic models) are seen in the plot of the empirical variogram. For example, in plots (e) and (f) in Fig. 4.10, the sill of the K_2O semivariogram is approximately 0.08 and its range approximately 50 m.

For example, using this guideline, a spherical model with sill 0.08, range 50 m and nugget effect 0.015 is fitted to the empirical K_2O semivariograms in plots (e) and (f) in Fig. 4.10.

All variogram models frequently used for the interpolation of spatial random functions are available in Gstat [105]. For example, when fitting a model to the empirical variogram of K_2O in the "Strassenacker" soil as calculated in plot (f) in Fig. 4.10, `variogram model` is selected in the menu of the Gstat interface for the variogram estimation introduced in Sect. 4.3.1: Gstat is now ready to accept `0.015 Nug(0) + 0.065 Sph(50)`, i.e., the notation used in Gstat for a spherical model with sill $0.08 = 0.015 + 0.065$, range 50 m and a nugget effect 0.015 (these "estimates" are obtained above by trial and error). When, however, the jump in $|h| = 0$ would be found to be due to an error in the measurements, a model `0.015 Err(0) + 0.065 Sph(50)` would rather be fitted. Gstat also provides for models with both a nugget effect and an error in the measurements, as introduced in Sect. 4.6.3.

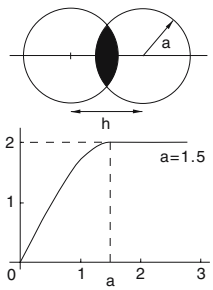
(i) power
$\gamma(h) = d|h|^b$ with $0 < b < 2$
with the special case of the
linear variogram for $b = 1$



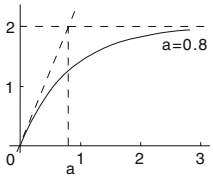
(ii) spherical

$$\gamma(h) = \begin{cases} c(0)F(|h|/a) & \text{for } |h| \le a \\ c(0) & \text{for } |h| > a \end{cases}$$

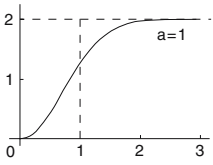
where $F(v) = (3/2)v - (1/2)v^3$
with a covariance $c(h) = c(0) - \gamma(h)$ being proportional to the volume of the intersection
of two spheres



(iii) exponential
$\gamma(h) = c(0)\left(1 - \mathrm{e}^{-|h|/a}\right)$
with a tangent in the origin that intersects the sill $c(0)$ in a limiting the interpolation distance:
do not interpolate for distances larger than $3a$
where $\gamma(h) \approx 0.95c(0)$



(iv) Gaussian
$\gamma(h) = c(0)\left(1 - \mathrm{e}^{-|h|^2/a^2}\right)$
with a limiting the interpolation distance: do
not interpolate for distances larger than $1.73a$,
where $\gamma(h) \approx 0.95c(0)$



(v) cubic

$$\gamma(h) = \begin{cases} c(0)F(|h|/a) & \text{for } |h| \le a \\ c(0) & \text{for } |h| > a \end{cases}$$

with $F(v) = 7v^2 - (35/4)v^3 + (7/2)v^5 - (3/4)v^7$
being smooth (like the Gaussian) in the origin

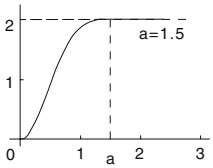


Fig. 4.11. Variogram models. On the horizontal axis $|h|$, the absolute value of translations as defined in (4.8) or (4.9), on the vertical axis the semivariogram $\gamma(|h|)$. The power variogram (i) is plotted for $d = 1$, the other models are bounded with sill $c(0) = 2$, a is the range in the spherical (ii) and cubic (v) models.

When fitting a variogram model by trial and error (`no fit` in Gstat), the parameters of the model are corrected until the best fit (as judged from the plot) is obtained. For example, a spherical model with sill $0.08 = 0.015 + 0.065$, range 50 m and nugget effect 0.015 is fitted to the K_2O empirical variogram (plot (a) in Fig. 4.12) provided that the estimates 0.12 for $7.5 \le |h| < 10$ m are neglected due to them stemming solely from observations at the left

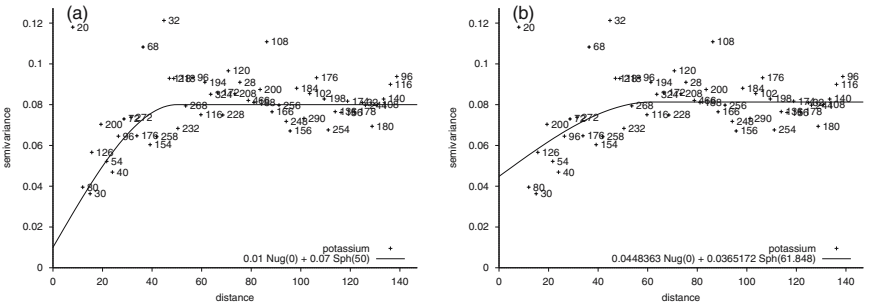


Fig. 4.12. Spherical models, both with a nugget effect, fitted to the empirical semivariogram (in $(\log(\text{mg in 100 g soil}))^2$) of K_2O in the "Strassenacker" soil as in plot (f) in Fig. 4.10: fitted by trial and error (a) and using the WLS method as implemented in Gstat (b).

and right margins of the "Strassenacker" plot. When, however, a fit is not obtained after a few attempts, the trial and error fitting is repeated using another model.

Fitting by trial and error is preferred when the values in the empirical variogram are given subjective weights, as seen above in the K_2O example. Usually, however, a variogram model is fitted in Gstat using iterative reweighted least squares (WLS) or restricted maximum likelihood (REML) estimates [105]. For example when, as shown in plot (b) in Fig. 4.12, a spherical model is fitted using WLS to the empirical variogram of K_2O, a nugget effect amounting to 0.045 is obtained, — much larger than in the model fitted by trial and error and neglecting the empirical variogram values for $7.5 \leq |h| < 10$ m (plot (a) in Fig. 4.12).

The spherical model with a nugget effect 0.015 fitted by trial and error is believed to be closer to the true variogram of K_2O in the "Strassenacker" soil than the model with a larger nugget effect fitted using WLS. Unfortunately, no better estimate for the variogram close to $|h| = 0$ can be obtained because no empirical variogram values are available for small $|h|$, due to the lack of observations with $|x_i - x_k| < 7.5$ m in the "Strassenacker" plot.

In Sects. 4.3.1 and 4.3.2, K_2O in the soil of the "Strassenacker" plot is used as an introductory example for variogram estimation. It is demonstrated that definitions (4.8) and (4.9) of stationary and intrinsically stationary random functions, as well as the definition of the empirical variogram in (4.11) and Fig. 4.8, are easily applied to the K_2O observations: it is obvious, in Fig. 4.9, that the mean of K_2O is approximately constant in the "Strassenacker" plot, and, in Fig. 4.10, that the empirical variogram of K_2O is isotropic and bounded. Hence, K_2O is assumed to be a globally (i.e., all over the "Strassenacker" plot) stationary spatial random function. Less straightforward, however, is the fitting of a model to the empirical variogram of K_2O:

the amount of the nugget effect remains uncertain because no variogram estimates are available for $|h| < 7.5$ m.

Often, a variogram estimation is more intricate, and it is recommended to proceed as proposed in (4.12).

When a variogram is estimated answers to the following questions have to be found:
1. *Is the logarithm of the observations taken?*
2. *Is the mean constant? Globally? Locally? Is there a trend?*
3. *How is the empirical variogram calculated:*
3.1 *for what maximal distance? for which directions?*
3.2 *using distance and direction classes how wide?*
3.3 *When the empirical variogram is not isotropic, is the anisotropy geometric?*
4. *Which analytical variogram is fitted applying which method?*

(4.12)

When estimating, in Sect. 4.3.3, the variogram of grain yield in the "Strassenacker" plot, a constant mean of the observations, as required in (4.8) and (4.9), is found. The task, however, becomes difficult when the empirical variogram is calculated and answers to the questions (4.12,3) have to be given. Even more intricate is, in Sect. 4.3.4, estimating the variogram of the yearly SWIR in Germany and parts of the adjacent countries owing to difficulties originating from a non-constant mean of the observations.

4.3.3 Empirical Variograms II: the Anisotropic Case

The measurements of the soil parameters and yields (grain and biomass) in the "Strassenacker" plot are introduced in the remarks to Figs. 4.9 and 4.10. The observed grain yields are plotted in Fig. 4.13 from which it is concluded that grain yield has a constant expectation function since no systematic increases or decreases are seen. Hence, the observed grain yields are assumed to stem from a globally stationary or intrinsically stationary spatial random function on condition that its second moment function is reconcilable with (4.8) or (4.9). Thus, answers to the questions (4.12,1,2) have been found. To give answers to the questions (4.12,3) empirical semivariograms are calculated for difference vectors $h = x_i - x_k$, with $|h| \leq 140$ m for directions 0^0 (positive y^* coordinate) and 90^0 (positive x^* coordinate). The empirical variogram for direction 0^0 (plot (a) in Fig. 4.14) remains constant for distances $|h| \leq 100$ m and thereafter decreases slightly, whereas the empirical variogram for direction 90^0 (plot (b) in Fig. 4.14) increases approximately linearly. Is grain yield a white noise process in direction 0^0 as suggested by plot (a) in Fig. 4.14 (this plot resembles a realisation of a variogram with a pure nugget effect as shown in plot (d) in Fig. 4.7)? Furthermore, since the variogram of a random walk process is shown to increase linearly in the remarks to (4.9), is grain yield a random walk process in direction 90^0 as suggested by plot (b) in Fig. 4.14?

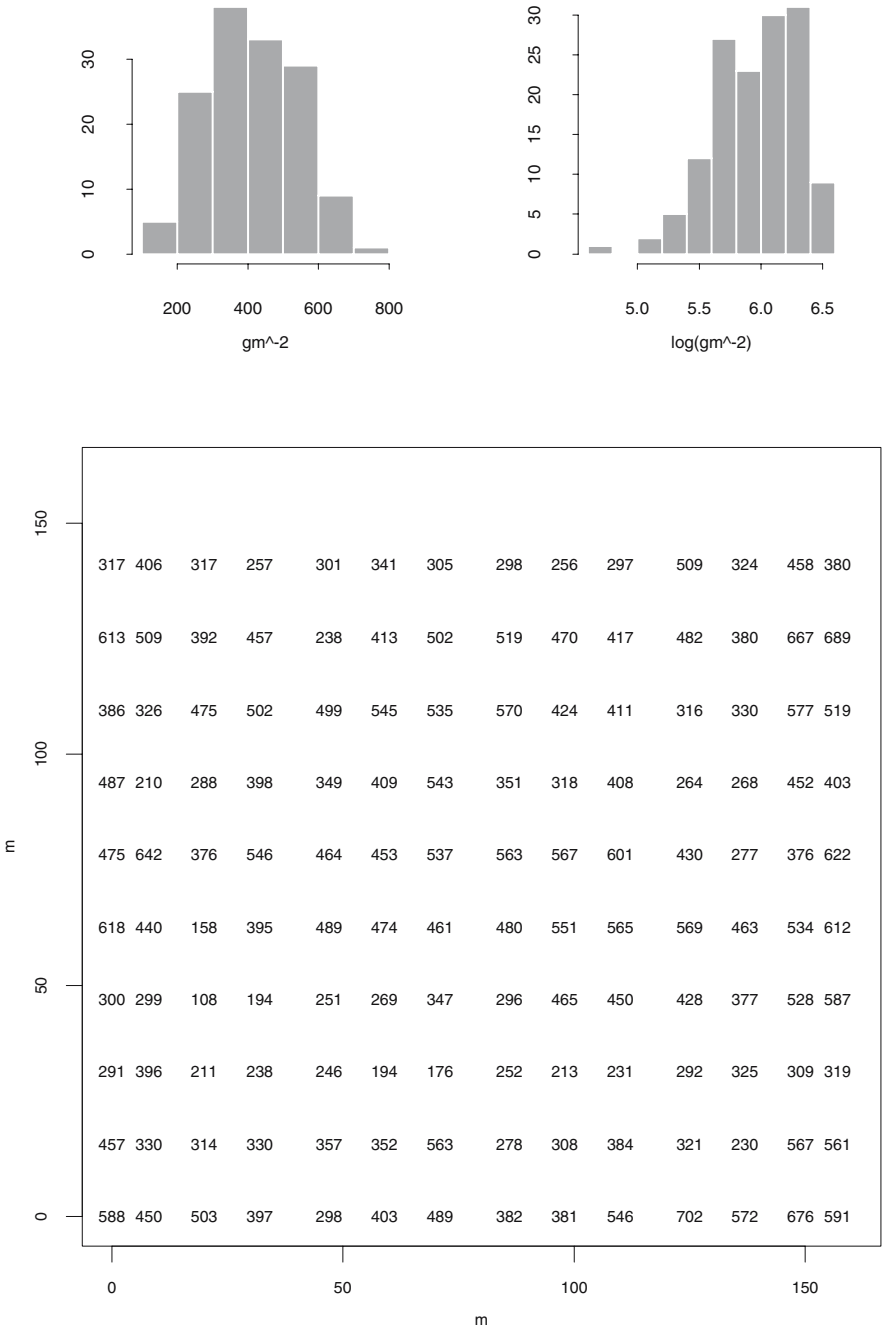


Fig. 4.13. Observations of grain yield (in gm^{-2}) in the "Strassenacker" plot. Above, histograms of the observations and of the logarithms of the observations plotted below.

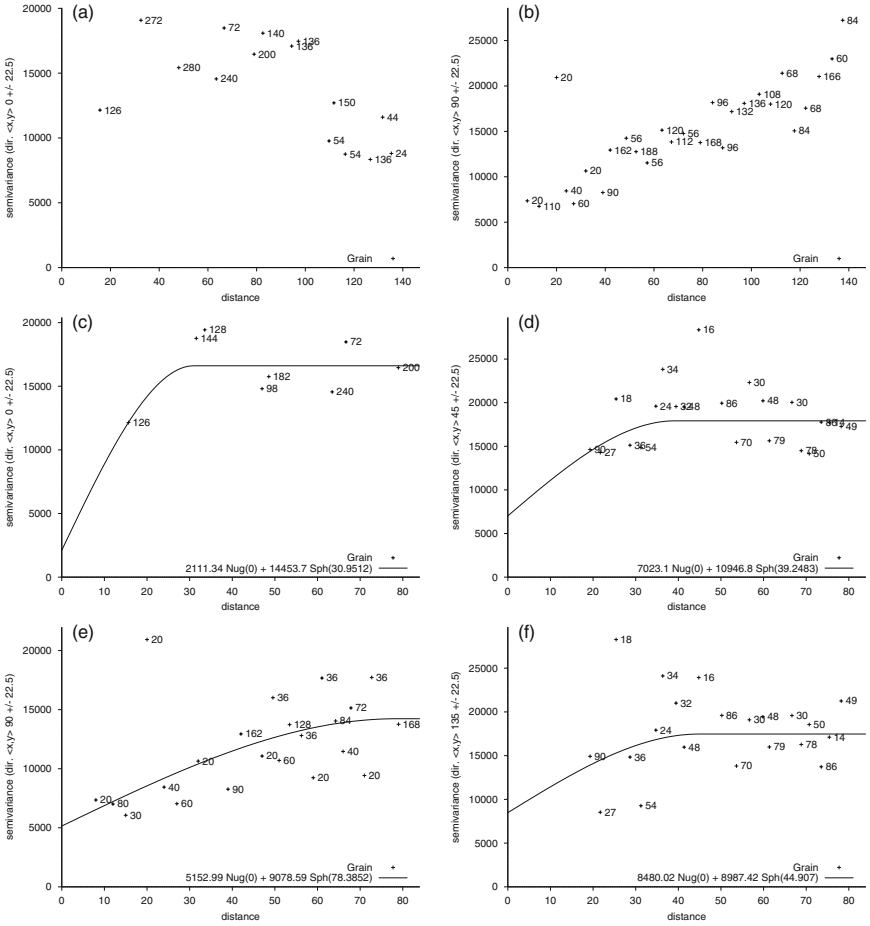


Fig. 4.14. Empirical semivariograms $\hat{\gamma}(h)$ in $(\mathrm{gm}^{-2})^2$ of the "Strassenacker" grain yield calculated with 2.5 m wide distance classes. (a) and (b): for $|h| \leq 140$ m for directions 0^0 and 90^0; (c), (d), (e) and (f): for $|h| \leq 80$ m with fitted spherical models for directions 0^0, 45^0, 90^0 and 135^0.

The empirical semivariogram of grain yield for direction 90^0 remains constant for $50 \leq |h| < 80$ m as can be seen in plot (b) in Fig. 4.14, but increases for larger $|h|$. Is this increase due to an anisotropy? An answer is found by calculating empirical semivariograms for $|h| = |x_i - x_k| \leq 80$ m and directions 0^0, 45^0, 90^0 and 135^0. These semivariograms are shown to arrive, in plots (c), (d), (e) and (f) in Fig. 4.14 emphasised by fitted spherical models, at their sills (between 14000 and 18000 $(\mathrm{gm}^{-2})^2$) within ranges of 30.95 (direction 0^0) and 78.39 m (direction 90^0). When these empirical variograms are compared with the prototype in Fig. 4.5 (c), it is concluded that the empirical semivariogram of grain yield is anisotropic.

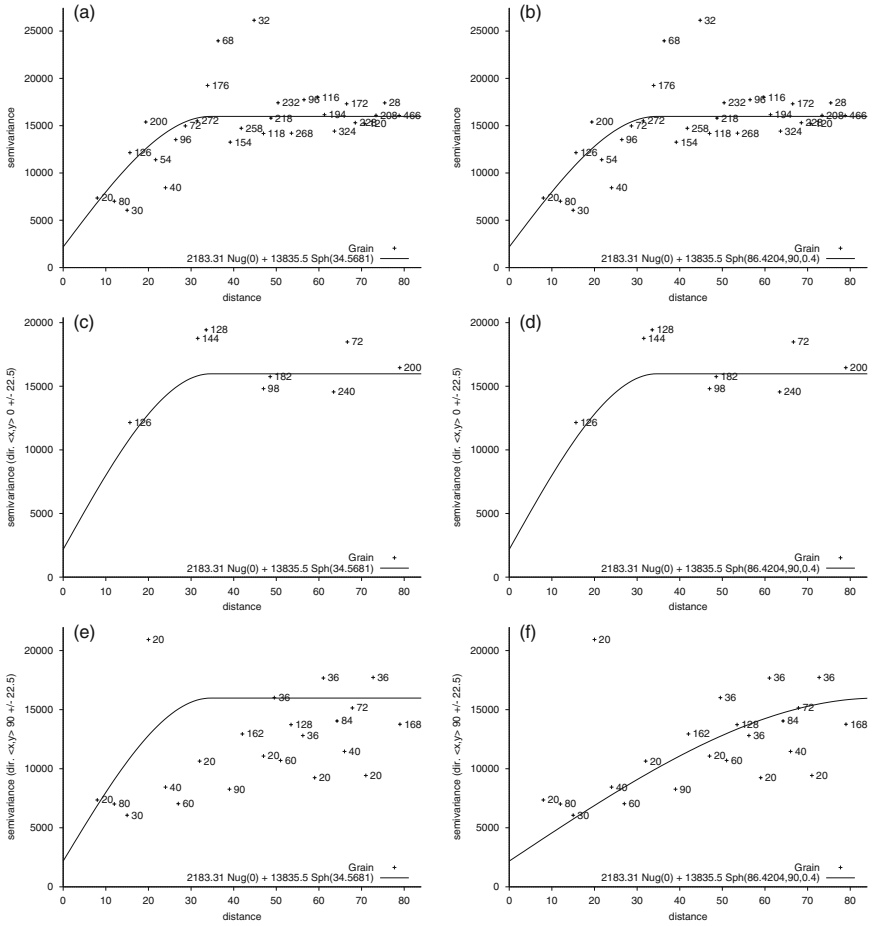


Fig. 4.15. Above, empirical semivariograms of the "Strassenacker" grain yield for all directions, (a) with a fitted isotropic model, (b) with a fitted anisotropic model as defined in (4.10) and Fig. 4.6. Below, empirical semivariograms for the directions 0^0 (c,d) and 90^0 (e,f), with the models in (a) and (b). The width of the distance classes is 2.5 m, the unit of the semivariograms is $(\mathrm{gm}^{-2})^2$.

In Gstat, an anisotropy is modelled geometrically using an ellipse, as described in (4.10) and Fig. 4.6. For example, in Fig. 4.14, the empirical variogram of grain yield in direction 0^0 arrives at its sill for $30 \leq |h| < 32.5$ m and, in direction 90^0, for $75 \leq |h| < 80$ m: the ratio of the ranges of the fitted spherical models being $0.4 \approx 30.95/78.39$ becomes the ratio (short axis/long axis) of an ellipse provided that the variogram is assumed to be geometrically anisotropic. A variogram model with such an anisotropy is fitted by selecting, in the Gstat interface for the variogram estimation introduced in Sect. 4.3.1, `direction` and typing in 0^0 for the direction and 180^0 for the width of the

direction classes thus urging Gstat to calculate a variogram without using direction classes. Then, a spherical model with a nugget effect and geometrical anisotropy is defined with `3000 Nug(0) + 13000 Sph(75,90,0.4)`. The sill of this model at $3000 + 13000 = 16000$ $(\mathrm{gm}^{-2})^2$ is arrived at translations of 75 m in direction 90^0 but 30 m in direction 90^0, the anisotropy ratio being 0.4. Thereafter, this model is fitted with the WLS method and plot (b) in Fig. 4.15 is obtained.

Once estimated, this model is plotted for diagnostic purposes in the empirical semivariograms for the grain yield for directions 0^0 and 90^0 (plots (d) and (f) in Fig. 4.15). These diagnostic plots show that the fit of the model is adequate in both directions, although slightly better fits are obtained, in plots (c) and (e) in Fig. 4.14, when using models fitted to the empirical variograms calculated for these directions solely. Diagnostic plots for directions 45^0 and 135^0 are generated in Problem 4.25.

As a second diagnostic, a spherical model with nugget effect but without geometrical anisotropy is fitted to the empirical semivariogram of grain yield calculated for all directions (plot (a) in Fig. 4.15). This model is then plotted in the empirical semivariograms for directions 0^0 and 90^0. In direction 0^0, this isotropic model fits, as does the geometrically anisotropic model (plots (c) and (d) in Fig. 4.15). In direction 90^0, however, the isotropic model does not fit (plot (e) in Fig. 4.15) whereas the geometrically anisotropic model shows a good fit (plot (f) in Fig. 4.15).

Hence, grain yield in plot "Strassenacker" is assumed to be globally stationary and anisotropic in its second moment function since a spherical model as shown in plot (b) in Fig. 4.15, with sill ≈ 16000 $(\mathrm{gm}^{-2})^2$ and geometrical anisotropy (range $a_0 \approx 32$ m in direction 0^0 and $a_{90} \approx 80$ m in direction 90^0), fits the empirical semivariogram of grain yield. This empirical semivariogram is calculated for $|h| = |x_i - x_k| \leq 80$ m using all pairs (x_i, x_k), x_i, x_k being the points with observations as plotted in Fig. 4.13.

Admittedly, some intuition is needed when trying empirical semivariograms for $|h| = |x_i - x_k| \leq 80$ m after having calculated the ones for $|h| = |x_i - x_k| \leq 140$ m (plots (a) and (b) in Fig. 4.14). Empirical variograms looking even more intimidating than the ones in plots (a) and (b) in Fig. 4.14 are calculated for the spatial random function analysed in Sect. 4.3.4.

4.3.4 Empirical Variograms III: the Case with Non-constant Drift

The spatial random function under analysis is SWIR (more precisely, the climatological means of the yearly values of SWIR) in the region bounded by 4^0 and 16^0 (eastern) longitude and by 48^0 and 54^0 (northern) latitude. In this region, there are 39 stations with at least 5 yearly values of SWIR measured in the period from 1948 through to 1993. From the yearly values of SWIR measured at these stations, the climatologies as plotted in Fig. 4.16 are calculated and made available in file `/path/swirnorde.eas`:

```
yearly climatologies of SWIR in Germany
```

```
7
longitude (degree)
latitude (degree)
number of yearly values used for the climatologies
mean (W/m2)
standard deviation (W/m2)
station number
name of station
4.033 49.300 11 123.27 9.62 1241 Reims
...
```

From Fig. 4.16 it is concluded that the surface (of the earth) receives more than 120 Wm^{-2} of SWIR in the mean during the period with measurements south of 50^0 and less than 110 Wm^{-2} north of 53^0 (except Norderney station on the East Friesian islands). This systematic increase from north to south is due to the increase of solar radiation incident on a horizontal surface at the top of the atmosphere. Being subject to a trend, the expectation function of SWIR in this region is not constant and, therefore, neither (4.8) nor (4.8) are reconcilable with SWIR.

The non-constant first moment function of SWIR is the first obstacle to overcome when the second moment function is estimated by means of empirical semivariograms. The second obstacle is due to a deficiency in the software used: Gstat does not provide for spherical distances on the surface of the earth since it calculates the differences $h = x_i - x_k$, x_i, x_k points with observations, assuming that x^* and y^* (i.e., longitude and latitude in the SWIR case) span an orthogonal coordinate system. The distortions resulting from the use of orthogonal coordinates can be disregarded provided that the region under analysis is not large or on higher latitudes, i.e., properties pertaining to the region with SWIR observations as plotted in Fig. 4.16. If, however, a spatial random function is analysed in a large region on higher latitudes, the coordinates are transformed prior to using Gstat. In [27], for example, before calculating empirical variograms of precipitation in Greenland using Gstat, the coordinates underwent an affine transformation with the pole in the centre of Greenland.

Bearing in mind that the expectation function of SWIR is not constant, empirical semivariograms of SWIR are calculated in Gstat without a prior coordinate transformation, and, as always in Gstat, using the residuals of the mean of the observations as plotted in Fig. 4.16, for maximal distances $|h| = |x_i - x_k| \leq 5^0$, distance classes 0.25^0 and 0.5^0 wide and all directions, as well as directions 0^0, 45^0, 90^0 and 135^0, i.e., north, north-east, east, and south-east. The results of all these calculations are plotted in Fig. 4.17. In direction 90^0 (Fig. 4.17, e), the empirical semivariogram arrives at a sill of $\approx 20\ (Wm^{-2})^2$ within a range of 2.5^0; in direction 0^0 (Fig. 4.17, c) at a sill of $\approx 80\ (Wm^{-2})^2$ within a range 3.5^0: the empirical variogram appears to be anisotropic when compared with the prototype (Fig. 4.5, c). For directions

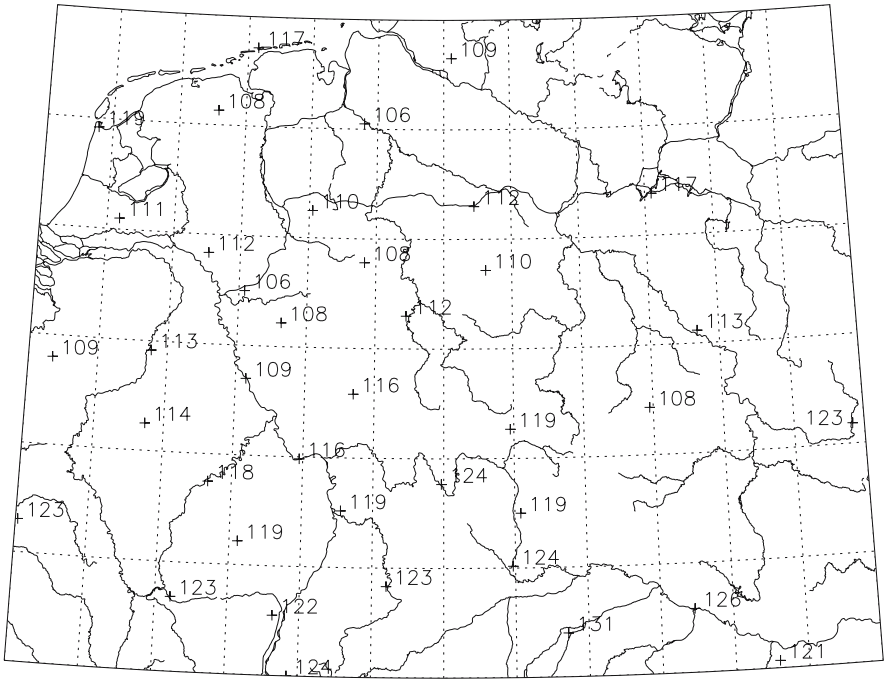


Fig. 4.16. Climatological means of SWIR in Wm^{-2} measured at 39 stations (plotted with +) in the region bounded by 4^0 and 16^0 (eastern) longitude and by 48^0 and 54^0 (northern) latitude.

45^0 and 135^0 (Fig. 4.17, d,f), however, the empirical semivariograms do not arrive at a possible sill within the maximal distance of 5^0. All empirical semivariograms are smooth close to $|h| = 0$, and, therefore, Gaussian models can be fitted where sills do exist. Are the differences in the sills and ranges of these empirical semivariograms due to an anisotropy or due to the trend in the observations as plotted in Fig. 4.16?

The examples in Figs. 2.23 and 3.1 demonstrate that a slowly decaying empirical covariance function (corresponding to a slowly increasing empirical variogram not arriving at a sill in a short distance, e.g., the semivariograms shown in plots (d) and (f) in Fig. 4.17) can be due to a trend. This suggests that the differences in the ranges and sills of the empirical semivariograms as plotted in Fig. 4.17 will become smaller when estimates are calculated from the residuals of a first order trend surface fitted to the observations. The decomposition of the data into a trend (the systematic increase of SWIR in southward direction) and (possibly spatially correlated) random residuals is also favoured, by climatological evidence.

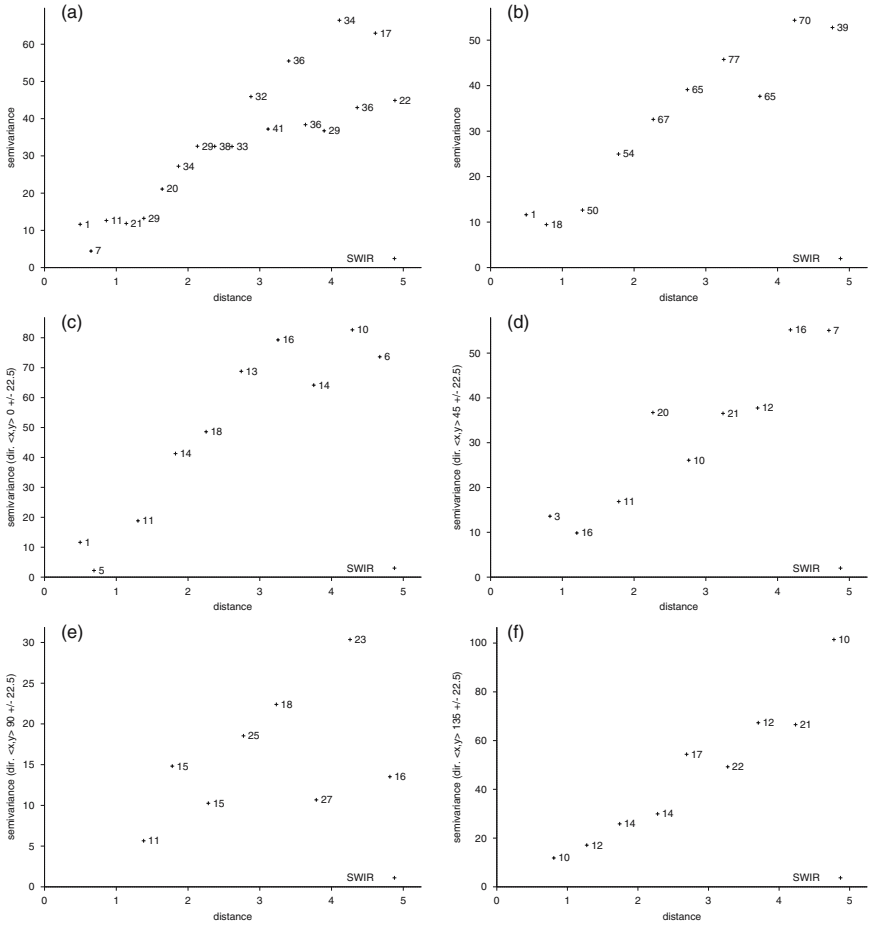


Fig. 4.17. Empirical semivariograms $\hat{\gamma}(h)$ in $(\mathrm{Wm}^{-2})^2$ of SWIR measured at the stations in Fig. 4.16 calculated from the residuals of the mean (116.67 Wm^{-2}) of all stations using distances in ^{0} of longitude and latitude. (a) and (b): for all directions, using distance classes 0.25^0 (a) and 0.5^0 (b) wide; (c), (d), (e) and (f): for directions 0^0, 45^0, 90^0 and 135^0 and distance classes 0.5^0 wide.

Gstat estimates trend surfaces as defined in (3.23) of order 0, (i.e., the mean of the observations), 1, 2 and 3, and then calculates the empirical variogram of the residuals. Fig. 4.18, for example, contains plots of empirical semivariograms calculated from the residuals of a first order trend surface fitted to the SWIR observations in Fig. 4.16. The empirical semivariograms in Fig. 4.18 are approximately isotropic and continuous (no longer smooth as are the semivariograms in Fig. 4.17) in $|h| = 0$, and, therefore, spherical models without nugget effect are fitted to those calculated for all directions (plots (a) and (b) in Fig. 4.18) by means of the WLS method. These estimates, i.e., first

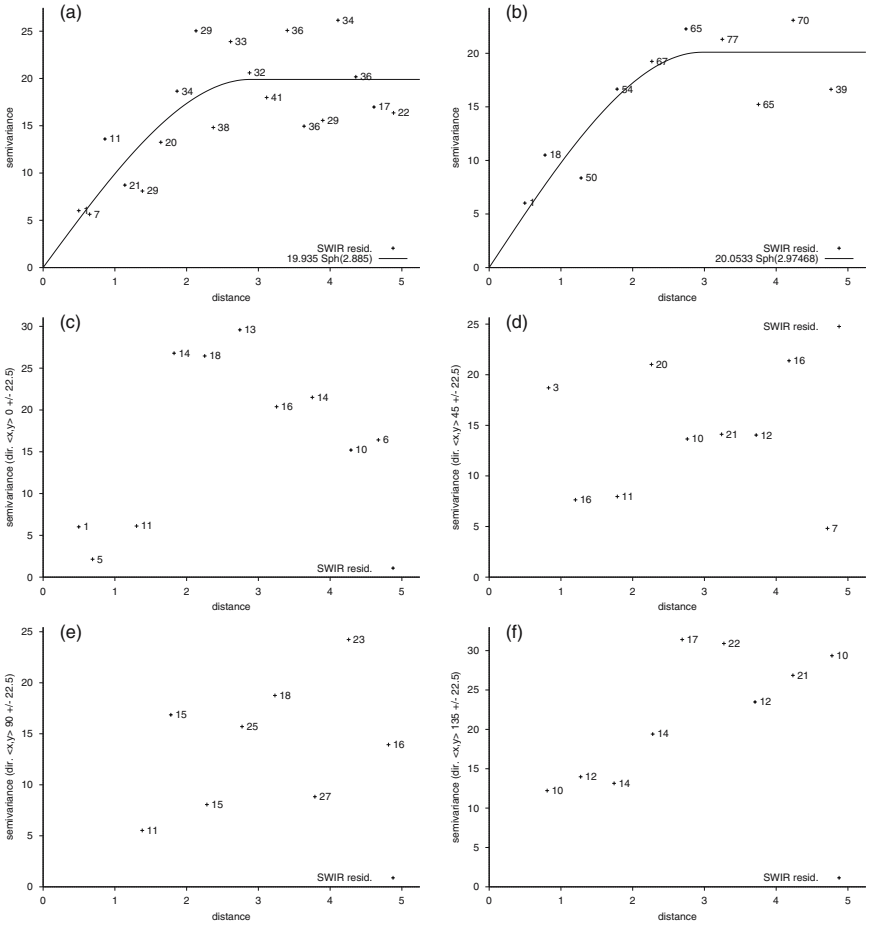


Fig. 4.18. Empirical semivariograms $\hat{\gamma}(h)$ in $(\mathrm{Wm}^{-2})^2$ of SWIR measured at the stations in Fig. 4.16 calculated from the residuals of a first order trendsurface fitted to the data using distances in ^{0} of longitude and latitude. (a) and (b): for all directions, with fitted spherical models, using distance classes 0.25^0 (a) and 0.5^0 (b) wide; (c), (d), (e) and (f): for directions 0^0, 45^0, 90^0 and 135^0 and distance classes 0.5^0 wide.

order trend surface, empirical variograms and fitted models, are calculated from the SWIR observations as demonstrated in the following paragraph.

Having calculated the empirical semivariogram as plotted in Fig. 4.17 (a), the work is saved and then the Gstat interface for the variogram estimation is quit: Gstat saves the parameters for the calculation of the empirical semivariogram by writing file `swir.cmd` (header lines are omitted)

```
data(SWIR): '/path/swirnorde.eas',x=1,y=2,v=4;
set cutoff = 5;
```

```
set width = 0.25;
```

with Gstat expressions (a description of the expressions available in Gstat is available in [105]) in format

```
expression(identifier): parameters as required;
```

separated by ;. When these Gstat expressions are read by Gstat

1. `data()` prompts Gstat to read the observations for spatial random function `SWIR` from file `/path/swirnorde.eas` with the coordinates in the first and second columns (`x=1,y=2`) and the observations in the fourth column (`v=4`) and, implicitly, to calculate the empirical semivariograms from the residuals of the mean of the observations
2. `set cutoff = 5` together with `set width = 0.25` define the partition of the $(\Delta x^*, \Delta y^*)$-plane into classes of distance and direction as required on the right, in Fig. 4.8: the empirical variogram is assumed to be isotropic (no direction classes are given) and is calculated for a maximal distance of 5^0 and distance classes 0.25^0 wide.

The expressions in file `swir.cmd` are then modified to prompt Gstat to calculate the empirical semivariograms using the residuals of a first order trend surface by inserting `d=1` (default is `d=0`) in the `data(SWIR)` command:

```
data(SWIR): '/path/swirnorde.eas',x=1,y=2,v=4,d=1;
```

File `swir.cmd` with a `data()` expression as modified above is read by Gstat when you type in

```
gstat swir.cmd
```

(recall that you typed `gstat -i` when invoking Gstat for the estimation of the empirical variograms in Fig. 4.17). This causes Gstat to open the user interface for the variogram estimation since nothing else (other than reading the data) is required. Now, with `d=1` in the `data()` expression, empirical semivariograms as plotted in Fig. 4.18 are calculated using the residuals of a first order trend surface.

A first order trend surface as calculated above is only a preliminary estimate using (3.9), i.e., ordinary least squares, under the assumption that the residuals are not correlated as required in (3.11). Since the residuals are correlated as shown in Fig. 4.18, these calculations are considered to be the first step in an iterated generalised least squares estimation as proposed in the remarks to (3.15). When further steps are performed in Sect. 4.5.3, they result in the empirical semivariogram and the fitted spherical model as plotted in Fig. 4.21, which is then used, together with the estimated trend surface, for the interpolation of the SWIR observations.

A successful estimation of a variogram pertaining to a spatial random function answers the questions in (4.12) and thus permits, in principle, an estimation of the spatial random function at points without observations using the methods introduced in Sect. 2.7. These methods are refined and adapted for the interpolation of spatial random functions in Sects. 4.4 and 4.5.

4.4 Optimal Linear Predictors for Spatial Random Functions

A spatial random function $Z(x)$ as defined in (4.6) is estimated, in Sect. 4.5, at an interpolation point x_0, provided that (i) the estimate belongs to the observed realisation, (ii) the variogram $\gamma_Z(h)$ of $Z(x)$ is known or has been estimated using the methods introduced in Sect. 4.3, (iii) the observations are assumed to be point observations as required in Sect. 4.6.2, and (iv) linear combinations of the observations as defined in (4.13) are available as estimators. Do linear combinations exist that can be used as estimators of a stationary $Z(x)$ as required in (4.8) and of a $Z(x)$ that satisfies the intrinsic hypothesis defined in (4.9)?

The optimal linear prediction of a stochastic process as defined in (2.62) is a linear combination of the observations. Being mathematically simple, linear combinations of the observations are also used for the interpolation of a spatial random function.

The interpolator $\widehat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i)$ *is linear in the observations* $Z(x_i)$ *(the observed values* $z(x_i)$ *stem from a realisation of* $Z(x)$*) and the unknown weights* λ_i*, depending on the points with observations* x_i *and the interpolation point* x_0*.* $\widehat{Z}(x_0)$ *is optimal provided that:*

1. $\mathrm{E}\big(\widehat{Z}(x_0) - Z(x_0)\big) = 0$*, i.e., it is not biased, and* (4.13)
2. $\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) = \mathrm{E}\big((\widehat{Z}(x_0) - Z(x_0))^2\big) = \min(\lambda_i)$*, i.e., its variance is minimal with respect to the weights* λ_i*. This variance is called variance of the interpolation error, interpolation variance or, as in definitions (2.62) and (2.61), mean square prediction error.*

As a prerequisite for the following definitions and derivations it is assumed that $Z(x)$ is a spatial random function with known moment functions $\mu_Z(x)$ and $\gamma_Z(h)$.

The optimal linear interpolator $\widehat{Z}(x_0)$ for a spatial random function $Z(x)$ as defined in (4.13) is one of many possible linear combinations $Y = \sum_{i=1}^{N} \lambda_i Z(x_i)$, namely the element $\widehat{Z}(x_0)$ in the set of the Y that delivers an interpolator with a minimal variance of the interpolation error $\widehat{Z}(x_0) - Z(x_0)$. Which Y are candidates for interpolators? Clearly, the Y must be constrained to those having a non-negative and finite variance because an interpolator with a variance that can be calculated is required.

If a spatial random function $Z(x)$ is stationary as defined in (4.8) then the variance of any linear combination $Y = \sum_{i=1}^{N} \lambda_i Z(x_i)$ is non-negative because the covariance function of $Z(x)$ exists and is positive definite, as proposed in (2.7). Hence, as shown in the remarks to (2.7), $\mathrm{Var}Y = \mathrm{Var}\big(\sum_{i=1}^{N} \lambda_i Z(x_i)\big) = \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \lambda_k c(x_i, x_k) = \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \lambda_k c(x_i -$

$x_k) = \sum_{i=1}^N \sum_{k=1}^N \lambda_i \lambda_k c(h_{ik}) \geq 0$ for arbitrary λ_i. Substituting the variogram for the covariance function $\gamma(h) = c(0) - c(h)$ as derived in Problem 4.4, (4.14) is arrived at.

$$\begin{aligned} &\textit{If } Y = \textstyle\sum_{i=1}^N \lambda_i Z(x_i) \textit{ is a linear combination of random} \\ &\textit{variables } Z(x_i) \textit{ from a spatial random function } Z(x) \textit{ being} \\ &\textit{stationary as defined in (4.8), then for arbitrary } \lambda_i \textit{ and } x_i\textit{:} \\ &\mathrm{Var}Y = c_Z(0) \textstyle\sum_{i=1}^N \sum_{k=1}^N \lambda_i \lambda_k - \sum_{i=1}^N \sum_{k=1}^N \lambda_i \lambda_k \gamma_Z(x_i - x_k) \geq 0. \end{aligned} \tag{4.14}$$

If, however, a spatial random function $Z(x)$ is intrinsically stationary as defined in (4.9) then not all possible linear combinations $Y = \sum_{i=1}^N \lambda_i Z(x_i)$ have a variance that is non-negative, as stipulated in (4.15)

$$\begin{aligned} &\textit{If } Y = \textstyle\sum_{i=1}^N \lambda_i Z(x_i) \textit{ is a linear combination of random} \\ &\textit{variables } Z(x_i) \textit{ from a spatial random function } Z(x) \\ &\textit{satisfying the intrinsic hypothesis as required in (4.9) and} \\ &\textit{with weights satisfying } \textstyle\sum_{i=1}^N \lambda_i = 0\textit{, then for arbitrary } x_i\textit{:} \\ &\mathrm{Var}Y = -\textstyle\sum_{i=1}^N \sum_{k=1}^N \lambda_i \lambda_k \gamma_Z(x_i - x_k) \geq 0. \end{aligned} \tag{4.15}$$

and derived below in (4.16), (4.17), (4.18), (4.19) and (4.20) by calculating the variance of the linear combination using the assumptions.

The condition constraining the weights λ_i, $\sum_{i=1}^N \lambda_i = 0$, is multiplied with $Z(x_0)$ to obtain $Z(x_0) \times \sum_{i=1}^N \lambda_i = Z(x_0) \times 0$ and further $\sum_{i=1}^N \lambda_i Z(x_0) = 0$ for arbitrary x_0. Thus,

$$\begin{aligned} \sum_{i=1}^N \lambda_i Z(x_i) &= \sum_{i=1}^N \lambda_i Z(x_i) - \sum_{i=1}^N \lambda_i Z(x_0) \\ &= \sum_{i=1}^N \lambda_i \big(Z(x_i) - Z(x_0)\big) = \sum_{i=1}^N \lambda_i \Delta(x_i) \end{aligned} \tag{4.16}$$

is arrived at. The differences $\Delta(x_i) = Z(x_i) - Z(x_0)$ are a stationary spatial random function as required in (4.9) since $Z(x)$ is assumed to be intrinsically stationary. Hence, the covariance function of the differences, $c_\Delta(x_i, x_k) = \mathrm{Cov}\big(\Delta(x_i), \Delta(x_k)\big)$, is positive semidefinite as required in (2.7,2).

In the next step, $c_\Delta(x_i, x_k)$ is written using the variogram $2\gamma_Z(h) = 2\gamma_Z(x - y)$ of the intrinsically spatial random function $Z(x)$.

$$\begin{aligned} 2\gamma_Z(x-y) &= \mathrm{Var}\big(Z(x) - Z(y)\big) \\ &= \mathrm{Var}\big(Z(x) - Z(x_0) - Z(y) + Z(x_0)\big) \\ &= \mathrm{Var}\big(Z(x) - Z(x_0)\big) - 2\mathrm{Cov}\big(Z(x) - Z(x_0), Z(y) - Z(x_0)\big) \\ &\qquad + \mathrm{Var}\big(Z(y) - Z(x_0)\big) \\ &= 2\gamma_Z(x - x_0) - 2c_\Delta(x, y) + 2\gamma_Z(y - x_0) \\ c_\Delta(x, y) &= \gamma_Z(x - x_0) + \gamma_Z(y - x_0) - \gamma_Z(x - y) \end{aligned} \tag{4.17}$$

Using these results the variance of the linear combination $\mathrm{Var}\sum_{i=1}^{N}\lambda_i Z(x_i)$ is easily calculated as variance of the differences $\mathrm{Var}\sum_{i=1}^{N}\lambda_i \Delta(x_i)$, since, with $\sum_{i=1}^{N}\lambda_i = 0$,

$$\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\lambda_k\gamma(x_k - x_0) = \sum_{i=1}^{N}\left(\lambda_i \times \sum_{k=1}^{N}\lambda_k\gamma(x_k - x_0)\right)$$
$$= \left(\sum_{k=1}^{N}\lambda_k\gamma(x_k - x_0)\right)\sum_{i=1}^{N}\lambda_i = 0 \qquad (4.18)$$

$$\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\lambda_k\gamma(y_k - x_0) = 0 \qquad (4.19)$$

are obtained for the variances, due to the first and second term in (4.17); however, the variance due to the third term remains:

$$\mathrm{Var}\sum_{i=1}^{N}\lambda_i\Delta(x_i) = \sum_{i=1}^{N}\sum_{i=1}^{N}\lambda_i\lambda_k c_\Delta(x_i, x_k)$$
$$= -\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\lambda_k\gamma_Z(x_i - x_k). \qquad (4.20)$$

The non-negative definite covariance function (in the remarks to (4.16)) of the $\Delta(x_i) = Z(x_i) - Z(x_0)$ implies a non-negative variance of a linear combination $\sum_{i=1}^{N}\lambda_i\Delta(x_i)$ and, further, $-\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\lambda_k\gamma_Z(x_i - x_k) \geq 0$, as proposed in (4.15). Variograms with this property are called *conditionally negative definite*.

When comparing (4.15) with (4.14) it becomes obvious that the set of possible interpolators $\sum_{i=1}^{N}\lambda_i Z(x_i)$ is constrained to the subset with weights satisfying $\sum_{i=1}^{N}\lambda_i = 0$, when the spatial random function to be interpolated is no longer stationary even though intrinsically stationary. This constraint is compensated by including all spatial random functions which satisfy the intrinsic hypothesis in the set of random functions that can be interpolated. (The set of all random functions satisfying the intrinsic hypothesis contains the set of the stationary random functions as a subset).

If the second moment function of the spatial random function $Z(x)$ to be interpolated is not known then $Z(x)$ is assumed to be intrinsically stationary and the empirical semivariogram $\hat{\gamma}_Z(h)$ is calculated using (4.11) and Fig. 4.8. $\hat{\gamma}_Z(h)$ is then used for the interpolation, and variances of the interpolator are calculated as required in (4.14) or (4.15), with $\hat{\gamma}_Z(h)$ substituted for $\gamma_Z(h)$. Negative variances can be avoided on condition that $\hat{\gamma}_Z(h)$, when substituted in (4.14) or (4.15), delivers non-negative variances for the x_i and arbitrary λ_i and stationary $Z(x)$ (for λ_i with the $\sum_{i=1}^{N}\lambda_i = 0$ and intrinsically stationary $Z(x)$). An (empirical) variogram with this property is called, as in the remarks to (4.20), conditionally negative definite.

Consequently, an empirical variogram is required to be conditionally negative definite. This is not easily proved subsequent to a variogram estimation and prior to an interpolation; however, the difficult task can be avoided by fitting a model known to be conditionally negative definite (an analytical variogram, examples being given in Fig. 4.11) to an empirical variogram estimated using (4.11) and Fig. 4.8, as demonstrated in Sect. 4.3.

Having fitted a conditionally negative definite variogram model $\gamma_Z(x_i - x_k)$, the variance of a linear combination of random variables $Z(x_i)$, satisfying the intrinsic hypothesis, is calculated using (4.15): $\mathrm{Var}\left(\sum_{i=1}^{N} \lambda_i Z(x_i)\right) = -\sum_{i=1}^{N}\sum_{k=1}^{N} \lambda_i\lambda_k\gamma_Z(x_i - x_k) \geq 0$, for all points with observations $x_1, \ldots, x_N$ and weights $\lambda_1, \ldots, \lambda_N$ with $\sum_{i=1}^{N} \lambda_i = 0$. When $Z(x)$ is stationary the variogram is bounded and the variance is calculated using (4.14). Due to (4.14) and (4.15), $\mathrm{Var}\left(\sum_{i=1}^{N} \lambda_i Z(x_i)\right)$ is non-negative for intrinsically stationary and stationary $Z(x)$.

The condition in (4.14), where $Z(x)$ is required to be stationary, is stronger than the condition stipulated in (4.15) where intrinsic stationarity is demanded: the set of all variograms $2\gamma_Z(h)$ contains bounded and not bounded variograms (prototypes (i) and (ii) in Fig. 4.5). If $2\gamma_Z(h)$ is bounded, then $\gamma(h) = c(0) - c(h)$ as derived in Problem 4.4, with the covariance function $c_Z(h)$ of $Z(x)$ being absolutely convergent as defined in (2.24). If $2\gamma_Z(h)$ is not bounded then $c_Z(h)$ does not converge absolutely.

4.5 Optimal Interpolation of Spatial Random Functions

A spatial random function $Z(x)$ as defined in (4.6) is estimated at an interpolation point x_0 with interpolators $\widehat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i)$ as defined in (4.13) being linear in their weights λ_i and observations $Z(x_i)$, provided that (i) the estimate belongs to the observed realisation and (ii) the second moment function of $Z(x)$ is known or has been estimated using the methods introduced in Sect. 4.3. Optimal interpolators with a non-negative variance are derived for three cases, i.e., when

1. $Z(x)$ is stationary as defined in (4.8) (in Sect. 4.5.1)
2. $Z(x)$ satisfies the intrinsic hypothesis (4.9), (in Sect. 4.5.2), and
3. $Z(x)$ has a non-constant first moment function (in Sect. 4.5.3).

4.5.1 The Stationary Case: Simple Kriging

In the most straightforward case, the spatial random function (4.6) to be interpolated is assumed to be stationary. Under this assumption and with an interpolator (4.13), $\sum_{i=1}^{N} \lambda_i = 1$ (as in (2.64)) is obtained when the interpolator is required to be not biased for all x in D. Then, as in Sect. 2.8.3, the mean square prediction error (4.13,2) is minimised with respect to the weights λ_i,

$$\begin{aligned}\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) &= \mathrm{Var}\widehat{Z}(x_0) - 2\mathrm{Cov}\big(\widehat{Z}(x_0), Z(x_0)\big) + \mathrm{Var}Z(x_0)\\ &= \sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i\lambda_k c_Z(x_i - x_k) - 2\sum_{i=1}^{N}\lambda_i c_Z(x_i - x_0) + c_Z(0) \qquad (4.21)\end{aligned}$$

under the constraint $\sum_{i=1}^{N}\lambda_i - 1 = 0$. Consequently, the partial derivatives $\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) - 2m(\sum_{i=1}^{N}\lambda_i - 1)$ with respect to λ_k, $k = 1, \ldots, N$ and m are calculated and set $= 0$. Thus, $2\sum_{i}^{N}\lambda_i c_{ik} - 2m = 2c_{k0}$, i.e., equation k, is obtained.

If $Z(x)$ is a stationary random function as defined in (4.8) and $\widehat{Z}(x_0) = \sum_{i=1}^{N}\lambda_i Z(x_i)$ an interpolator as defined in (4.13), then the weights λ_i are the solution of the system:

1. $\sum_{i=1}^{N}\lambda_i c_{ik} - m = c_{k0}$, *for* $k = 1, \ldots, N$, $c_{i,k} = c_Z(x_i - x_k)$, *and*
2. $\sum_{i=1}^{N}\lambda_i = 1$,
3. *or, written with matrices*

$$\begin{pmatrix} c_{11} & c_{12} & \ldots & c_{1N} & -1\\ c_{21} & c_{22} & \ldots & c_{2N} & -1\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ c_{N1} & c_{N2} & \ldots & c_{NN} & -1\\ 1 & 1 & \ldots & 1 & 0 \end{pmatrix}\begin{pmatrix}\lambda_1\\ \lambda_2\\ \vdots\\ \lambda_N\\ m\end{pmatrix} = \begin{pmatrix} c_{10}\\ c_{20}\\ \vdots\\ c_{N0}\\ 1\end{pmatrix}$$

4. $\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) = c_Z(0) - \sum_{k=1}^{N}\lambda_k c_{k0} + m$
5. *The equations as in (4.22,1,2) or (4.22,3) are called simple kriging system.*

(4.22)

$2m$ is the Lagrange multiplier used for the minimisation of the mean square prediction error with respect to the weights λ_i under $\sum_{i=1}^{N}\lambda_i = 1$. This constraint ensures that the interpolator is not biased for all x in D, D being the parameter domain as defined in (4.6), and, therefore, it is called *universality condition*. Lagrange multiplier methods allow for calculating maxima and minima of a function under constraints (Problem 4.10).

The matrix in (4.22,3) is non-negative definite due to (2.7,2); if positive definite then its inverse exists which is used for writing, as in (2.75), the solution of the equations. In Problem 4.11, the solution is substituted in (4.21) to obtain the mean square prediction error (4.22,4). This prediction error is the variance of the process (the first term on the right side in (4.22,4)) reduced by the weighted covariances of $Z(x)$ between the interpolation point x_0 and the points with observations x_i (the second term) and augmented by half of the Lagrange multiplier (the third term).

In Geostatistics, the interpolation of a spatial random function is called kriging, and therefore, to be more specific, the equations as in (4.22,2) or (4.22,3) are called simple kriging system and the mean squared prediction error is called kriging variance.

The interpolator $\widehat{Z}(x_0)$ behaves, close to points with observations x_i, as the covariance function $\mathrm{Cov}\big(Z(x), Z(y)\big)$ for $y \to x$. $\widehat{Z}(x_0) = z(x_i)$ on condition that $x_0 = x_i$, i.e., $\widehat{Z}(x_0)$ is exact as required in (4.1,6), because the mean square prediction error (4.21) becomes identically zero for $x_0 = x_k$ and is therefore minimal. Consequently, $\lambda_k = 1$ for $x_0 = x_i$ and $\lambda_i = 0$ for $i \neq k$ are obtained, with $i = 1, \ldots, N$ and $k = 1, \ldots, N$.

The kriging variance in (4.21) can be written using the variogram because, for stationary random functions, $\gamma_Z(h) = c_Z(0) - c_Z(h)$ as shown in Problem 4.4. Substituting and using the universality condition $\sum_{i=1}^{N} \lambda_i = 1$

$$\begin{aligned}\mathrm{Var}(\widehat{Z}(x_0) - Z(x_0)) &= \sum_{i=1}^{N}\sum_{k=1}^{N} \lambda_i \lambda_k \big(c_Z(0) - \gamma_Z(x_i - x_k)\big) \\ &\quad -2\sum_{i=1}^{N} \lambda_i \big(c_Z(0) - \gamma_Z(x_i - x_0)\big) + c_Z(0) \\ &= -\sum_{i=1}^{N}\sum_{k=1}^{N} \lambda_i \lambda_k \gamma_Z(x_i - x_k) + 2\sum_{i=1}^{N} \lambda_i \gamma_Z(x_i - x_0) \qquad (4.23)\end{aligned}$$

is arrived at. From a comparison of (4.23) and (4.21) it is concluded that, in kriging system (4.22), $c(h)$ can be substituted with $-\gamma(h)$ (for stationary $Z(x)$ and universality condition $\sum_{i=1}^{N} \lambda_i = 1$).

When substituting the variogram for the covariance function, it is not required to estimate the constant first moment function $\mu_Z(x) = \mu_Z = \mu$ of $Z(x)$ prior to estimating the second moment function. μ_Z is obtained, subsequent to the variogram estimation, using the kriging system (4.24) which is derived in Problem 4.12. In addition to this derivation, in Problem 4.12, the properties of both estimators for the expectation function of a stationary random function, as proposed in (4.24) and (2.56), are compared.

If $Z(x)$ is a stationary spatial random function as in (4.8) and $\hat{\mu} = \sum_{i=1}^{N} \lambda_i' Z(x_i)$ an estimator as defined in (4.13), then:

1. $\sum_{i=1}^{N} \lambda_i' \gamma(x_i - x_k) + m' = 0$, *for* $k = 1, \ldots, N$ (4.24)
2. $\sum_{i=1}^{N} \lambda_i' - 1 = 0$, *the universality condition, and*
3. $\mathrm{Var}(\hat{\mu} - \mu) = m'$, *the kriging variance.*

Usually, when interpolating a stationary spatial random function, the covariances in (4.22) are not known and, therefore, a covariance model $c_Z(h) = c_Z(0) - \gamma_Z(h)$ is substituted, $\gamma_Z(h)$ being an analytical variogram (examples are given in Sect. 4.3.2) fitted to the empirical variogram that has been calculated as required in (4.11) and Fig. 4.8.

For example, a spherical model with sill 0.08, range 50 m and a nugget effect 0.015 (plot (a) in Fig. 4.12) is fitted to the empirical semivariograms estimated for K_2O in the "Strassenacker" soil in Sect. 4.3.1. Using this model

and simple kriging as defined in (4.22), Gstat interpolates K_2O when invoked with the following command file

```
data(Potassium): 'stracker.eas',x=2,y=3,v=10,log,sk_mean=1.16;
variogram(Potassium): 0.015 Nug(0) + 0.065 Sph(50);
data(): 'stracker.grid.eas', x=1, y=2;
set cutoff = 140;
set width = 2.5;
set output = 'potassiumstracker.rep.eas';
```

The above command file is generated by inserting option `sk_mean=1.16` in the `data(Potassium)` command and lines

```
data(): 'stracker.grid.eas', x=1, y=2;
set output = 'potassiumstracker.rep.eas';
```

in the Gstat command file, obtained when quitting Gstat after having fitted the model in Fig. 4.12 (a). The inserted lines describe the files with the interpolation points and the results, and option `sk_mean=1.16` gives an estimate for the unknown expectation function which is assumed to be constant. The value given is the mean of the observations calculated by Gstat when reading the data for the variogram estimation in Sect. 4.3.1. This option omitted, Gstat estimates the expectation function as required in (4.24).

A grid with interpolation points (an interpolation grid) for plot "Strassenacker" is generated by calculating the cross product of R vectors

```
#interpolation grid with distances 5 m for 'Strassenacker' plot
xnull <- c(0,5,10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,
        90,95,100,105,110,115,120,125,130,135,140,145,150,155)
ynull <- c(0,5,10,15,20,25,30,35,40,45,50,55,60,65,70,75,80,85,
        90,95,100,105,110,115,120,125,130,135,140,145,150,155)
...   #as demonstrated in Sect. 3.6 when the grid for the estima-
...   #tion of the trend surfaces for the tree line is calculated
stracker.grid <- data.frame(x,y)
write(t(stracker.grid),file="stracker.grid",
                ncol = ncol(stracker.grid))
```

and then writing file `stracker.grid`. This file is renamed in `stracker.grid.eas` once the `.eas` header lines (an example is given in Sect. 4.3.1) have been inserted.

Executing the Gstat command file above, Gstat interpolates with the results plotted in Figs. 4.19 and 4.20. These plots are generated from result file `potassiumstracker.rep.eas` from which, after having removed the `.eas` header lines, an R data frame is generated. Then, the optimal interpolation of K_2O and its kriging variance are plotted using R expressions similar to those in the remarks to Fig. 3.11.

The optimal interpolation of K_2O, as plotted in Fig. 4.19, resembles a plot obtained by way of freehand contouring the observations in Fig. 4.9, a deterministic interpolation method (cf. the introduction to this chapter). In contrast to deterministic methods, optimal interpolation estimates the mean

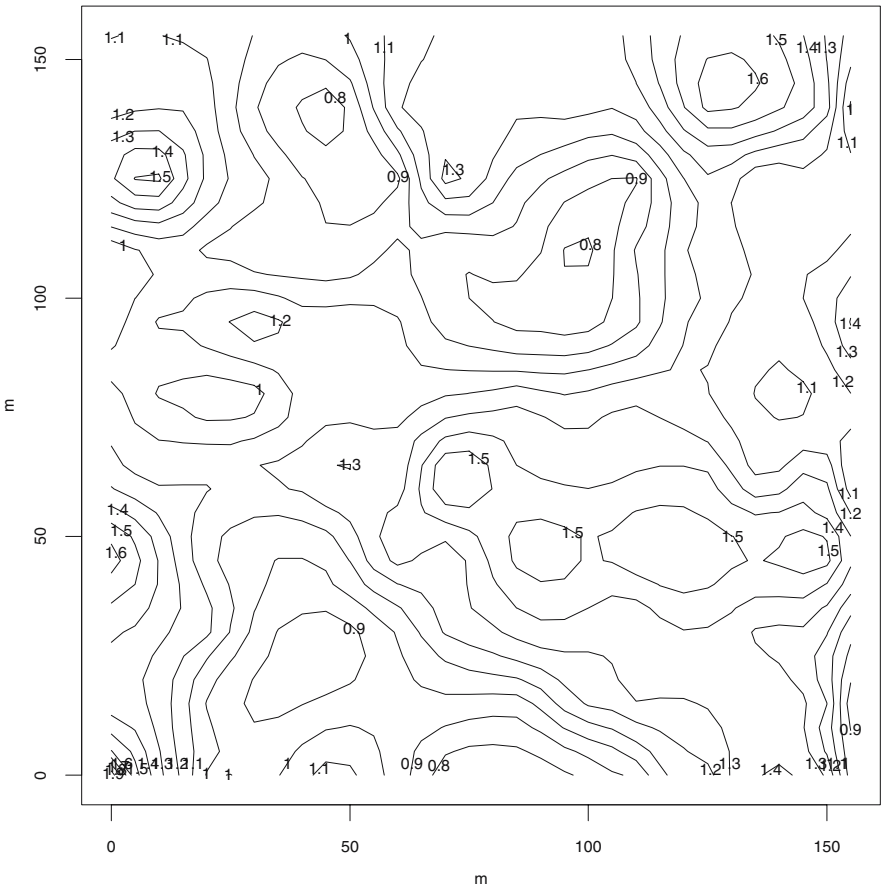


Fig. 4.19. K_2O (log(mg in 100 g soil)) in plot "Strassenacker" interpolated with simple kriging using the variogram model as fitted in Fig. 4.12 (a).

square prediction error: the kriging variance of K_2O in the "Strassenacker" soil as plotted in Fig. 4.20 is (i) less than 0.025 (log(mg in 100 g of soil))^2 close to the interpolation points, (ii) between 0.025 and 0.03 in the interior of the plot (except a few points), however (iii), increases rapidly at the lower and upper margins of the plot since, there, the distances between the interpolation points and the points with observations increase.

The interpolation points used for the estimation of K_2O in the "Strassenacker" soil and the results obtained from Gstat are transferred from R to the Gstat stand-alone executable (and also in opposite direction) using `.eas` table files as introduced in Sect. 4.3.1. Besides `.eas` table files, Gstat reads and writes gridded data from and to files in the formats frequently used in Geographical Information Systems (GIS) and other software systems for the

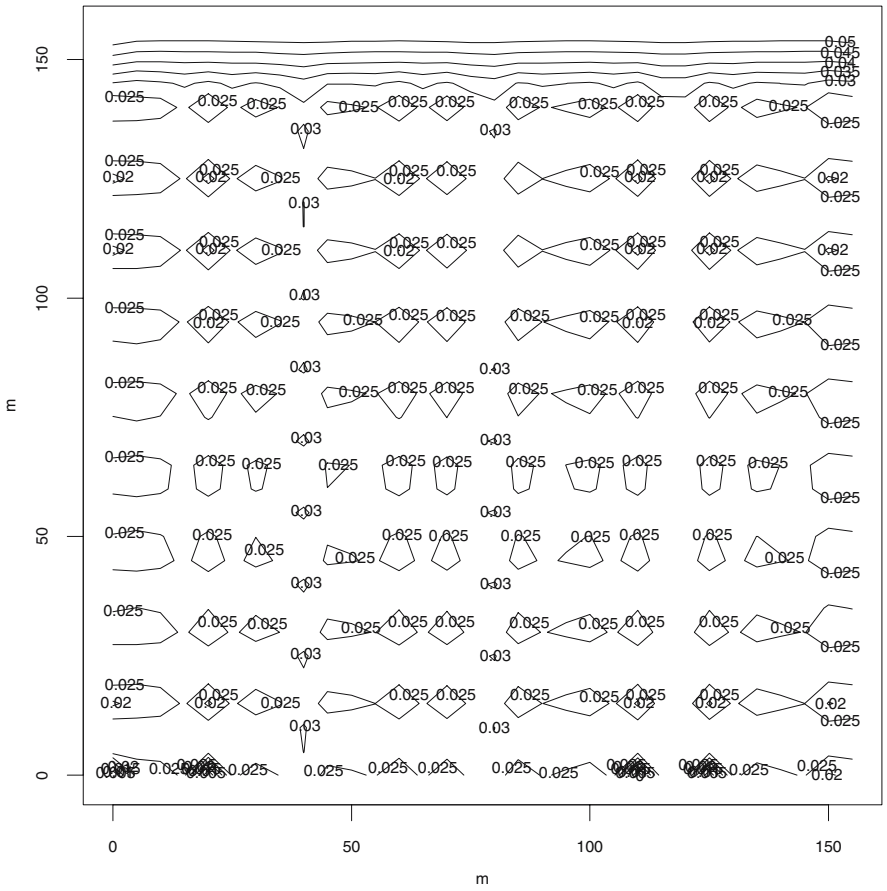


Fig. 4.20. Variance $\left((\log(\text{mg in 100 g of soil}))^2\right)$ pertaining to the interpolation of K_2O in Fig. 4.19.

analysis of spatially referenced data. It is recommended to use a GIS (or another suitable software) together with Gstat in projects involving large amounts of spatially referenced data (cf. link "R spatial projects" available under [114]). Using such a software system,

1. masks can be easily defined to limit an interpolation to the area under study: for example, the area with the observed tree-lines as plotted in Fig. 3.10 is easily defined in a GIS; in R, however, the area with the observations as plotted in Fig. 3.11 is constructed as the convex hull of the points with observations
2. contour maps (including base maps and in the usual projections, examples being given in Figs. 4.22 and 4.34) are readily produced from the Gstat results.

4.5.2 The Intrinsic Case: Ordinary Kriging

An optimal linear interpolator $\widehat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i)$ as defined in (4.13), with x_0 being in the parameter domain D of a spatial random function $Z(x)$ satisfying the intrinsic hypothesis defined in (4.9), is derived here. From the interpolation error $\widehat{Z}(x_0) - Z(x_0) = \sum_{i=1}^{N} \lambda_i Z(x_i) - Z(x_0) = \sum_{i=0}^{N} \lambda_i Z(x_i)$ with $\lambda_0 = -1$, the interpolation variance in (4.25) is obtained using (4.15).

$$\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) = -\sum_{i=0}^{N}\sum_{k=0}^{N} \lambda_i \lambda_k \gamma_Z(x_i - x_k) \tag{4.25}$$

(4.15) also implies that the variance in (4.25) is non-negative provided that $\sum_{i=0}^{N} \lambda_i = 0$, i.e., provided that $\sum_{i=1}^{N} \lambda_i = 1$ since $\lambda_0 = -1$. Further, the universality condition in (4.26,2) is obtained, because $\mathrm{E}\sum_{i=0}^{N} \lambda_i Z(x_i) = \mathrm{E}\sum_{i=1}^{N} \lambda_i \big(Z(x_i) - Z(x_0)\big) = \sum_{i=1}^{N} \lambda_i \mathrm{E}\big(Z(x_i) - Z(x_0)\big)$ using (4.15). This sum is concluded, from (4.9,1), to be identically zero: the estimator has no bias.

As in the stationary case in Sect. 4.5.1, the mean square prediction error in (4.25) is minimised under the universality condition using a Lagrange multiplier: in Problem 4.13, the partial derivatives of $-\sum_{i=0}^{N}\sum_{k=0}^{N} \lambda_i \lambda_k \gamma_Z(x_i - x_k) - 2m(\sum_{i=1}^{N} \lambda_i - 1)$ with respect to λ_k, $k = 1, \ldots, N$ and m are calculated, set $= 0$ to obtain $2\sum_{i=1}^{N} \lambda_i \gamma_Z(x_i - x_k) + 2\lambda_0 \gamma_Z(x_k - x_0) + 2m = 0$, and, with $\lambda_0 = -1$, equation k in kriging system (4.26) is arrived at.

If $Z(x)$ is an intrinsically stationary spatial random function as in (4.9) and if $\widehat{Z}(x) = \sum_{i=1}^{N} \lambda_i Z(x_i)$ an interpolator as in (4.13), then the weights λ_i are the solution of the following equations:

1. $\sum_{i=1}^{N} \lambda_i \gamma(x_i - x_k) + m = \gamma(x_k - x_0)$, *for* $k = 1, \ldots, N$ *and*
2. $\sum_{i=1}^{N} \lambda_i = 1$, *the universality condition;*
3. *or, written with matrices*

$$\begin{pmatrix} 0 & \ldots & \gamma(x_1 - x_N) & 1 \\ \gamma(x_2 - x_1) & \ldots & \gamma(x_2 - x_N) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(x_N - x_1) & \ldots & 0 & 1 \\ 1 & \ldots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ m \end{pmatrix} = \begin{pmatrix} \gamma(x_1 - x_0) \\ \gamma(x_2 - x_0) \\ \vdots \\ \gamma(x_N - x_0) \\ 1 \end{pmatrix}$$

4. $\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) = \sum_{k=1}^{N} \lambda_k \gamma(x_i - x_0) + m$, *the mean square prediction error, is called kriging variance.*
5. *The equations as in (4.26,1,2) or (4.26,3) are called ordinary kriging system.*

(4.26)

If $2\gamma(h)$ is conditionally negative definite (in the remarks to (4.20)) then $\mathrm{Var}\sum_{i=1}^{N} \lambda_i Z(x_i) \geq 0$ is obtained in (4.15) and the inverse of the matrix with the variograms in (4.26) exists. However, the variogram-matrix is not

positive semidefinite, as its diagonal elements are identically zero. The kriging variance in (4.26,4) is derived in Problem 4.14.

Formally, the kriging systems (4.26) and (4.22) are very similar when, subsequent to a comparison of (4.21) and (4.23), $-\gamma(h)$ is substituted for $c(h)$ in (4.22). Mathematically, however, the kriging systems are different: in (4.22), the spatial random function $Z(x)$ is assumed to be stationary; in (4.26), however, its differences $Z(x+h)-Z(x)$ are assumed to be stationary. For example, universality condition (4.22,2) is obtained from the constant expectation function of a stationary random function, whereas universality condition (4.26,2) is obtained from the first moment function of differences of a random function satisfying the intrinsic hypothesis, i.e., the expectation function being identically zero by definition (4.9).

The intrinsic hypothesis is weaker than the stationarity assumption and the variogram is more general than the covariance functions as shown in the remarks concluding Sect. 4.4: the variogram describes the second moment functions in cases where no useful covariance function can be estimated. This advantage is paid for with the disadvantage of more restricted operational possibilities: the set of all linear combinations acting as interpolators for a stationary spatial random function in (4.14) is constrained, in (4.15), to those linear combinations with $\sum_{i=1}^{N} \lambda_i = 0$ when an intrinsically stationary random function is interpolated.

4.5.3 The Case with a Non-constant Drift: Universal Kriging

Kriging systems for stationary or intrinsically stationary spatial random functions $Z(x)$ as proposed in (4.22) and (4.26) allow for interpolating $Z(x)$ with a constant expectation function (drift) $\mu_Z(x) = \mu_Z$, as is concluded from definitions (4.8) and (4.9). There are, however, $Z(x)$ with $\mu_Z(x)$ that are not constant: for example, the SWIR climatologies as plotted in Fig. 4.16 stem from a realisation of $Z(x)$ with $\mu_Z(x) \neq \mu_Z$. Such a spatial random function $Z(x)$ has a drift $\mathrm{E}Z(x) = \mu_Z(x) = \mu(x)$ which is a function of x and a covariance $\mathrm{Cov}\big(Z(y), Z(x)\big) = \mathrm{E}\big((Z(y) - \mu(y))(Z(x) - \mu(x))\big)$ which is a function of x and y, i.e., both moment functions depend on parameter x. Under this assumption, estimators $\hat{\mu}(x_0)$ for the non-constant drift and $\widehat{Z}(x_0)$ for the random function are to be calculated in Sect. 4.5.3, x_0 being in the parameter domain D. This estimation is called *universal kriging*.

When kriging universally, it is further assumed that the spatial random function $Z(x)$ can be decomposed into a rather systematic and smooth drift $\mu(x)$ and the residual component $Y(x)$ with random fluctuations. Such a decomposition is often used, for example, where a random function is decomposed into a systematic and a random component in (1.14), (3.1), (3.8) and for the estimation of the variogram in Fig. 4.18 of the SWIR climatologies as plotted in Fig. 4.16. The decomposition assumption for spatial random functions with a non-constant drift is usually formulated as in (4.27).

When interpolating a spatial random function $Z(x)$ with a non-constant drift $\mathrm{E}Z(x) = \mu(x)$, it is assumed that:

1. $Z(x) = Y(x) + \mathrm{E}Z(x) = Y(x) + \mu(x)$, *with*
2. $\mu_Y(x) = 0$ *and* $c_Y(y, x) = c_Y(x + h, x) = c_Y(h) = c_Y(y - x)$ *as well as*
3. $\mathrm{E}Z(x) = \mu(x) = \sum_{i=1}^{p} b_i f_i(x)$ *at least locally as defined in (2.55), with b_i being unknown possibly random coefficients, and $f_i(x)$ being base functions, usually monomials in the coordinates, e.g., 1, x^*, y^*, z^*, x^{*2}, ...*

(4.27)

To model the drift $\mu(x)$, a function (usually a polynomial in the coordinates) is chosen, as in Chap. 3, that fits the observations. The coefficients b_j in the model for the drift are not known and neither is the covariance function. As a consequence, when interpolating a spatial random function $Z(x)$ with universal kriging, (i) $Z(x)$, (ii) its non-constant drift $\mathrm{E}Z(x) = \mu(x)$, and (iii) the stationary covariance function $c_Y(h)$ of the random component $Y(x)$ as defined in (4.27,1,2) are to be estimated simultaneously. When posed in this form, the problem cannot be solved.

Hence, in the following derivation of the system for universal kriging as proposed in (4.33) and (4.34), or alternatively written with matrices in (4.35), $Z(x)$ and $\mathrm{E}Z(x)$ are estimated under the assumption that the stationary covariance function $c_Y(h)$ is known. Then, in (4.39), an additional assumption is postulated under which $\mathrm{E}Z(x)$ and $c_Y(h)$ can be estimated using generalised least squares as introduced in Sect. 3.2.4.

The interpolator used in universal kriging is, as in simple and ordinary kriging, (i) linear in the observations $z(x_i)$ and the unknown weights λ_i, (ii) universal (i.e., free from bias for all x in D albeit the drift), and (iii) optimal. By substituting (4.27) in (4.13,1) and then applying the linearity and the no-bias property,

$$\begin{aligned} 0 = \mathrm{E}\big(\widehat{Z}(x_0) - Z(x_0)\big) &= \mathrm{E}\sum \lambda_i Z(x_i) - \mathrm{E}Z(x_0) \\ &= \sum_{i=1}^{N} \lambda_i \sum_{j=1}^{p} b_j f_j(x_i) - \sum_{j=1}^{p} b_j f_j(x_0) \\ &= \sum_{j=1}^{p} b_j \left(\sum_{i=1}^{N} \lambda_i f_j(x_i) - f_j(x_0) \right) \end{aligned} \tag{4.28}$$

$$\sum_{i=1}^{N} \lambda_i f_j(x_i) = f_j(x_0) \qquad j = 1, \ldots, p \tag{4.29}$$

(4.28) is obtained, and further the universality conditions in (4.29) are met because (4.28) holds, due to (4.27,3), for arbitrary $b_j \neq 0$.

As the interpolator is required to be not only linear and universal but also optimal, the universal kriging equations are derived from the optimality, the

property which has not yet been applied. In the first step, the interpolation error in (4.30) is calculated using (4.28) and (4.29),

$$\begin{aligned}\Delta &= \sum_{i=1}^{N} \lambda_i Z(x_i) - Z(x_0) \\ &= \sum_{i=1}^{N} \lambda_i \left(Y(x_i) + \sum_{j=1}^{p} b_j f_j(x_i) \right) - \left(Y(x_0) + \sum_{j=1}^{p} b_j f_j(x_0) \right) \\ &= \sum_{i=1}^{N} \lambda_i Y(x_i) - Y(x_0) + \sum_{j=1}^{p} b_j \left(f_j(x_0) - \sum_{i=1}^{N} \lambda_i f_j(x_i) \right) \\ &= \sum_{i=1}^{N} \lambda_i Y(x_i) - Y(x_0) \end{aligned} \tag{4.30}$$

then, in the second step, its variance, i.e., the kriging variance or the mean square prediction error, is arrived at in (4.31).

$$\begin{aligned}\mathrm{Var}\Delta &= \mathrm{Var} \sum_{i=1}^{N} \lambda_i Y(x_i) - Y(x_0) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \lambda_k c_Y(x_i - x_k) - 2 \sum_{i=1}^{N} \lambda_i c_Y(x_i - x_0) + c_Y(0) \\ &= \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \lambda_k c_{ik} - 2 \sum_{i=1}^{N} \lambda_i c_{i0} + c_0 \\ &\qquad \text{with } c_{ik} = c_Y(x_i - x_k) \end{aligned} \tag{4.31}$$

The kriging variance in (4.31) is minimised using the universality conditions (4.29) as constraints in order to arrive at the weights λ_i pertaining to an optimal interpolator: using a Lagrange multiplier m_j for any constraint,

$$\sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \lambda_k c_{ik} - 2 \sum_{i=1}^{N} \lambda_i c_{i0} + c_0 - 2 \sum_{j=1}^{p} m_j \left(\sum_{i=1}^{N} \lambda_i f_j(x_i) - f_j(x_0) \right) \tag{4.32}$$

(4.32) is obtained, with unknowns λ_i and m_j. Then, the partial derivatives of (4.32) with respect to λ_i, $\mathrm{d}/\mathrm{d}\lambda_i(\ldots)$ for $i = 1, \ldots, N$, and m_j, $\mathrm{d}/\mathrm{d}m_j(\ldots)$ for $j = 1, \ldots, p$, are calculated, and set $= 0$ in Problem 4.17. The result of these straightforward but rather lengthy calculations are the equations for the universal kriging in (4.33) and (4.34):

$$\sum_{i=1}^{N} \lambda_i c_{ik} - \sum_{j=1}^{p} m_j f_j(x_k) = c_{k0}, \quad k = 1, \ldots, N \tag{4.33}$$

$$\sum_{i=1}^{N} \lambda_i f_j(x_i) = f_j(x_0), \quad j = 1, \ldots, p \tag{4.34}$$

which contain N optimality conditions and p universality conditions, one for each b_j in (4.27,3). Hence, the possibility to interpolate spatial random functions with a non-constant drift is paid for by more coefficients having to be estimated as compared with the equations for the simple and ordinary kriging (4.22) and (4.26). Using matrices, equations (4.33) and (4.34) become

$$\begin{pmatrix} \begin{pmatrix} c_{11} & \dots & c_{1N} \\ c_{21} & \dots & c_{2N} \\ \vdots & \ddots & \vdots \\ c_{N1} & \dots & c_{NN} \end{pmatrix} & \begin{pmatrix} -f_1(x_1) & \dots & -f_p(x_1) \\ -f_1(x_2) & \dots & -f_p(x_2) \\ \vdots & \ddots & \vdots \\ -f_1(x_N) & \dots & -f_p(x_N) \end{pmatrix} \\ \begin{pmatrix} f_1(x_1) & \dots & f_1(x_N) \\ f_2(x_1) & \dots & f_2(x_N) \\ \vdots & \ddots & \vdots \\ f_p(x_1) & \dots & f_p(x_N) \end{pmatrix} & \begin{pmatrix} 0 & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ m_1 \\ m_2 \\ \vdots \\ m_p \end{pmatrix} = \begin{pmatrix} c_{10} \\ c_{20} \\ \vdots \\ c_{N0} \\ f_1(x_0) \\ f_2(x_0) \\ \vdots \\ f_p(x_0) \end{pmatrix}$$

or, more concisely,

$$\begin{pmatrix} \boldsymbol{C}_{ik} & -\boldsymbol{F} \\ \boldsymbol{F}^T & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_0 \\ \boldsymbol{f}_0 \end{pmatrix} \tag{4.35}$$

with $\boldsymbol{f}_0 = \big(f_1(x_0), \dots, f_p(x_0)\big)^T$ being the column vector of the base functions as in (4.27) for interpolation point x_0, $\boldsymbol{f}_i = \big(f_1(x_i), \dots, f_p(x_i)\big)^T$ the column vector of the base functions for points with observations x_i, $\boldsymbol{F} = \big(\boldsymbol{f}_1^T, \dots, \boldsymbol{f}_N^T\big)^T$ the matrix of the base functions for all x_i, $\boldsymbol{C}_{ik} = (c_{ik})$ the matrix of the covariances between the observations, $\boldsymbol{c}_0 = \big(c_{10}, \dots, c_{N0}\big)^T$ the column vector of the covariances between the x_0 and the x_i, $\boldsymbol{\lambda} = \big(\lambda_1, \dots, \lambda_N\big)^T$ the column vector of the kriging weights, and $\boldsymbol{m} = (m_1, \dots, m_p)^T$ the column vector of the Lagrange multipliers.

Solving the equations for the universal kriging,

$$\begin{pmatrix} \boldsymbol{\lambda} \\ \boldsymbol{m} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}_{ik} & -\boldsymbol{F} \\ \boldsymbol{F}^T & \boldsymbol{0} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{c}_0 \\ \boldsymbol{f}_0 \end{pmatrix} \tag{4.36}$$

is obtained, and furthermore, the interpolator

$$\widehat{Z}(x) = (\boldsymbol{\lambda}^T, \boldsymbol{m}^T) \begin{pmatrix} \boldsymbol{z}_i \\ \boldsymbol{f}_0 \end{pmatrix} = \left(\begin{pmatrix} \boldsymbol{C}_{ik} & -\boldsymbol{F} \\ \boldsymbol{F}^T & \boldsymbol{0} \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{c}_0 \\ \boldsymbol{f}_0 \end{pmatrix} \right)^T \begin{pmatrix} \boldsymbol{z}_i \\ \boldsymbol{f}_0 \end{pmatrix} \tag{4.37}$$

with $\boldsymbol{z}_i = (z_1, \dots, z_N)^T$ being the column vector of the observations. Thus, $\widehat{Z}(x)$ in (4.37) is a linear combination of functions $c_{i0} = c_Y(x_i - x_0) = \mathrm{Cov}\big(Y(x_i), Y(x_0)\big)$ and $f_j(x_0)$ with constant coefficients. The solution (4.36) is unique, provided that the covariance matrix is positive definite and the base functions $\boldsymbol{f}_0$ are linearly independent with respect to the points with observations x_i.

As a direct consequence of (4.37), the interpolator is exact: if $x = x_i$, x an interpolation point, and x_i, $i = 1, \ldots, N$, the points with observations, then $\lambda_i = 1$ and $\lambda_k = 0$, $k = 1, \ldots, N$, $k \neq i$, and $m_j = 0$, $j = 1, \ldots, p$, is a solution (and the only one) of the system. For this solution, the interpolation error becomes identically zero, which implies a zero (i.e., minimal) kriging variance. Exact interpolators are also obtained when interpolating a stationary or intrinsically stationary spatial random function by solving the simple or ordinary kriging systems in (4.22) or (4.26). This property of optimal interpolators is reconcilable with (i) the optimality of the conditional expectation as proposed in (2.82) because the conditional expectation of an observed random variable is the observed value and (ii) the intuitive notion that the best interpolation of $Z(x)$ for $x = x_i$ is the observation $z(x_i)$.

The interpolation variance as in (4.38) is calculated in Problem 4.18.

$$\mathrm{Var}\big(\widehat{Z}(x) - Z(x)\big) = c_Y(0) - \sum_{k=1}^{N} \lambda_k c_Y(x_i - x_0) + \sum_{j=1}^{p} m_j f_j(x) \qquad (4.38)$$

The universal kriging equations (4.33) and (4.34) are derived, above, by assuming decomposition $Z(x) = Y(x) + \mu(x)$ in (4.27) and a known covariance function $c_Y(h)$. Consequently, the equations can be solved in order to obtain the interpolator $\widehat{Z}(x)$ in (4.37) provided that $c_Y(h)$ is known or, as in the more realistic case, an estimate $\hat{c}_Y(h)$ is substituted for $c_Y(h)$. An empirical covariance function $\hat{c}_Y(h)$ can be calculated from the residuals $z_i - \hat{\mu}(x_i)$, $i = 1, \ldots, N$, obtained by subtracting the drift from the observations. An estimator for the drift is arrived at, in Problem 4.19, by applying universal kriging with a slight modification in (4.33) and (4.34): 0 is substituted for c_{i0} and $f_j(x_i)$ is substituted for $f_j(x_0)$ on the right side of the equations.

Consequently, if $c_Y(h)$ is known then $\mu(x)$ can be estimated, and, if $\mu(x)$ is known then $c_Y(h)$ can be estimated. This is the setting encountered when a linear model with correlated residuals is estimated using generalised least squares as introduced in Sect. 3.2.4: a joint estimation of both $\mu(x)$ and $c_Y(h)$ is not possible. However, an iterated estimation as proposed in the remarks to (3.14) and (3.15) is feasible provided that $Z(x)$ is not only in agreement with assumptions (4.27) but also with assumption (4.39).

Let $Z(x) = Y(x) + \mu(x)$ be a spatial random function as in (4.27) having a variogram $2\gamma_Y(h)$ with a range a that is small when compared with D. Then $c_Y(h) \approx c_Z(h)$ can be assumed and, under this assumption, generalised least squares estimates (Sect. 3.2.4) can be calculated for $\mu(x)$ and $c_Y(h)$. (4.39)

In a first step, a linear model for the drift $\mu(x)$ is estimated as required in (3.7) and (3.8) and the empirical covariance function $\hat{c}_Y^{(1)}(h)$ of the residuals is calculated as a first estimate for $c_Y(h)$. In a second step, the drift is again estimated but now the covariances in the residuals are accounted for using

(3.14), with $\hat{c}_Y^{(1)}(h)$ substituted for $c_Y(h)$, i.e., using generalised least squares. These generalised least squares estimates are repeated until the estimates for $c_Y(h)$ become stable, i.e., until $\hat{c}_Y^{(k-1)}(h) \approx \hat{c}_Y^{(k)}(h)$.

A generalised least squares estimation is, however, not feasible in the following cases: (i) when $Z(x)$ is not reconcilable with (4.39) because the range of the empirical variogram calculated from the observations is approximately the same size as the area under study, and (ii), when $Z(x)$ cannot be decomposed as required (4.27) because no base functions can be found. These difficulties can be circumvented when local (as defined in (2.55)) estimates are calculated provided that enough observations are available.

For example, the climatological means of SWIR as plotted in Fig. 4.16 are assumed to be in agreement with assumption (4.39): the range of the empirical variograms calculated from the residuals of a first order trend surface (i.e., monomials 1, x^*, y^* and $x^* \times y^*$ in (4.27,3)) is approximately 3^0 as shown in Fig. 4.18, and thus is small when compared with the region under study bounded by 4^0 and 16^0 (eastern) longitude and by 48^0 and 54^0 (northern) latitude. Hence, the spherical model fitted to the empirical variogram of SWIR in Fig. 4.18 (a) is refitted using generalised least squares as implemented in Gstat.

Assuming that estimates in plot (a) in Fig. 4.18 are saved in a Gstat command file subsequent to the calculations as demonstrated in the remarks to Fig. 4.18, the line for a generalised least squares estimation is inserted

```
data(SWIR): '/pfad/swirnorde.eas',x=1,y=2,v=4,d=1;
variogram(SWIR): 19.935 Sph(2.885);
set cutoff = 5;
set fit = 1;
set width = 0.25;
set gls_residuals = 1;  #(generalised least squares, gls)
```

and the variogram estimation is repeated resulting, with only a few iterations, in the fitted spherical model as plotted Fig. 4.21.

The generalised least squares estimate of the base functions (a first order trend surface) for the drift of SWIR deviates slightly, as shown in Problem 4.26, from the estimate obtained when using ordinary least squares and, therefore, also the empirical residuals of both estimates are not identical. Because in Gstat, a semivariogram is calculated from the residuals, the empirical semivariograms in plot (a) in Fig. 4.18 and in Fig. 4.21 are not identical.

Since a generalised least squares estimate as introduced in Sect. 3.2.4 accounts for correlations in the residuals, it has, on condition that the residuals are not iid. as required in (3.11), a smaller residual variance than an ordinary least squares estimate (Problem 4.26): the spherical model in Fig. 4.21 shows a better fit than the one in Fig. 4.18 (a). Consequently, using generalised least squares under the assumption that the SWIR climatologies in Fig. 4.16 are reconcilable with (4.27) and (4.39), the estimates for the drift $\mu(x)$ and the covariances $c_Y(h)$ are improved.

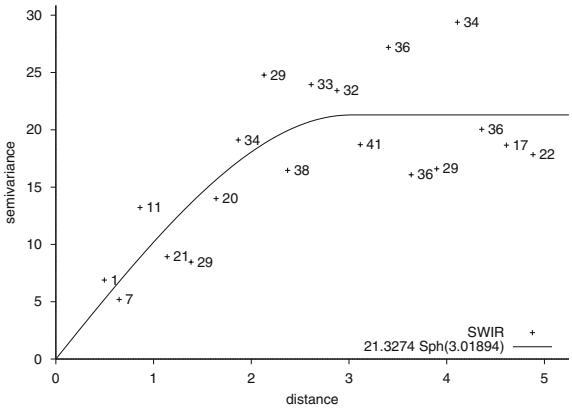


Fig. 4.21. Empirical semivariogram in $(\mathrm{Wm}^{-2})^2$ of the residuals of a trend surface as defined in (3.23) for the universal kriging of SWIR climatologies as plotted in Fig. 4.16. The trend surface and the covariances (the variogram) are estimated using generalised least squares under the assumption (4.39) and for all directions, a maximal distance of 5^0 and distance classes 0.25^0 wide.

The variogram model as estimated in Fig. 4.21 is saved in a Gstat command file which then is used (with lines inserted describing the interpolation grid and the result file) for interpolating the SWIR climatologies by universal kriging. This interpolation results in the maps plotted in Figs. 4.22 and 4.23.

In Fig. 4.22, SWIR increases, in the mean, from North to South. The kriging variance, in Fig. 4.23, is less than 5 $(\mathrm{Wm}^{-2})^2$ near the points with observations x_i, from whence it increases with increasing distance from the x_i. Interpolation variances of approximately 25 $(\mathrm{Wm}^{-2})^2$ (this corresponds to a relative error of approximately 5%) are obtained in the north-east of the area under study, where all x_i are on the West of the interpolation points. In the interior of the area with observations, however, the interpolation variance is less than 10 $(\mathrm{Wm}^{-2})^2$, owing to abundant observations.

4.5.4 Summary

Kriging is the optimal prediction of a (spatial) random function. Hence, when interpolating with kriging methods, the basic assumption known from Sect. 2.7 is made: the observations stem from a realisation of a random function and the estimate belongs to the observed realisation.

In most cases, only one realisation has been observed. This imposes additional assumptions reconcilable with the observations when the moment functions of the random function under analysis are estimated. Usual assumptions are (i) the stationarity as defined in (4.8), (ii) the intrinsic hypothesis as defined in (4.9), and (iii), applied to random functions with a non-constant expectation function, also the assumptions (4.27) and (4.39), which require

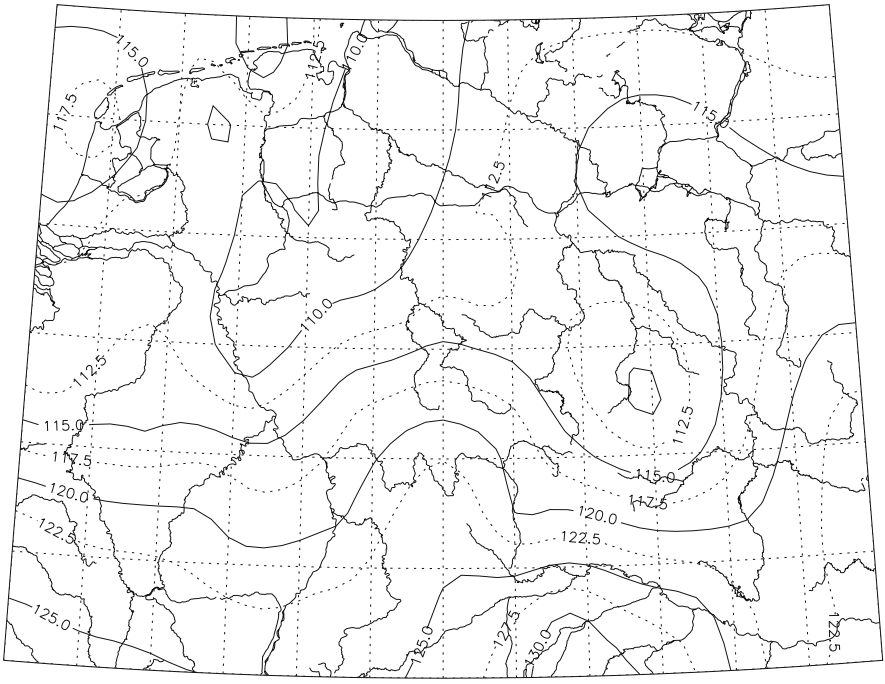


Fig. 4.22. SWIR climatology in Wm^{-2} universally kriged from SWIR climatological means as plotted in Fig. 4.16 using the variogram model estimated in Fig. 4.21.

a decomposition of the random function into a relatively smooth first moment function and a residual component assumed to be random. Further, it can be assumed that the random function under study is intrinsic of order k, i.e., an IRFk as proposed in [92]. IRFks include random functions with a non-constant drift. An introduction to IRFks is found in [41] and [35].

Kriging has some advantages when compared with deterministic interpolation methods, for example, those introduced in Sect. 4.1:

1. The assumptions used for the interpolation are stated explicitly.
2. The interpolator is calculated from covariances (variograms) of differences $x_0 - x_i$, x_0 being an interpolation point, and x_i, $i = 1, \ldots, N$, the points with observations (on the right side of the kriging systems (4.22), (4.26) and (4.35)), and also from differences $x_i - x_k$ (in the matrices in these kriging systems). On the one hand, x_i in short distances of each other (as compared with their average distance to x_0) are "downweighted"; on the other hand, no x_i is neglected.
3. The functions used for the interpolation, i.e., the expectation and the covariance function (the variogram) are estimated from the observations:

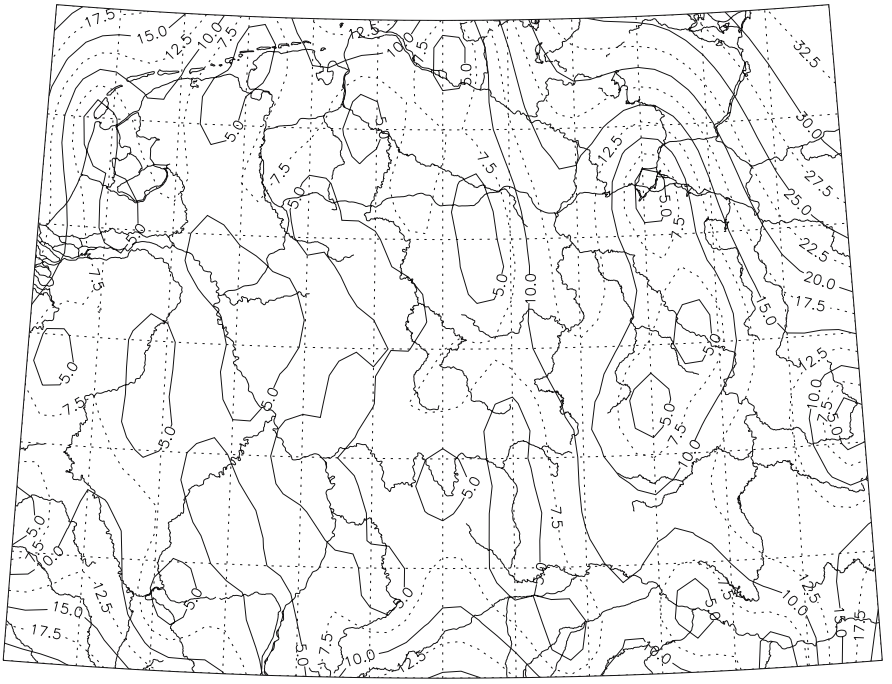


Fig. 4.23. Kriging variance in $(\mathrm{Wm}^{-2})^2$ of the interpolated SWIR climatology in Fig. 4.22.

fitting a model to an empirical variogram is less arbitrary than the selection of a weight function.

4. The interpolator is linear and optimal: the interpolation weights are calculated such that the mean square prediction error (the interpolation variance) becomes minimal. The interpolation variance is estimated together with the random function.
5. The equations in (4.22), (4.26) and (4.35) depend on the covariance function (the variogram) of differences $x_0 - x_i$ and $x_i - x_k$ as shown in item 2., above, and not on the observations (or only indirectly on the observations when an empirical variogram or covariance function is substituted for the theoretical second moment function).
6. The kriging interpolators as obtained from (4.22), (4.26) and (4.35) are point estimates. Areal estimates are introduced in Sect. 4.6.5.
7. A measurement error can be accounted for, as shown in Sect. 4.6.3.

4.6 Supplements

Isotropic empirical semivariograms are calculated, in Sect. 4.6.1, without partitioning the (x^*, y^*)-plane (on the right in Fig. 4.8) into distance classes. In Sect. 4.6.2 it is argued that estimates for the second moment function calculated from observations of a random function depend on the sampling interval. Hence, an interpolation calculated using a variogram or covariance estimate will also depend on the sampling interval. In Sect. 4.6.3 it is shown that a measurement error that satisfies certain conditions can be accounted for in optimal interpolation. In Sec. 4.6.4 and 4.6.5, an example for both a local interpolation and an areal estimate are given.

4.6.1 Isotropic Empirical Variograms

An empirical semivariogram is calculated, as required in (4.11) and Fig. 4.8, by partitioning the (x^*, y^*)-plane (on the right, in Fig. 4.8) into direction as well as distance classes which are then used to aggregate the differences $h = x_i - x_k$ for computing the empirical semivariogram, x_i and x_k being points with observations. In the isotropic case, no direction classes are required to aggregate the $|h| = |x_i - x_k|$, as is demonstrated in Sects. 4.3.1 and 4.3.4, and, as demonstrated in this subsection, an empirical semivariogram can be calculated even without using distance classes.

An estimator for an isotropic empirical variogram $2\hat{\gamma}_Z(h)$ *of an intrinsically stationary random function* $Z(x)$ *can be calculated by plotting* $(z(x_i) - z(x_k))^2$ *versus* $|x_i - x_k|$ *as scatterplot and subsequently fitting an analytical variogram.* (4.40)

For example, plot (f) in Fig. 4.24 shows an empirical semivariogram that is estimated, without distance classes as required in (4.40), from the residuals of a first order trend surface for the SWIR climatologies in Fig. 4.16. This result is obtained in Gstat under the assumptions (4.27) and (4.39) using generalised least squares and with very small distance classes containing only one $|x_i - x_k|$. Under the same assumptions and using the same method but with distance classes 0.25^0 wide, the empirical semivariogram (but not the fitted model) in Fig. 4.28 is obtained.

Often, when calculating an empirical semivariogram as required in (4.40), the plot contains a large amount of strongly scattering points and it becomes impossible to guess an analytical variogram using the recommendations in Sect. 4.3.2: in Fig. 4.24, a model cannot be devined from plot (f) despite this empirical variogram having been calculated from only 39 observations shown in Fig. 4.16.

An isotropic empirical semivariogram calculated using (4.40) is compared with estimates calculated using (4.11) and Fig. 4.8 in the following experiment. Isotropic empirical semivariograms are calculated from the residuals

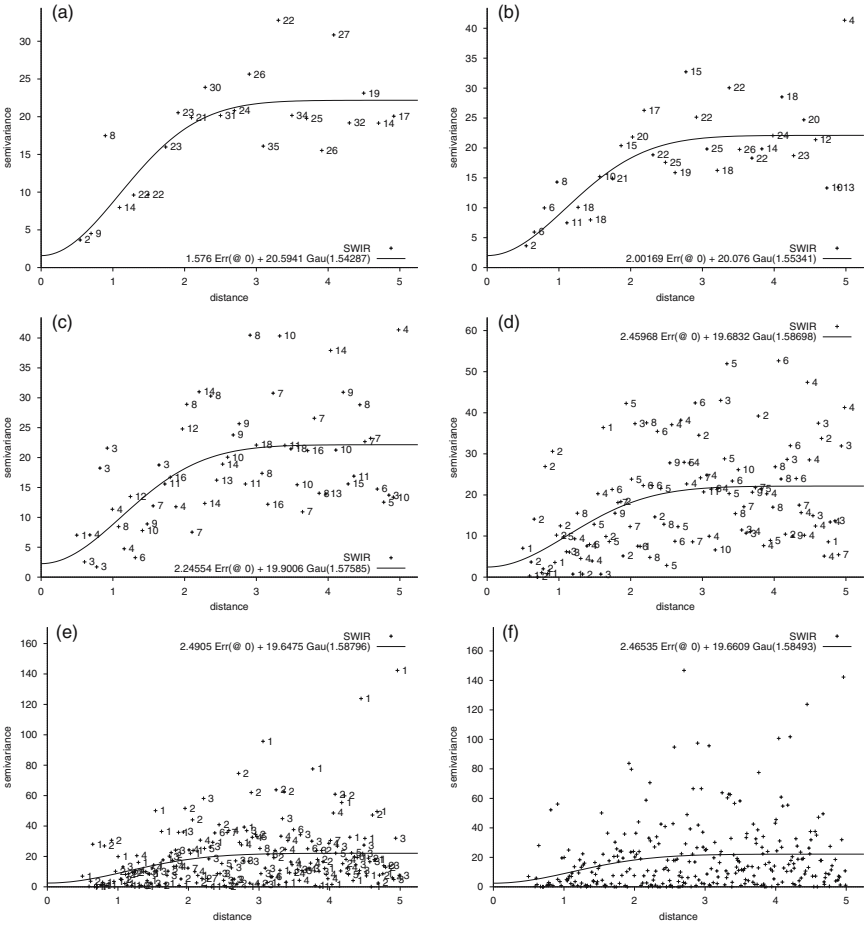


Fig. 4.24. Empirical semivariograms $\hat{\gamma}(h)$ and Gaussian models in $(\mathrm{Wm}^{-2})^2$ estimated as in Fig. 4.28 but using using widths 0.20^0 (a), 0.15^0 (b), 0.08^0 (c), 0.04^0 (d), 0.02^0 (e) and 0.1^0 (f) for the distance classes.

of a first order trend surface for the SWIR climatologies in Fig. 4.16, and Gaussian analytical variograms are fitted using generalised least squares for distance classes 0.25^0 (in Fig. 4.28) as well as 0.2^0, 0.15^0, 0.08^0, 0.04^0, 0.02^0 and 0.01^0 (in Fig. 4.24) wide. The fitted models have very similar parameter estimates. The small differences in the estimated parameters suggest that the model is reconcilable with the covariance function of the spatial random function from which a realisation has been observed. This experiment also demonstrates that an analytical variogram can be guessed from empirical variograms calculated for distance classes wider than 0.04^0. Consequently, in case of the SWIR climatologies in Fig. 4.16, estimates for the empirical

variogram calculated using (4.11) and Fig. 4.8 are preferred to the estimates obtained applying (4.40).

As a general conclusion, it is recommended to calculate an empirical semivariogram as required in (4.11) by partitioning the (x^*, y^*)-plane, as plotted on the right in Fig. 4.8, into classes of direction and distance with different widths. Thereafter, for these partitions, variogram models are guessed and estimated. Approximately identical estimates, if obtained at all, are then assumed to be reconcilable with the second moment function of the random function under analysis.

4.6.2 Support

The observations for the soil parameters in the "Strassenacker" plot (e.g., the K_2O values as plotted in Fig. 4.9) are obtained from a chemical analysis of probes with diameter 5 cm in the first 20 cm of soil, and, consequently, the observed values are averages over approximately 400 cm^3 of soil (in three dimensions) or over approximately 20 cm^2 of the surface (in two dimensions). The measurements of the "Strassenacker" yields, both grain and biomass, (e.g., the grain yields as plotted in Fig. 4.13) are averages over a square with side 1 m. The precipitation values measured on October 5, 1992 at the stations shown in Fig. 4.2 are sums accumulated over 24 hours and covering the catchment areas of the rain gauges having a diameter of a few cm. The climatological yearly means of SWIR at the stations in Fig. 4.16 are means over the period with measurements and spanning the active plate of the pyranometer with a diameter of a few mm. However, these observations are assumed to be point (in an idealised sense) observations for the variogram estimation (in the remarks concluding Sect. 4.3.1) and for the interpolation of these random functions (in the introduction to Sect. 4.4). Does neglecting the physical dimensions associated with a measurement have an effect on the estimations?

Observations of a (spatial) random function are usually obtained from a physical sample associated with a period (in time) and a length, area or volume (in space). This period, length, area or volume is called support in Geostatistics. (4.41)

Using definition (4.41), it is demonstrated in Fig. 4.25 that the histogram, empirical variance and empirical covariance functions of the turbulent wind speed as plotted in Fig. 2.1 undergo changes when the support is increased: the empirical variance decreases and, accordingly, the sill of the variogram decreases as the time series is assumed to be stationary. Consequently, the support has an influence on an interpolation using an estimated variogram. This influence can be accounted for in the kriging interpolators, if required, as described in [35].

For example, when the supports are compared with the ranges of the variogram estimates, assuming point observations is:

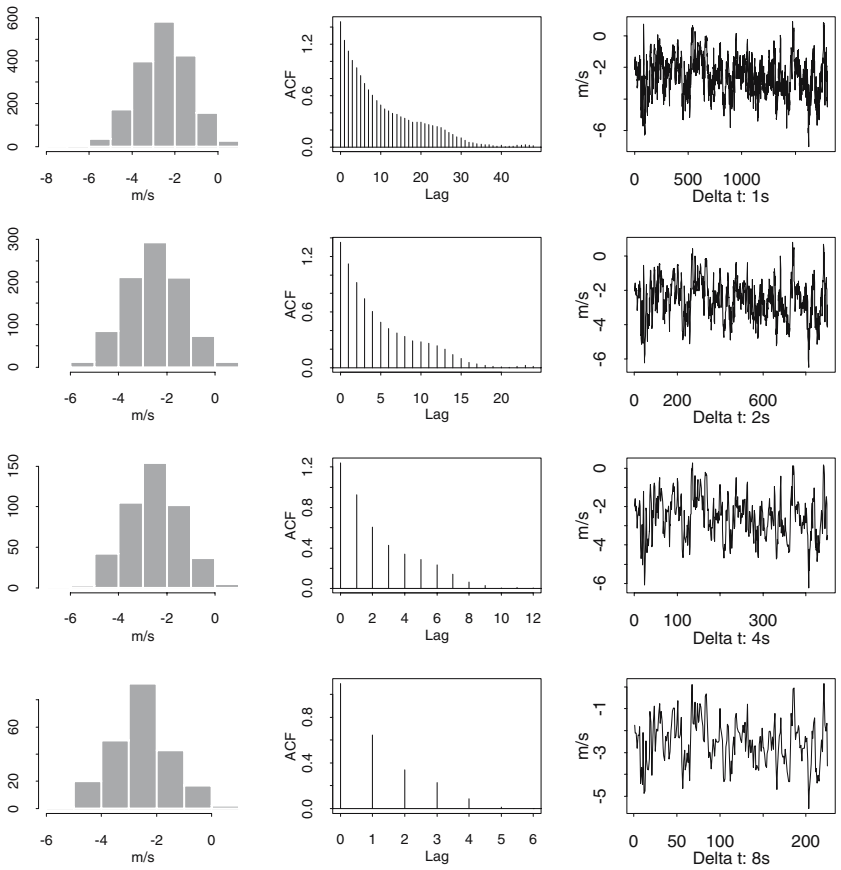


Fig. 4.25. Second moment and support. Histograms, empirical covariance functions and plots of wind speeds sampled from the wind speeds in Fig. 2.1 (on the top) for supports (i.e., time intervals) $\Delta t = 2\ s$, $\Delta t = 4\ s$ and $\Delta t = 8\ s$ (from top to bottom). (The time series for supports $\Delta t = 2, 4, 8$ are averages calculated from the original time series with $\Delta t = 1$). Wind speeds in ms^{-1} (meter per second) and time in s (seconds).

1. an excellent approximation when the empirical semivariogram for the SWIR climatologies is estimated in Sects. 4.3.4, 4.5.3 and 4.6.3
2. a very good approximation when the empirical semivariogram for the precipitation measurements is calculated in Sect. 4.6.5
3. a good approximation when the empirical semivariogram for K_2O in the "Strassenacker" plot is calculated in Sect. 4.3.1, and
4. a sufficing approximation when the empirical semivariogram for the grain yield in the "Strassenacker" plot is calculated in 4.3.3.

4.6.3 Measurement Errors

Every measurement is subject to error. There are random and systematic errors as defined in Sect. 1.1. Random errors cannot be corrected, but their statistical structure can be taken into account when the data are used. For example, when interpolating a spatial random function $Z(x)$ using observations $z(x_i)$ which are afflicted with errors $e(x_i)$, then $e(x_i)$ having certain advantageous properties can be incorporated in the kriging systems. Properties of $e(x_i)$ as given in (4.42) are discussed in this subsection.

Let $Z(x)$ be a spatial random function and $z(x_i)+e(x_i)$ be observations of a realisation of $Z(x)$ at points x_i in D, $i=1,\dots,N$, $e(x_i)$ being an error in the measurement: the sum $Z(x)+e(x)$ of the random variables $Z(x)$ and $e(x)$ is observed. The measurement error $e(x)$ is:

1. *random, provided that* $\mathrm{E}e(x_i)=0$ *(as in (1.14)),*
2. *not correlated with $Z(x)$ (except at the x_i), provided that* $\mathrm{Cov}\big(Z(x),e(x_k)\big)=0$, $x\neq x_k$,
3. *stationary, provided that* $\mathrm{Cov}\big(e(x_i),e(x_k)\big)=c_e(x_i-x_k)$, *and*
4. *white noise, provided that* $\mathrm{Cov}\big(e(x_i),e(x_k)\big)=0$ *for* $x_i\neq x_k$. (4.42)

When a spatial random function $Z(x)$ is interpolated from observations $z(x_i)+e(x_i)$, as defined in (4.42), then the error has an influence on the interpolator: when $z(x_i)+e(x_i)$ (and not x_i) is measured, then the interpolator becomes $\widehat{Z}(x_0)=\sum_{i=1}^{N}\lambda_i\big(Z(x_i)+e(x_i)\big)$. Since $\widehat{Z}(x_0)$ is required to be not biased in (4.13), $\mathrm{E}\widehat{Z}(x_0)=\sum_{i=1}^{N}\lambda_i\mathrm{E}Z(x_i)+\sum_{i=1}^{N}\lambda_i\mathrm{E}e(x_i)=\mathrm{E}Z(x_0)$ is obtained.

The second term in this sum becomes identically zero, $\sum_{i=1}^{N}\lambda_i\mathrm{E}e(x_i)=0$, provided that the measurement error is random as required in (4.42,1). Consequently, under this condition, $\widehat{Z}(x_0)$ is not biased, and, making use of the favourable properties (e.g., intrinsic stationarity) of $Z(x)$, the interpolation weights λ_i can be obtained by solving the kriging equations. These equations (e.g., as in (4.26)) contain either the covariances or the variograms between the interpolation point x_0 and the points with observations x_i. If the measurement error is reconcilable with (4.42,2) then, in Problem 4.20, (4.43) and (4.44) as well as (4.45) and (4.46) are arrived at.

$$\begin{aligned}
2\gamma_{Z+e}(x_i,x_k) &= 2\gamma\big(Z(x_i)+e(x_i),Z(x_k)+e(x_k)\big)\\
&= \big(\mathrm{Var}Z(x_i)+\mathrm{Var}Z(x_k)+\mathrm{Var}e(x_i)+\mathrm{Var}e(x_k)\big)\\
&\quad +2\mathrm{Cov}\big(Z(x_i),e(x_i)\big)+2\mathrm{Cov}\big(Z(x_k),e(x_k)\big)\\
&\quad -2\mathrm{Cov}\big(Z(x_i),Z(x_k)\big)-2\mathrm{Cov}\big(e(x_i),e(x_k)\big) \qquad (4.43)\\
2\gamma_{Z+e}(x_i,x_0) &= 2\gamma\big(Z(x_i)+e(x_i),Z(x_0)\big)\\
&= \big(\mathrm{Var}Z(x_i)+\mathrm{Var}Z(x_0)+\mathrm{Var}e(x_i)\big)
\end{aligned}$$

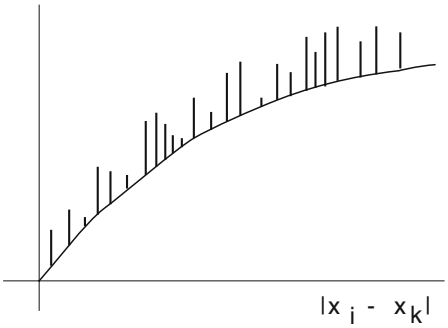


Fig. 4.26. Variogram $2\gamma_{Z+e}(x_i, x_k)$ of a stationary random function $Z(x)$ whose observations $z(x_i)$ are afflicted with a non-stationary error $e(x_i)$. The variation due to the measurement error is superposed to the variogram of $Z(x)$.

$$+2\mathrm{Cov}\big(Z(x_i), e(x_i)\big) - 2\mathrm{Cov}\big(Z(x_i), Z(x_0)\big) \quad (4.44)$$

$$\begin{aligned} c_{Z+e}(x_i, x_k) &= \mathrm{Cov}\big(Z(x_i) + e(x_i), Z(x_k) + e(x_k)\big) \\ &= \mathrm{Cov}\big(Z(x_i), Z(x_k)\big) + \mathrm{Cov}\big(e(x_i), e(x_k)\big) \quad (4.45) \\ c_{Z+e}(x_i, x_0) &= \mathrm{Cov}\big(Z(x_i) + e(x_i), Z(x_0)\big) = \mathrm{Cov}\big(Z(x_i), Z(x_0)\big) \quad (4.46) \end{aligned}$$

If, in addition to assumptions (4.42,1,2) it is assumed that $Z(x)$ is stationary, then the terms $\mathrm{Cov}\big(Z(x_i), Z(x_k)\big) = c_Z(x_i - x_k)$ and $\mathrm{Cov}\big(Z(x_i), Z(x_0)\big) = c_Z(x_i - x_0)$ in the above expressions for the variograms and covariances of the sum $Z(x_i) + e(x_i)$ in (4.43), (4.44), (4.45) and (4.46) become stationary; the remaining covariances in these expressions are, however, not stationary. Consequently, the empirical variogram is not an estimator for the variogram $2\gamma_{Z+e}(x_i, x_k)$ since the measurement error is non-stationary, as shown in Fig. 4.26. This result also applies to the empirical covariance function.

In Problem 4.21 it is easily derived that, under assumptions (4.42,1,2,3), $2\gamma_{Z+e}(x_i, x_k) = 2\gamma_Z(x_i - x_k) + 2\gamma_e(x_i - x_k) = 2\gamma_Z(h) + 2\gamma_e(h)$.

If the measurement error is reconcilable with all assumptions as in (4.42), and if $Z(x)$ is stationary, then $c_{Z+e}(x_i, x_k) = c_Z(x_i - x_k) + c_e(0)$. Under these assumptions, the variogram and the covariances of the spatial random function, together with the measurement error, behave as shown in Fig. 4.27.

Under assumptions (4.42,1,2), the variograms in (4.43) and (4.44) as well as the covariances in (4.45) and (4.46) have been arrived at. Hence, a kriging system for the interpolation of observations afflicted with an error as defined in (4.42) can be obtained, provided that (i) $Z(x)$ is stationary, i.e., $\mathrm{E}Z(x) = \mu_Z(x) = \mu_Z$ and $c_Z(y, x) = c_Z(y - x) = c_Z(h)$ for all x and y in D, and (ii) the measurement error is random and not correlated with the observations as required in (4.42,1,2). Under these assumptions and proceeding as in the derivation of the simple kriging system (4.22), the universality condition $\sum_{i=1}^{N} \lambda_i = 1$ is obtained, which is then used as a constraint to minimise the

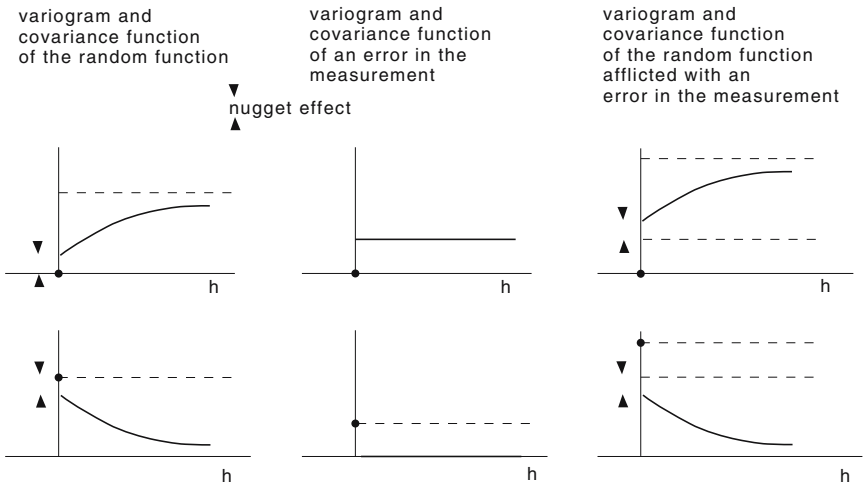


Fig. 4.27. Variogram $2\gamma_{Z+e}(x_i, x_k)$ and covariance function $c_{Z+e}(x_i, x_k)$, on condition that all properties as defined in (4.42) apply to a measurement error $e(x_i)$ and that $Z(x)$ is stationary.

mean square prediction error $\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) = \sum_{i=1}^{N} \sum_{k=1}^{N} \lambda_i \lambda_k \big(c_Z(x_i - x_k) + \mathrm{Cov}(e(x_i), e(x_k))\big) - 2\sum_{i=1}^{N} c_Z(x_k - x_0) + \mathrm{Var}Z(x_0)$.

Let $Z(x)$ be a stationary random function with observations $z(x_i) + e(x_i)$ as required in (4.42,1,2) and let $\widehat{Z}(x_0) = \sum_{i=1}^{N} \lambda_i \big(Z(x_i) + e(x_i)\big)$ be an interpolator as in (4.13). Then the weights λ_i are obtained when solving:

1. $\sum_{i=1}^{N} \lambda_i \Big(c_Z(x_i - x_k) + \mathrm{Cov}\big(e(x_i), e(x_k)\big)\Big) - m = c_Z(x_k - x_0)$, $k = 1, \ldots, N$, *the optimality conditions, and* (4.47)
2. $\sum_{i=1}^{N} \lambda_i = 1$, *the universality condition.*
3. $\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) = \mathrm{Var}Z(x) - \sum_{k=1}^{N} \lambda_k c_Z(x_k - x_0) + m$ *is the mean square prediction error (interpolation variance).*

An interpolator obtained from the solution of (4.47) is not exact, i.e., it is not implied that $\widehat{Z}(x_k) = z(x_k) + e(x_k)$. This is shown by assuming that the interpolator is exact and then, from this assumption, deriving a contradiction. If $x_0 = x_k$, then $\lambda_i = 0$ for all $i \neq k$ and $\lambda_k = 1$ and, further, for equation k in (4.47), $1 \times \mathrm{Cov}\big(Z(x_k), Z(x_k)\big) + 1 \times \mathrm{Cov}\big(e(x_k), e(x_k)\big) - m = \mathrm{Cov}\big(Z(x_k), Z(x_k)\big)$ is obtained which implies $m = \mathrm{Cov}\big(e(x_k), e(x_k)\big)$. Under these (identical) assumptions, for equation i in (4.47), $i \neq k$, $1 \times \mathrm{Cov}\big(Z(x_k), Z(x_i)\big) + 1 \times \mathrm{Cov}\big(e(x_k), e(x_i)\big) - m = \mathrm{Cov}\big(Z(x_k), Z(x_i)\big)$ is obtained implying $m = \mathrm{Cov}\big(e(x_k), e(x_i)\big)$. $m = \mathrm{Cov}\big(e(x_k), e(x_i)\big)$ is, however, in contradiction to $m = \mathrm{Cov}\big(e(x_k), e(x_k)\big)$ derived above.

Gstat interpolates as required in (4.47) provided that an error in the measurements (i) satisfies (4.42,1,2) and (ii) is known for all points with observations x_i. The intricacy of the covariance matrix in (4.47) is strongly reduced when the error is assumed to stem from a white noise process as required in (4.42,4): (4.47) becomes (4.22) except for the diagonal of the covariance matrix, where $\mathrm{Var}Z(x) + \mathrm{Var}e(x)$ is substituted for $\mathrm{Var}Z(x)$. In this case, the measurement error is accounted for in the variogram model when interpolating with Gstat.

For example, the SWIR climatologies plotted in Fig. 4.16 are afflicted with errors due to deficiencies in the maintenance of the instruments and/or the data acquisition systems, as discussed in Sect. 1.1. These climatologies are means of the time series of yearly values of SWIR for each station and are available, together with their empirical standard deviations, in file `/pfad/globradnorde.eas` as described in Sect. 4.3.4. The empirical variances of a few stations are also given in Table 3.4. These estimates for the second moment, however, are not estimates for the measurement error, because, as demonstrated in Sect. 3.1.1, the variability in SWIR yearly values is owed to three sources: a possible trend, the interannual variability and errors in the measurement.

A fourth source of variability is taken into account when time series at neighbouring stations in a region are analysed. In Fig. 3.8 and Table 3.6 it is shown, for example, that the variability of SWIR yearly values due to station effects are not to be neglected. In this example, the total variation of yearly SWIR values measured at stations in grid box $\big((10^0\mathrm{E}, 50^0\mathrm{N}), (15^0\mathrm{E}, 52.5^0\mathrm{N})\big)$ amount to $(65.9\ \mathrm{Wm}^{-2})^{-2}$; of these, approximately $(45\ \mathrm{Wm}^{-2})^{-2}$ are shown to be due to both the interannual variability and the measurement error in the comments to Table 3.6, and consequently, assuming independent sources of variability, $(20\ \mathrm{Wm}^{-2})^{-2}$ are due to the trend and station effects.

Approximately $(20\ \mathrm{Wm}^{-2})^{-2}$ is also obtained at the sill of the empirical semivariogram for the SWIR climatologies in Fig. 4.21, from which approximately $(40\ \mathrm{Wm}^{-2})^{-2}$ is obtained as an estimate for the variance of the SWIR climatologies plotted in Fig. 4.16. Obviously, this variance is due to the trend, the part of the interannual variability that had not been captured when the climatologies were calculated, the station effects and the measurement error. For simplicity, this variance is assumed to be due solely to the measurement error and the station effects, and both sources are assumed to be independent of each other for the following variogram estimation and interpolation. It is further assumed that the error is reconcilable with (4.42,1,2,3,4). In this case, the SWIR climatologies are interpolated by applying universal kriging with the measurement error incorporated in the variogram model.

In Gstat, a measurement error as required in (4.42,1,2,3,4) is incorporated in a variogram estimation by specifying an error term in a variogram model. For example, when an error term is included in the variogram model as plotted in Fig. 4.18 (a), then $2\gamma(0) = 0$, enforced when estimating the model

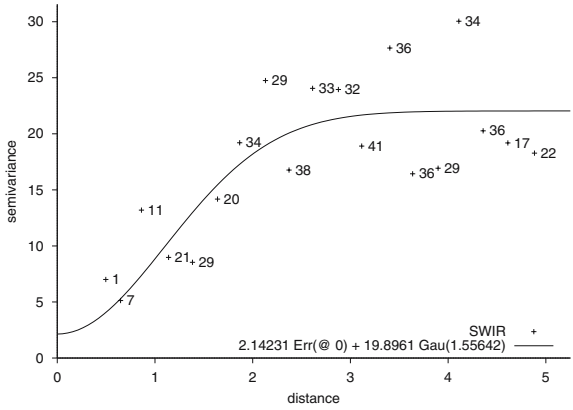


Fig. 4.28. Empirical semivariogram in $(\mathrm{Wm}^{-2})^2$ as estimated in Fig. 4.21 but using a Gaussian model with a term for the measurement error (in $(\mathrm{Wm}^{-2})^2$) for the universal kriging of SWIR climatological yearly means at stations in Fig. 4.16.

as plotted Fig. 4.21 with generalised least squares, is no longer required: the model is changed to

```
variogram(SWIR): 1.0 Err() + 19.935 Sph(2.885)
```

and generalised least squares estimations are repeated until the estimate becomes stable: with

```
variogram(SWIR): -4.185 Err(@ 0) + 25.531 Sph(2.839)
```

a negative variance for the measurement error is estimated. The negative error term is due to the spherical model not being well enough suited for these data. A better model for SWIR that undergoes very small changes within short distances (e.g., identical radiation climates are assumed for the pairs of stations in short distances of each other given in Problem 1.11 when checking for homogeneity in Fig. 2.19) is an analytical variogram that is smooth in $|h| = 0$. When a Gaussian model, for example, is fitted using generalised least squares, the result plotted in Fig. 4.28 is obtained.

Under the assumptions stipulated above and with a Gaussian variogram model, the measurement error in the SWIR climatologies plotted in Fig. 4.16 is estimated to be 4.3 $(\mathrm{Wm}^{-2})^2$, close to the 5 $(\mathrm{Wm}^{-2})^2$ obtained for the measurement error in SWIR yearly values at the stations in Fig. 4.16 using the relative error (2% of the mean) as estimated in [58]. Admittedly, to account for the above simplification, the result thus obtained includes contributions due to the trend and that part of the interannual variability which had not been captured when the climatologies were calculated.

Using the Gaussian variogram model with an error term as estimated in Fig. 4.28 and universal kriging, the SWIR climatologies plotted in Fig. 4.16 are interpolated with the Gstat command file

```
data(SWIR): '/path/swirnorde.eas',x=1,y=2,v=4,d=1;
```

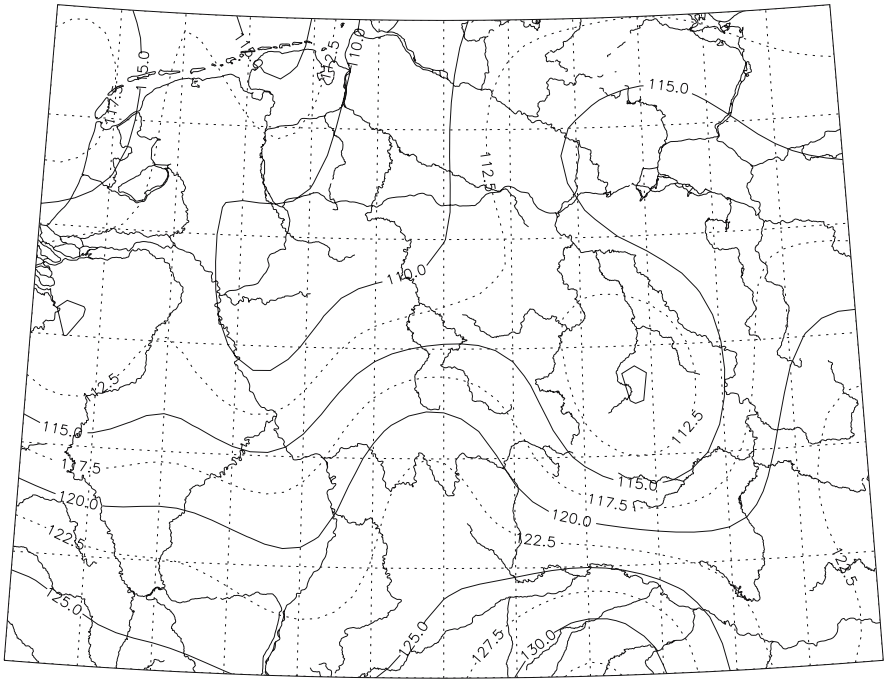


Fig. 4.29. SWIR climatology in Wm^{-2} interpolated from the observations in Fig. 4.16 using universal kriging and the Gaussian model with the measurement error term as plotted in Fig. 4.28.

```
variogram(SWIR): 2.14231 Err(@ 0) + 19.896 Gau(1.55642);
set cutoff = 5;
set fit = 1;
set gls_residuals = 1;
set width = 0.25;
data(): 'swirnorde.grid.eas', x=1, y=2;
set output = 'swirnordeerr.rep.eas';
```

and the results in Figs. 4.29 and 4.30 are obtained.

These estimates are then compared with the estimates in Figs. 4.22 and 4.23 which were obtained by universal kriging with the variogram model without error term as plotted in Fig. 4.21. The estimates in Figs. 4.29 (using a variogram model with an error term) and 4.22 (using a variogram model without error term) are close to each other in the interior of the region with observations. Outside this region, however, small differences are seen when the estimates are compared.

The interpolation variances in Fig. 4.30 (using a variogram model with an error term) are smaller than those in Fig. 4.23 (using a variogram model

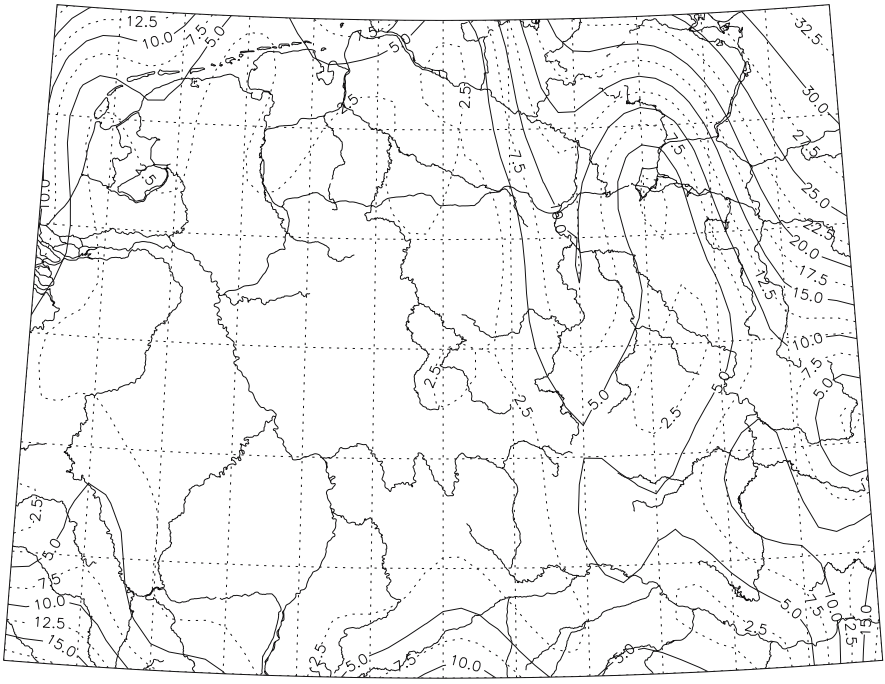


Fig. 4.30. Variance in $(\mathrm{Wm}^{-2})^2$ of the interpolation of SWIR climatologies in Fig. 4.29.

without error term): when interpolating using `2.14231 Err(@ 0) + 19.896 Gau(1.55642)` as plotted in Fig. 4.28, the measurement error is not included in the interpolation variance. Consequently, when the error is added to the interpolation variance in Fig. 4.30, an interpolation variance very close to that plotted in Fig. 4.23 is obtained.

When interpolating the SWIR climatology in Fig. 4.29 using the Gaussian variogram model with an error term as estimated in Fig. 4.28, the estimate is calculated, at every interpolation point in the region bounded by 4^0 and 16^0 (eastern) longitude and by 48^0 and 54^0 (northern) latitude, using all observations plotted in Fig. 4.16, and, therefore, the distances between the interpolation points and the points with observations (the interpolation distances) are often larger than 5^0, being (approximately) the maximal distance for which an empirical semivariogram value is calculated, as shown in Fig. 4.28. Thus, the interpolation distances are twice as large as the maximal distance recommended, in Fig. 4.11, when interpolating with a Gaussian model ($1.73 \times a = 2.69^0$ longitude or latitude, $a = 1.55462^0$ in Fig. 4.28).

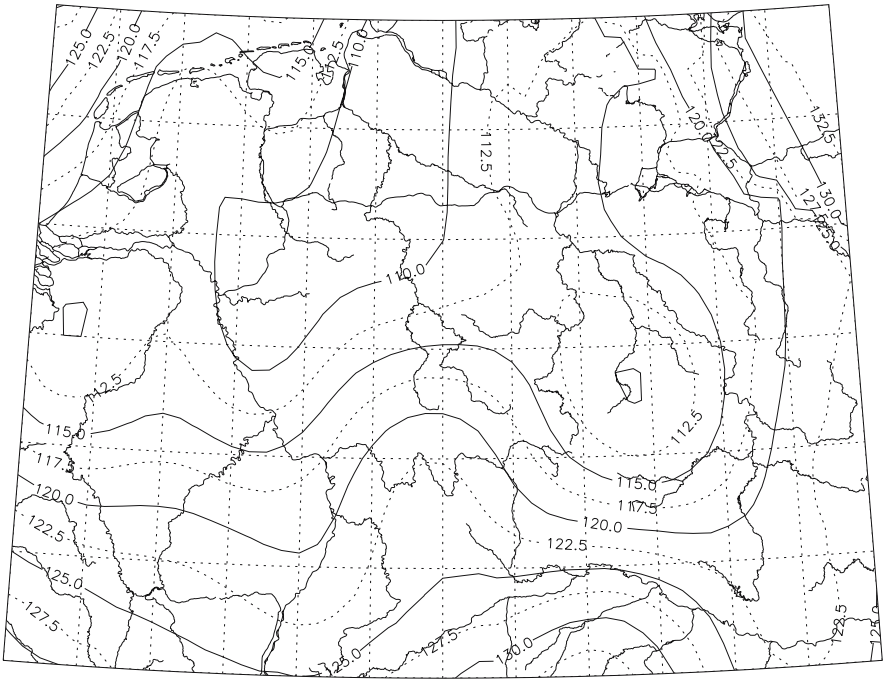


Fig. 4.31. SWIR climatology in Wm^{-2} locally ($|h| \leq 4^0$) interpolated from the observations in Fig. 4.16 using universal kriging and the Gaussian model with the measurement term as plotted in Fig. 4.28.

4.6.4 Local Interpolation

An interpolation with too large interpolation distances (as found to be inadequate, in the remarks concluding Sect. 4.6.3, for the universal kriging of the SWIR climatoloy with the Gaussian model in Fig. 4.28) can be circumvented by a local interpolation using a moving interpolation neighbourhood. For example, results as in Figs. 4.31 and 4.32 are obtained when the SWIR climatologies are interpolated, applying universal kriging with the variogram model in Fig. 4.28, at each interpolation point x_0, from those observations x_i with interpolation distance $|h| = |x_0 - x_i| \leq 4^0$ longitude or latitude, i.e., performing a local (as defined in (2.55), more specifically: not calculated from all observations) interpolation with a circular interpolation neighbourhood of radius 4^0 longitude or latitude.

For this interpolation, the `data()` expression in the Gstat command file as used in Sect. 4.6.3 is modified by adding the `radius` option

```
data(SWIR): '/path/swirnorde.eas', x=1,y=2,v=4,d=1,radius=4;
```

prior to interpolating. Despite this constraint, the estimates in Figs. 4.31 and 4.32 are calculated using larger interpolation distances than recommended

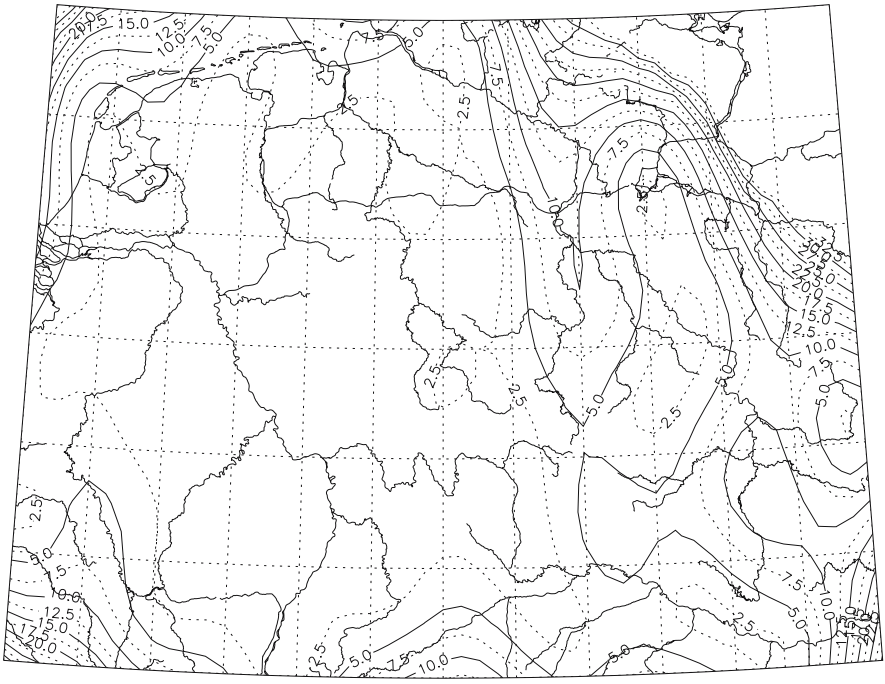


Fig. 4.32. Variance in $(\mathrm{Wm}^{-2})^2$ of the interpolation of SWIR climatologies in Fig. 4.31.

in Fig. 4.11 for a Gaussian model ($1.73 \times a = 2.69^0$ longitude or latitude, $a = 1.55462^0$ in Fig. 4.28). The Gaussian model in Fig. 4.28 is however safely within the estimates for the empirical variogram for $|h| \leq 5^0$, and, therefore, interpolation distances up to 4^0 longitude or latitude are arguable.

When the interpolations in Figs. 4.29 and 4.31 are compared, only small deviations in the interpolated SWIR climatologies are noticed in the interior of the area with observations; at the margins however, the deviations become substantial. Over the North Sea and close to the Oder river in the northeast, a too large SWIR climatology results, in Fig. 4.31, from the local interpolation: a SWIR climatology as obtained in Fig. 4.29 is more realistic.

If a stationary spatial random function $Z(x)$ or a function with a non-constant expectation function $\mu_Z(x)$ and stationary residuals is estimated, using simple or ordinary kriging, at interpolation points x_0 in larger distances from the points with observations x_i (usually at the margins of the area with observations), then, with an increasing interpolation neighbourhood, the interpolator $\widehat{Z}(x)$ comes closer to $\mu_Z(x)$. Hence, at these x_0, the interpolation variance of a local interpolation is larger than those of a global interpolation using all x_i. In Fig. 4.32 for example, at the margins of the area

with observations as seen in Fig. 4.16, the interpolation variance of the locally interpolated SWIR climatology (in Fig. 4.32) is larger than the variance of the globally interpolated SWIR climatology (in Fig. 4.30).

If the SWIR climatology is universally kriged in interpolation neighbourhoods with radius 4^0 longitude and latitude, then a variogram model (as plotted in Fig. 4.28) and a linear model for the trend are used that are obtained from global generalised least squares estimates. These estimates are global (as defined in (2.55), more specifically: calculated from all observations) because no interpolation neighbourhood is defined in the `data(SWIR): '/path/swirnorde.eas',x=1,y=2,v=4,d=1;` line in the Gstat command file in Sect. 4.6.3. In this case, one faces the choice between global interpolation (with results in Figs. 4.29 and 4.30) or local interpolation (with results in Figs. 4.31 and 4.32).

Only a local interpolation is, however, feasible in cases where the moment functions are estimated for distances $|h| \leq d$ solely, with d much smaller than the shorter side of the area with observations, though globally using all pairs of observations $\big(z(x_i), z(x_k)\big)$ with $|x_i - x_k| = |h| \leq d$. In Sect. 4.3.3, for example, an empirical variogram for the grain yield in the "Strassenacker" plot is estimated for $|h| = |x_i - x_k| \leq 80$ m, being only approximately half of 140 m, i.e., the shorter side of the area with observations as plotted in Fig. 4.13, to obtain the estimates as plotted in Fig. 4.15. This is to be accounted for in Problem 4.25 where the grain yield in the "Strassenacker" plot is interpolated locally using the variogram model obtained in Fig. 4.15.

The choice between local or global interpolation becomes easier with the guideline given in (4.48).

A local (as defined in (2.55), more specifically: not using all observations in the interpolator) interpolation of a spatial random function $Z(x)$ with expectation function $\mu_Z(x)$ is recommended when:

1. *$Z(x)$ is only locally (intrinsically) stationary,*
2. *a non-constant $\mu_X(x)$ cannot be estimated globally using generalised least squares,*
3. *enough observations are available, and*
4. *a large number of observations is available and the kriging system becomes too large for a global estimation.*

(4.48)

Gstat allows for defining of interpolation neighbourhoods that depend on the interpolation point x_0 and the points with observations x_i. A local interpolation in neighbourhoods depending on x_0 and x_i allows for challenging unfavourable constellations of the interpolation point and the points with observations which occur, for example, when the density of the x_i changes. Hence, areal precipitation is estimated, in Figs. 4.34 and 4.34, using local simple kriging with an interpolation neighbourhood that depends on the x_i that are neighbours to x_0.

4.6.5 Spatial Averages

Under the assumptions stipulated in the introduction to Sect. 4.5, with simple, ordinary or universal kriging as defined in (4.22), (4.26) or (4.35) interpolators that are linear in their weights and the observations are obtained. For the derivation of these kriging systems it is assumed implicitly that the interpolation points and the points with observations are idealised points in the mathematical sense. Hence, the estimation of areal means is straightforward, on condition that the means are linear.

Let $Z(x)$ be a spatial random function as defined in (4.6) with x being in the domain D and let $V(x) \subset D$ with centre in x. Then:

1. $Z_V(x) = \big(1/V(x)\big) \int_{V(x)} Z(x')\mathrm{d}x'$ *is the mean of* $Z(x)$ *in* $V(x)$.
2. $Z_V(x) \approx \big(1/n(V)\big) \sum_{j=1}^{n(V)} Z(x'_j)$, x'_j *being* $n(V)$ *points in* $V(x)$.

(4.49)

An areal mean is approximated, in (4.49,2), by interpolating $Z(x)$ for $n(V)$ points x'_j in $V(x)$ and then calculating the arithmetic mean of the estimated values. If the estimates for all interpolation points x'_j are computed from the same points with observations x_i then, in the kriging systems (4.22), (4.26) or (4.35), the matrices remain unchanged as only the right sides undergo changes: the covariances or variograms (in (4.35) together with the values of the base functions for the x'_j) between the x'_j and x_i change. These point-to-point covariances are averaged, in (4.50),

$$\bar{c}_V(x) = \frac{1}{n(V)} \sum_{j=1}^{n(V)} \left(\frac{1}{N(V)} \sum_{i=1}^{N(V)} c(x'_j - x_i) \right) \tag{4.50}$$

to a point-to-area covariance which is then substituted for the right side in the kriging systems (together with, in (4.35), the averages of the base functions for the x'_j). The systems thus modified are then solved to obtain directly the areal mean. This technique is restricted to the estimation of linear functionals (as introduced in the remarks to (1.1)) of a spatial random function.

For example, areal precipitation means for October 5, 1992 as seen in Fig. 4.1 are calculated, using the SYMAP algorithm (a deterministic interpolation method introduced in Sect. 4.1.1), from the observations at the stations as plotted in Fig. 4.2. What is obtained when areal estimates are calculated from the same observations but using optimal interpolation?

Applying (4.48,4) to Fig. 4.2 showing the rain gauge stations, it becomes obvious that a local interpolation is called for. For simplicity, the areas for estimating the mean precipitation are assumed to be cells with sides $\Delta x^* = 0.3^0$ and $\Delta y^* = 0.22^0$ centred in the interpolation points shown in Fig. 4.3. Additionally, precipitation for October 5, 1992 is assumed to be a spatial random function satisfying locally the intrinsic hypothesis.

Then, empirical semivariograms are calculated for $|h| = |x_i - x_i| \leq 3^0$ using distance classes 0.1^0 wide and for directions 0^0, 45^0, 90^0 and 135^0

(not shown). From these variograms it is concluded that precipitation for October 5, 1992 is not isotropic and that the anisotropy can be captured geometrically using an ellipse with the longer axis in direction 90^0, i.e., from west to east, and the shorter axis half as long, i.e., an anisotropy ratio 0.5 is obtained. These variograms also jump in $|h| = 0$, accounted for by fitting a preliminary spherical model with a nugget effect. This model is considered to be preliminary because, in Problem 4.22, the jump is assumed to be due to a measurement error. Under these assumptions, the anisotropic empirical semivariogram as shown in plot (f) in Fig. 4.33 is estimated for $|h| = |x_i - x_i| \leq 2^0$ given that this distance is assumed to be the largest interpolation distance.

Thereafter, the spherical model with nugget effect and geometric anisotropy obtained above is plotted in empirical semivariograms calculated for directions 0^0, 45^0, 90^0 and 135^0, and direction classes 45^0 wide, to obtain the diagnostics in Fig. 4.33, plots (a), (b) (c) and (d). Obviously, the relatively good overall fit (plot (e) in Fig. 4.33) becomes unfavourable for directions 0^0 and 135^0 (plots (a) and (d) in Fig. 4.33). Neglecting those directions with a lack of fit (item 1. below), the variogram model as in plot (e) in Fig. 4.33 is used to estimate the areal precipitation for October, 5, 1992.

In Gstat, the precipitation observations are interpolated with the following Gstat command file obtained by editing the results from the variogram estimation, i.e., the model in Fig. 4.33 (e). The following changes are made:

1. In the line with `data(precipit)`, the interpolation neighbourhood is defined: observations from at least 10 and at most 30 stations, being in a generalised distance (as defined in the variogram model) of less than 2^0 from the interpolation point, are used to calculate the interpolator. Due to these restrictions, the generalised interpolation distances are smaller than 0.5^0 in regions with a large density of rain gauge stations (Figs. 4.2 and 4.3), and the estimates are calculated using a model that fits well for all directions as is obvious from Fig. 4.33. In regions with a small density of rain gauge stations, the above restrictions ensure that an estimate is computed from at least 10 observations with generalised distances smaller than 2^0 from the interpolation point, otherwise no estimate is obtained.
2. The file with the interpolation grid as specified in the line with `data()` defines the interpolation points as shown in Fig. 4.3.
3. The results are written to file `precipit.rep`.
4. In Geostatistics, a block is a volume in the form of a parallelepiped in a deposit that is to be mined (with the ore extracted to be predicted). Hence, "block" is used in connection with areal means, e.g., the optimal prediction of spatial means (in areas or volumes) is called *block kriging*. In Gstat, areas with means to be estimated are defined in the line with `blocksize`: the areas have sides 0.3^0 in west-east direction, 0.22^0 in south-north direction and their centre in the interpolation points.

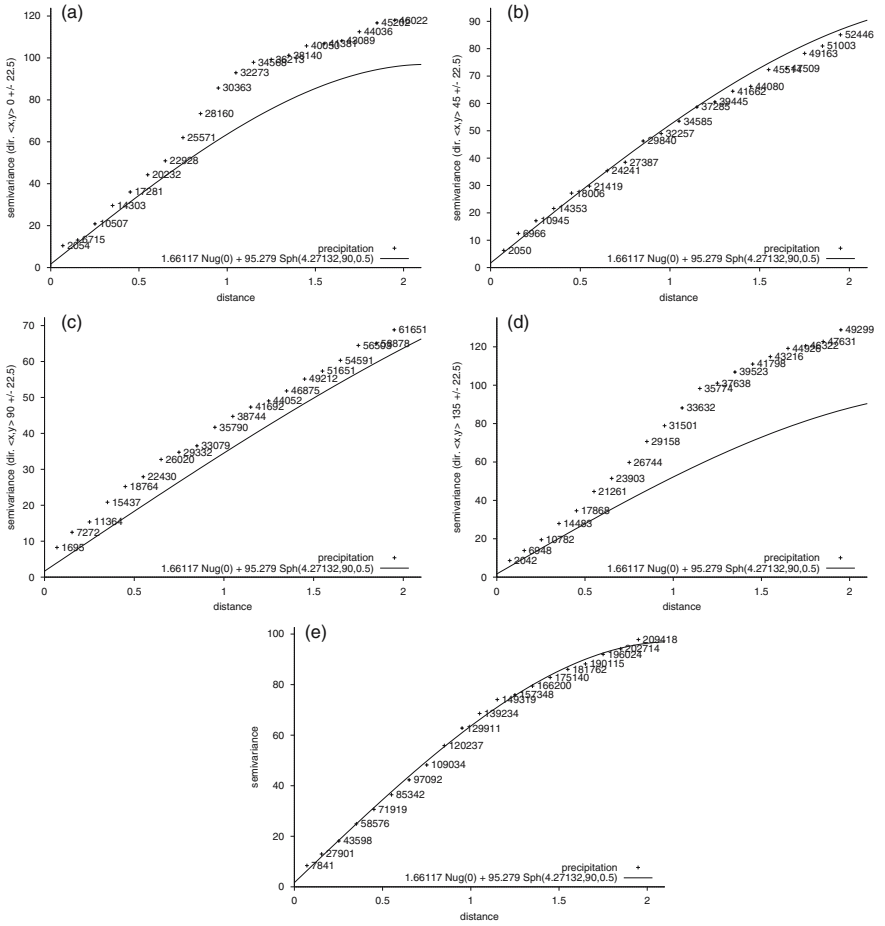


Fig. 4.33. Empirical semivariograms $\hat{\gamma}(h)$ in $(\text{mm})^2$ of the rain gauge measurements for October 5, 1992 at the stations in Fig. 4.2 with cutoff 2^0 and width 0.1^0. (e): $\hat{\gamma}(h)$ for all directions and fitted spherical model with nugget effect and geometric anisotropy; (a), (b), (c) and (d): the same model as in (e) plotted in $\hat{\gamma}(h)$ for the directions 0^0, 45^0, 90^0 and 135^0.

5. `set nblockdiscr=10` describes the discretisation of the areas to be estimated. The mean variogram for each area is calculated as required in (4.50) using 100 (10 in each direction) x'_j. Alternative schemes as described in [105] are also available in Gstat.

so as to obtain Gstat commandfile

```
data(precipit): '/path/ppitoct592.dat', x=2, y=1, v=3, average,
                        radius=2, max=30, min=10;
variogram(precipit): 1.66117 Nug(0) + 95.279 Sph(4.27132,90,0.5);
```

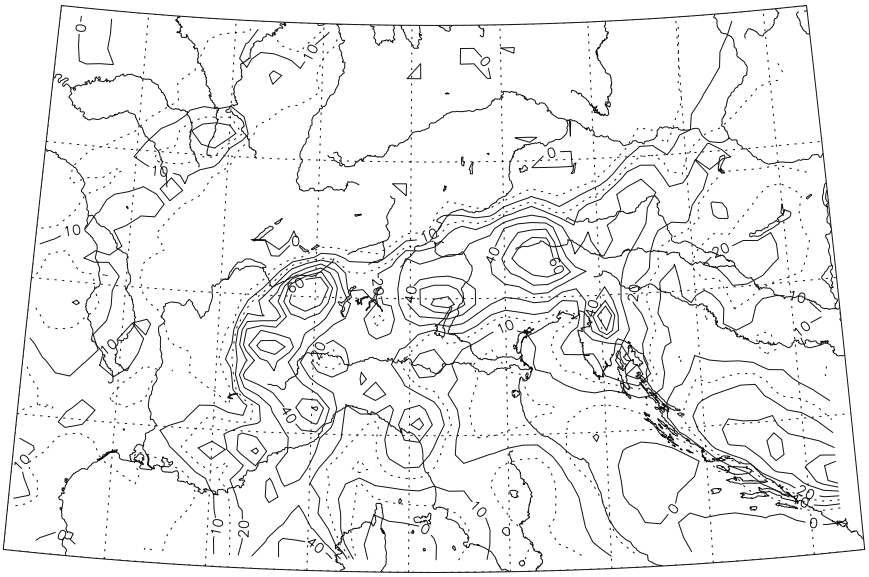


Fig. 4.34. Precipitation in mm for October 5, 1992 and region $\big((2^0\text{E}, 42^0\text{N}), (19^0\text{E}, 50.0^0\text{N})\big)$ interpolated from the observations at the stations in Fig. 4.2 using local block kriging and the variogram model in Fig. 4.33 (e).

```
set cutoff = 2;
set fit = 1;
set width = 0.1;
data(): 'ppitoct592.grid', x=1, y=2;
set output = 'precipit.rep';
blocksize: dx=0.3, dy=0.22; #sides of the area
set nblockdiscr = 10;        #n(V) = 10×10, n(V) as in (4.50)
```

for estimating the areal precipitation for October 5, 1992.

Using this command file, Gstat estimates, for each interpolation point, the mean variogram with a discretisation as defined above from the variogram model in Fig. 4.33 (e), substitutes the estimate for the right side in the ordinary kriging system (4.26), and solving the system, obtains the mean precipitation and its interpolation variance for an area as defined above, as shown in Figs. 4.34 and 4.35.

The block kriging estimate as shown in Fig. 4.34 resembles the SYMAP interpolation as seen in Fig. 4.1 in those regions where both results are available. The interpolation variance, as in Fig. 4.35, depends on the density of the stations with rain gauge observations. The interpolation variance is smaller than 1 $(\text{mm})^2$ in regions with a large density of observations and increases to approximately 5 $(\text{mm})^2$ in northern Italy where observations are sparse. At

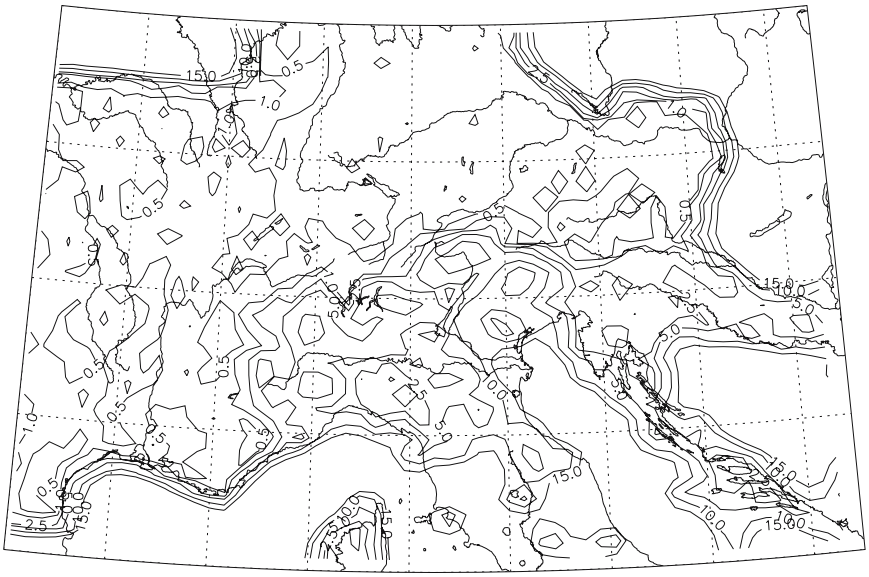


Fig. 4.35. Variance in $(\text{mm})^2$ of the areal precipitation estimates as obtained in Fig. 4.34.

the margin of the area with observations which is shown in Fig. 4.2, interpolation variance increases rapidly and is no longer plotted when larger than 15 $(\text{mm})^2$.

4.6.6 Optimal Interpolation: Summary

When a spatial random function is interpolated using simple, ordinary or universal kriging then anwers to questions (4.51) must be given.

Prior to an interpolation of a spatial random function $Z(x)$ answers to the following questions are required:

1. *Is the interpolation local or global?*
2. *In case of a local interpolation: which interpolation neighbourhood is used?*
3. *In the case of a non-constant drift: Is a random function with a non-constant drift in agreement with assumptions (4.27) and (4.39)?* (4.51)
4. *When estimating spatial means: how is the mean variogram calculated?*
5. *Can a possible measurement error be incorporated in the kriging systems?*
6. *Is a coordinate transform required prior to the interpolation?*

These answers depend often on those given to questions (4.12), prior to estimating the moment functions of the spatial random function under analysis, and also on those to questions (4.51). For example, the empirical variograms in Fig. 4.33 are estimated for generalised distances $|h| \leq 2^0$, 2^0 being the maximal interpolation distance, which is a part of the answer to question (4.51,2). If an uncertainty due to these interdependencies remains then it is recommended to validate an interpolation using resampling methods.

For example, the variogram $2\gamma_Z(h)$ (and, when the expectation function $\mu_Z(x)$ is not constant and universal kriging is called for, the drift $\mu_Z(x)$) is (are) estimated without using observation $z(x_k)$. Thereafter the interpolator $\widehat{Z}(x_k)$ is compared with $z(x_k)$. This procedure is repeated for each observation. It is also possible to construct two subsets from the set of observations. In this case, using one subset, the spatial random function is predicted for the points with observations in the other subset, and vice-versa. The estimates are then compared with the observations.

4.7 Problems

4.1. Interpolate the climatological means of SWIR measured at stations as given in Table 4.1 using weight functions alternative to $g(r) = r^{-1}$, for example $g(r) = r^{-2}$, $g(r) = \mathrm{e}^{-2r}$, and $g(r) = \mathrm{e}^{-r^2}$.

4.2. For Splus users only: in Sect. 4.1.2, the tree-line in the Valais alps is interpolated with Akima's method using $m = 5$, with the result given in Fig. 4.4. Try $m = 7$, $m = 10$, $m = 3$, etc.

4.3. For Splus users only: Interpolate the precipitation measurements for October 5, 1992 in file `/path/ppitoct592.dat` using Akima's method as demonstrated in Sect. 4.1.2. Akima's method fails in this case. Why?

4.4. Show that $\gamma_Z(h) = c_Z(0) - c_Z(h)$, provided that $Z(x)$ is stationary.

4.5. Estimate the empirical variogram of the differences in the pyranometer measurements for 1991 as obtained in the comparison experiment described in Sect. 1.1. The empirical covariances of this time series are estimated in Problem 2.32. The pyranometer differences are available in file `/path/difpyr.eas`.

4.6. Sediment samples with a dry weight of about 20 kg were taken in a gravel deposit [80]. Based on grain size analysis, the hydraulic conductivity of the sediment was estimated using Kozeny's model, which was calibrated with the parameters of two undisturbed samples. These estimates are available in file `/path/gravel2.dat`. Is the hydraulic conductivity (intrinsically) stationary? Estimate its second moment with Gstat.

A spatial sampling interval Δx can be associated with the hydraulic conductivity values because the samples are taken at points in identical distances on a straight line. Are there alternatives to an estimation using Gstat?

4.7. Simulate a realisation of an AR[1] process (X_t) with $N = 1000$ and $a = 0.9$ using (2.13), and plot, as in Fig. 2.11, the empirical covariance function estimated from the simulation. Then draw random samples $x_t^{(i)}$, $i = 1, 2, \ldots, n$, $n = 10, 30, 50, 100, 200, 300, \ldots$ from the simulated values and estimate the empirical variogram of (X_t) using these samples. A random sample of the simulated values obtained in R vector `ar1samp1` is written to file `ar1samp1.dat` using

```
sink("ar1samp1.dat")
for (i in (1:length(ar1samp1)) {cat(i, ar1samp1[i], "#n")}
sink()
```

and, from this file, subsequent to adding the `.eas` header lines, an empirical variogram can be calculated with Gstat.

4.8. Repeat this, i.e., the performance in Problem 4.7, but this time for a random walk process as defined in (2.20).

4.9. Fit an analytical variogram to the empirical variograms obtained in Problems 4.7 and 4.8.

4.10. Lagrange multipliers. The position of the extreme value of a function changes under an additional assumption (acting as a constraint) as shown in Fig. 4.36. Solving equations (4.52) the extreme of function $F(x, y)$ under constraint $f(x, y) = 0$ is obtained, x and y being real.

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}x}\Big(F(x,y) - m\big(f(x,y)=0\big)\Big) &= 0 \\ \frac{\mathrm{d}}{\mathrm{d}y}\Big(F(x,y) - m\big(f(x,y)=0\big)\Big) &= 0 \\ \frac{\mathrm{d}}{\mathrm{d}m}\Big(F(x,y) - m\big(f(x,y)=0\big)\Big) &= 0 \end{aligned} \tag{4.52}$$

Write $\mathrm{d}F(x,y)/\mathrm{d}x = F_x$, $\mathrm{d}F(x,y)/\mathrm{d}y = F_y$, $\mathrm{d}\big(f(x,y)=0\big)/\mathrm{d}x = f_x$ as well as $\mathrm{d}\big(f(x,y)=0\big)/\mathrm{d}y = f_y$, and thus (4.52) becomes (4.53)

$$F_x = m f_x \quad F_y = m f_y \quad \tfrac{\mathrm{d}}{\mathrm{d}m}\Big(F(x,y) - m\big(f(x,y)=0\big)\Big) = 0 \tag{4.53}$$

$$\frac{F_y}{F_x} = \frac{f_y}{f_x} \tag{4.54}$$

The condition for an extreme under constraint $f(x, y) = 0$ in (4.54) is obtained from the first two equations in (4.53) in a few steps.

4.11. (4.22,4) is obtained by substituting the solution of (4.22), the interpolation variance as written in (4.21). Let $\boldsymbol{C}_{ik}$ be the matrix with the covariances in (4.22), $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N, m)^T$ the column vector with the solution of the system, and $\boldsymbol{c}_{i0} = (c_{10}, c_{20}, \ldots, c_{N0}, 1)^T$ the column vector of its right

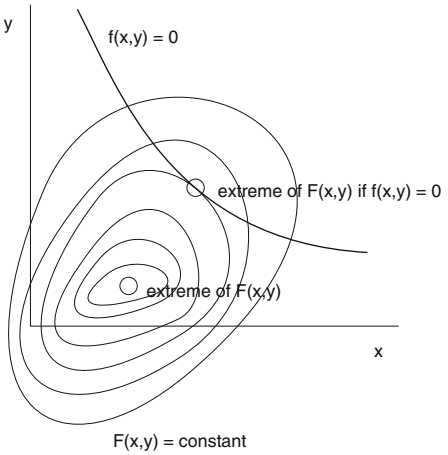


Fig. 4.36. Lagrange multipliers: the position of the extreme value of a function $F(x, y)$ changes under the additional assumption $f(x, y) = 0$.

side. Using this matrix and these vectors, (4.21) is written in the first line of the following derivation, and, in the second line, the solution $\boldsymbol{\lambda} = \boldsymbol{C}_{ik}^{-1}\boldsymbol{c}_{i0}$ is substituted.

$$\begin{aligned}\mathrm{Var}\big(\widehat{Z}(x_0) - Z(x_0)\big) &= \boldsymbol{\lambda}^T\boldsymbol{C}_{ik}\boldsymbol{\lambda} - 2\boldsymbol{\lambda}^T\boldsymbol{c}_{i0} + 2m + c_Z(0) \\ &= (\boldsymbol{C}_{ik}^{-1}\boldsymbol{c}_{i0})^T\boldsymbol{C}_{ik}\boldsymbol{C}_{ik}^{-1}\boldsymbol{c}_{i0} - 2(\boldsymbol{C}_{ik}^{-1}\boldsymbol{c}_{i0})^T\boldsymbol{c}_{i0} + 2m + c_Z(0)\end{aligned}$$

The result is obtained in a few steps.

4.12. Derive the kriging system in (4.24). The expression being minimised to obtain the system is derived as follows: (i) Because $\hat{\mu}$ is required to be not biased, universality condition $\sum_{i=1}^{N}\lambda_i' - 1 = 0$ is obtained from $\mathrm{E}(\hat{\mu} - \mu) = 0$ as in (2.64), and (ii) because $\hat{\mu}$ is further required to be optimal, the interpolation variance $\mathrm{Var}(\hat{\mu} - \mu) = \sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i'\lambda_k' c_Z(x_i - x_k) - 2\mathrm{Cov}(\hat{\mu}, \mu) + \mathrm{Var}\mu = \sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i'\lambda_k' c_Z(x_i - x_k)$ is minimised under the constraint $\sum_{i=1}^{N}\lambda_i' - 1 = 0$. The derivatives $\sum_{i=1}^{N}\sum_{k=1}^{N}\lambda_i'\lambda_k' c_Z(x_i - x_k) - 2m'\big(\sum_{i=1}^{N}\lambda_i' - 1\big)$ with respect to λ_k', $k = 1, \ldots, N$ and m' are calculated and set $= 0$.

You are to estimate the expectation function of a stationary random function. When do you solve the kriging system in (4.24)? When do you calculate the arithmetic mean? The arithmetic mean of a time series as defined in (2.1) has the properties (2.56).

4.13. Derive the kriging system in (4.26).

4.14. Derive (4.26,4). As in Problem 4.11: Let $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_N, \mu)^T$ be the column vector with the solution of (4.26), $\boldsymbol{\gamma}_{ik}$ be the variogram matrix in

the system, and $\boldsymbol{\gamma}_{i0} = (\gamma(x_1 - x_0), \ldots, \gamma(x_N - x_0), 1)^T$ the column vector of its right side. Then (4.26,4) is obtained subsequent to substituting $\boldsymbol{\lambda} = \boldsymbol{\gamma}_{ik}^{-1}\boldsymbol{\gamma}_{i0}$ in $-\boldsymbol{\lambda}^T\boldsymbol{\gamma}_{ik}\boldsymbol{\lambda} + 2\lambda^T\boldsymbol{\gamma}_{i0}$, the interpolation variance as in (4.25).

4.15. When an empirical semivariogram is calculated using Gstat, as demonstrated in Sect. 4.3.1, the analyst is asked whether duplicate observations should be averaged. Write a kriging system as in (4.22) for a point with observations (for example, x_1) occurring twice. A system with two identical equations is obtained that does not have a unique solution on condition that the system with only one equation for x_1 does (have a unique solution).

4.16. Calculate the kriging weights and the interpolation variance for $\mathrm{E}Z(x) = \mu_Z$ (being constant) and

a) $\mathrm{Cov}\big(Z(x), Z(y)\big) = c_0$ for $x = y$
   $\mathrm{Cov}\big(Z(x), Z(y)\big) = 0$ for $x \neq y$
b) $\mathrm{Cov}\big(Z(x), Z(y)\big) = c_0$ for $x = y$
   $\mathrm{Cov}\big(Z(x), Z(y)\big) = kc_0,\ 0 \leq k \leq 1$ for $x \neq y$
c) Assuming (b), calculate the kriging weights and the the variance of the estimate for μ_Z.

Hints: a) Write equation k in the kriging system (4.22): a covariance matrix for white noise is a diagonal matrix with c_0. The results are obtained using the universality condition and the symmetry of the system. b) The diagonal of the covariance matrix contains c_0, the other elements are kc_0.

4.17. Derive the universal kriging system in (4.35).

4.18. Derive the variance of the universal kriging system in (4.38).

4.19. The system (4.24) for optimally estimating a constant expectation is derived in Problem 4.12. Now derive the system for the optimal estimation of an expectation function that is not constant by applying

1. $\sum_{i=1}^N \lambda_i' \sum_{j=1}^p b_j f_j(x_i) = \sum_{j=1}^p b_j f_j(x_i)$, $j = 1, \ldots, p$, i.e., universality conditions as in (4.29), and
2. $\mathrm{Var}\big(\hat{\mu}(x) - \mu(x)\big) = \sum_{i=1}^N \sum_{k=1}^N \lambda_i' \lambda_k' c_{ik}$, i.e., optimality conditions as in (4.31).

From these, $\sum_{i=1}^N \sum_{k=1}^N \lambda_i' \lambda_k' c_{ik} - 2\sum_{j=1}^p m_j' \big(\sum_{i=1}^N \lambda_i' f_j(x_i) - f_j(x_i)\big)$ is arrived at using Lagrange multipliers. Thereafter, equating the partial derivations with zero, the system:

$$\sum_{i=1}^N \lambda_i c_{ik} - \sum_{j=1}^p m_j f_j(x_k) = 0, \qquad k = 1, \ldots, N \tag{4.55}$$

$$\sum_{i=1}^N \lambda_i f_j(x_i) = f_j(x_i), \qquad j = 1, \ldots, p \tag{4.56}$$

is obtained.

4.20. Derive (4.43), (4.44), (4.45) and (4.46) using (1.15) and (1.16).

4.21. Show that $2\gamma_{Z+e}(x_i, x_k) = 2\gamma_Z(x_i - x_k) + 2\gamma_e(x_i - x_k) = 2\gamma_Z(h) + 2\gamma_e(h)$, provided that the measurement error is as required in (4.42,1,2,3).

4.22. The precipitations observed for October 5, 1992 at the stations plotted in Fig. 4.2 are afflicted with a random error. Estimate this error. The precipitation observations are available in file `/path/ppitoct592.dat` and are locally interpolated, using the variogram in Fig. 4.33, to obtain the areal precipitation as shown in Fig. 4.34.

4.23. Interpolate the tree-line observations plotted in Fig. 3.10. Select a trend surface as a model for the drift and then estimate the drift and the variogram of the residuals using generalised least squares.

4.24. Compare the result obtained in Problem 4.23 with the results in Sect. 3.6.

4.25. Interpolate the observations of "Strassenacker" grain yield as plotted in Fig. 4.13 using the variogram model in Fig. 4.15.

4.26. Estimate a first order linear model for expectation function of the SWIR climatologies in Fig. 4.16 using generalised least squares. Calculate the residuals and then estimate the variance of the residuals. Repeat these estimations using ordinary least squares, i.e., without incorporating the correlations of the residuals in the estimator. Hint: Gstat allows for estimating the random function and its expectation function.

5 Linear Processes

In Sect. 2.3.2, an AR[1] model is fitted to the wind speed values of the turbulent atmospheric flow plotted in Fig. 2.1, and, in Sect. 2.3.4, an MA[1] model is fitted to the differences of the yearly values in the Basel temperature series plotted above in Fig. 2.14. Both models have been fitted by trial and error and both are linear processes generated from a white noise process as defined in (2.25), i.e., they are linear combinations of a sequence of independent and identically distributed random variables. In the closing remarks in Sect. 2.3.5, it is supposed that there are autoregressive and moving average models with two or more parameters, AR[p] and MA[q] models, and even combined models; and, to circumvent the shortcomings of the fitting by trial and error, a systematic fitting procedure is wished for.

This wish will be granted once you work through this chapter in which you will find an introduction to the usual procedure for fitting a linear process to a time series. This procedure involves the steps in (5.1), i.e.,

1. *the identification of a small number of candidate models*
2. *the estimation of the parameters in the candidate models*
3. *the diagnostic of the estimated models, and*
4. *the selection of a model.*

(5.1)

It should be noted that the sequence of the steps is not fixed: often one advances one step and is then forced to retreat one or even several steps.

The result of procedure (5.1) is a model that can later be used to predict the time series under analysis. With this third possibility to calculate predictions, the desire declared when concluding Sect. 2.7 is fulfilled: a stochastic process can be predicted using (i) linear models for its first moment function (in Chap. 3), (ii) analytical variograms for its second moment function (in Chap. 4), or (iii) models for the process itself (in Sect. 5.5 of this chapter).

Sects. 5.1 and 5.2 introduce estimators for AR[p] models and Sect. 5.3 introduces estimators for both ARMA[p, q] and ARIMA[p, d, q] models. Thereafter, in Sect. 5.4, the systematic estimation of a linear process as proposed in (5.1) is demonstrated using three example time series: AR[p], MA[q] and ARMA[p, q] models are fitted to the stationary examples, and an ARIMA[p, d, q] model is fitted to an example with a non-constant mean.

In Sect. 5.5, a fitted model is used to calculate predictions for time points with no observations. The supplements are given in Sect. 5.6, and the chapter closes with the problems in Sect. 5.7.

Introductions to linear processes are given, e.g., in [16], [21], [17] and [22], and a summary is given in [79].

5.1 AR[p] Models

In Sect. 2.3.2, it is assumed that the time series (x_t), $t = 1, \ldots, 1800$, x_t being the wind speed values measured in a turbulent atmospheric flow and plotted in Fig. 2.1, is a time slice from a realisation of a stationary stochastic process (X_t) as defined in (2.8). Under this assumption, an AR[1] process $(X_t - \mu_X) = a(X_{t-1} - \mu_X) + W_t$, (W_t) being a white noise process with $\mu_W = 0$ and $\sigma_W^2 = 1$ as well as $a = 0.9$, is then found to be a suitable model for this process. The autogressive coefficient $a = 0.9$ is obtained by comparing the plots in Fig. 2.11 with the plots in Figs. 2.1 and 2.3, i.e., by trial and error. As $|a| < 1$, the fitted AR[1] model is stationary if sufficient time has elapsed since the process (simulation) started. This is concluded from (2.15), (2.16), (2.17) and (2.18) or, alternatively, from $X_t = \sum_{u=0}^{\infty} a^u W_{t-u}$, i.e., the representation of (X_t) as a linear process (2.25), which is obtained since X_t is the stationary (the general solution is given in (2.27)) solution of the linear difference equation pertaining to the AR[1] model.

The AR[2] model $(X_t - \mu_X) = a_1(X_{t-1} - \mu_X) + a_2(X_{t-2} - \mu_X) + W_t$, (W_t) a white noise process with $\mu_W = 0$ and σ_W^2, is defined in (2.48), and the solution of its difference equation is given in (2.49). From (2.49) it is concluded that an AR[2] model is stationary provided that (i) the moments of the W_t exist, (ii) the roots z_1 and z_2 of the characteristic polynomials pertaining to the model coefficients $A(z) = 1 - a_1 z - a_2 z^2 = 0$ are not in the unit circle, $1 < |z_1, z_2|$, as required in (2.46), and (iii) $t \to \infty$. These conditions fulfilled, the general solution of $(X_t - \mu_X) - a_1(X_{t-1} - \mu_X) - a_2(X_{t-2} - \mu_X) = 0$ becomes identically zero and the particular solution of $(X_t - \mu_X) - a_1(X_{t-1} - \mu_X) - a_2(X_{t-2} - \mu_X) = W_t$ then converges in mean square, as defined in (7.11).

The difference equation of the AR[2] model is written alternatively as convolution $(1, -a_1, -a_2) * (X_t - \mu_X) = W_t$ in (2.48). In the case of an AR[1] model, $(1, -a) * (X_t - \mu_X) = W_t$ is obtained. If these difference equations are written as convolutions, then their stationary solutions are convolution sums $\sum_{u=0}^{\infty} b_u W_{t-u} = (b_u) * (W_t)$ of a white noise process with sequences $(b_u) = (1, -a_1 - a_2)^{-1}$ or $(b_u) = (1, -a_1)^{-1}$ being the inverses as defined in (2.41) to the sequences of the model coefficients. These convolution sums converge in mean square, provided that (i) the moments of the W_t exist, (ii) the roots of the characteristic polynomials pertaining to the model coefficients, i.e., $(1, -a_1 - a_2)$ or $(1, -a_1)$, are not in the unit circle, as is required in (2.46), and (iii) a very long time has elapsed since the initialisation of the models.

The second condition is satisfied for $|a| < 1$ in the case of an AR[1] model, and a_1 and a_2 within the triangular region of convergence plotted in Fig. 2.17 in the case of an AR[2] model.

In the remarks closing Sect. 2.3.5,

1. it is supposed that there are autoregressive and moving average models with two or more parameters, i.e., AR[p] and MA[q] models
2. the absence of a procedure is felt that would allow to assess the order of the model from the observations, and
3. the need for an estimator for the coefficients in the models arises.

A first procedure for the estimation of an autoregressive model is the solution of the Yule-Walker equations in (2.53) in the case of a stationary AR[2] model (or $c_X(1) = ac_X(0)$ directly from (2.16) in the case of a stationary AR[1] model) subsequent to substituting the empirical correlations (covariances) for the theoretical ones.

In this section, the AR[p] model is defined in Sect. 5.1.1 as being a general model for stationary autoregressive processes, and the pertaining Yule-Walker equations are derived. The Yule-Walker equations connect, as in the case of the AR[1]- and AR[2] models, the model coefficients with the covariances of the process. Once the equations are solved, with empirical covariances substituted for the theoretical ones, the Yule-Walker estimates for the coefficients are obtained. A solution of the Yule-Walker equations is calculated very efficiently with the Levinson-Durbin recursion which is introduced together with the partial correlation function of a stationary AR[p] model in Sect. 5.1.2.

5.1.1 The Yule-Walker-Equations of an AR[p] Model

The AR[p] model is defined by giving its difference equation in (5.2,2,3).

If (W_t) is a white noise process with $\mu_W = 0$ and σ_W^2, and if the solutions to $A(z) = 1 - a_{1p}z - \ldots - a_{pp}z^p$ are outside the complex unit circle, then:

1. $X_t - \mu_X = (1, -a_{1p}, \ldots, -a_{pp})^{-1} * (W_t)$
   $= (b_u) * (W_t) = \sum_{u=0}^{\infty} b_u W_{t-u}$
   is the stationary solution to the linear difference equation
2. $(X_t - \mu_X) - a_{1p}(X_{t-1} - \mu_X) - a_{2p}(X_{t-2} - \mu_X) - \ldots$
   $-a_{pp}(X_{t-p} - \mu_X) = W_t$
   or, written as a convolution,
3. $(1, -a_{1p}, -a_{2p}, -\ldots, -a_{pp}) * (X_t - \mu_X) = (W_t)$
   pertaining to the autoregressive process of order p (AR[p] process, AR[p] model).

(5.2)

Due to (2.46), the inverse $(1, -a_{1p}, \ldots, -a_{pp})^{-1} = (b_u)$ to the sequence of the model coefficients converges absolutely, and, consequently, the convolution sum (5.2,3) converges in mean square to the limiting random variable $X_t - \mu_X$.

A proof is constructed using the template in Sect. 7.2.1 where the convergence of an AR[1] model with $|a| < 1$ is shown. As a second consequence, $(X_t - \mu_X)$ becomes a linear process as defined in (2.25) and thus the AR[p] model (5.2) is reconcilable with a stationary process: the model as defined in (5.2) is called a stationary AR[p] model.

If the expectation function μ_X of a stationary process (X_t) is not identically zero then the de-meaned process is modelled, as required in (5.2,2). μ_X is estimated with $\hat{\mu}_X$ defined in (2.1,1), the properties of this estimate being summarised in (2.56).

Unlike in definitions (2.13) and (2.48), the coefficients of the AR[p] model are written, in definition (5.2,2), using two indices, the second index being the order of the model. You will find many forms of the AR[p] model in articles and books: often the model is written using only one index for the coefficients and not as difference equation (in difference equation (5.2,2), all process variables are on the left side, and the white noise variable is on the right side of the equation), often ϕ (not a) is used for the coefficients, and sometimes the signs are inverted. (In R expressions the signs are used as in (5.2,2,3)). It is straightforward and comfortable, however not in use elsewhere, to write an AR[p] model as a convolution sum in (5.2,3).

An AR[p] model is called causal on condition that W_t is not correlated with X_s, $s < t$. As a consequence, the influence of the past to the present is contained in $\sum_{j=1}^{p} a_{jp} X_{t-j}$ solely, i.e., in the linear (autoregressive) part of the model. (5.3)

An AR[p] model as defined in (5.2) is causal since $\mathrm{Cov}((X_{t-j} - \mu_X), W_t) = \mathrm{Cov}\big(\sum_{u=0}^{\infty} b_u W_{t-j-u}, W_t\big) = 0$, $j = 1, 2, \ldots$, as $\mathrm{Cov}(W_s, W_t) = 0$ for $s < t$. Hence, the W_t are often called *innovations.*. The causality of an AR[p] model is desirable with respect to its use for predictions: it asserts the orthogonality conditions (2.79) of an optimal linear prediction.

The Yule-Walker equations (2.53) pertaining to an AR[2] model are derived from (2.51). (2.51) is obtained, for its part, by multiplying two equations: (i) the stationary solution of the AR[2] difference equation for the model variable with lag τ, $X_{t-\tau} = W_{t-\tau} + \sum_{u=1}^{\infty} b_u W_{t-\tau-u}$, and (ii) the difference equation of the model, and then taking the expectations. Proceeding as demonstrated in the AR[2] case and substituting k for τ (the lag), (5.4) is obtained for an AR[p] model with $\mu_X = 0$.

$$
\begin{aligned}
&(X_t - a_{1p} X_{t-1} - \ldots - a_{pp} X_{t-p}) X_{t-k} = W_t \Big(W_{t-k} + \sum_{u=k+1}^{\infty} b_{u-k} W_{t-u} \Big) \\
&\mathrm{E}(X_t X_{t-k}) - a_{1p} \mathrm{E}(X_{t-1} X_{t-k}) - \ldots - a_{pp} \mathrm{E}(X_{t-p} X_{t-k}) \\
&\qquad = \mathrm{E}(W_t W_{t-k}) + \sum_{u=k+1}^{\infty} b_{u-k} \mathrm{E}(W_t W_{t-u}) = \begin{cases} 0 & \text{for } k > 0 \\ \sigma_W^2 & \text{for } k = 0 \end{cases}
\end{aligned}
\tag{5.4}
$$

Assuming $\mu_X = 0$ and since $c_X(k) = c_X(-k)$, the Yule-Walker equations (5.5) pertaining to an AR[p] model are derived by substituting $p = 1$ and $k = 1$, $p = 2$ and $k = 1, 2$, $p = 3$ and $k = 1, 2, 3, \ldots$, $p = p$ and $k = 1, 2, \ldots, p$ in (5.4). The Yule-Walker equations also hold for de-meaned processes, being in agreement with (5.2).

$$
\begin{array}{ll}
c_X(1) = a_{11} c_X(0) & p = 1 \\
\begin{array}{l} c_X(1) = a_{12} c_X(0) + a_{22} c_X(1) \\ c_X(2) = a_{12} c_X(1) + a_{22} c_X(0) \end{array} & p = 2 \\
\begin{array}{l} c_X(1) = \; a_{1p} c_X(0) \quad + a_{2p} c_X(1) \quad + \ldots + a_{pp} c_X(p-1) \\ c_X(2) = \; a_{1p} c_X(1) \quad + a_{2p} c_X(0) \quad + \ldots + a_{pp} c_X(p-2) \\ \quad \ldots \\ c_X(p) = \; a_{1p} c_X(p-1) + a_{2p} c_X(p-2) + \ldots + a_{pp} c_X(0) \end{array} & p = p
\end{array}
\tag{5.5}
$$

With $\boldsymbol{c}_X(p) = \big(c_X(1), c_X(2), \ldots, c_X(p)\big)^T$ and $\boldsymbol{a}(p) = (a_{1p}, a_{2p}, \ldots, a_{pp})^T$, and with $\boldsymbol{C}_X(p)$ being the covariance matrix of (X_t) including the correlations until (but without) lag p, the Yule-Walker equations and their solutions are obtained in (5.6):

$$
\boldsymbol{c}_X(p) = \boldsymbol{C}_X(p)\boldsymbol{a}(p) \qquad \boldsymbol{a}(p) = \big(\boldsymbol{C}_X(p)\big)^{-1} \boldsymbol{c}_X(p) \qquad p = 1, 2, \ldots \tag{5.6}
$$

since $\big(\boldsymbol{C}_X(p)\big)^{-1}$ exists for positive definite $\boldsymbol{C}_X(p)$ (in the remarks to (2.7,2)).

Assuming that an AR[p] model as defined in (5.2) is reconcilable with the stationary stochastic process (X_t) and that a time series (x_t) is a time slice from a realisation of (X_t), the empirical covariances with the properties (2.58,1) are substituted in (5.6) and the Yule-Walker estimates (5.7,1) for the coefficients in (5.2) are obtained. The variance of the innovations (W_t) is estimated by substituting $k = 0$ in (5.4): $c_X(0) - a_{1p} c_X(1) - \ldots - a_{pp} c_X(p) = \sigma^2_W(p)$ follows and, by substituting the empirical covariances and the estimated coefficients, the estimate (5.7,2) is arrived at.

Let (X_t) be an AR[p] model, $\hat{\boldsymbol{c}}_X(p)$ and $\hat{\boldsymbol{C}}_X(p)$ the empirical versions of the covariances in (5.6), $\hat{\boldsymbol{a}}(p)$ the estimates for $a_{1p}, \ldots, a_{pp}$ and $\hat{\sigma}^2_W(p)$ the estimate for σ^2_W. Then:

1. $\hat{\boldsymbol{a}}(p) = \big(\hat{\boldsymbol{C}}_X(p)\big)^{-1} \hat{\boldsymbol{c}}_X(p)$
2. $\hat{\sigma}^2_W(p) = \hat{c}_X(0) - \big(\hat{\boldsymbol{a}}(p)\big)^T \hat{\boldsymbol{c}}_X(p)$

(5.7)

When comparing the Yule-Walker equations (5.6) with the equations (2.75) for an optimal linear prediction of a stationary stochastic process, the following question may be asked: are the coefficients of an AR[p] model (X_t) as defined in (5.2) the weights of an optimal linear prediction for X_t, given observations for X_{t-1}, X_{t-2}, ..., X_{t-p}? If the answer is yes, then the right side W_t of the difference equation (5.2,3) is, using the terms defined

in (2.61), the prediction error $X_t - \widehat{X}_t$. The answer to the above question is given in the following paragraphs.

The optimal linear prediction of a stationary process (X_t) for the time point $N + \tau$ as defined in (2.62) is a linear combination of the observed values for $(X_1, \ldots, X_N)$. In the actual case, however, a de-meaned ($\mu_X = 0$) stationary process (X_t) is to be predicted for time point t with a linear combination of k (k acting as a substitute for p in (5.2)) values for the past $X_{t-1}, \ldots, X_{t-k}$.

If (X_t) is stationary with $\mu_X = 0$ then:
1. $\widehat{X}_t^{(v)}(k) = a_{1k}X_{t-1} + a_{2k}X_{t-2} + \ldots + a_{kk}X_{t-k}$ *is a linear forward prediction for X_t, given X_{t-1}, …, X_{t-k},*
2. $X_t - \widehat{X}_t^{(v)}(k) = \hat{e}_t^{(v)}(k)$ *is the prediction error, and*
3. $\mathrm{PVar}\widehat{X}_t^{(v)}(k) = \mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k)\big)^2\Big) = \mathrm{E}\Big(\big(X_t - \widehat{X}_t^{(v)}(k)\big)^2\Big)$ *becoming minimal is required to obtain an optimal prediction.*

(5.8)

The superscript $^{(v)}$ in (5.8) denotes that X_t is predicted looking forward (into the future) and using past X_{t-j}, $j = 1, \ldots, k$. (5.8,3) is then evaluated using the template (2.78) to arrive at optimality condition j in (5.9):

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}a_j}\mathrm{PVar}\widehat{X}_t^{(v)}(k) &= \frac{\mathrm{d}}{\mathrm{d}a_j}\mathrm{E}\Big(\big(X_t - (a_{1k}X_{t-1} + \ldots + a_{kk}X_{t-k})\big)^2\Big) \\ &= -2\mathrm{E}\Big(\big(X_t - (a_{1k}X_{t-1} + \ldots + a_{kk}X_{t-k})\big)X_{t-j}\Big) = 0 \qquad (5.9)\end{aligned}$$

The optimality conditions are evaluated and the covariances are substituted for the expectations of the products with two variables to obtain, in (5.10), the equations for the optimal linear forward prediction $\widehat{X}_t^{(v)}(k)$ for X_t, given observed values for $(X_1, \ldots, X_N)$.

$$\begin{aligned} c_X(1) &= a_{1k}c_X(0) + a_{2k}c_X(1) + \ldots + a_{kk}c_X(k-1) \\ c_X(2) &= a_{1k}c_X(1) + a_{2k}c_X(0) + \ldots + a_{kk}c_X(k-2) \\ &\ldots \\ c_X(k) &= a_{1k}c_X(k-1) + a_{2k}c_X(k-2) + \ldots + a_{kk}c_X(0) \end{aligned} \qquad (5.10)$$

For positive definite $\boldsymbol{C}_X(p)$ (introduced in the remarks to (2.7,2)), the equations (5.10) deliver the weights of the optimal linear prediction $\widehat{X}_t^{(v)}(k)$ in (5.8,1). The mean square prediction error (5.12) is derived in Problem 5.2.

$$\mathrm{PVar}\widehat{X}_t^{(v)}(k) = \mathrm{E}\big(\hat{e}_t^{(v)}(k)X_t\big) = \mathrm{Cov}\big(\hat{e}_t^{(v)}(k), X_t\big) \qquad (5.11)$$

$$= c_X(0) - a_{1k}c_X(1) - \ldots - a_{kk}c_X(k) \qquad (5.12)$$

(5.11) implies that the mean square prediction error is the covariance of the prediction error and the variable to be predicted. (5.10) and (5.5) as well

as (5.12) and (5.7,2) are identical if the empirical correlations are substituted for the theoretical ones and if, in (5.5) and (5.7,2), k is substituted for p.

Thus, the answer to the question in the remarks to (5.7) is yes, i.e., the coefficients of an AR[p] model are the weights of an optimal linear prediction for X_t, given observations for X_{t-1}, X_{t-2}, ..., X_{t-p}, and the Yule-Walker equations become Janus-faced as summarised in (5.13):

1. *If (X_t) is an AR[p] model (5.2) then the Yule-Walker equations connect the covariances with the coefficients. When substituting empirical covariances for the theoretical ones and by solving the equations, estimates for the coefficients can be obtained.*
2. *If (X_t) is a stationary stochastic process with $\mu_X = 0$ and $c_X(\tau)$ then the Yule-Walker equations connect the covariances with the weights of an optimal linear forward prediction $\widehat{X}_t^{(v)}(k)$ for X_t, given observed values for $X_{t-1}, \ldots, X_{t-k}$.*
3. *The variance of the innovations in the AR[p] model equals the mean square prediction error: $\sigma_W^2(p) = \mathrm{PVar}\widehat{X}_t^{(v)}(k)$, $k = p$.*

(5.13)

5.1.2 Levinson-Durbin Recursion, Partial Correlation Function

The Yule-Walker equations are solved, as a first possibility, using one of the usual algorithms for solving systems of linear equations. More expeditious however, is a solution calculated using the Levinson-Durbin recursion as proposed in (5.14) and being derived in Sect. 5.6.1.

The Yule-Walker equations (5.6) can be solved using the Levinson-Durbin recursion:

1. $a_{kk} = \theta_k = \Big(c_X(k) - \sum_{j=1}^{k-1} a_{j,k-1} c_X(k-j)\Big) \Big/ \mathrm{PVar}(k-1)$
2. $a_{jk} = a_{j,k-1} - \theta_k a_{k-j,k-1}, \qquad j = 1, \ldots, k-1$
3. $\mathrm{PVar}(k) = c_X(0) - \sum_{j=1}^{k} a_{jk} c_X(j) = \big(\mathrm{PVar}(k-1)\big)(1 - a_{kk}^2)$

(5.14)

Recursion k uses $a_{j,k-1}$, $j = 1, \ldots, k-1$ and $\mathrm{PVar}(k-1)$ obtained in the preceding recursion $k-1$ to calculate a_{kk}, a_{jk}, $j = 1, \ldots, k-1$, and $\mathrm{PVar}(k)$. The recursion is initialised with $a_{11} = c_X(1)/c_X(0)$ and $\mathrm{PVar}(1) = c_X(0) - a_{11} c_X(1)$ as derived in (5.5).

Having substituted the empirical covariances $\hat{c}_x(1), \ldots, \hat{c}_X(p)$ for the model covariances in (5.5) or (5.6), the coefficients of an AR[p] model can be estimated using the Levinson-Durbin recursion as follows: Start with $k = 1$ and calculate $\hat{a}_{11} = \hat{c}_X(1)/\hat{c}_X(0)$ as well as $\hat{\sigma}_W^2(1) = \hat{c}_X(0) - \hat{a}_{11}\hat{c}_X(1) = \hat{c}_X(0)(1-\hat{a}_{11}^2) = \mathrm{PVar}(1)$. Then, for $k = 2$, $\hat{a}_{22} = \big(\hat{c}_X(2) - \hat{a}_{11}\hat{c}_X(1)\big)/\hat{\sigma}_W^2(1)$, $\hat{a}_{12} = \hat{a}_{11} - \hat{a}_{22}\hat{a}_{11}$, and $\mathrm{PVar}(2) = \hat{c}_X(0) - \hat{a}_{12}\hat{c}_X(1) - \hat{a}_{22}\hat{c}_X(2) = \mathrm{PVar}(1)(1 - \hat{a}_{22}^2) = \hat{\sigma}_W^2(2)$ are obtained. Proceed until $k = p$ to obtain the Yule-Walker estimates $\hat{a}_{1p}, \hat{a}_{2p}, \ldots, \hat{a}_{pp}$ and $\hat{\sigma}_W^2(p)$.

If an AR[p] model is fitted to a time series (x_t), $t = 1, 2, \ldots, N$, by calculating the Yule-Walker estimates, then the covariances of the fitted model are

identical with the empirical covariances up to lag p. The bias of the empirical covariances is small provided that N is large, as concluded in (2.58,1), and this property propagates to the Yule-Walker estimates. For small N however, the estimates can be biased.

In the Yule-Walker equations pertaining to a stationary stochastic process (X_t) and also in the Levinson-Durbin recursion, the sequence (a_{kk}) occurs. (a_{kk}) is the partial correlation function of (X_t) as defined in (5.15):

The sequence of the $a_X(k,k) = (a_{kk})$ occurring in (5.5) and in (5.14) is called the sequence of partial correlations or the partial correlation function of (X_t), a_{kk} being the partial correlation of X_t and X_{t-k}:

$$a_X(k,k) = \frac{\mathrm{Cov}\big(X_t - \widehat{X}_t^{(v)}(k-1), X_{t-k} - \widehat{X}_{t-k}^{(r)}(k-1)\big)}{\Big(\mathrm{Var}\big(X_t - \widehat{X}_t^{(v)}(k-1)\big)\mathrm{Var}\big(X_{t-k} - \widehat{X}_{t-k}^{(r)}(k-1)\big)\Big)^{1/2}} \tag{5.15}$$

(5.15) is derived in Sect. 5.6.2. $a_X(k,k)$ is the covariance of the differences $\hat{e}_t^{(v)}(k-1) = X_t - \widehat{X}_t^{(v)}(k-1)$ and $\hat{e}_{t-k}^{(r)}(k-1) = X_{t-k} - \widehat{X}_{t-k}^{(r)}(k-1)$, scaled by the variances.

$\hat{e}_t^{(v)}(k-1)$ and $\hat{e}_{t-k}^{(r)}(k-1)$ are the differences of X_t and X_{t-k} as well as the pertaining predictions calculated from the variables between time points $t-k$ and t, as is obvious in Fig. 5.19, since $\widehat{X}_t^{(v)}(k-1)$ estimates X_t using $k-1$ preceding variables as required in (5.8), and $\widehat{X}_{t-k}^{(r)}(k-1)$ estimates X_{t-k} using $k-1$ following variables as required in (5.72). Consequently, a_{kk} is the correlation of X_t and X_{t-k} corrected for the influence exerted by $X_{t-1}, \ldots, X_{t-k+1}$.

$-1 \le a_X(k,k) = a_{kk} = \theta_k \le 1$, obtained in the remarks to (5.79), is a desired property of a partial correlation.

The sequence $(\hat{a}_{kk}) = \big(\hat{a}_X(k,k)\big)$, k small as compared to N in order to account for (2.58), is called the empirical partial correlation function of a time series (x_t), $t = 1, \ldots, N$, being a time slice from a realisation of the stationary stochastic process (X_t). From (x_t), $(\hat{a}_{kk})$ is calculated, without using the Levinson-Durbin recursion, by solving the empirical Yule-Walker equations (5.5) for $p = 1$, $p = 2$, $p = 3$, ..., to obtain $\hat{a}_{11}$; $\hat{a}_{12}$, $\hat{a}_{22}$; $\hat{a}_{13}$, $\hat{a}_{23}$, $\hat{a}_{33}$, etc. Usually however, $(\hat{a}_{kk})$ is calculated much faster with the Levinson-Durbin recursion, e.g., applying R function `acf(...,type="partial", ...)`.

The Levinson-Durbin recursion is, in general, an algorithm for solving efficiently a system of linear equations containing a Toeplitz matrix. The values in a Toeplitz matrix are a function of only the differences of the indices implying identical values in each subdiagonal. For instance, the covariance matrix of a stationary stochastic process is a symmetric Toeplitz matrix, as concluded from (2.9). From both properties, i.e., the covariances being a function of the lag only and the symmetry, the Levinson-Durbin recursion (5.15) is derived in Sect. 5.6.2.

The partial correlations a_{kk} of an AR[p] model (X_t) become identically zero for lags k larger than the order p: $a_X(k,k) = 0$, for $k > p$. This is obtained from (5.8) and (5.72) once the numerator in (5.15) is written for $k > p$,

$$\mathrm{Cov}\left(X_t - \sum_{j=1}^{p} a_{jp} X_{t-j},\ X_{t-k} - \sum_{j=1}^{p} a_{jp} X_{t-k+j}\right) = \mathrm{Cov}\big(W_t, W_{t-k}\big) = 0 \quad (5.16)$$

because the coefficients $(\ldots, 0, 0, 1, -a_{1p}, -a_{2,p}, \ldots, -a_{pp}, 0, 0, \ldots)$ in a stationary AR[p] model, as defined in (5.2), are identically zero for lags $k > p$ (this property is shared by the sequence $(\ldots, 0, 1, -b, 0, 0, \ldots)$ with its right-inverse $(\ldots, 0, 1, b, b^2, b^3, \ldots)$ given as an example in (2.42)) and because the model is causal, as defined in (5.3).

Property (5.16) allows for assessing the order p of an autoregressive model to be fitted to a time series stemming from a stationary process: the empirical correlation and empirical partial correlation functions are calculated and p is assumed to be lag, after which the empirical partial correlations become approximately zero.

5.1.3 Examples

The time series of wind speed values measured in a turbulent atmospheric flow and plotted in Fig. 2.1, with its empirical correlation function plotted in Fig. 2.3, is assumed to be stationary in the remarks to (2.8) and in Sect. 2.5. If `wind` is the R time series object generated from the wind speed data in the remarks to Fig. 2.1, then the de-meaned time series (x_t) and its empirical second moment functions $\hat{c}_X(\tau)$ and $\hat{a}_X(k,k)$, as plotted in Fig. 5.1, are obtained with the R expressions

```
wmb <- wind - mean(wind)
plot(wmb,type="l",xlab="s", ylab="m/s")
acf(wmb,lag.max=60,type="correlation",plot=T)
acf(wmb,lag.max=60,type="partial",plot=T)
```

In Fig. 5.1, the empirical partial correlations $\hat{a}_X(k,k)$ become approximately zero for lags $k > 2$ and the empirical correlations decay approximatively exponentially until lag 30.

A stationary AR[p] model for the de-meaned wind speed series can be estimated, as proposed in the remarks to (5.14), by solving the Yule-Walker equations. The order p of the model is assessed by applying (5.16) to the plots below in Fig. 5.1: with $\hat{a}_X(2,2)$ being outside and $\hat{a}_X(3,3)$, $\hat{a}_X(4,4)$, ..., inside the .95 confidence interval, $p = 2$ is obtained. However, to allow for a comparison with the AR[1] model fitted to the wind speed observations using trial and error in Sect. 2.3.2, an AR[1] model is estimated using the R expression

```
wmb.ar1ywfit <- ar.yw(wmb, aic=F, order=1)   #.yw for Yule-Walker
```

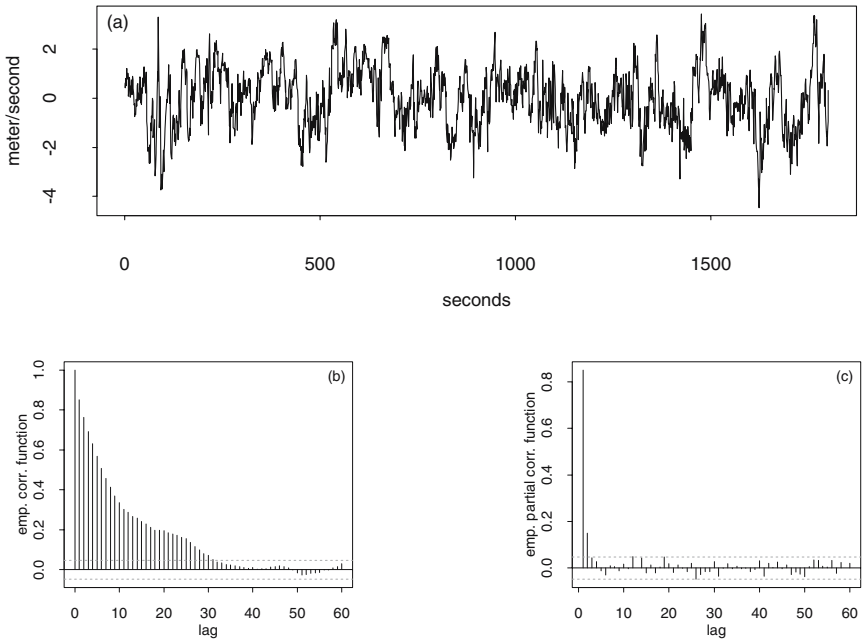


Fig. 5.1. De-meaned horizontal winds speeds (a) calculated from the original time series as plotted in Fig. 2.1 together with estimates of its second moment functions: empirical correlation (b) and empirical partial correlation (c) functions.

to obtain $\hat{a}_{11} = 0.8503$ for the autoregressive coefficient and $\hat{\sigma}^2_W(1) = 0.405$ for the variance of the empirical residuals of the model. $\hat{a}_{11}$ is close to $a^* = 0.9$ but $\hat{\sigma}^2_W(1)$ is smaller than $\sigma^{*2}_W = 1$, a^* and σ^{*2}_W being found by trial and error in Sect. 2.3.2 by comparing the plots in Figs. 2.1 and 2.3 with the plots in Fig. 2.11. An AR[1] model with a^* and σ^{*2}_W, however, produces simulations with too large fluctuations as shown in Fig. 2.11 (on the top, to the left), and thus $\hat{\sigma}^2_W(1) = 0.405$ is a better estimate.

With `...,aic=T,...` in `ar.yw()`, the estimate is calculated taking into account the AIC. The AIC is introduced in (5.61) as a diagnostic tool for autoregressive (and other) models for linear processes.

The estimates computed by `ar.yw()` are made available as R matrices (introduced in the remarks to Fig. 2.1) since `ar.yw()` allows for fitting autoregressive models to multivariate stochastic processes (the random variables are multi-dimensional, in the remarks to (2.2)): the empirical residuals, for example, become available as `wmb.ar1ywfit$resid[,1]`.

The empirical residuals of a fitted AR[p] model are, as can be concluded from (5.13), the empirical innovations of the fitted model and also the estimates for the prediction errors in (5.8). If the fitted model is suitable for the time series, then its empirical residuals are not (or only very weakly)

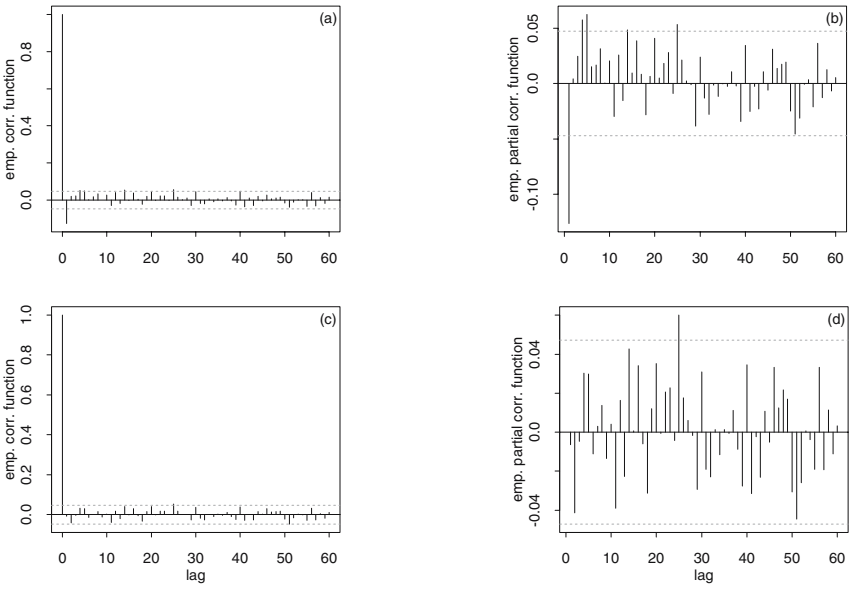


Fig. 5.2. Sequences of the empirical and partial empirical correlation functions of the residuals obtained from an AR[1] model (above, plots (a) and (b)) as well as from an AR[2] model (below, plots (c) and (d)) for the de-meaned wind speed series plotted at the top, in Fig. 5.1.

correlated, seeing that an AR[p] model (5.2) is causal, as stipulated in (5.3). Hence, the empirical correlation and the empirical partial correlation functions calculated from the empirical residuals of the AR[1] model fitted to the wind speed time series are plotted above in Fig. 5.2, plots (a) and (b). Since these plots show empirical correlations and partial correlations outside the .95 confidence intervals, it becomes obvious that an AR[1] model is not reconcilable with the observations. However, since the estimates outside the .95 confidence intervals are small, it is concluded that the tentatively fitted AR[1] model is not too far off from being the best suited model and therefore, an AR[2] model is fitted with

```
wmb.ar2ywfit <- ar.yw(wmb, aic=F, order=2)
```

to obtain $\hat{a}_{12} = 0.7236$, $\hat{a}_{22} = 0.1492$ and $\hat{\sigma}^2_W(2) = 0.396$. The empirical correlation and empirical partial correlation functions calculated from the empirical residuals of this AR[2] model are plotted below in Fig. 5.2, plots (c) and (d). Comparing plot (c) with the example plots in Fig. 2.22, it is concluded from (2.59) that the empirical residuals of the AR[2] model could be a time slice from a realisation of a white noise process and, thus, the AR[2] model fits the wind speed data.

Since (i) an AR[2] model is reconcilable with the wind speed data, (ii) in Fig. 5.1 (c), $\hat{a}_X(k,k) \approx 0$, for $k > 2$, i.e., the empirical partial correlation

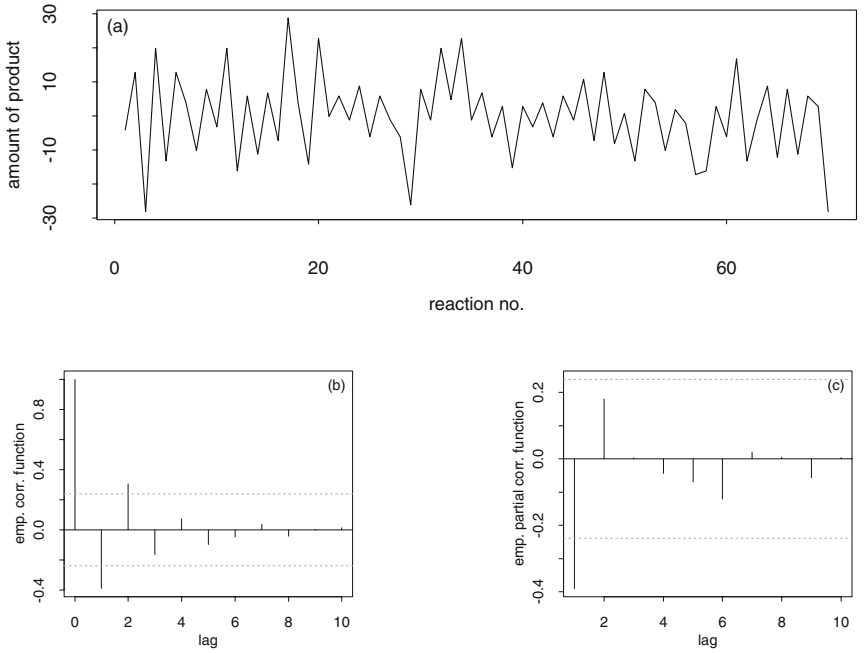


Fig. 5.3. De-meaned time series of the amount of a product formed during a chemical reaction occurring in a batch reactor (plot a) together with estimates of its second moment functions: empirical correlation and empirical partial correlation functions (plots b and c).

function cuts off after lag 2, and (iii) in Fig. 5.1 (b), the empirical correlation function decays exponentially, it is concluded that the order p of an AR[p] model to be fitted to a time series is chosen such that $p = k$, k being the lag after which the empirical partial correlation function cuts off. This choice is in line with (5.16) which proposes that the partial correlations of an autoregressive process become identically zero for lags larger than the order of the process, and also in line with the properties of the correlation function of an AR[2] process in the remarks to (2.52).

As a second example, the amount of a product formed during a chemical reaction occurring in a batch reactor is analysed. Using the R expressions

```
chemrea <- ts(scan("/path/chemrea.dat"))
chemreamb <- chemrea - mean(chemrea)
plot(chemreamb,type="l",xlab="batch no.", ylab="product")
acf(chemreamb,lag.max=10,type="correlation",plot=T)
acf(chemreamb,lag.max=10,type="partial",plot=T)
```

plots of the de-meaned time series and plots of the pertaining empirical correlation and partial correlation functions are obtained in Fig. 5.3. Applying

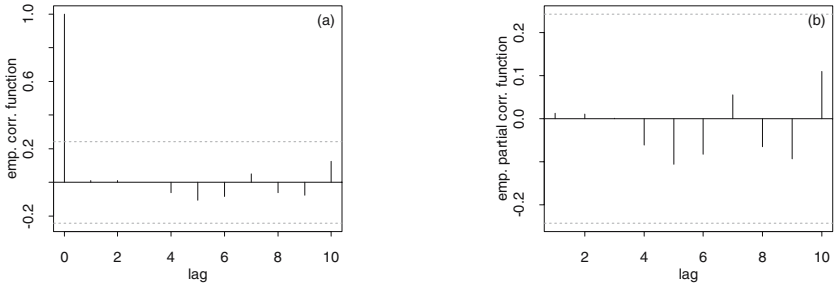


Fig. 5.4. Empirical second moment functions of the residuals of an AR[2] model fitted to the amounts of the product obtained in a chemical reaction as plotted in Fig. 5.3 (a).

the diagnostics (2.54,4) to the plots in Fig. 5.3, the process is assumed to be stationary.

These data are available as time series F in the set of example time series given in [16]. When the chemical reaction terminates, the reactor is discharged, the substance produced is weighted, the reactor is re-filled and the reaction re-initiated. The observed random variable is the amount of the product obtained per batch (the unit is not given in [16] being considered unimportant when fitting an AR[p] model), and the parameter of this time series is the batch number.

Since (i) in Fig. 5.3 (b), the empirical correlation function decays exponentially and (ii) in Fig. 5.3 (c), the empirical partial correlation function cuts off after lag 2, an AR[2] model for the amounts of the product obtained in a chemical batch reactor is estimated using

```
chemreamb.ar2ywfit <- ar.yw(chemreamb, aic=F, order=2).
```

The estimates are $\hat{a}_{12} = -0.3198$, $\hat{a}_{22} = 0.1797$ and $\hat{\sigma}^2_W(2) = 115.88$, and the empirical correlation and partial correlation functions calculated from the empirical residuals of this model are plotted in Fig. 5.4. Comparing these plots with the example plots in Fig. 2.22, it is concluded that they stem from a white noise process and thus that the AR[2] model fitted above is suitable for the product obtained in a chemical batch reactor.

The example time series analysed above, i.e., the wind speeds measured in a turbulent atmospheric flow and the amounts of the product obtained in a chemical batch reactor, demonstrate that

1. the order of an AR[p] model reconcilable with the observations can be found by analysing plots of the empirical correlation and empirical partial correlation functions calculated from the time series, and
2. diagnostics can be obtained by analysing its empirical residuals.

The diagnostics for an AR[p] model are introduced systematically in Sect. 5.4.

5.2 Estimating an AR[p] Model

In Sect. 5.1, an AR[p] model is estimated by applying the Levinson-Durbin recursion to solve the Yule-Walker equations (with empirical covariances substituting the model covariances). A solution is obtained if the empirical covariance function is positive definite (introduced in the remarks to (2.7,2)). (The empirical covariance function calculated from a time slice of a realisation of a stationary stochastic process is non-negative definite as proposed in the remarks to (2.58) and derived in Problem 2.34.) The Yule-Walker estimates inherit the properties (2.58) of the empirical covariance function and thus are not unbiased when calculated from a short time series.

Estimates superior to the Yule-Walker ones are obtained with the following methods: Burg's algorithm as defined in Sect. 5.2.1, regression techniques as proposed in Sect. 5.2.2, and maximum-likelihood procedures as introduced in Sect. 5.2.3. For a Burg estimate the symbol ˜ is used, for a regression estimate the ˇ, and for a maximum-likelihood estimate the ˘.

5.2.1 Burg's Algorithm

Burg's algorithm is a version of the Levinson-Durbin recursion. In the original version as defined in (5.14,1) and (5.79), the $\theta_k = a_{kk}$ are estimated such that the expectations (5.80) and (5.81) become identically zero, a precondition for applying the recursion to obtain the error of the forward prediction (5.78). Since (5.80) and (5.81) hold for arbitrary $\theta_k = a_{kk}$ and $j = 1, \ldots, k-1$, it is feasible to estimate a_{kk} such that (i) the prediction errors in (5.80) and (5.81) become minimal, and (ii) the other properties of the Levinson-Durbin recursion still apply [26]. In the following derivation of Burg's algorithm it is shown that (5.78) still holds when a_{kk} is not estimated as required in (5.14,1), and, thereafter, Burg's estimator is defined.

Given the time slice (x_t), $t = 1, \ldots, N$, from a realisation of a stationary stochastic process (X_t), the parameters in an AR[p] model (5.2) for (X_t) are estimated by solving the pertaining empirical Yule-Walker equations using the Levinson-Durbin recursion. Assume that $\acute{a}_{1,k-1}, \ldots, \acute{a}_{k-1,k-1}$ is an arbitrary estimate (the Yule-Walker estimate (5.7,1) is not required) for the coefficients $a_{1,k-1}, \ldots, a_{k-1,k-1}$ in a AR[$k-1$] model for (X_t), provided that $2 \le k \le p$. Then substitute $\acute{a}_{k-1,k-1}$ in (5.8) and (5.72) to obtain the prediction errors for $2 \le k \le p$:

$$\acute{e}_t^{(v)}(k-1) = X_t - \acute{X}_t^{(v)}(k-1) = X_t - \sum_{j=1}^{k-1} \acute{a}_{j,k-1} X_{t-j}, \quad k \le t \le N \tag{5.17}$$

$$\begin{aligned} \acute{e}_{t-k}^{(r)}(k-1) &= X_{t-k} - \acute{X}_{t-k}^{(r)}(k-1) \\ &= X_{t-k} - \sum_{j=1}^{k-1} \acute{a}_{j,k-1} X_{t-k+j}, \quad k+1 \le t \le N+1 \end{aligned} \tag{5.18}$$

In the next recursion k, an arbitrary estimate $\acute{a}_{kk}$ is used (it is not necessary to use (5.14,1)); the $\acute{a}_{jk}$, $j = 1, \ldots, k-1$, however, is calculated as required in (5.14,2). The error of the forward prediction at time point t for $2 \le k \le p$, $\acute{e}_t^{(v)}(k)$ obtained in (5.19),

$$\begin{aligned}
\acute{e}_t^{(v)}(k) &= X_t - \sum_{j=1}^{k} \acute{a}_{jk} X_{t-j} \\
&= X_t - \sum_{j=1}^{k-1} \big(\acute{a}_{j,k-1} - \acute{a}_{kk}\acute{a}_{k-j,k-1}\big) X_{t-j} - \acute{a}_{kk} X_{t-k} \\
&= X_t - \sum_{j=1}^{k-1} \acute{a}_{j,k-1} X_{t-j} - \acute{a}_{kk}\left(X_{t-k} - \sum_{j=1}^{k-1} \acute{a}_{k-j,k-1} X_{t-j}\right) \\
&= \acute{e}_t^{(v)}(k-1) - \acute{a}_{kk}\acute{e}_{t-k}^{(r)}(k-1) \qquad k+1 \le t \le N+1 \qquad (5.19) \\
\acute{e}_{t-k}^{(r)}(k) &= \acute{e}_{t-k}^{(r)}(k-1) - \acute{a}_{kk}\acute{e}_{t-k}^{(v)}(k-1) \qquad \text{derived as } \acute{e}_t^{(v)}(k-1) \qquad (5.20)
\end{aligned}$$

is the reconstruction of an empirical version of (5.78) for arbitrary estimates $\acute{a}_{kk}$. Both errors (5.19) and (5.20) apply to arbitrary estimates for a_{kk}, provided that (i) a_{jk}, $j = 1, \ldots, k-1$, is calculated as required in (5.14,2) and (ii) $2 \le k \le p$. Thus, in recursion k, other estimates for a_{kk} than those given in (5.14,1) are possible. In [26], Burg proposes the estimator (5.21):

If (x_t), $t = 1, \ldots, N$, is a time slice from a realisation of an AR[p] model (X_t) as defined in (5.2), then Burg's estimator $\tilde{a}_{kk}$ for a_{kk} in (5.5), (5.14) and (5.15) minimises the sum of the squared errors of the forward and backward predictions:

$$1.\ \sum_{t=k+1}^{N} \left(\left(\tilde{e}_t^{(v)}(k)\right)^2 + \left(\tilde{e}_{t-k}^{(r)}(k)\right)^2\right) \quad \textit{minimal for } 2 \le k \le p \qquad (5.21)$$

$$2.\ \tilde{a}_{kk} = \frac{2\sum_{t=k+1}^{N}\left(\tilde{e}_t^{(v)}(k-1)\right)\left(\tilde{e}_{t-k}^{(r)}(k-1)\right)}{\sum_{t=k+1}^{N}\left(\left(\tilde{e}_t^{(v)}(k-1)\right)^2 + \left(\tilde{e}_{t-k}^{(r)}(k-1)\right)^2\right)}$$

(5.21,2) is derived from (5.21,1) in Problem 5.8. Further, $|\tilde{a}_{kk}| \le 1$ is a desirable property of an estimator for the partial correlations (5.15), since $0 \le \big(\tilde{e}_t^{(v)}(k-1) \pm \tilde{e}_{t-k}^{(r)}(k-1)\big)^2 = \big(\tilde{e}_t^{(v)}(k-1)\big)^2 + \big(\tilde{e}_{t-k}^{(r)}(k-1)\big)^2 \pm 2\tilde{e}_t^{(v)}(k-1)\tilde{e}_{t-k}^{(r)}(k-1)$ implies $|2\tilde{e}_t^{(v)}(k-1)\tilde{e}_{t-k}^{(r)}(k-1)| \le \big(\tilde{e}_t^{(v)}(k-1)\big)^2 + \big(\tilde{e}_{t-k}^{(r)}(k-1)\big)^2$.

The recursion is initialised by minimizing (possibly subsequent to demeaning the time series), the sum $\sum_{t=2}^{N}\big((\tilde{e}_t^{(v)}(1))^2 + (\tilde{e}_{t-k}^{(r)}(1))^2\big) = \sum_{t=2}^{N}\big((X_t - \tilde{a}_{11}X_{t-1})^2 + (X_{t-1} - \tilde{a}_{11}X_t)^2\big)$ to obtain $\tilde{a}_{11} = 2\sum_{t=2}^{N}\big((X_tX_{t-1})/(X_t^2 + X_{t-1}^2)\big)$. The variance of the innovations is estimated by applying (5.14,3) to obtain $\tilde{\sigma}_W^2(k) = \tilde{\sigma}_W^2(k-1)(1-\tilde{a}_{kk})$, where $\tilde{\sigma}_W^2(0) = (1/N)\sum_{t=1}^{N} X_t^2$ is the estimator for the variance of the de-meaned time series as defined in (2.1,2).

Using Burg's algorithm to solve the Yule-Walker equations, an AR[p] model as defined in (5.2) is obtained, i.e., with the roots of $A(z) = 1 - a_{1p}z - \ldots - a_{pp}z^p$ (with estimates substituted for the model coefficients) being outside the unit circle in the complex plane. Consequently, the estimated model is stationary. Stationary estimates are also obtained when the Yule-Walker equations are solved using another algorithm, e.g., the Levinson-Durbin recursion. However, Burg's estimates are superior to the Yule-Walker estimates obtained solving (5.7,1) since (5.7,1) contains empirical covariances which are possibly biased when the time series is short, cf. the warnings in the remarks to (2.58). Using

`chemrea.ar2burgfit <- ar.burg(chemrea, order=2) #Burg's estimate,`

an AR[2] model for the de-meaned amounts of the product obtained in a chemical batch reactor in Fig. 5.3 is estimated to obtain $\tilde{a}_{12} = -0.3324$, $\tilde{a}_{22} = 0.1831$ and $\tilde{\sigma}^2_W(1) = 115.80$ ($\tilde{}$ for Burg's estimate), being slightly different from the Yule-Walker estimates obtained in Sect. 5.1.3 since the time series is short.

5.2.2 Regression Estimates

For an estimation with regression techniques, the AR[p] model as defined in (5.2) is re-written below, in (5.22), with e_t substituting W_t and using only one index for the coefficients.

$$X_t - \mu_X = a_1(X_{t-1} - \mu_X) + a_2(X_{t-2} - \mu_X) + \ldots + a_p(X_{t-p} - \mu_X) + e_t \quad (5.22)$$

(5.22) is a linear model, however without parameter b_0, as defined in (3.7), since the innovations (W_t) as defined in (5.2) share the properties of the residuals e_t of a linear model as required in (3.11).

The coefficients $a_1, \ldots, a_p$ of the linear model in (5.22), and the variance of the residuals σ_e^2, are estimated using regression methods, as introduced in Sect. 3.2, from a time series (x_t), $t = 1, \ldots, N$. The predictor variables in (5.22) are the de-meaned random variables $X_{t-1} - \mu_X, X_{t-2} - \mu_X, \ldots, X_{t-p} - \mu_X$. Hence, the estimates are conditional, given the realisations $x_{t-1} - \hat{\mu}_X, x_{t-2} - \hat{\mu}_X, \ldots, x_{t-p} - \hat{\mu}_X$.

To arrive at the regression estimates $\breve{\boldsymbol{a}}$ of the coefficients in (5.22), the time slice $t = 1, \ldots, N$ from the model (X_t) to be estimated, and its copies with lags $1, 2, \ldots, p$ (i.e., displaced backwards by $1, 2, \ldots, p$ time steps), are written in matrix $\left(\boldsymbol{X}_t, \boldsymbol{X}_p^{(\uparrow)}\right)$ in (5.23). $\left(\boldsymbol{X}_t, \boldsymbol{X}_p^{(\uparrow)}\right)$ contains $N + p$ rows and p columns, NA for missing values.

For the part of $\left(\boldsymbol{X}_t, \boldsymbol{X}_p^{(\uparrow)}\right)$ without NAs the symbol $(\boldsymbol{X}_t, \boldsymbol{X}_p)$ is used, and the matrix of the explanatory variables $\boldsymbol{X}_p$ is defined as $(\boldsymbol{X}_t, \boldsymbol{X}_p)$ without the column containing the variable $\boldsymbol{X}_t$ to be predicted. $(\boldsymbol{X}_t, \boldsymbol{X}_p)$ can be used to obtain least squares estimates $\breve{\boldsymbol{a}} = \left(\boldsymbol{X}_p^T \boldsymbol{X}_p\right)^{-1} \boldsymbol{X}_p^T \boldsymbol{X}_t$ as required in (3.9). These are the *regression estimates* for the coefficients in an AR[p] model.

$$
\begin{array}{l|cccc}
t & X_t & X_{t-1} & X_{t-2} & \ldots \; X_{t-p} \\
-p & \mathrm{NA} & \mathrm{NA} & \mathrm{NA} & X_1 \\
\vdots & & & & \\
-1 & \mathrm{NA} & \mathrm{NA} & X_1 & X_{p-1} \\
0 & \mathrm{NA} & X_1 & X_2 & X_p \\
1 & X_1 & X_2 & X_3 & X_{p+1} \\
\vdots & & & & \\
N-p & X_{N-p} & X_{N-p+1} & X_{N-p+2} & X_N \\
\vdots & & & & \\
N-1 & X_{N-1} & X_N & \mathrm{NA} & \mathrm{NA} \\
N & X_N & \mathrm{NA} & \mathrm{NA} & \mathrm{NA}
\end{array}
= \left(\boldsymbol{X}_t, \boldsymbol{X}_p^{(\uparrow)}\right) \tag{5.23}
$$

When the estimates are calculated from a de-meaned time series then $\big(1/(N-p+1)\big)\boldsymbol{X}_p^T\boldsymbol{X}_p = \check{\boldsymbol{C}}_X(p)$ is an estimate for the covariance matrix until (but without) lag p, obtained from $\boldsymbol{X}_p$ with lagged copies of (X_t), i.e., copies of (X_t) displaced in backward direction. Using $\boldsymbol{X}_p^T\boldsymbol{X}_p$, where $\boldsymbol{X}_p$ is a matrix containing time slices $t = 1, \ldots, N$ of the process (X_t) displaced in forward direction by $1, 2, \ldots, p$ time steps, estimates close to $\check{\boldsymbol{C}}_X(p)$ are obtained. Both estimates are close to each other in values since, although the covariance matrix of a stationary process is a symmetric Toeplitz matrix (introduced in the remarks to (5.15)), $\boldsymbol{X}_p$ does not contain exactly identical values when calculated from copies of a time slice of a realisation (i) displaced backwards and (ii) displaced forwards. Does the direction influence the estimates in the case of the de-meaned wind speed series as plotted in Fig. 5.1 (a)? An answer is given in Problem 5.10.

$\check{\boldsymbol{C}}_X(p)$ is, however, not identical with $\hat{\boldsymbol{C}}_X(p)$ in (5.7), since $\hat{\boldsymbol{C}}_X(p)$ contains the empirical covariances calculated as required in (2.1,3). The difference between $\hat{\boldsymbol{C}}_X(p)$ and $\check{\boldsymbol{C}}_X(p)$ is to be analysed further on in Problem 5.9.

In applications, $\check{\boldsymbol{C}}_X(p)$ depends on the construction of the matrix $(\boldsymbol{X}_t, \boldsymbol{X}_p)$ in (5.23): in R for example, using `ts.intersect()` introduced in the remarks to Fig. 2.2.

For instance, an AR[2] model for the de-meaned wind speeds in a turbulent atmospheric flow plotted in Fig. 5.1 is estimated using the regression methods available in R. From the de-meaned R time series `wmb` generated in the remarks to Fig. 5.1 and its lagged (i.e., displaced backwards) copies, using

```
wmbb1 <- lag(wmb,1)
wmbb2 <- lag(wmb,2)
wmb.frame <- data.frame(ts.intersect(wmb,wmbb1,wmbb2))
```

the R dataframe `wmb.frame` is constructed. Then, applying the linear model

```
wmb.ar2lmmod <- wmb ~ wmbb1 + wmbb2 -1
wmb.ar2lmfit <- lm(wmb.ar2lmmod, wmb.frame)
```

estimates $\check{a}_1 = 0.7240$, $\check{a}_2 = 0.1488$ and $\check{\sigma}_\epsilon^2 = 0.3951$ are obtained, $\check{}$ denoting a regression estimate for an AR[p] model. The regression estimates are, for practical purposes, identical with the Yule-Walker estimates $\hat{a}_{12} = 0.7236$, $\hat{a}_{22} = 0.1492$ and $\hat{\sigma}_W^2(2) = 0.396$ computed in the remarks to Fig. 5.2.

When an AR[p] model is estimated with regression methods as proposed in the remarks to (5.23), i.e., when a linear model for a de-meaned time series with its lagged copies as explanatory variables is estimated using R function `lm()` as demonstrated above, then

1. no special software for the analysis of time series is needed
2. $(\boldsymbol{X}_t, \boldsymbol{X}_p)$ can be constructed from copies of the time series under analysis, displaced either forwards or backwards
3. the usual regression statistics (including the residuals) are calculated from which, as demonstrated in Sect. 3.3, diagnostics for the estimated linear model can be obtained; furthermore, confidence intervals for the estimated autoregressive coefficients can be calculated provided that the residuals are found to be normally distributed.

The analysis of the residuals plays a prominent role in the diagnostics of an estimated AR[p] model, as is demonstrated in the remarks to Figs. 5.2 and 5.4. If the residuals are found to be a time slice from the realisation of a white noise process then the stochastic part of the model fitted is in agreement with (i) the properties of the innovations as required in (5.2) and (5.3) in the case of a Yule-Walker or Burg's estimate, and (ii) the properties of the residuals as required in (3.11) in the case of a regression estimate. The diagnostics introduced in Sect. 5.4.4 can be systematically applied to the estimates for an AR[p] model, calculated using whatever method selected (not only regression estimates).

Contrary to the Yule-Walker or Burg's estimates, regressions methods do not always result in an stationary model as required in (5.2). Therefore, having estimated an AR[p] model using R function `lm()`, apply R function `polyroot()` to ensure that the estimated model is stationary. `polyroot()` calculates the roots of the characteristic polynomial pertaining to (the sequence of) the estimated coefficients as defined in (2.46). For example,

```
plot(polyroot(c(1, -0.7240, -0.1488)))       #Problem 2.21
```

shows that the regression estimates $\check{a}_1 = 0.7240$ and $\check{a}_2 = 0.1488$, obtained above for the coefficients in the AR[2] model for the de-meaned wind speeds, result in a stationary model.

5.2.3 Maximum Likelihood Estimates

An AR[p] model (5.2) can also be estimated using the maximum likelihood (ML) method. When the likelihood function for the covariances of an AR[p] process is calculated, the very same difficulties are encountered as when the likelihood function for the covariances of an stationary process in (5.24)

1. *Let* $(X_1, \ldots, X_N)^T = \boldsymbol{X}$ *be a time slice from a stationary (as defined in (2.8)) stochastic process* (X_t) *with* $\mu_X = 0$ *as well as* $c_X(\tau)$ *and let* $(x_1, \ldots, x_N)$ *be observations of* $\boldsymbol{X}$.
2. *In addition, let* (X_t) *be a normal process as defined in (2.3) and thus* $\boldsymbol{X}$ *be multivariate normally distributed with covariance matrix* $\boldsymbol{C}_X(N)$ *with determinant* $|\boldsymbol{C}_X(N)|$. *Then*
3. $$\mathrm{L}\big(\boldsymbol{C}_X(N)|\boldsymbol{X}\big) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}_X(N)|^{1/2}} \exp\Big(-(1/2)\boldsymbol{X}^T\big(\boldsymbol{C}_X(N)\big)^{-1}\boldsymbol{X}\Big)$$
   is the likelihood function of $\boldsymbol{C}_X(N)$, *given* $\boldsymbol{X}$.

(5.24)

is calculated, i.e., the time needed for the computation of the determinant and the inverse of $\boldsymbol{C}_X(N)$ in (5.24,3) is prohibitively large. To make a computation of the likelihood function $\mathrm{L}\big(\boldsymbol{C}_X(N)|\boldsymbol{X}\big)$ in (5.24,2) feasible, algorithms had to be invented. A first possibility is to use the properties of $\boldsymbol{C}_X(N)$ (the covariance matrix of a stationary stochastic process is a symmetric Toeplitz matrix as introduced in the remarks to (5.15)) and to compute $\mathrm{L}\big(\boldsymbol{C}_X(N)|\boldsymbol{X}\big)$ using a Cholesky decomposition of $\boldsymbol{C}_X(N)$. A description of these algorithms can be found in [21]. As a second possibility, the AR[p] process is written as a state space model ([85], [45]) and its likelihood function is computed using Kalman recursion as described in [55]. Using Kalman recursion, the likelihood function $\mathrm{L}\big(\boldsymbol{C}_X(N)|\boldsymbol{X}\big)$ in (5.24,2) can also be computed from a time series with missing values as shown in [78].

Taking for granted that the likelihood function $\mathrm{L}\big(\boldsymbol{C}_X(N)|\boldsymbol{X}\big)$ in (5.24,2) can be computed quickly enough for the usual applications, it is assumed that (i) an AR[p] model is reconcilable with a stationary stochastic process (X_t) from whence a realisation (x_t) is available for the time slice $t = 1, \ldots, N$ and (ii) the innovations of the AR[p] model are normally distributed. Under these assumptions, $\boldsymbol{X}$ in (5.24) is, as concluded from the remarks to (1.34), multivariate normally distributed and, since the Yule-Walker equations connect the covariances with the coefficients in an AR[p] model as proposed in (5.13,1), $\boldsymbol{C}_X(N)$ in (5.24) is a function of the parameters $\boldsymbol{a}(p)$ and $\sigma_W^2(p)$ to be estimated. Consequently, under the above assumptions, the likelihood function (5.25) is a function of the parameters to be estimated:

$$\mathrm{L}\big(\boldsymbol{a}(p), \sigma_W^2(p)|\boldsymbol{X}\big) = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}_X(N)|^{1/2}} \exp\Big(-(1/2)\boldsymbol{X}^T\big(\boldsymbol{C}_X(N)\big)^{-1}\boldsymbol{X}\Big) \tag{5.25}$$

The ML estimate (denoted with $\breve{}$) of the AR[p] model are those values $\breve{\boldsymbol{a}}(p)$ and $\breve{\sigma}_W^2(p)$ for which $\mathrm{L}\big(\breve{\boldsymbol{a}}(p), \breve{\sigma}_W^2(p)|\mathbf{X}\big)$ in (5.25) becomes as large as possible. A necessary condition for a maximum in $\mathrm{L}\big(\breve{\boldsymbol{a}}(p), \breve{\sigma}_W^2(p)|\mathbf{X}\big)$ is that its partial derivatives with respect to $\breve{\boldsymbol{a}}(p)$ become identically zero. However, the likelihood function in (5.25) is not linear in $\breve{\boldsymbol{a}}(p)$ and therefore, using the partial derivatives, the equations obtained are not a linear system with a straightforward solution. Nevertheless, if initial values for $\breve{\boldsymbol{a}}(p)$ and $\breve{\sigma}_W^2(p)$ are available for which (5.25) is close to the maximum, then the estimates

$\breve{\boldsymbol{a}}(p)$ and $\breve{\sigma}_W^2(p)$ can be obtained iteratively. Initial values become available by (i) solving the Yule-Walker equations using Burg's algorithm, (ii) applying regression methods, or (iii) calculating least squares estimators for the conditional model in (5.25) as described in [16]. These estimators are usually, e.g., in [62], known as conditional sum of squares estimators.

In R, ML estimates of an AR[p] model are calculated using `arima()`. `arima()` initialises the parameters with conditional sum of squares estimates and then computes the likelihood function via a state space representation of the model and Kalman recursion. An AR[p] model estimate computed using `arima(...,transform.pars=T,...)` is, due to constraints applied to the estimated coefficients, always stationary.

R function `arima(...,order=c(p,d,q),...)` is designed for estimating the ARIMA[p, d, q] model defined in (5.55). From this definition as well as from the remarks to (5.27) it is learned that an AR[p] model corresponds to an ARIMA[$p, 0, 0$] model. `arima()` can be applied to stationary time series having a non-zero mean; when doing so, an ML estimate for the mean of the time series under analysis is calculated. `arima()` is used, in Sects. 5.3.4 and 5.4.3, to compute ML estimates for both an ARIMA[0,1,1] and an ARIMA[1,0,1] model.

For example, from the original time series `wind` as plotted in Fig. 2.1 (`wind` is de-meaned to obtain `wmb` in the remarks to Fig. 5.1), R expression

```
windar2mlefit <- arima(wind, order=c(2,0,0),transform.pars=T)
```

computes the following ML estimates (for the variance of the innovations, two autoregressive coefficients and the mean) with their variances, as in the following covariance matrix:

```
> windar2mlefit$sigma2
[1] 0.3950631
windar2mlefit$coef
      ar1       ar2  intercept
0.7231758 0.1489845 -2.5368581
windar2mlefit$var.coef
                     ar1            ar2     intercept
ar1        5.432651e-04 -4.626752e-04 -6.041764e-07
ar2       -4.626752e-04  5.436351e-04 -2.848889e-06
intercept -6.041764e-07 -2.848889e-06  1.331152e-02.
```

In the case of the wind speeds in a turbulent atmospheric flow, the ML estimates obtained above are very close to the Yule-Walker estimates obtained in Sect. 5.1.3 as well as to those estimates calculated with Burg's algorithm and regression methods in Sects. 5.2.1 and 5.2.2.

When the likelihood function is computed as required in (5.25) then $\breve{\boldsymbol{C}}_X(N)$ is obtained as a by-product. Since $\breve{\boldsymbol{C}}_X(N)$ is an estimate for $\boldsymbol{C}_X(N)$, i.e., the covariance matrix of the process, $\breve{\boldsymbol{C}}_X(p)$ becomes available. $\breve{\boldsymbol{C}}_X(p)$ is an estimate for $\boldsymbol{C}_X(p)$, i.e., the covariance matrix (until but without lag p) in (5.6). From $\breve{\boldsymbol{C}}_X(p)$, confidence intervals for the estimates can be calculated

using (5.26).

The asymptotic distributions of the ML estimates for an AR[p] model are derived in, e.g., [21]: For large N,

1. *a multivariate normal distribution with expectations* $(a_{1p}, \ldots, a_{pp})^T = \boldsymbol{a}(p)$ *and covariances* $(1/N)\sigma_W^2(p)\big(\boldsymbol{C}_X(p)\big)^{-1}$, $\boldsymbol{C}_X(p)$ *as in (5.6), is an approximation for the distribution of the maximum likelihood estimates for* $\boldsymbol{a}(p)$,
2. *a normal distribution with expectation* $\sigma_W^2(p)$ *and variance* $2\sigma_W^2(p)/N$ *is an approximation for the distribution of a maximum likelihood estimate for* $\sigma_W^2(p)$, *and*
3. *both maximum likelihood estimates are approximately independent.*

(5.26)

Approximation (5.26,2) strongly depends on (X_t) being a normal process as required in (5.24). (X_t) is a normal process on condition that the innovations are normally distributed, as argued in the remarks to (5.25). Hence, approximation (5.26,2) is applicable only if the innovations are found to be normally distributed from diagnostic plots of the empirical residuals as demonstrated in Sect. 5.4.4 by the examples in Figs. 5.16, and 5.17.

5.2.4 Summary of Sects. 5.1 and 5.2

Proceeding as recommended in (5.1), in Sects. 5.1 and 5.2, AR[2] models are proposed for the de-meaned wind speeds plotted in Fig. 5.1 and also for the de-meaned amounts of the product obtained from a chemical reaction plotted in Fig. 5.3. The order of the model is found by applying (5.16) and then comparing the correlation and partial correlation functions of the empirical residuals of the fitted models with the plots in Fig. 2.22.

The models are estimated using the methods introduced, i.e., by (i) solving the Yule-Walker equations with the Durbin-Levinson recursion, (ii) applying Burg's algorithm, (iii) using regression techniques, and (iv) calculating the maximum-likelihood estimates. When the Yule-Walker equations are solved (or when the estimates are computed using Burg's algorithm), a stationary AR[p] model as required in (5.2) is obtained. Regression and ML estimates are, however, not always stationary: `arima(...,transform.pars=T,...)` produces stationary estimates due to built-in constraints. When confidence intervals for the estimated coefficients are required, regression or ML estimates can be computed provided that, following the recommendations in the remarks to (5.26), the innovations can be shown to be normally distributed using diagnostic plots of the residuals. Then, the assumptions in (3.17) (in the case of a regression estimate) or in (5.24) (in the case of an ML estimate) apply.

The approximations as proposed in (5.26) for the distributions of the ML estimates for the parameters in an AR[p] model do not only apply to ML

Table 5.1. Estimated AR[2] models for the de-meaned time series of the wind speeds in a turbulent atmospheric flow and of the amounts of the product of a chemical reaction as plotted above, in Figs. 5.1 and 5.3.

time series	parameter	Yule-Walker `ar.yw()`	Burg `ar.burg()`	regression `lm()`	maximum-likelihood `arima()`
wind speeds	a_{12}	0.7236	0.7237	0.7240	0.7232
in a turbu-	$\sigma_{a_{12}}$			0.0233	0.0233
lent atmos-	a_{22}	0.1492	0.1489	0.1488	0.1490
pheric flow	$\sigma_{a_{22}}$			0.0233	0.0233
$N = 1800$	σ_W^2	0.3959	0.3951	0.3951	0.3951
product	a_{12}	−0.3198	−0.3324	−0.3422	−0.3407
of a	$\sigma_{a_{12}}$			0.1204	0.1218
chemical	a_{22}	0.1797	0.1831	0.1761	0.1873
reaction	$\sigma_{a_{22}}$			0.1165	0.1223
$N = 70$	σ_W^2	115.88	115.80	108.78	112.72

estimates but also, as is shown in [21], to the other estimates (Yule-Walker, Burg's and regression estimates) introduced in Sects. 5.1 and 5.2, provided that the innovations are normally distributed. Hence, these estimates are close in value to the ML ones if the observed time series is long and the empirical residuals of the model are approximately normally distributed. Yule-Walker estimates, calculated from a short time series, are inferior to those obtained using the other methods since the empirical covariances obtained from a short time series are possibly biased, as argued in the remarks to (2.58).

Table 5.1 contains four estimated AR[2] models for the de-meaned time series of (i) the wind speed measurements in a turbulent atmospheric flow as plotted in Fig. 5.1, and (ii) the amounts of the product obtained from a chemical reaction as plotted in Fig. 5.3. The estimates for the long time series are considered to be identical for practical purposes, e.g., the calculation of predictions. The estimates for the short times series are, however, not close in value for the reason given above.

Thereafter, as proposed in (5.1,3,4), diagnostics are applied to the estimates and a model is selected. Prior to further demonstrations of (5.1) in Sect. 5.4, ARMA$[p, q]$ and ARIMA$[p, d, q]$ models are introduced in Sect. 5.3.

5.3 MA[q] , ARMA[p, q] and ARIMA[p, d, q] Models

In Sect. 2.3.4, an MA[1] process (Y_t) as defined below, in (5.27), is found to be a suitable model for the differences of the Basel yearly temperature values, as plotted above in Fig. 2.14.

$$\begin{array}{l}\Delta^1(X_t) = X_t - X_{t-1} = Y_t = W_t - 0.9 \times W_{t-1},\ (W_t)\ \textit{being}\\ \textit{a white noise process with}\ \mu_W = 0\ \textit{and}\ \sigma_W = 0.7^0\text{C}.\end{array} \tag{5.27}$$

The model (5.27) is an example of an ARIMA[p, d, q] process with (i) $d = 1$: the first differences $\Delta^1(X_t)$ of (X_t), as defined in(2.39,2), are required to be (ii) $q = 1$: an MA[1] process with $b = 0.9$, as defined in (2.21), and, (iii) $p = 0$: the model does not contain any autogressive terms as defined in (5.2). ARIMA[p, d, q] models are introduced in Sect. 5.3.3.

With $d = 0$ and $p > 0$ and $q > 0$, the process becomes an ARIMA[$p, 0, q$] model or an ARMA[p, q] model, i.e., a combination of an AR[p] with an MA[q] model, as defined in Sect. 5.3.2. The existence of ARMA[p, q] models is supposed in the remarks closing Sect. 2.3.5.

With $d = 0$ and $p = 0$ and $q > 0$, the process reduces to an MA[q] model as defined in Sect. 5.3.1.

5.3.1 MA[q] Model

The moving-average process is constructed as a weighted sum of a white noise process using a one-sided sequence of weights as defined in (5.28):

$$\begin{array}{l}(Y_t)\ \textit{is called a moving average process of order}\ q,\\ \textit{(MA[q] process, MA[q] model) provided that:}\\ 1.\ Y_t - \mu_Y = W_t - b_1 W_{t-1} - \ldots - b_q W_{t-q} = \sum_{u=0}^{q} b_u W_{t-u} = (b_t) * (W_t)\\ 2.\ (b_t) = \begin{cases} 1, -b_1, -b_2, \ldots, -b_q & \textit{for}\ t = 0, 1, \ldots, q\\ 0 & \textit{otherwise}\end{cases}\\ 3.\ (W_t)\ \textit{is a white noise process with}\ \mu_W = 0\ \textit{and}\ \sigma_W^2.\end{array} \tag{5.28}$$

If the expectation function μ_Y of a stationary stochastic process (Y_t) is not identically zero, $\mu_Y \neq 0$, then the MA[q] model is fitted to the de-meaned process, as required in (5.28,1), using $\hat{\mu}_X$ as defined in (2.1,1) with the properties (2.56) as an estimate for μ_Y.

The moment functions of an MA[q] model are calculated using (2.26) with $b_0 = -1$, i.e., the sequence of weights becomes $(b_t) = (-b_0, -b_1, \ldots, -b_q)$):

$$\mathrm{E}(Y_t - \mu_Y) = \mathrm{E}\left(\sum_{u=0}^{q} b_u W_{t-u}\right) = \mu_W \sum_{u=0}^{q} (b_u) = 0 \tag{5.29}$$

$$c_Y(\tau) = \sum_{v=0}^{q} \sum_{w=0}^{q} b_v b_w c_W(\tau + w - v) \tag{5.30}$$

$$= \begin{cases} 0 & \text{for } \tau > q \\ \sigma_W^2 \sum_{u=0}^{q-\tau} b_u b_{u+\tau} & \text{for } 0 \le \tau \le q \\ c_Y(-\tau) & \text{for } \tau < 0 \end{cases} \tag{5.31}$$

From (5.29) and (5.31) it is concluded that an MA[q] model is stationary for weights (b_t) being absolutely convergent as defined in (2.24). The covariance function cuts off after lag q: $c_Y(\tau) = 0$ for $\tau > q$ and is thus, with σ_W^2, proportional to the autocorrelation as defined in (6.103) of the weights $(1, -b_1, \ldots, -b_q)$. The covariance function $c_Y(\tau)$ is scaled with the variance $\sigma_Y^2 = c_Y(0) = \sigma_W^2 \sum_{u=0}^{q} (b_u)^2$ to obtain the correlation function $\varrho_Y(\tau)$.

(5.30) relates the covariance function of an MA[q] model to its weights. Given the weights (b_u), q equations are obtained with the substitution $q = 1, 2, \ldots, q$. These equations can be solved for the weights (b_u) with iterative procedures (except for $q = 1$ in (2.23)) since they are, unlike the Yule-Walker equations introduced in Sect. 5.1.1, not linear. Besides being non-linear, these equations have multiple solutions, e.g., the quadratic equation pertaining to the MA[1] model in the remarks to (2.23) has two solutions. Only by assuming an additional property, i.e., the invertibility of the MA[q] model, the number of solutions for each coefficient in (b_t) can be be restricted to one.

A MA[q] model $Y_t - \mu_Y = \sum_{u=0}^{q} b_u W_{t-u} = (b_t) * (W_t)$ *with* $(b_t) = (1, -b_1, \ldots, -b_q)$ *is invertible provided that* $(1, -b_1, \ldots, -b_q)^{-1}$ *converges absolutely.* (5.32)

If $Y_t - \mu_Y$ is assumed to be an invertible MA[q] model then W_t, the innovation at time point t, can be calculated from $Y_t,\ Y_{t-1},\ Y_{t-2},\ \ldots$, and, if the complete past of the process is known, the sum $\big((1, -b_1, \ldots, -b_q)^{-1}\big) * (Y_t - \mu_Y)$ can be calculated. This favourable property allows for calculating predictions, as proposed in Sect. 5.5, from a finite past of the process, as stipulated in (5.33,1).

The following properties belong to an invertible MA[q] model $(Y_t) = \sum_{u=0}^{q} b_u W_{t-u} = (b_t) * (W_t)$*:*

1. W_t *can be calculated from* $Y_t,\ Y_{t-1},\ Y_{t-2},\ \ldots$ *for use in a prediction.*
2. *Its covariance function* $c_Y(\tau)$ *and thus also its correlation function* $\rho_Y(\tau)$ *cut off after lag* q*.*
3. *Its partial correlation function* $a_Y(k, k)$ *damps out.*

(5.33) (5.33,1)

(5.33,1) is derived using the AR[∞] representation (an AR[p] model as defined in (5.2) with $p \to \infty$) of an invertible MA[q] model: convolve $Y_t - \mu_Y = (b_t) * (W_t)$ on both sides with $(a_t) = (b_t)^{-1}$ to obtain $(a_t) * (Y_t - \mu_Y) = W_t$, because the convolution sum converges in mean square to the limiting random variable W_t. For a derivation, use the template in Sect. 7.2.1 where the convergence of the convolution sum (2.19) is shown. Further, using the AR[∞] representation of an invertible MA[q] model together with (5.15)

and (5.16), property (5.33,3) is arrived at, i.e., it is shown that the partial correlation function of the model does not cut off, as $(b_t)^{-1}$ contains an infinite number of weights. Property (5.33,2) directly follows from (5.31).

For instance the MA[1] model (5.27) is invertible since $1/0.9 > 1$. If the model contains two or more weights, then the characteristic polynomom $B(z) = 1 - b_1 z - b_2 z^2 - \ldots - b_q z^q$ pertaining to its coefficients $(b_t) = (1, -b_1, \ldots, -b_q)$ is constructed, and (2.46) is applied using R function `polyroot()`, as demonstrated in Problem 2.21 and in the remarks concluding Sect. 5.2.2.

The covariance and correlation functions pertaining to an invertible MA[1] model are calculated in (2.23). In (5.34), the partial correlation function for lag 1 of this model is obtained using the Levinson-Durbin recursion (5.14), and, for lag k, the partial correlation function is derived in (5.86). It is justified to calculate the partial correlation function with the Levinson-Durbin recursion in the above case because, due to (5.13), the Levinson-Durbin recursion applies when calculating the partial correlation function of any stationary stochastic process.

$$a_Y(1,1) = c_Y(1)/c_Y(0) = \frac{-b}{1+b^2} \tag{5.34}$$

$$a_Y(k,k) = \frac{-b^k(1-b^2)}{1-b^{2(k+1)}} \tag{5.35}$$

It can be seen that the partial correlation function $a_Y(k,k)$ of an MA[1] model (Y_t), obtained in both (5.34) and (5.35), decays exponentially. All $a_Y(k,k)$ are negative for $0 < b < 1$ but $a_Y(k,k)$ changes its signs for $-1 < b < 0$.

The MA[2] model $Y_t - \mu_Y = W_t - b_1 W_{t-1} - b_2 W_{t-2}$ is invertible on condition that the roots z_1 and z_2 of $B(z) = 1 - b_1 z - b_2 z^2 = 0$ are outside the unit circle in the complex plane. This condition also applies to the roots of $A(z) = 1 - a_1 z - a_2 z^2 = 0$, a_1 and a_2 being the weights of a stationary AR[2] process. Therefore, an MA[2] model with weights (b_1, b_2) inside the triangle in Fig. 2.17 (substituting b_1 and b_2 for a_1 and a_2) is invertible and its AR[∞] representation $\big((1, -b_1, -b_2)^{-1}\big) * (Y_t - \mu_Y)$ converges.

The covariance function $c_Y(\tau)$ of the MA[2] model is calculated from (5.31) to obtain $c_Y(0) = \sigma_W^2(1 + b_1^2 + b_2^2)$, $c_Y(1) = \sigma_W^2(-b_1(1 - b_2))$, $c_Y(2) = \sigma_W^2(-b_2)$, $c_Y(3) = 0$, $c_Y(4) = 0$, ... and thus, $c_Y(\tau)$ cuts off after lag 2. The partial correlation function $a_Y(k,k)$, however, cannot be arrived at as easily as in the case of the MA[1] model. The properties of $a_Y(k,k)$, (Y_t) an invertible MA[2] model, are therefore demonstrated in the simulation experiments to be performed in Problem 5.11. There, some realisations of an invertible MA[2] model (Y_t) are simulated, and for each simulation the empirical partial correlation function $\hat{a}_Y(k,k)$ is calculated. The $\hat{a}_Y(k,k)$ are then compared with the empirical correlation functions $\hat{c}_X(\tau)$ obtained, in Problem 2.21, from simulated realisations of a stationary AR[2] model (X_t). From these comparisons, it is concluded that the empirical partial correlation

Table 5.2. Properties of the correlation and partial correlation functions of a stationary AR[p] model as defined in (5.2), of an invertible MA[q] model as defined in (5.28) and (5.32) and of a stationary and invertible ARMA[p, q] model as defined in (5.36).

model	correlation function covariance function	partial correlation function
AR[p]	decays exponentially with or without oscillation	cuts off after lag p
MA[q]	cuts off after lag q	decays exponentially with or without oscillation
ARMA[p, q]	exponential decay or damped oscillation after lag q	exponential decay or damped oscillation after lag p

functions calculated from simulations of an invertible MA[2] model behave in the same manner as the empirical correlation functions calculated from simulations of a stationary AR[2] model, i.e., as described in the remarks to (2.52) where the covariance and correlation functions of a stationary AR[2] model are discussed.

As a second example, in Problem 5.11, empirical correlation and empirical partial correlation functions are calculated from simulated realisations of an invertible MA[4] and a stationary AR[4] model and then plotted. As concluded from the plots, the empirical partial correlation functions obtained from the MA[4] model behave identically to those obtained from the AR[4] model.

If the MA[q] model (5.28) is compared with the linear process (2.25), it is then concluded that the linear process is a MA[q] model. However, definition (2.25) includes all stationary AR[p] models as argued in the introduction to Sect. 5.1. Thus, a stationary AR[p] model (X_t) as defined in (5.2), with correlation and partial correlation functions $c_X(\tau)$ and $a_X(k,k)$, has an MA[∞] representation, and, vice-versa, an invertible MA[q] model (Y_t) as defined in (5.28) and (5.32) with correlation and partial correlation functions $c_Y(\tau)$ and $a_Y(k,k)$, has an AR[∞] representation, as shown in the remarks to (5.32). Therefore it becomes plausible that (i) $c_X(\tau)$ behaves as $a_Y(k,k)$ and (ii) $a_X(k,k)$ behaves as $c_Y(\tau)$. This behaviour is deduced, for AR[1]- and MA[1] models on the one hand, from a comparison of (2.16) and (2.23) as well as (5.16) and (5.35); for AR[2] and MA[2] as well as for AR[4] and MA[4] models on the other hand, it is demonstrated by the simulations in Problem 5.11. Thus, the first and second lines in Table 5.2 are obtained.

5.3.2 ARMA[p, q] Model

If both representations, the autoregressive and the moving average one, are used for a linear process, then the properties of the process can be derived from the more suitable representation: in Sect. 2.4.4, the expectation and covariance functions of a stationary AR[2] model are calculated from its MA[∞] representation $X_t = (1, -a_1, -a_2)^{-1} * W_t = \sum_{u=0}^{\infty} b_u W_{t-u}$, i.e., the stationary solution of its difference equation, and in the remarks to definition (5.32), the AR[∞] representation of an invertible MA[q] model is used to derive that the partial correlation function of an MA[q] model decays exponentially, as proposed in (5.33,3). Both models are linear processes and thus can also be joined to an ARMA[p, q] model, as defined in (5.36).

Let (W_t) be a white noise process with $\mu_W = 0$ and σ_W^2, and let $(a_i) = (1, -a_1, \ldots, -a_p)$ as well as $(b_i) = (1, -b_1, \ldots, -b_p)$ both be real-valued sequences with the solutions of their characteristic polynomials outside the complex unit circle. Then

$$\begin{aligned}(X_t - \mu_X) - a_1(X_{t-1} - \mu_X) - \ldots - a_p(X_{t-p} - \mu_X) \\ = W_t - b_1 W_{t-1} - \ldots - b_q W_{t-q}\end{aligned} \tag{5.36}$$

is called an autoregressive moving average process of order p and q (ARMA[p, q] process or model, mixed process or model).

If the expectation function μ_X of a stationary stochastic process (X_t) is not identically zero, $\mu_X \neq 0$, then an ARMA[p, q] model is fitted to the de-meaned process, as required in (5.36).

Usually, an ARMA[p, q] model is parsimonious and is thus preferably fitted to a short time series, as argued in the introduction to Sect. 2.3.5. If a model with m parameters is fitted to a fixed number N of observations, then N/p is larger for a parsimonious model containing a small number of parameters as compared to a non-parsimonious one containing a large number of parameters. Hence, in general, if models with a varying number of parameters are fitted to a time series (examples are given in Table 5.3) then the variances of the estimated parameters will be smaller for parsimonious models, and, as a further consequence, predictions are preferably calculated using that model having the minimal number of parameters (an example is given in Sect. 5.5.2).

The stationary solution of the equation defining the ARMA[p, q] model is a linear process as defined in (2.25), as is concluded from (5.38). (5.38) is obtained using the assumptions in (5.36): subsequent to writing the model as in (5.37), both sides are convolved with the inverse to (a_t):

$$(a_t) * (X_t - \mu_X) = (b_t) * (W_t) \tag{5.37}$$

$$\begin{aligned}(a_t)^{-1} * (a_t) * (X_t - \mu_X) &= (a_t)^{-1} * (b_t) * (W_t) \\ X_t - \mu_X &= \big((a_t)^{-1} * (b_t)\big) * (W_t) \\ &= (1, -\psi_1, -\psi_2, \ldots) * (W_t) = \sum_{u=0}^{\infty} \psi_u W_{t-u}\end{aligned} \tag{5.38}$$

to arrive, in (5.38), at the MA[∞] representation of the ARMA[p, q] model. (5.38) converges since $(\psi_t) = (a_t)^{-1} * (b_t)$, being the convolution of two absolutely convergent sequences, converges absolutely with the roots of $\Psi(z) = (1 - \psi_1 z - \psi_2 z^2 - \ldots)$ outside the unit circle in the complex plane. (5.38) also implies that an ARMA[p, q] model (5.36) is causal, as defined in (5.3), since its autoregressive part is causal.

From (5.38) the expectation and the covariance functions pertaining to an ARMA[p, q] model can be calculated using (2.26).

$$\mathrm{E}(X_t - \mu_X) = \mathrm{E}\left(\sum_{u=0}^{\infty} \psi_u W_{t-u}\right) = \mu_W \sum_{u=0}^{\infty} \psi_u = 0 \tag{5.39}$$

$$\begin{aligned} c_X(\tau) &= \sum_{v=0}^{\infty} \sum_{w=0}^{\infty} \psi_v \psi_w c_W(\tau + w - v) \\ &= \begin{cases} \sigma_W^2 \sum_{u=0}^{\infty} \psi_u^2 & \text{for } \tau = 0 \\ \sigma_W^2 \sum_{u=0}^{\infty} \psi_u \psi_{u+\tau} & \text{for } \tau > 0 \\ c_X(-\tau) & \text{for } \tau < 0 \end{cases} \end{aligned} \tag{5.40}$$

(ψ_t) is obtained as in the derivation of (5.38) or, alternatively, using the characteristic polynomials (2.45) pertaining to the sequences of weights: with $A(z) = 1 - a_1 z^1 - \ldots - a_p z^p$ being the characteristic polynomial pertaining to (a_t) and $B(z) = 1 - b_1 z^1 - \ldots - b_q z^q$ the characteristic polynomial pertaining to (b_t), using $(\psi_t) = (a_t)^{-1} * (b_t)$ in the derivation of (5.38), $\Psi(z) = B(z)/A(z)$, or $A(z)\Psi(z) = B(z)$ are obtained:

$$\begin{aligned} &(\psi_0 + \psi_1 z + \psi_2 z^2 + \ldots)(1 - a_1 z - \ldots - a_p z^p) \\ &\qquad = (1 - b_1 z - \ldots - b_q z^q - 0 \times z^{q+1} - \ldots) \end{aligned}$$

provided that $b_i = 0$ for $i > q$. In Problem 5.12, the left side of this equation is evaluated to obtain the equations for each z^i, $i = 0, 1, \ldots$, in the left column in (5.41). Substituting their solutions sequentially, the weights in the centre column in (5.41)

$$\begin{array}{lll} \psi_0 = 1 & \psi_0 = 1 & i = 0 \\ \psi_1 - \psi_0 a_1 = -b_1 & \psi_1 = a_1 - b_1 & i = 1 \\ \psi_2 - \psi_1 a_1 - \psi_0 a_2 = -b_2 & \psi_2 = a_2 + a_1^2 - a_1 b_1 - b_2 & i = 2 \\ \ldots & \ldots & \ldots \end{array} \tag{5.41}$$

as well as the general solution in (5.42)

$$\begin{array}{ll} \psi_i - \sum_{j=1}^{p} \psi_{i-j} a_j = 0 & i = q+1, q+2, \ldots \\ \psi_i - \sum_{j=1}^{i} \psi_{i-j} a_j = -b_i & i = 1, \ldots, p \text{ and } p \le q \\ \psi_i - \sum_{j=1}^{p} \psi_{i-j} a_j = -b_i & i = p+1, \ldots, q \text{ and } p \le q \\ \psi_i - \sum_{j=1}^{i} \psi_{i-j} a_j = -b_i & i = 1, \ldots, q \text{ and } p > q \end{array} \tag{5.42}$$

are arrived at. (5.42) contains, in the second and third lines, the weights ψ_i, $i = 1, \ldots, q$, being a function of (b_i) and (a_i) for $p \le q$. In the first and fourth

lines in (5.42), the weights ψ_i, $i = 1, \ldots, q$, are a function of (b_i) and (a_i) for $p > q$; however, the ψ_i, $i = q, q+1, \ldots, p$, are only a function of the (a_i). The dependences of the ψ-weights are inherited by the covariances $c_X(\tau)$ of an ARMA[p, q] model (X_t), $c_X(\tau)$ being calculated from the ψ-weights in (5.42).

For the case $p > q$, this result is also arrived at by multiplying both sides of definition (5.36) with X_{t-k} and then taking the expectations (for simplicity, $\mu_X = 0$ is assumed):

$$\begin{aligned} &\mathrm{E}\Big(\big(X_t - a_1 X_{t-1} - \ldots - a_p X_{t-p}\big) X_{t-k}\Big) \\ &\quad = \mathrm{E}\Big(\big(W_t - b_1 W_{t-1} - \ldots - b_q W_{t-q}\big) X_{t-k}\Big) \qquad (5.43) \\ &\mathrm{E}(X_t X_{t-k}) - a_1 \mathrm{E}(X_{t-1} X_{t-k}) - \ldots - a_p \mathrm{E}(X_{t-p} X_{t-k}) \\ &\quad = \mathrm{E}(W_t X_{t-k}) - b_1 \mathrm{E}(W_{t-1} X_{t-k}) - \ldots - b_q \mathrm{E}(W_{t-q} X_{t-k}) \qquad (5.44) \end{aligned}$$

Since $X_{t-k} = a_1 X_{t-k-1} - \ldots - a_p X_{t-k-p} + W_{t-k} - b_1 W_{t-k-1} - \ldots - b_q W_{t-k-1}$ only depends on the innovations until and including time point $t - k$, the covariances on the right side become identically zero for lags $k > q$, and

$$c_X(k) - a_1 c_X(k-1) - \ldots - a_p c_X(k-p) = 0 \quad k = q+1, \ldots p, \; p > q \qquad (5.45)$$

is obtained. The equations in (5.45) connect the covariances of an ARMA[p, q] model (5.36) with its autoregressive weights (a_t) for lags $k > q$, provided that $p > q$. For $q = 0$, however, the equations (5.45) become the Yule-Walker equations as defined in (5.5): the covariance function of an ARMA[p, q] model behaves, for lags $k > q$, as the covariance function of an AR[p] model (Table 5.2).

An ARMA[p, q] model as defined in (5.36) is not only stationary, as concluded from (5.39) and (5.40), but also invertible, because the sum (5.47) converges in mean square, as shown below. This sum, i.e., the AR[∞] representation of an ARMA[p, q] model, is derived from the model written, as in in (5.37), as a convolution in (5.46):

$$\begin{aligned} (1, -a_1, \ldots, -a_p) * (X_t - \mu_X) &= (1, -b_1, \ldots, -b_q) \; * \; (W_t) \qquad (5.46) \\ (1, -b_1, \ldots, b_q)^{-1} * (1, -a_1, \ldots, -a_p) * (X_t - \mu_X) &= (W_t) \\ (1, -\pi_1, -\pi_2, -\pi_3, \ldots) * (X_t - \mu_X) &= (W_t) \\ (X_t - \mu_X) - \pi_1 (X_{t-1} - \mu_X) - \pi_2 (X_{t-2} - \mu_X) - \ldots &= W_t \qquad (5.47) \end{aligned}$$

The π-weights, $(1, -\pi_1, -\pi_2, -\pi_3, \ldots)$, converge absolutely, with the solutions of its characteristic polynomial $\Pi(z) = (1 - \pi_1 z - \pi_2 z^2 - \pi_3 z^3 - \ldots)$ being outside $|z| \leq 1$, since $(1, -\pi_1, -\pi_2, -\pi_3, \ldots)$ is generated from $(1, -b_1, \ldots, b_q)^{-1}$ and $(1, -a_1, \ldots, -a_p)$, both being absolutely convergent as required in (5.36). Hence, the sum (5.47) converges in mean square to the limiting random variable W_t.

From the AR[∞] representation of the ARMA[p, q] model it is concluded, using (5.16), that the pertaining partial correlation function decays exponentially (and thus does not cut off) since there is an infinite number of

π-weights. Can the partial correlation function $a_X(k,k)$ of an ARMA$[p,q]$ model (X_t) be calculated from its covariance function using the Levinson-Durbin recursion (5.14)? This calculation is successfully performed in (5.35) and (5.86) in the case of an MA[1] model, but be warned that very complicated expressions result after a few steps even in the case of the most simple mixed process, i.e., the ARMA[1,1] model. Nevertheless, it is plausible that $a_X(k,k)$ behaves, for lags $k > p$, as the partial correlation function of an MA$[q]$ model does (Table 5.2).

For example, the properties of an ARMA[1,1] model are deduced from its correlation and partial correlation functions calculated below, in (5.50), (5.51) and (5.53) as well as in (5.54). The ARMA[1,1] model (X_t),

$$(X_t - \mu_X) - a(X_{t-1} - \mu_X) = W_t - bW_{t-1} \qquad \text{or} \tag{5.48}$$

$$(1, -a) * (X_t - \mu_X) = (1, -b) * (W_t) \tag{5.49}$$

as written in (5.48) or (5.49) (for simplicity, $\mu_X = 0$ is assumed), is stationary and invertible for $|a| < 1$ and $|b| < 1$. Its covariance function $c_X(\tau)$ is calculated in (5.50), (5.51) and (5.53) using (5.40), the sequence (ψ_i) being calculated as required in (5.41) and (5.42): $\psi_0 = 1$, $\psi_1 = a - b$, $\psi_2 = (a-b)a$, $\psi_3 = (a-b)a^2$, ...; or by convolving $(1,-a)^{-1} * (1,-b) = (1, a, a^2, a^3, \ldots) * (1,-b) = \big(1, a-b, a(a-b), a^2(a-b), \ldots\big)$.

$$\begin{aligned} c_X(0)/\sigma_W^2 &= 1^2 + (a-b)^2 + a^2(a-b)^2 + a^4(a-b)^2 + \ldots \\ &= 1 + (a-b)^2\big(1 + a^2 + a^4 + \ldots\big) \\ c_X(0) &= \sigma_W^2 \times \frac{1 - 2ab + b^2}{1 - a^2} \qquad (5.50) \\ c_X(1) &= \sigma_W^2 \times \frac{(a-b)(1-ab)}{1-a^2} \qquad (5.51) \\ c_X(2) &= \sigma_W^2 \times \frac{a(a-b)(1-ab)}{1-a^2} = ac_X(1) \qquad (5.52) \\ &\ldots \\ c_X(\tau) &= ac_X(\tau - 1), \qquad \text{for } \tau \geq 2. \qquad (5.53) \end{aligned}$$

Thus, $c_X(\tau)$ depends on weights a and b and decays exponentially. This exponential decay starts after lag 1, whereas decay of the covariance function of an AR[1] process, as proposed in (2.16), starts with lag 1. The rate of decay and the possibly alternating signs of the covariances depend on both coefficients a and b.

From $c_X(\tau)$, using the Levinson-Durbin recursion (5.14), the partial correlation function $a_X(k,k)$ of the ARMA[1,1] model for the first lag is calculated from the covariances in (5.50) and (5.51):

$$a_X(1,1) = c_X(1)/c_X(0) = \frac{(a-b)(1-ab)}{1 - 2ab + b^2} \tag{5.54}$$

For the following lags, however, the calculations become intricate and therefore, to circumvent these calculations, the exponential decay of $a_X(k, k)$ after the first lag is assessed in the simulation experiment to be performed in Problem 5.13. The partial correlation function of an MA[1] model, as obtained in (5.34) and (5.35), also decays exponentially, however, starting with lag 1.

5.3.3 ARIMA[p, d, q] Model

Fluctuations in the expectation function of a stochastic process generate covariances which disappear when the fluctuations are removed and the expectation function becomes constant as required in (2.8,1). Hence, an empirical correlation function calculated from a time series changes when trends and/or other fluctuations are removed. Examples are given in Figs. 2.23 and 3.1 and also in Figs. 4.17, 4.18 and 4.21.

As a first possibility, a non-stationarity in the first moment function can be removed by subtracting a linear model for a trend estimated using the regression techniques introduced in Chap. 3. When a trend is estimated, (3.1,3) and (3.11) require non-correlated residuals. Correlations remaining in the residuals can be accounted for by calculating a generalised least squares estimate as proposed in (3.14).

As a second possibility, often to be preferred to the first one, a non-stationarity in the first moment function can be removed by calculating the first, second, etc. differences. The first differences are defined in (2.39,2), and differences of a higher order in the remarks to (2.40). For example, in (5.27), an ARIMA[0,1,1] model is proposed for the yearly values in the Basel temperature series.

(X_t) is called an ARIMA[p, d, q] process or an ARIMA[p, d, q] model (ARIMA: autoregressive integrated moving average), provided that $(Y_t) = \Delta^d(X_t)$ is an ARMA[p, q] model as defined in (5.36) with $(Y_t) = \Delta^d(X_t)$, $d = 1, 2, \ldots$, being the differences of order d of (X_t). $\Delta^d(X_t)$ is defined in (2.40). (5.55)

Substituting (5.55) in (5.36), the difference equation of an ARIMA[p, d, q] model is obtained. If $d = 0$, then (X_t) becomes an ARMA[p, q] model with its stationary solution as in (5.38). If $d \geq 1$, then (X_t) is not stationary in the first moment function and therefore, an ARIMA[p, d, q] model is suitable for a time series with a non-constant first moment function.

An ARIMA[p, d, q] model with $d = 1$ is, however, also suitable for stochastic processes satisfying the intrinsic hypothesis as defined in (4.9) (but with one-dimensional integer parameter) since (5.55) requires $\Delta^d(X_t)$ (and not (X_t)) to be stationary. For example, the random walk process (X_t) as defined in (2.20) is an intrinsically stationary process with a parameter being one-dimensional and integer. In passing, as a very short excercise, re-define the random walk process using (5.55).

5.3.4 Estimators for an ARMA[p, q] Model

In the remarks to definition (1.34) it is shown that linear combinations of multivariate normally distributed random variables are multivariate normally distributed as well. Hence, the MA[∞] representation (5.38) of the ARMA[p, q] model (X_t) defined in (5.36) implies proposition (5.56):

Let $(X_1, \ldots, X_N)^T = \mathbf{X}$ *be a sequence of random variables in an ARMA[p, q] model as defined in (5.36) with normally distributed innovations* (W_t). *Then* $\mathbf{X}$ *is multivariate normally distributed.* (5.56)

from which it is concluded below, in the remarks to (5.57), that the maximum-likelihood estimates of an ARMA[p, q] model are normally distributed subject to the condition that the model innovations are normally distributed.

Maximum-likelihood estimates for the parameters in MA[q], ARMA[p, q] and ARIMA[p, d, q] models are obtained using R function `arima()`, which is introduced in Sect. 5.2.3. In `arima()`, however, an ARMA[p, q] model is written as $(X_t - \mu_X) - a_1(X_{t-1} - \mu_X) - \ldots - a_p(X_{t-p} - \mu_X) = W_t + b_1 W_{t-1} + \ldots + b_q W_{t-q}$ with moving-average coefficients differing in sign from definition (5.36). `arima()` calculates efficiently $\mathrm{L}\big(\boldsymbol{C}_X(N)|\boldsymbol{X}\big)$ as defined in (5.24), i.e., the likelihood function of the covariance matrix $\boldsymbol{C}_X(N)$ of a stationary stochastic process (X_t), given $\boldsymbol{X} = (X_1, \ldots, X_N)^T$. If (X_t) is an ARMA[p, q] process as defined in (5.36) then $\boldsymbol{C}_X(N)$ is, due to (5.40) and (5.38), a function of (i) the model coefficients $(a_1, a_2, \ldots, a_p)^T = \boldsymbol{a}(p)$ and $(b_1, b_2, \ldots, b_q)^T = \boldsymbol{b}(q)$, as well as (ii) the variance of the innovations $\sigma^2_W(p, q)$. Consequently, the likelihood function of an ARMA[p, q] model is arrived at in (5.57):

$$\begin{aligned} &\mathrm{L}\big(\boldsymbol{a}(p), \boldsymbol{b}(q), \sigma^2_W(p, q)|\boldsymbol{X}\big) \\ &\quad = \frac{1}{(2\pi)^{N/2}|\boldsymbol{C}_X(N)|^{1/2}} \exp\Big(-(1/2)\boldsymbol{X}^T\big(\boldsymbol{C}_X(N)\big)^{-1}\boldsymbol{X}\Big) \end{aligned} \tag{5.57}$$

The maximum likelihood estimates of the parameters in the ARMA[p, q] model are those $\breve{\boldsymbol{a}}(p)$, $\breve{\boldsymbol{b}}(q)$ and $\breve{\sigma}^2_W(p, q)$, for which (5.57) reaches a maximum.

If an ARIMA[p, d, q] model is reconcilable with $\boldsymbol{X}$ then the differences (of order d) of $\boldsymbol{X}$ are assumed be a time slice from an ARMA[p, q] model.

If the innovations (W_t) in the ARMA[p, q] model are normally distributed then the estimates $\breve{\boldsymbol{a}}(p)$, $\breve{\boldsymbol{b}}(q)$ and $\breve{\sigma}^2_W(p, q)$ are multivariate normally distributed since $\boldsymbol{X}$ is multivariate normally distributed as proposed in (5.56) and thus, the properties enumerated in (5.25) also apply to the maximum-likelihood estimates of an ARMA[p, q] model. If, however, the innovations (W_t) are not normally distributed, then the estimates are not multivariate normally distributed either, and the calculation of confidence intervals becomes hazardous: caution is recommended, as proposed in the remarks to (5.26).

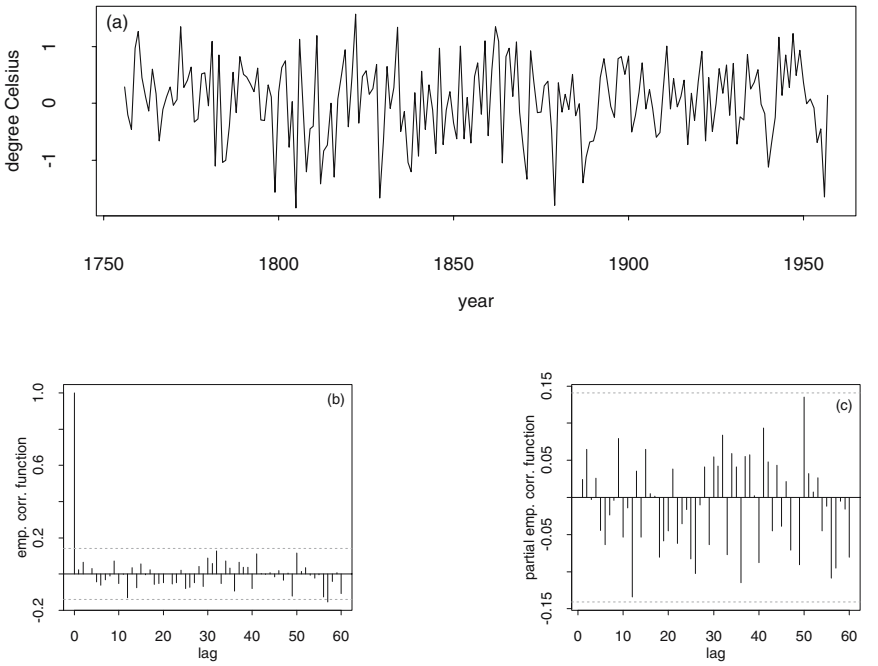


Fig. 5.5. Above, the empirical residuals of an ARIMA[0,1,1] model fitted to the Basel yearly temperature values plotted in Fig. 2.13 (a); below, the moment functions of the empirical residuals.

For example, the ARIMA[0,1,1] process given in (5.27) as an introductory example is found, in Sect. 2.3.4, to be a suitable model for the yearly means in the Basel temperature series as plotted in Fig. 2.14. Can the estimates $\sigma_W^* = 0.7^0$C and $b^* = 0.9$, obtained by trial and and error in Sect. 2.3.4, be improved upon using the theory and methods now available?

Assuming that an ARIMA[0,1,1] model is suitable for Basel yearly temperature values being stored in R time series `tsbtemp` as constructed in Problem 2.7:

```
bstemp.fit <- arima(tsbtemp,order=c(0,1,1),transform.pars=T)
```

calculates the maximum-likelihood estimates $\breve{b} = 0.881$ and $\breve{\sigma}_W^2 = 0.469$, both close to the estimates assessed by trial and error in Sect. 2.3.4.

The empirical residuals of this model are plotted together with their correlation and partial correlation functions in Fig. 5.5 using

```
plot(bstemp.fit$resid,type="l",xlab="year",ylab="degree Celsius")
acf(bstemp.fit$resid,lag.max=60,type="correlation",plot=T)
acf(bstemp.fit$resid,lag.max=60,type="partial",plot=T).
```

When plots (a) and (b) in Fig. 5.5 are compared with the templates in Fig. 2.22, it is concluded from (2.59) that the empirical residuals are a time slice from a realisation of a white noise process as required in (5.55), (5.36) and (5.28). Thus, the Basel yearly temperature values are found to be a time slice from a realisation of an ARIMA[0,1,1] process with parameters $b = 0.881$ and $\sigma_W^2 = 0.469$.

Further diagnostic plots and statistics to be applied, as required in (5.1,3), to an estimated ARIMA$[p, d, q]$ model can now be obtained as demonstrated in Sect. 5.4.

5.4 Fitting an ARIMA$[p, d, q]$ Model

In Sects. 5.1 and 5.2, AR[2] models for the de-meaned times series of the wind speed measurements in a turbulent atmospheric flow and the amounts of a product formed during a chemical reaction occurring in a batch reactor, as plotted in Figs. 5.1 and 5.3, are estimated using four methods, with the results given in Table 5.1. In Sect. 5.3.4, an IMA[1,1], i.e., an ARIMA[0,1,1] model, for the Basel yearly temperature values plotted in Fig. 2.13 is estimated using the maximum likelihood method. Since the empirical residuals of these models, as plotted in Figs. 5.2, 5.4 and 5.5, are not correlated, it is concluded that the models are in agreement with the observations.

However, these models are not arrived at by systematically applying the steps as proposed in (5.1) since the models are found by trial and error using only one diagnostic: the fit is claimed to be successful once the second moment functions of the model residuals suggest that the residuals stem from a white noise process. The systematic approach to the identification, estimation, diagnostics and selection of an ARIMA$[p, d, q]$ model therefore remains to be introduced in this section. Comprehensive descriptions of these methods are given in, e.g., [16], [61], [36], [17], and [22].

The first step required in (5.1) is the identification of candidate models suitable for the process from which the time series under analysis stems. If it stems from a stationary process then possible values for p and q in an ARMA$[p, q]$ model as defined in (5.36) are determined with the procedures introduced in Sect. 5.4.2, else the non-stationarity is removed using the techniques introduced in Sect. 5.4.1. In the steps following the first one in (5.1), candidate models are estimated in Sect. 5.4.3 to which the diagnostics introduced in Sect. 5.4.4 are then applied.

5.4.1 Removing Trends or other Fluctuations from a Time Series

If the time series under analysis stems from a process with a non-constant expectation function, the trend or the fluctuations in the first moment function is (are) removed by calculating (i) the residuals of a linear model for the expectation function or (ii) its differences.

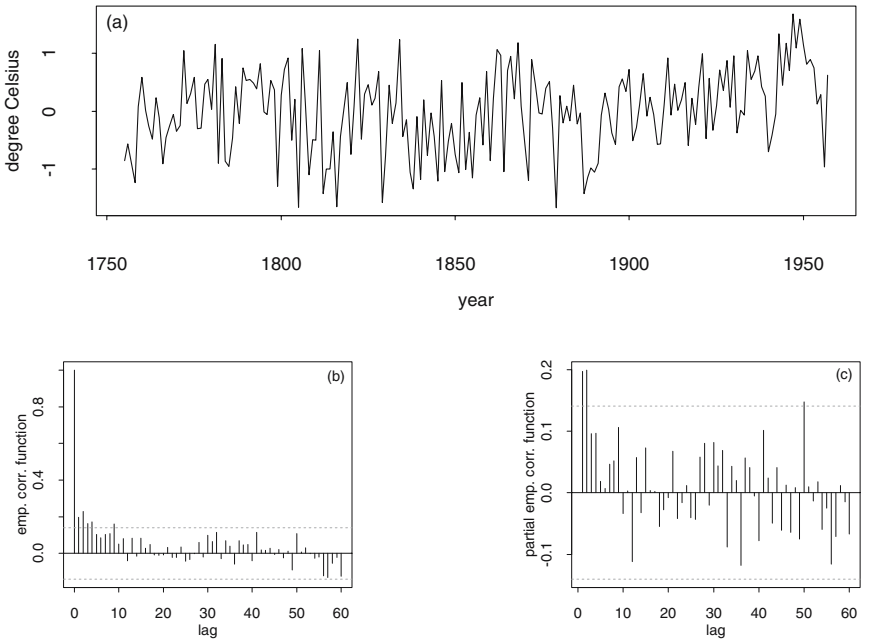


Fig. 5.6. Basel yearly temperature values originally plotted in Fig. 2.13: de-meaned time series (above) with its empirical second moment functions (below).

When using the diagnostics introduced in Sect. 2.5.1, a stochastic process is assumed to be stationary or non-stationary in its expectation function. A time series stemming from a process with a non-constant expectation function becomes possibly stationary with the properties required in (2.54) and the time series is said to be de-trended on condition that:

1. *its differences as defined in (2.39) and (2.40) are calculated or*
2. *the residuals of a linear model (including the mean) for the expectation function are calculated. The linear model is fitted using the regression techniques introduced in Chap. 3.*

(5.58)

The removal of a non-stationarity in the expectation function, as proposed in (5.58,1,2), is demonstrated in Figs. 5.6, 5.7 and 5.8 using the Basel yearly temperature values as a straightforward example. More can be learned from a second example time series: in Figs. 5.9, 5.10 and 5.11, the exponential increase and the annual cycle are removed from the monthly values of the atmospheric carbon dioxide measured at the Mauna Loa Observatory (Hawaii) and downloaded in Problem 3.23.

The Basel yearly temperature values are not stationary in the expectation function as can be concluded from the plots in Figs. 2.13, 2.14 and 3.3. The

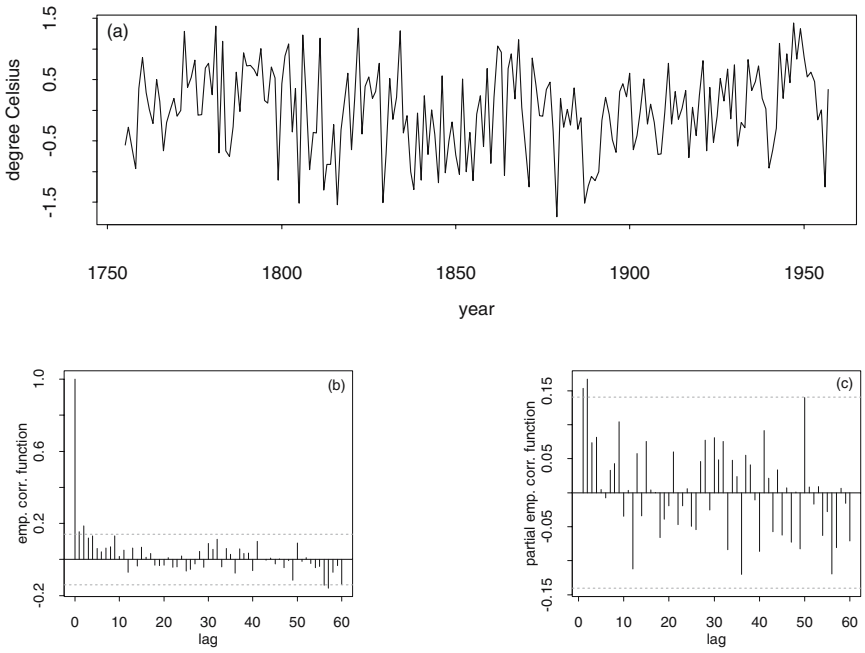


Fig. 5.7. Basel yearly temperature values originally plotted in Fig. 2.13: time series de-trended using a linear model of order 1 (above) with its empirical second moment functions (below).

same result is obtained by comparing the plots shown in Figs. 5.6, 5.7 and 5.8. The de-trended time series as plotted in Fig. 5.7 (a) lacks the secular trend shown in Fig. 2.13 which is still present in the de-meaned time series as plotted in Fig. 5.6 (a). Hence, the empirical correlations and empirical partial correlations plotted in Fig. 5.7 (b,c) are somewhat smaller than those plotted in Fig. 5.6 (b,c). The deviations in the empirical second moment functions are, however, small since the decadal fluctuations in the annual Basel temperature values are not captured by the first order linear model for the secular trend.

These decadal fluctuations, however, can no longer be seen in Fig. 5.8 which contains plots of the first order differences calculated, as required in (2.39), from the yearly values in the Basel temperature series. Here, the first moment function of the differences remains constant, as is seen in plot (a), and the pertaining empirical second moment functions shown in plots (b) and (c) behave, as concluded from Table 5.2, as the correlation and partial correlation function of an MA[1] model: the empirical correlation function (plot b) cuts off after lag 1 and the empirical partial correlation function (plot c) decays exponentially. This points to a stationary second moment function since an MA[1] model is stationary. Consequently, in the case of the

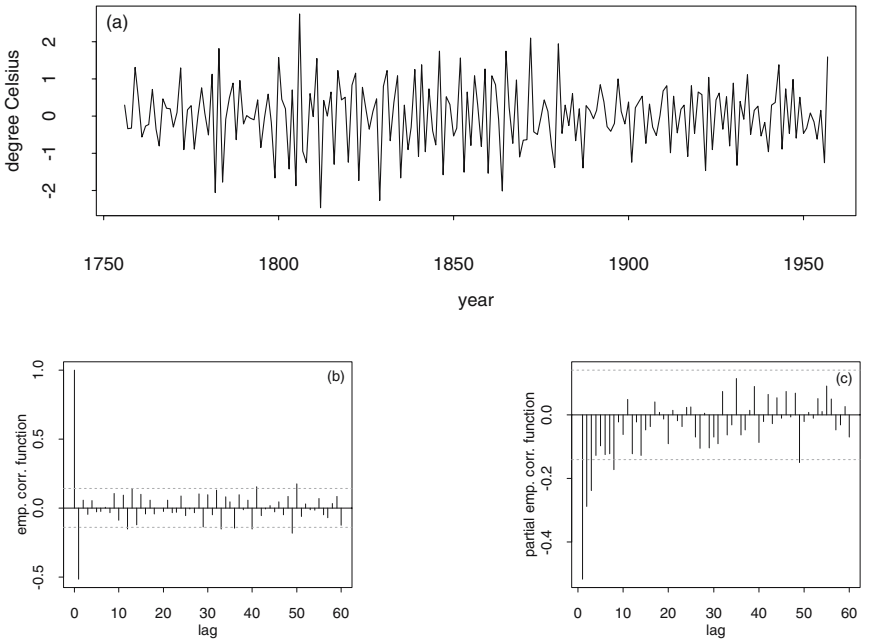


Fig. 5.8. Basel yearly temperature values originally plotted in Fig. 2.13: time series of the first order differences (above) with its empirical second moment functions (below).

yearly means in the Basel temperature series, (5.58,1), i.e., calculating the first order differences, is preferred to (5.58,2), i.e., de-trending the time series, for removing the fluctuations in the expectation function. De-trending using linear models of higher order will not produce a better result, as can be seen from the plots in Figs. 3.3 and 3.5 since these models, like the first order linear model, do not capture the decadal fluctuations.

Are the first differences of the Basel temperature yearly values as plotted above in Fig. 5.8, a stationary time series with properties as required in (2.54)? These differences oscillate around an approximately constant expectation function and thus are assumed to be stationary in the first moment function. From the plots of the empirical correlation functions calculated from the entire series of the differences, as well as from the first and second half-series plotted below, in Fig. 2.14, it is concluded that the differences are stationary in their covariance function, and, consequently, also in their variance function.

As a second example, trends and other fluctuations are removed from the expectation function of the atmospheric carbon dioxide (CO_2) monthly values measured at Mauna Loa Observatory (MLO) [84]. MLO is located at a high altitude (approximately 3400 m) on the north slope of the Mauna Loa Vol-

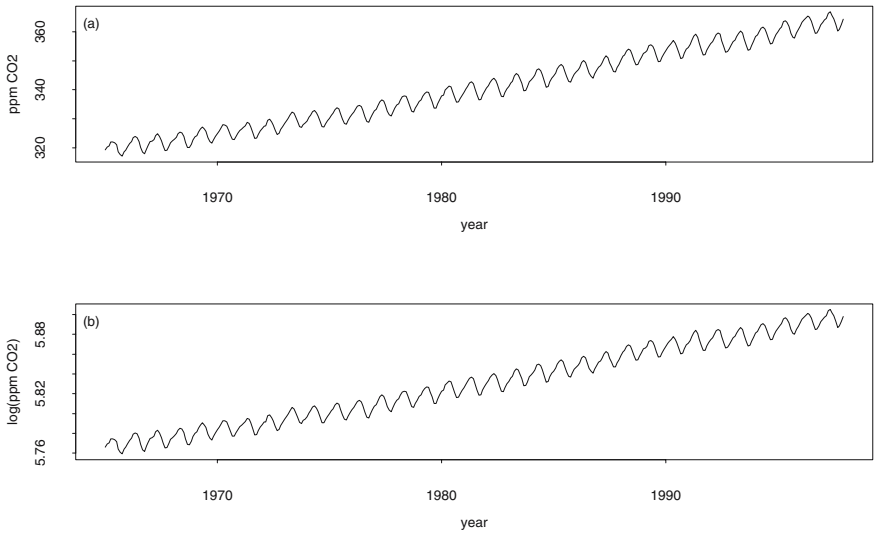


Fig. 5.9. Monthly values of the atmospheric carbon dioxide (CO_2) measured at Mauna Loa Observatory (MLO) in ppm (parts per million), before (above) and after (below) taking the logarithm.

cano, Hawaii. Since the mid 1950's MLO has been collecting data related to changes in the atmosphere. The MLO record, as downloaded in Problem 3.23, is considered to be representative for the global mean of atmospheric CO_2.

The exponential growth of the atmospheric CO_2, shown in Fig. 5.9 (a), becomes linear after taking the logarithms. The logarithmic transformation also stabilises the variance (remarks to (3.16) and (3.19), Problems 3.22 and 3.23): the resulting time series increases approximately linearly and the amplitudes of its annual fluctuations remain approximately constant, as can be seen in Fig. 5.9 (b).

In Problem 3.23, a component model (3.19) for the trend and yearly cycle of the logarithm of the atmospheric CO_2 is estimated. The empirical residuals of this component model are plotted together with their second moment functions in Fig. 5.10. Figs. 5.10 and 5.11 contain plots (b) and (c) of the empirical correlation and empirical partial correlation functions with the lags in years, e.g., lag 1 corresponds to a displacement amounting to 12 months.

In Fig. 5.10, the residuals of the component model decrease until 1977 and then increase in the period from 1978 through to 1990 since the decadal fluctuations are not captured by the component model (a linear model for the trend and the seasonal cycle). Therefore, due to the decadal fluctuations in the logarithm of the atmospheric CO_2, (i) the residuals of the component

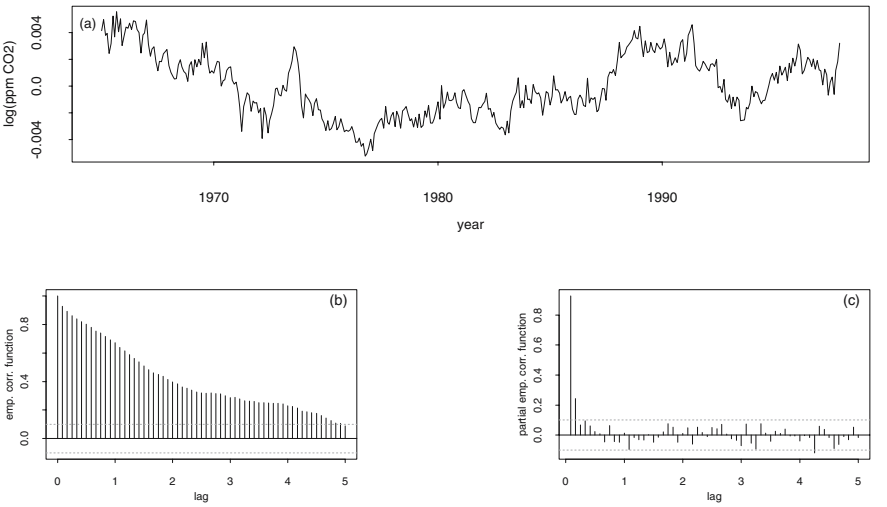


Fig. 5.10. Logarithm of the atmospheric CO_2: residuals of a component model as defined in (3.19) for the decadal trend and the yearly cycle (above) with empirical second moment functions (below, lag in years).

model are non-constant in the first moment function and thus are not stationary, and (ii) the correlation function decays very slowly (the lag is given in years). The empirical partial correlation function, however, cuts off after lag 0.166 years, i.e., two months.

The second possibility for removing fluctuations in the first moment function is, as proposed in (5.58,1), to calculate the differences of a time series. The first differences with lag 12 of the logarithm of the atmospheric CO_2 are obtained using

```
mauna <- ts(scan("/path/maunaloa.datshort"),
            start=1965,frequency=12,)
lmauna <- log(mauna)
d1lmauna <- diff(lmauna, lag=12, differences=1)
```

and then are plotted together with their empirical second moment functions in Fig. 5.11.

The first order differences of the logarithm of the atmospheric CO_2 calculated with lag 12 are, as can be seen in the plot above in Fig. 5.11, no longer afflicted with decadal fluctuations and thus are assumed to be stationary in their first moment function.

Is the time series of these differences also stationary in their second moment function? Since plots (not shown) of the empirical correlation functions calculated from the first and second half-series are damped sinusoidal oscillations similar to the empirical correlation function obtained from the entire series (plot (b) in Fig. 5.11), it is concluded that the difference time series

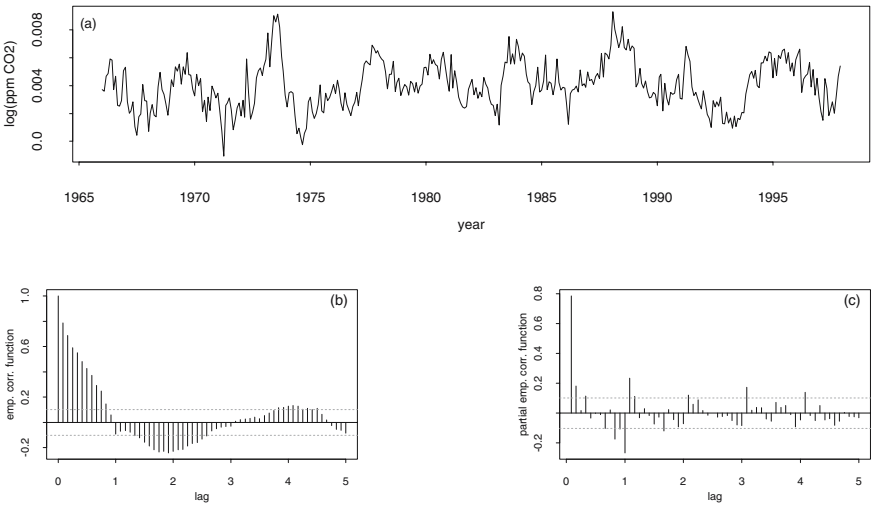


Fig. 5.11. Logarithm of the atmospheric CO_2: first order differences calculated with lag 12 (above) together with the pertaining empirical second moment functions (below, lag in years).

is stationary in the second moment function. Hence, having used the diagnostics proposed in (2.54), the first order differences of the logarithm of the atmospheric CO_2 calculated with lag 12, as plotted above in Fig. 5.11, are assumed to be a time slice from the realisation of a stationary stochastic process.

From a comparison of Figs. 5.7 with 5.8 as well as Figs. 5.10 with 5.11, it is concluded that the decadal fluctuations have been removed from both example time series when the first differences are calculated but are still present in the residuals of the linear models for the expectation function fitted to these time series. It is therefore concluded from these examples, that both methods can be combined: if a linear model for the expectation function is estimated (e.g., because an estimate for a trend is required), then fluctuations still present in its residuals can be removed by calculating the differences of the residuals (Problem 5.14).

Using the yearly values in the Basel temperature series and the MLO atmospheric CO_2 monthly values as example time series, it is demonstrated above that a trend and/or less systematic fluctuations in the first moment function of a time series can be removed when the methods proposed in (5.58) are applied to obtain a de-trended time series. Sometimes, however, these methods need to be combined to arrive at the desired result.

A de-trended time series is supposed to be constant in its mean function and thus is possibly stationary in its first and second moment functions. Then, using the diagnostics introduced in Sect. 2.5.1, the time series under analysis

can possibly be assumed to stem from a stochastic process that becomes stationary after it has undergone the same operations (i.e., transformations, calculating the residuals of a linear model, or differencing) as the time series. However, the assumption that a stochastic process is stationary remains a mere approximation for the reasons given in Sect. 2.5.1.

5.4.2 Identification

Candidate AR$[p]$, MA$[q]$ or ARMA$[p, q]$ models for a stationary time series, or a time series that has become stationary after applying the methods introduced in Sect. 5.4.1, are found using plots of the empirical correlation and the empirical partial correlation functions calculated from the time series under analysis as diagnostic tools:

Candidate AR[p], MA[q] or ARMA[p, q] models for a stationary process from which a time series is available are those models whose correlation and partial correlation functions behave, in accordance with Table 5.2, similarly to the empirical correlation and partial correlation functions calculated from the observations. (5.59)

The diagnostics proposed in (5.59) exploit the properties of the correlation and partial correlation functions of a linear process as derived in Sect. 5.3 and summarised in Table 5.2.

Using (5.59), a candidate model is easily identified on condition that the observations stem from a stationary stochastic process reconcilable with an AR$[p]$ or MA$[q]$ model because these models imply a cutoff in either the correlation or partial correlation functions calculated from the observations. In the following cases, using:

1. an AR[2] model for the wind speed measurements in the turbulent atmospheric flow as plotted (together with its empirical second moment functions) in Fig. 5.1,
2. an AR[2] model for the product formed in a chemical reaction as plotted in Fig. 5.3, and
3. an IMA[1,1] model for the Basel yearly temperature values as plotted in Fig. 2.13, i.e., an MA[1] model for the first order differences of this time series as plotted in Figs. 2.14 and 5.8,

an identification is straightforward and only one candidate model is obtained.

More challenging is the identification of stationary candidate models if both, the empirical correlation and the partial correlation functions, do not cut off. This behaviour pertains only, as is learned from Table 5.2, to the second moment functions of an ARMA$[p, q]$ model. Two examples for such processes are given: only one candidate model is found to be suitable, in the remarks to Figs. 5.12 and 5.13, for the first order differences calculated with lag 12 of the MLO atmospheric CO_2 monthly values, whereas more than 10

candidate models are found to be suitable, in the remarks to Fig. 5.14, for the monthly values of the Southern Oscillation (SO) index.

The monthly values of the atmospheric CO_2, plotted above in Fig. 5.9, first undergo a logarithmic transformation to obtain time series (y_u), whereafter the first differences with lag 12 are calculated to obtain time series $(x_u) = \Delta_{12}^1(y_u)$, as plotted above in Fig. 5.11.

(x_u) is found to be stationary in the remarks to Fig. 5.11 and is thus assumed to stem from a linear process (X_u). Can candidate models for (X_u) be identified using the diagnostics proposed in (5.59)? Does a model with second moment functions similar to the empirical correlation and partial correlation functions, $\hat{c}_X(\tau)$ and $\hat{a}_X(k,k)$, plotted below in Fig. 5.11, exist?

$\hat{c}_X(\tau)$ (plot (b) in Fig. 5.11) is a damped sinusoidal oscillation, and $\hat{a}_X(k,k)$ (plot (c) in Fig. 5.11) amounts to 0.8 and 0.2 for lags 0.0833 and 0.1666 years (1 and 2 months) and does not cut off after lag 0.1666, as values outside the .95 confidence interval for no correlation are calculated for lags of approximately 1, 2, 3, and 4 years. No model implying this behaviour of the empirical second moment functions is found in Table 5.2.

To find a way out, the partial correlations after the lag 0.1666 years are ignored, and with R expressions

```
maunaar2 <- arima(d1lmauna,order=c(2,0,0),transform.pars=T) #fit
acf(maunaar2$residuals,type="partial",lag.max=60)     #diagnostics
acf(maunaar2$residuals,type="corr",lag.max=60)
```

an AR[2] model is fitted to (x_u) and the residuals of this model are estimated. These estimates are plotted together with their empirical second moment functions in Fig. 5.12.

The empirical correlations (plot (b) in Fig. 5.12) remain, except for lag 1 year, inside the .95 confidence interval for no correlation, and the empirical partial correlations (plot (c) in Fig. 5.12) decay exponentially for lags 1, 2, 3, and 4 years; for the lags inbetween, they remain approximately inside the .95 confidence interval for no correlation. Comparing these plots with the descriptions of the behaviour of the second moment functions in Table 5.2, an MA[1] model (but with time step $\Delta_t = 12\Delta_u$, i.e., one year, since $\Delta_u = u_i - u_{i-1}$, i.e., one month) is found to be in agreement with the correlations in the residuals of the AR[2] model.

Thus, from the plots in Figs. 5.11 and 5.12 it is concluded that (x_u), i.e., the time series of the first order differences with lag 12 of the logarithm of the atmospheric CO_2 monthly values, is in agreement with a combination of an AR[2] model and an MA[1] model, on condition that the time step is calculated in months in the case of the AR[2] model but in years in the case of the MA[1] model.

An ARIMA[0,1,1] model with time step one year for (y_u), i.e., the logarithm of the atmospheric CO_2 monthly values with $(x_u) = \Delta_{12}^1(y_u)$, is a joint MA[1] model for the first differences of the seasonal processes, resulting from splitting (y_u) as defined in Table 3.2. This model is called a seasonal

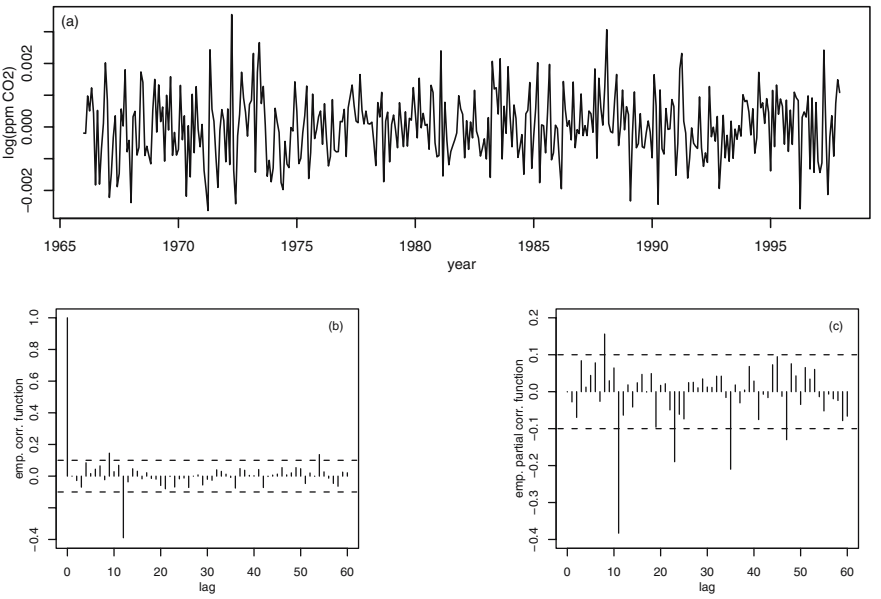


Fig. 5.12. Empirical residuals of an AR[2] model fitted to the first order differences with lag 12 of the logarithm of the atmospheric CO_2, as plotted above in Fig. 5.11, (above), with the pertaining empirical second moment functions (below, lag in months).

ARIMA$[0, 1, 1]_{12}$ model. In general, a seasonal linear process is called either a seasonal ARIMA$[P, D, Q]_s$ model, a SARIMA$[P, D, Q]_s$ model or an ARIMA$[P, D, Q]_s$ model, with s being the number of the seasonal processes and P, D and Q the parameters of the joint model for the seasonal processes. Such a model is often fitted to a time series with a seasonal cycle, e.g., the seasonal cycle and trend in the logarithms of the atmospheric CO_2 monthly values, as shown below in Fig. 5.9, are captured with an ARIMA$[0, 1, 1]_{12}$ in Problem 5.15.

If an ARIMA$[P, D, Q]_s$ model is fitted to the process (X_u), then

1. (X_u) is split into, using the template in Table 3.2, s seasonal processes $\left(X_t^{(r)}\right)$, $r = 1, \ldots, s$ the season and t the period,
2. the innovations (V_u) are split into s seasonal processes $\left(V_t^{(r)}\right)$, $r = 1, \ldots, s$, such that the $\left(V_t^{(r)}\right)$ are white noise processes; for both splittings into seasonal processes, it is assumed that
3. the differences of order D of each seasonal process are reconcilable with an ARMA$[P, Q]$ model, and that
4. these models are approximately identical such that only one parameter set has to be estimated.

Often, however, the residuals of an ARIMA$[P, D, Q]_s$ model are still correlated for lags $\tau \neq ns$, $n = 1, 2, \ldots$, since the time step of the process (X_u) under analysis is $\Delta u = u_i - u_{i-1}$. For example, plots of the empirical residuals (together with their second moment functions) of the ARIMA$[0, 1, 1]_{12}$ model fitted in Problem 5.15 to (y_u), (i.e., the logarithm of the atmospheric CO_2,) show that y_u is also correlated with $y_{u-\tau}$, $\tau = 1, 2, \ldots$. This suggests that the correlations in the residuals of the ARIMA$[0, 1, 1]_{12}$ model are reconcilable with an AR[2] model. Therefore, both models are combined to constitute a multiplicative ARIMA$[2, 0, 0] \times [0, 1, 1]_{12}$ model, i.e., the candidate model for (y_u).

A multiplicative ARIMA$[p, d, q] \times [P, D, Q]_s$ model is introduced in [16], where an ARIMA$[0, 1, 1] \times [0, 1, 1]_{12}$ model (today called a *Box-Jenkins airline model*) is fitted to the number of passengers transported by the international airlines in the years 1952, 1953 and 1954.

The multiplicative ARIMA$[p, d, q] \times [P, D, Q]_s$ model is estimated, as are any of the others models for a linear process, using R functions `arima()`. The candidate ARIMA$[2, 0, 0] \times [0, 1, 1]_{12}$ model for (y_u), i.e., the logarithm of the atmospheric CO_2, is estimated using

```
maunaseasonal <- arima(lmauna, seasonal=list(order=c(0,1,1),
     period=12), order=c(2,0,0), transform.pars=T)  #seasonal model
#diagnostics: do not use tsdiag(maunaseasonal) here:
res <- window(maunaseasonal$residuals,c(1966,1),c(1998,12))
plot(res)
acf(res,type="corr",lag.max=60)
acf(res,type="partial",lag.max=60)
```

and the time series of the empirical residuals is plotted together with its second moment functions in Fig. 5.13.

Comparing the plots in Fig. 5.13 with the template plots in Fig. 2.22, it is concluded that the residuals of the multiplicative ARIMA$[2, 0, 0] \times [0, 1, 1]_{12}$ model are a time slice from the realisation of a white noise process, and thus this model is found to be suitable for (y_t). This result surprises: a very parsimonious model with only five parameters (the variance of the innovations, the order of the seasonal differences, two AR coefficients and one seasonal MA coefficient) is found to account for the trend, seasonal cycle and correlations with lags in the monthly scale in the logarithm of the MLO atmospheric CO_2 monthly values, which, being representative for the global atmospheric CO_2, are the result of coupled physical, chemical, biological, geological and anthropogenic processes in the atmosphere, the oceans and on the surface of the earth (Problem 5.31).

As a second example, the Southern Oscillation (SO) index is analysed in search of candidate ARIMA$[p, d, q]$ models. The SO index is the difference in the standardised atmospheric pressure measured at Pepeete (Tahiti) and Darwin (Australia) stations. It measures the strength of the so-called Walker circulation in the tropical Pacific atmosphere. The Walker circulation is an

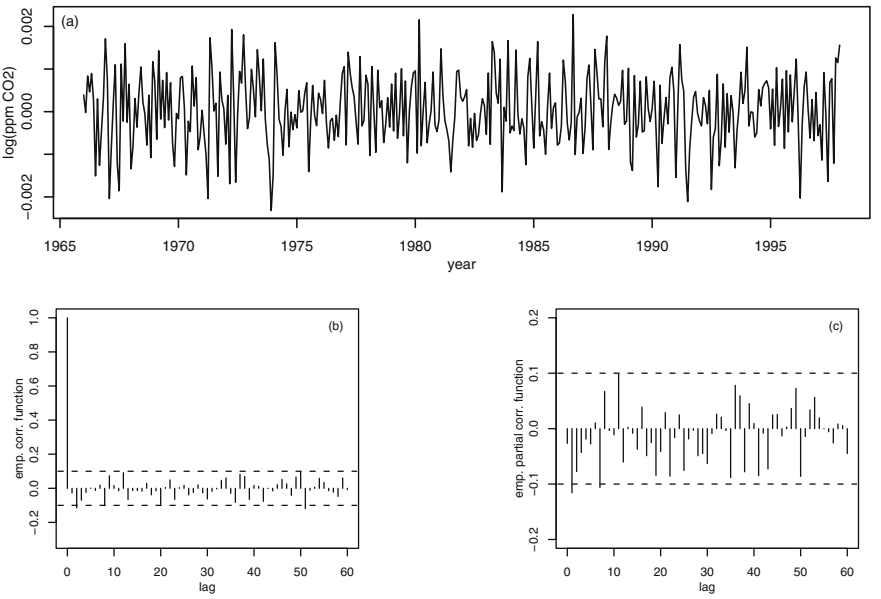


Fig. 5.13. Empirical residuals of a multiplicative ARIMA$[2, 0, 0] \times [0, 1, 1]_{12}$ model for the logarithm of the atmospheric CO_2, as plotted below in Fig. 5.9, (above) together with their empirical second moment functions (below, lag in months).

east-west oriented circulation consisting of low-level convergence and rising motion over the western tropical Pacific (low pressure at Darwin), eastward flow at upper-levels, sinking motion over the eastern tropical Pacific (high pressure at Tahiti and east thereof) and westward flow at the surface along the equator as trade winds.

The westward flow drives divergent oceanic surface currents, producing a pronounced west-east gradient in tropical Pacific sea surface temperatures due to upwelling of cold water. These anomalies tend to reinforce the atmospheric flow. Changes in the Walker circulation, as measured by the SO index, have a considerable effect on precipitation in the tropical Pacific region, and influence the climate over large distances in other parts of the world ([98], [99], [110], [43], [24]).

Plots (a) and (b) of the SO index in Fig. 5.14 show that relatively long periods with large differences in the atmospheric pressure between the southeastern Pacific and Indonesia are interrupted by shorter periods with smaller differences. In these shorter periods, changes in the sea surface temperatures, precipitation, etc. are observed. A prominent change is the increasing sea surface temperature at the coast of South America, inducing changes in ecology (e.g., the fish become less abundant) and economy. A traditional fishermen's name for this increase is El Niño. The connection between El Niño and the

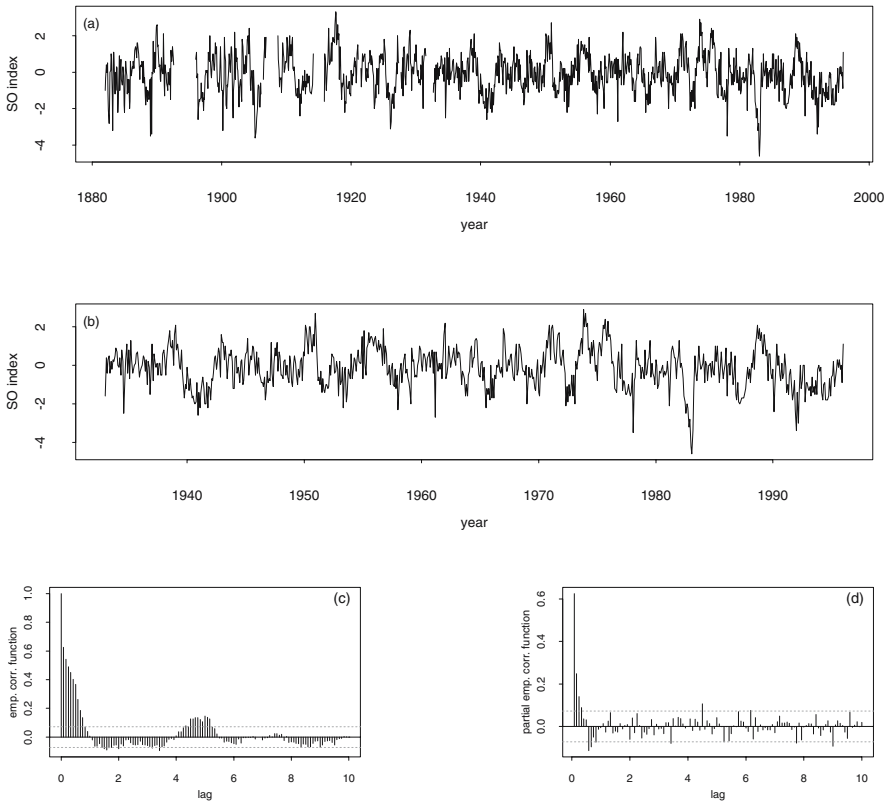


Fig. 5.14. SO index: monthly values from January 1882 through to January 1996 (a), monthly values from January 1933 through to January 1996 (b), as well as empirical correlation (c), and partial correlation (d) functions calculated from the observations plotted in (b).

Southern Oscillation has been known for decades: El Niño - Southern Oscillation (ENSO) is a term often used.

The strength of the SO is measured using several indices. Among these is the SO index introduced above. There are slight variations in the SO index values, obtained from various sources. The SO index, as plotted in Fig. 5.14 and available as file `/path/soi.dat`, was downloaded in 1999 from [86]. It is recommended to repeat the analyses made in this section as well as in Sects. 5.4.3 and 5.4.4, with time slices from an SO index time series extended further back using the methods described in [116]. This SO index can be obtained from the addresses given in [115]. The R expressions

```
sofmt <- list(month="",year=0,so=0)
sofn <- "/path/soi.dat"
```

```
soframe <- scan(sofn,sofmt)
so <- ts(soframe$so,start=1882,frequency=12,)
soshort <- window(so,c(1933,1),c(1996,1))
```

generate the R time series `so` and `soshort` shown in plots (a) and (b) in Fig. 5.14. From plot (a) in Fig. 5.14 it is concluded that the last period where no measurements were recorded occurred from August 1931 through to August 1932 and therefore, the empirical correlation and partial correlation functions (plots (c) and (d) in Fig. 5.14) are calculated from observations made since January 1933.

Using (5.59), it is concluded from the plots in Fig. 5.14 that

1. AR$[p]$ models with $3 \leq p \leq 10$ (with emphasis on the empirical partial correlation function being small for lags larger than 0.25 years and remaining practically inside the .95 confidence interval for no correlation for lags larger than 0.833 years) and
2. ARMA$[p, q]$ models with $p, q < 3$ (with emphasis on the damped sinusoidal oscillations in both empirical second moment functions, using the remark to (5.45))

are candidate models for the SO index. From this fairly large number of candidate models, a few are selected as being suitable for the SO index when the diagnostics introduced in Sect. 5.4.4 are applied to the model estimates obtained as demonstrated in Sect. 5.4.3.

5.4.3 Estimation

The first step of procedure (5.1) results in a small number of candidate models (AR$[p]$, MA$[q]$, ARMA$[p, d, q]$,ARIMA$[p, d, q]$ or seasonal models) to be estimated for the time series under analysis, as demonstrated in Sects. 5.4.1 and 5.4.2. These models are estimated, in the second step, using `arima()` as introduced in Sects. 5.2.3 and 5.3.4. In the ARMA$[p, d, q]$ definition underlying `arima()` (given in Sect. 5.3.4), the sign of the moving-average coefficients is $+$, whereas in definitions (5.28), (5.36) and (5.55), the sign is $-$. Therefore, the sign of the moving-average estimates obtained from `arima()` is changed when used in connection with definitions (5.28), (5.36) and (5.55).

For instance, ARMA$[p, q]$ models (with $p \leq 2$ and $q \leq 2$) and AR$[p]$ models (with $3 \leq p \leq 10$) are estimated for the SO index as shown in plot (b) in Fig. 5.14. These models are the candidate ones proposed in the remarks to Fig. 5.14. In addition, for the reasons given in the remarks to Table 5.3, AR[11] and AR[12] models are estimated.

Using Burg's algorithm (5.21) (`ar.burg()` de-means prior to estimating)

```
sosh.ar12burgfit <- ar.burg(soshort,aic=F, order.max=12)   #AR[12]
sosh.ar12burgfit$ar
[1]   0.405185999  0.156836645  0.103714905  0.095897832
[5]   0.058366544  0.108357691 -0.056266337 -0.060071795
[9]  -0.016843724 -0.069007548 -0.009667824 -0.008072675
```

```
soshres <- window(sosh.ar12burgfit$resid,c(1934,1),c(1996,1))
mean(soshres)     #possible, since first 12 missing residuals are
[1] -0.001216903  #cut above (sosh.ar12burgfit$resid begins
var(soshres)      #with 1933, so does R time series soshort)
[1] 0.5820025
```

the estimates for the autoregressive coefficients as well as the mean and variance of the empirical residuals as estimates for the moments of the innovations are obtained.

The ML estimates (5.25) for the autoregressive candidate models are calculated using `arima()` (below, for the AR[12] model)

```
sosh.ar12mlefit <- arima(soshort,order=c(12,0,0),transform.pars=T)
sosh.ar12mlefit$coef
         ar1          ar2          ar3          ar4          ar5
 0.404335405  0.156808030  0.103423985  0.096245233  0.057768104
         ar6          ar7          ar8          ar9         ar10
 0.108804279 -0.055641957 -0.059951813 -0.016378461 -0.068732625
        ar11         ar12    intercept
-0.009575356 -0.008043253 -0.143235509
sosh.ar12mlefit$sigma2
[1] 0.5788384
```

to arrive at the same values (at least for practical purposes, i.e., for the calculation of predictions) as those obtained with Burg's algorithm. The ML estimated AR[p] models for the SO index are stationary, due to constraints applied when calculating with `arima( ..., transform.pars=T, ...)`. Diagnostics for these models are to be found in Table 5.3.

The ML estimates (5.57) for the candidate ARMA[p, q] models, $p, q < 3$, are then calculated, also using R function `arima()` (below, for the ARMA[1,1] model)

```
sosh.arma11fit <- arima(soshort,order=c(1,0,1),transform.pars=T)
sosh.arma11fit$coef                 #change sign of moving average
      ar1        ma1  intercept     #coefficients when used with
0.8828684 -0.4648837 -0.1411459     #(5.55), (5.36) or (5.28)
sosh.arma11fit$sigma2
[1] 0.5997365
sosh.arma11fit$var.coef    #covariance matrix of estimates:
#change sign of covariances involving moving average coefficients
#when used with (5.55), (5.36) or (5.28)
                    ar1           ma1     intercept
ar1        5.434013e-04 -6.913596e-04  7.151109e-06
ma1       -6.913596e-04  1.865348e-03 -7.658224e-06
intercept  7.151109e-06 -7.658224e-06  1.624915e-02
-2*sosh.arma11fit$loglik    #dignostic (5.61)
[1] 1761.944
```

to obtain the results and diagnostics in Table 5.3. The ARMA$[p, q]$ models with the ML estimated parameters are stationary and invertible due to the constraints applied with option `arima( ..., transform.pars=T, ...)`.

5.4.4 Diagnostics

Is a candidate model in agreement with the observed time series? Answers to this question are found by applying diagnostics to the estimated models for a linear process in the third step proposed in (5.1).

As a first diagnostic, asymptotic .95 confidence intervals for the estimated model coefficients are approximated in (5.60):

$$\hat{a}_i \pm 1.96 \times \hat{\sigma}_{a_i},\ i = 1, \ldots, p \qquad \hat{b}_j \pm 1.96 \times \hat{\sigma}_{b_j},\ j = 1, \ldots, q \tag{5.60}$$

If an interval as proposed in (5.60) does not contain zero, it is then assumed that the model coefficient (i.e., the true coefficient) is not identically zero. If intervals for all coefficients in a model do not contain zero, it is then assumed that all model coefficients are not identically zero and are, thus, essential. The standard deviations for the calculation of the .95 confidence intervals as proposed in (5.60) are obtained from the diagonal in the covariance matrix of the estimated coefficients, which is among the results obtained from `arima()`, as demonstrated in Sect. 5.2.3.

The approximations for the asymptotic .95 confidence intervals as proposed in (5.60) are preferably combined with the second diagnostic, i.e., the AIC [1] and BIC [122] values as proposed in (5.61) being calculated in `arima()`. This is demonstrated in Table 5.3 using the SO index example. The AIC and BIC values in (5.61) are calculated

$$\begin{aligned} (\text{arg}) &= \big(\breve{\boldsymbol{a}}(p), \breve{\boldsymbol{b}}(q), \breve{\sigma}^2_W(p,q)|\boldsymbol{X}\big) \\ \text{AIC}(\text{arg}) &= -2\log\big(\text{L}(\text{arg})\big) + 2(p+q) \\ \text{BIC}(\text{arg}) &= -2\log\big(\text{L}(\text{arg})\big) + \big(\log(N)\big)(p+q) \end{aligned} \tag{5.61}$$

by doubling the negative logarithm of $\text{L}\big(\boldsymbol{a}(p), \boldsymbol{b}(q), \sigma^2_W(p,q)|\boldsymbol{X}\big)$, i.e., the likelihood function in (5.57) which is maximised when an ARMA$[p, q]$ model is fitted to a stationary time series $(x_t), t = 1, \ldots, N$, with a correction added.

Small AIC and BIC values (calculated from a large likelihood function) point to a good fit of the model. Since a model with a large number of coefficients usually fits better than a parsimonious one, corrections are applied to the value obtained from the likelihood function. These corrections depend on the number of parameters estimated (and also, in case of a BIC value, on the length N of the time series) and thus, acting in response to the remarks to (5.36): recompense parsimonious and punish non-parsimonious models!

The diagnostics as proposed in (5.60) and (5.61) are applied to the estimated candidate models for the SO index in Table 5.3. The AIC values

decrease from 1770 (as calculated for the ARMA$[p, q]$ models, $p, q < 3$, and also for the AR[3] and AR[4] models) to a mininum of 1755 (as calculated for the AR[10] model) and then increase for models with more than 10 autogressive coefficients. To demonstrate this increase, AR[11] and AR[12] models are estimated in Sect. 5.4.3, though, using the diagnostics (5.59), it is concluded from the plots in Fig. 5.14 that AR$[p]$ models, $3 \leq p \leq 10$, are candidate ones for the SO index.

The BIC values have a minimum for the ARMA[1,1] model and then increase for the models with a larger number of coefficients, except the AR[3] model which is comparable to the AR[4] one in the BIC value.

From the AIC values in Table 5.3, it is concluded that the AR[10] model is most suitable for the SO index; from the BIC values, however, the ARMA[2,1], ARMA[1,1] or the ARMA[1,2] models, possibly even the AR[3] or AR[4] models, are judged to be better suited. To arrive at a further reduction of the number of candidate models, diagnostic (5.60) is applied. Except for the AR[3], AR[4] and ARMA[1,1] models, all candidate models have estimated coefficients with .95 confidence intervals including zero and thus contain non-essential parameters. The ARMA[2,1] model, for example, has an estimate for coefficient a_2 being close to 0 and all its coefficients are estimated with relatively large (as compared with the AR[4] and ARMA[1,1] models) variances, as is concluded from the standard deviations of these estimates given below, in Table 5.3.

Thereafter, the AIC and BIC values pertaining to the remaining candidates, i.e., the AR[3], AR[4] and ARMA[1,1] models containing only essential coefficients, are compared. From this comparison it is concluded that, from all candidate models in Table 5.3, only the AR[4] and ARMA[1,1] models are in agreement with both diagnostics (5.60) and (5.61). This result is confirmed using properties of the AIC and BIC values obtained from simulation experiments. From such experiments, it becomes evident that, using the AIC value, the order of the model is often overestimated, whereas, using the BIC value, one comes closer to the order of the model, but risks an underestimation.

The diagnostics (5.60) and (5.61), i.e., the calculation of asymptotic .95 confidence intervals for the estimated coefficients and the comparison of the AIC and BIC values, often help to reduce the number of estimated candidate models for the time series under analysis. An additional reduction can then be arrived at by applying further diagnostics using the empirical residuals of the models estimated.

However, the empirical residuals of a fitted model should always (not only in order to reduce the number of candidate models) be analysed for the following reason. The definitions of the AR$[p]$, MA$[q]$ and ARMA$[p, q]$ models in (5.2), (5.28) and (5.36), and also those of the ARIMA$[p, d, q]$ model in (5.55), contain two parts: linear formulae with the coefficients to be estimated in the first part, and the properties of the innovations in the second part. Since (5.60) and (5.61) apply to the first part only, additional diagnostics must be

Table 5.3. Diagnostics of the candidate models for the SO index in the period from January 1933 through to January 1996 (plot (b) in Fig. 5.14) using (5.60) and (5.61) and ML estimates for the AR[4], ARMA[1,1] and ARMA[2,1] models. $-2\log\big(\mathrm{L}(\mathrm{arg})\big)$ in (5.61) is obtained from `arima()`, $N = 757$.

candidate model	$-2\log\big(\mathrm{L}(\mathrm{arg})\big)$	AIC	BIC	parameter with a confidence interval calculated using (5.61), that includes 0
AR[12]	1735.37	1759.37	1814.92	a_{12}, a_{11}, a_{10}, a_9, a_8, a_7, a_5
AR[11]	1737.42	1757.42	1808.34	a_{11}, a_{10}, a_9, a_8, a_7, a_5
AR[10]	1735.54	1755.54	1808.84	a_9, a_8, a_7, a_5
AR[9]	1739.82	1757.82	1799.48	a_9, a_7, a_5
AR[8]	1741.74	1757.74	1794.78	a_7, a_5
AR[7]	1748.62	1762.62	1795.02	a_5
AR[6]	1758.68	1770.68	1798.46	a_6, a_5
AR[5]	1759.39	1769.39	1792.53	a_5, a_4
AR[4]	1760.27	1768.66	1786.78	
AR[3]	1766.19	1772.15	1786.04	
ARMA[1,1]	1761.94	1765.94	1775.20	
ARMA[2,1]	1761.82	1767.82	1781.71	a_2
ARMA[1,2]	1761.80	1767.80	1781.69	b_2
ARMA[2,2]	1761.63	1769.63	1788.15	a_1, a_2, b_1, b_2

model	estimator	variance of innovations	coefficients	estimates for coefficients with standard deviations			
AR[4]	(5.21)	0.5976	$a_1, \ldots, a_4$	0.4283	0.1640	0.1029	0.0899
	(5.25)	0.5977	$a_1, \ldots, a_4$	0.4307	0.1610	0.1049	0.0899
				0.0364	0.0394	0.0363	0.0337
ARMA[1,1]	(5.57)	0.5997	a_1, b_1	0.8829	0.4649		
				0.0233	0.0432		
ARMA[2,1]	(5.57)	0.5982	a_1, a_2, b_1	0.9105	−0.0187	0.4841	
				0.0920	0.0640	0.0730	

applied to assert that the properties of the empirical residuals are not too far from the properties of the innovations as required in the definitions.

The residuals of a model are the differences of the estimated and true values. Estimates for the residuals are the empirical residuals, i.e., the differences $r_u^* = x_u - X_u^*$ of the observations x_u and the values X_u^* calculated using an estimated model. Assuming that

1. X_u^* is an estimate (Yule-Walker with Levinson-Durbin recursion, Burg's, regression, or ML) introduced in Sects. 5.2 and 5.3.4

2. $(x_1, \ldots, x_N)$ is a stationary time series in case of an ARMA$[p, q]$ model (however, in case of an ARIMA$[p, d, q]$ model, the stationary differences of order d of the original time series)
3. $(a_1^*, \ldots, a_q^*)$ and $(b_1^*, \ldots, b_q^*)$ are the estimated coefficients of the fitted model, and
4. x_u and r_u^* are identically zero for the time points prior to the period with observations,

the empirical residuals are arrived at using the difference equation pertaining to the ARMA$[p, q]$ model in (5.36), with x_u substituting X_t and past r_{u-j}^* substituting past W_{t-j}, by applying (5.62) recursively for $u = 1, 2, \ldots, N$,

$$x_u = 0 \text{ and } r_u^* = 0 \text{ for } u < 1; \qquad \mathrm{E}W_u = \mu_W = 0$$
$$r_u^* = x_u - \sum_{i=1}^{p} a_i^* x_{u-i} + \sum_{j=1}^{q} b_j^* r_{u-j}^*, \quad u = 1, 2, 3, \ldots, N \tag{5.62}$$

provided that, as a last assumption, the present innovations W_t in (5.36) have been substituted by their expectations μ_W (μ_W is shown to be, in (2.83), the optimal prediction for W_t). This last assumption is required, as W_u cannot be estimated from past r_{u-j}^*, $j > 0$, the W_t being non-correlated.

In applications, the empirical residuals are usually calculated by the algorithm implemented in the software used for estimation. For example, Table 5.4 contains the empirical residuals pertaining to the AR[4] and ARMA[1,1] models as estimated in Table 5.3, using `ar.burg()` to obtain Burg's estimates for the AR[4] model and `arima()` to obtain the ML estimates for the ARMA[1,1] model. In addition, Table 5.4 contains in its last column the empirical residuals of the ARMA[1,1] model. These are to be calculated, as required in (5.62) and using the estimates in Table 5.3, in Problem 5.17. The empirical residuals as required in (5.62) come closer, with increasing time having elapsed since January, 1933, to the empirical residuals as obtained using `arima()`. The deviations are due to the fairly rough initialisation in (5.62).

The empirical residuals, being estimates for the theoretical ones, should have the following properties given in (5.63)

The empirical residuals of an AR$[p]$, MA$[q]$, ARMA$[p, q]$ or ARIMA$[p, d, q]$ candidate model are approximately stationary and not correlated and stem, in the case of a maximum likelihood estimation, from a distribution being approximately normal. (5.63)

since the innovations in the causal models defined in (5.2), (5.28) and (5.55) are a (normal) white noise process. The properties stipulated in (5.63) are comparable to those claimed for the empirical residuals of a linear model in Sects. 3.2.3 and 3.2.5, for which diagnostics are introduced in Sect. 3.3.

Diagnostics for the empirical residuals of an estimated AR$[p]$, MA$[q]$, ARMA$[p, q]$ or ARIMA$[p, d, q]$ model are obtained, in this chapter, from plots

Table 5.4. Empirical residuals of the AR[4] and ARMA[1,1] models in Table 5.3 for the de-meaned (using $\check{\mu}_X = -1.1411459$ as estimated for the ARMA[1,1] model in Sect. 5.4.3) SO index. NA: not available.

	time	SO index de-meaned	AR[4] model Burg	AR[4] model ML	ARMA[1,1] model ML	ARMA[1,1] model (5.62)
1933	Jan.	−1.4588	NA	−1.0870	−1.0897	−1.4588
1933	Feb.	0.5411	NA	1.3905	1.3860	1.1509
1933	March	−0.2588	NA	−0.1540	−0.1199	−0.2016
1933	April	0.4411	NA	0.6570	0.6132	0.5759
1933	May	0.6411	0.5688	0.5702	0.5360	0.5194
1933	June	−0.2588	−0.6254	−0.6247	−0.5758	−0.5834
1933	July	0.3411	0.3252	0.3252	0.3020	0.2985
1933	August	−0.0588	−0.2662	−0.2653	−0.2196	−0.2213
1933	September	0.4411	0.3805	0.3806	0.3910	0.3902
1933	October	0.3411	0.1508	0.1518	0.1334	0.1331

of the empirical residuals as well as their correlation and partial correlation functions. Examples are given in Figs. 5.2, 5.4, 5.12 and 5.13. (5.63) requires, basically, that the empirical residuals are assessed as being (i) stationary (using the diagnostics introduced in Sect. 2.5.1) and (ii) not correlated (using the diagnostics in the remarks to (2.59) and Fig. 2.22). These diagnostics involve the inspection of plots of the empirical residuals, their correlation function and, compulsory in the case of an ML estimate, their normal probabilility plot as introduced in Fig. 2.18.

Plots of the partial correlation function of the empirical residuals are not required here, in the third step proposed in (5.1), however, they often contain useful hints for the model identification in the first step, as demonstrated in Sects. 5.1.3 and 5.4.2.

Plots of the empirical residuals of the AR[4] and ARMA[1,1] models for the SO index, with the estimates given in Table 5.3, are available, together with plots of their empirical correlation functions and their normal probability plots, in Figs. 5.15, 5.16 and 5.17. These are the diagnostics used to judge whether or not the empirical residuals stem from a normally distributed white noise process, as required in (5.63).

The empirical residuals obtained from Burg's estimate for the AR[4] model are plotted in Fig. 5.15. Figs. 5.16 and 5.17 contain plots of the empirical residuals obtained from ML estimates for an AR[4] and an ARMA[1,1] model using `arima()`.

The time series of the empirical residuals, as shown in plots (a) in Figs. 5.15, 5.16, and 5.17, are found to be stationary using the diagnostics introduced in Sect. 2.5.1. From plots (b) in these figures, it is concluded

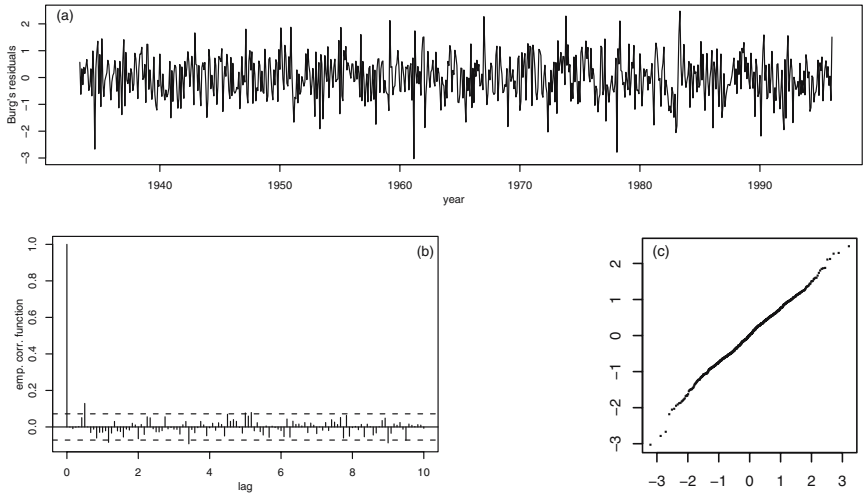


Fig. 5.15. SO index from January, 1933, through to January 1966. (a) Empirical residuals of an AR[4] model estimated using Burg's algorithm together with their (b) correlation function and (c) normal probability plot. (The normal probability plot is introduced in Fig. 2.18).

that the empirical residuals are afflicted, in lag 0.5 years, i.e., six months, with a small correlation being, however, outside the .95 confidence interval for no correlation: the empirical residuals do not stem from a white noise process. From plots (c) in these figures, it is concluded that the empirical residuals are approximately normally distributed. Therefore, the AR[4] and ARMA[1,1] estimates in Table 5.3 are only approximately reconcilable with (5.63). The performance of both models is comparable when all diagnostics, (5.60), (5.61) and (5.63), are applied.

In Problem 5.18, the models in Table 5.3 are re-assessed to arrive possibly at a model with un-correlated and normally distributed residuals. Models other than the AR[4], AR[3] and ARMA[1,1] ones, however, will have coefficients with asymptotically .95 confidence intervals that contain zero, as is obvious from Table 5.3. Consequently, no candidate model in Table 5.3 has (i) coefficients with .95 invervals not containing zero as required in (5.60), (ii) minimal AIC and BIC values as required in (5.61) and also (iii) empirical residuals stemming from a normally distributed white noise process as required in (5.63). The ARMA[1,1] and the AR[4] models, however, are preferred as they are the most parsimonious.

A model chosen by applying diagnostics (5.60), (5.61) and (5.63) as demonstrated above, should undergo a last diagnostic. This diagnostic is a version of the one used in Sect. 2.3 to compare AR[1] and MA[1] models fitted to stationary time series using trial and error: in Fig. 2.15, for example,

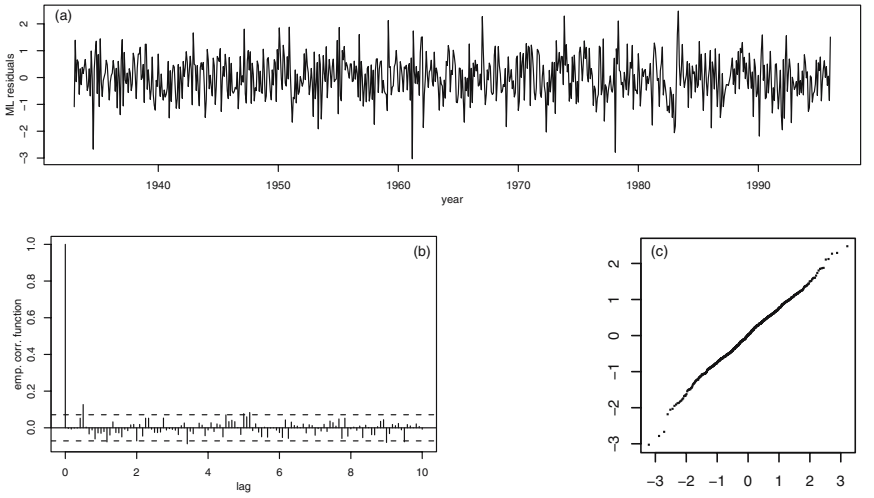


Fig. 5.16. SO index from January, 1933, through to January 1966. (a) Empirical residuals of an AR[4] model estimated using the maximum likelihood (ML) method together with their (b) correlation function and (c) normal probability plot.

the first differences of the yearly values in the Basel temperature series are plotted together with simulations obtained from three MA[1] models in order to find a model reconcilable with the observations. One notices the strength of the diagnostic increasing when the time series is compared with simulations obtained from only one model, as required in (5.64).

Using the estimated model, several realisations are simulated and, among their plots, a plot of the observations is hidden. If a person not initiated detects these observations, then an alternative model should be estimated. (5.64)

Is the ARMA[1,1] model for the SO index as estimated in Table 5.3 in agreement with the diagnostic proposed in (5.64)? Realisations from an estimated ARMA$[p, q]$ model are calculated with `arima.sim()`. This R function in introduced in Problems 2.22 and 5.11. From the ARMA[1,1] model as estimated in Table 5.3, with

```
wnlength <- 757  #length of time series
wnvar <- 0.6009  #variance of innovations
w <- rnorm(wnlength,mean=0,sd=sqrt(wnvar))  #white noise
wts <- ts(w,start=1933,frequency=12,)       #as time series in R
soarma11sim <- arima.sim(model=list(order=c(1,0,1),
          ar=c(-0.8829), ma=c(0.4649)), n=wnlength, innov=wts)
```

seven simulated realisations are generated and then plotted in Fig. 5.18. Fig. 5.18 also contains a plot of the observed time series.

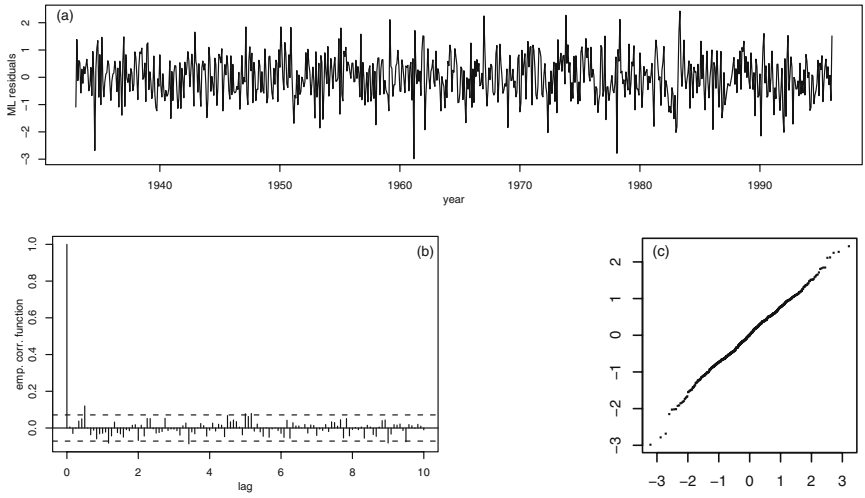


Fig. 5.17. SO index from January, 1933, through to January 1966. (a) Empirical residuals of an ARMA[1,1] model estimated using the ML method together with their (b) correlation function and (c) normal probability plot.

In Fig. 5.18, the time series hidden in the simulations can hardly be detected without consulting Fig. 5.14. Therefore, an ARMA[1,1] model as estimated in Table 5.3 is in agreement with the SO index plotted in Fig. 5.18. The same result is obtained for an AR[4] model as estimated in Table 5.3 applying diagnostic (5.64) in Problem 5.30.

Additional diagnostics for ARMA$[p, q]$ models are proposed in [16], [21], [17], and [22]. However, by increasing the number of diagnostics applied to a set of candidate models, the probability increases that no candidate model is reconcilable with the observations. This situation is arrived at, in the case of the SO index, applying the diagnostics (5.60), (5.61), (5.63) and (5.64) to the candidate models in Table 5.3.

5.4.5 Selection and Summary

When two or more candidate models are found to be reconcilable with the diagnostics proposed above, in Sect. 5.4.4: which one is preferred? An answer to this question is to be found in the fourth and last step in (5.1).

Often this step is straightforward, i.e., if a clear and non-ambigouous result has already been arrived at when identifying the model in the first step (5.1,1). For example, only one candidate model is found, in Sect. 5.1.3, for each of the example time series, i.e., the wind speed values measured in a turbulent atmospheric flow and the amounts of a product formed during a chemical reaction and, in Sects. 5.4.2 and 5.4.1, for the yearly values in the

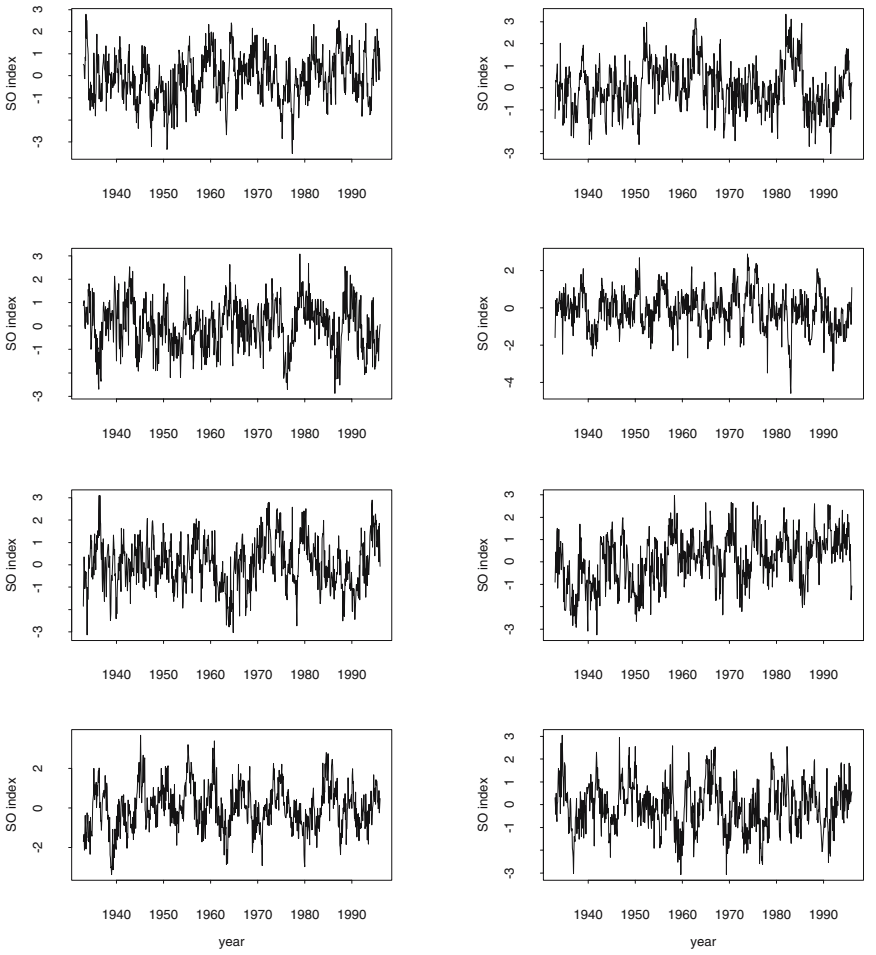


Fig. 5.18. SO index from January, 1933, through to January 1966. The observed time series is hidden in seven simulations calculated using the ARMA[1,1] model as estimated in Table 5.3.

Basel temperature series and the monthly values in the MLO atmospheric CO_2 series. These models can be shown to be in agreement with the observed time series by applying the diagnostics (5.60), (5.61), (5.63) and (5.64).

Far more difficult, however, is the selection of a model for the SO index from those estimated in Table 5.3. The AR[4] and ARMA[1,1] models have empirical residuals that are correlated with lag 0.5 years as seen in Figs. 5.15, 5.16 and 5.17, whereas the less parsimonious ones have coefficients with .95 confidence intervals including zero. These ambiguities can be avoided when the diagnostics are weighted. If a very parsimonious model is

preferred and the small correlations in the empirical residuals are ignored then the ARMA[1,1] model is selected. However, if a model with a larger number of possibly non-essential parameters can be used then an autoregressive model of higher order but with no correlations in its empirical residuals is selected.

In this section, proceeding stepwise guided by (5.1), the following models are found to be reconcilable with the example time series:

1. an ARIMA$[0, 1, 1]$ with the Basel temperature yearly values
2. an ARIMA$[2, 0, 0] \times [0, 1, 1]_{12}$ with the atmospheric CO_2 monthly values
3. an ARMA[1,1] with the SO index monthly values.

These models do not reveal much of the properties of the natural phenomena involved (Problem 5.31). Making use of them, however, predictions can be calculated very efficiently as demonstrated in Sect. 5.5.

5.5 Predictions

Using an ARIMA$[p, d, q]$ model reconcilable with an observed time series, predictions can be calculated assuming that the predicted values belong to the realisation observed. The very same assumption is made in Sect. 2.7 where optimal linear predictions for stationary processes are derived. The problems mentioned in the remarks concluding Sect. 2.7 can be solved when a prediction is calculated for a linear process as defined in (2.25): the reduction of the set of processes for which predictions are possible is compensated by efficient algorithms for their calculation. A linear process is written parsimoniously as an ARMA$[p, q]$ model (5.36). Assuming that the true coefficients in an ARMA$[p, q]$ model are known, two prediction formulae are derived in Sect. 5.5.1. Thereafter, in Sect. 5.5.2, predictions are calculated with estimates substituted for the true coefficients.

5.5.1 Predictions Using a Theoretical ARMA$[p, q]$ Model

In this subsection, two predictions using an ARMA$[p, q]$ model as defined in (5.36) are derived from (i) the AR$[\infty]$ representation and (ii) the difference equation of the model. From the AR$[\infty]$ representation proposed in (5.47) a prediction is derived in order to show that the innovation W_t can be calculated from X_t, X_{t-1}, ..., i.e., both the present and past model variables. Consequently, provided that the model reduces to an MA$[q]$ process, (5.33,1) holds. The results of this derivation are given in (5.65), (5.66) and (5.68). For practical purposes, however, predictions are calculated from the ARMA$[p, q]$ model as defined in (5.36), i.e., from its difference equation. The result is the recursion (5.70).

A derivation of a prediction using the AR$[\infty]$ representation (5.47) of an ARMA$[p, q]$ model is motived by propositions (5.7), (5.8), (5.9) and (5.10),

from which it is concluded that the coefficients in an AR[p] model are the weights of an optimal linear prediction for X_t, provided that X_{t-1}, X_{t-2}, ..., X_{t-p} are known. The result obtained in Sect. 5.6.3 is summarised in (5.13): there, an optimal linear prediction for an AR[p] model has been found provided that there is only one time step between the random variable to be predicted and the last random variable known.

Are the π-weights in the AR[∞] representation (5.47) of an ARMA[p, q] model (5.36), $X_t - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \pi_3 X_{t-3} - \ldots = W_t$, the weights of an optimal linear prediction X_t provided that X_{t-1}, X_{t-2}, ... are known?

If (X_t) is an ARMA[p, q] model as defined in (5.36) with its AR[∞] representation in (5.47), then:
1. *the one-step prediction forward from time point t for X_{t+1}, $\widehat{X}_{t+1}(\infty) = \pi_1 X_t + \pi_2 X_{t-1} + \ldots$, is linear and optimal for all t,*
2. *the errors $\big(\hat{e}_{t+1}(\infty)\big)$ of a sequence of one-step predictions are a white noise process (W_t), and*
3. *the mean square prediction error is the variance of (W_t): $\mathrm{PVar}\widehat{X}_{t+1}(\infty) = \sigma_W^2$.*

(5.65)

In (5.65,1), as well as in (5.13), there is only one time step between the random variable to be predicted and the last random variable known: a one-step prediction $\widehat{X}_{t+1}(\infty)$ for X_{t+1} is calculated. The prediction is a linear combination of the π-weights in the AR[∞] representation and of the present and past random variables. The errors $\hat{e}_{t+j+1}(\infty)$ pertaining to a sequence of one-step predictions forward from time points $t+j$, $j = 0, 1, 2 \ldots$, are not correlated, and the variance of these errors is the mean square prediction error. The properties of the one-step prediction in (5.65) are derived in Sect. 5.6.3.

Optimal k-step predictions forward from time point t can be calculated using optimal one-step predictions:

$$\begin{aligned}
\widehat{X}_{t+1}(\infty) &= \pi_1 X_t + \pi_2 X_{t-1} + \pi_3 X_{t-2} + \ldots \\
\widehat{X}_{t+2}(\infty) &= \pi_1 \widehat{X}_{t+1}(\infty) + \pi_2 X_t + \pi_3 X_{t-1} + \pi_4 X_{t-2} \ldots \\
\widehat{X}_{t+3}(\infty) &= \pi_1 \widehat{X}_{t+2}(\infty) + \pi_2 \widehat{X}_{t+1}(\infty) + \pi_3 X_t + \pi_4 X_{t-1} \ldots \\
&\ldots
\end{aligned} \tag{5.66}$$

In the first one-step prediction given in (5.66), $\widehat{X}_{t+1}(\infty)$ is calculated as a substitute for X_{t+1}. In the second one-step prediction, $\widehat{X}_{t+2}(\infty)$ is obtained from $\widehat{X}_{t+1}(\infty)$, X_t, X_{t-1}, ..., as a substitute for X_{t+2}. In one-step prediction no. k, the optimal k-step prediction $\widehat{X}_{t+k}(\infty)$ is arrived at using the prior predictions for $\widehat{X}_{t+k-1}(\infty)$, ..., $\widehat{X}_{t+1}(\infty)$ and the known variables X_t, X_{t-1}, X_{t-2}, X_{t-3}, ..., assuming that the infinite past of the process is known.

Using (5.65) and (5.66), optimal predictions for $t+1$, $t+2$, $t+3$, ... are obtained. These predictions as linear combinations of (i) the variables (in applications, the observations) in an ARMA[p, q] model for t, $t-1$, $t-2$, ...

and (ii) the π-weights in the AR[∞] representation of the model. The linear combinations converge, provided that the ARMA[p, q] model is invertible. The predictions are, as required in (5.87,3), optimal in the mean over all possible realisations of the linear process provided that the ARMA[p, q] model is the true model. If this is the case then there is no prediction (possibly not even a non-linear one) with constant weights (i.e., weights not varying in time) to which a smaller mean square error pertains since the prediction is calculated using the definition of the linear process.

The error and the mean square error of a k-step prediction in (5.67) and (5.68)

$$\hat{e}_{t+k}(\infty) = X_{t+k} - \widehat{X}_{t+k}(\infty) = \sum_{u=0}^{k-1} \psi_u W_{t+k-u} \tag{5.67}$$

$$\mathrm{PVar}\widehat{X}_{t+k}(\infty) = \sigma_W^2 \sum_{u=0}^{k-1} \psi_u^2 \tag{5.68}$$

are obtained, in (5.94) and (5.95), using the k-step prediction derived in (5.94) from the MA[∞] representation of the ARMA[p, q] model.

$\hat{e}_{t+k}(\infty)$ in (5.67) is an MA[$k-1$] model, with the ψ-weights as obtained in the MA[∞] representation of the ARMA[p, q] model. Thus, due to (5.40), the prediction errors $\hat{e}_{t+1}(\infty)$, ..., $\hat{e}_{t+k-1}(\infty)$, $\hat{e}_{t+k}(\infty)$ are correlated when predicting in forward direction from a fixed time point t.

Despite (5.65), (5.66) and (5.68) only being applicable when the infinite past of the process (i.e., the variables X_{t-j} and the innovations W_{t-j}, $j = 1, 2, \ldots$) is known, these formulae can still be used to calculate predictions in applications, i.e., with observations limited to a finite past, provided that

1. the π weights decay rapidly since only a finite past is available, and
2. the estimated coefficients in the ARMA[p, q] model are assumed to be the true ones.

The π weights usually decay very rapidly in applications and thus allow for the calculation of a prediction from a relatively short past. The error originating from substituting estimates for the true coefficients can be accounted for, as demonstrated in Sect. 5.5.2.

For example, an ARMA[1,1] model as estimated in Table 5.3 for the SO index is reconcilable with the first condition stipulated above and therefore, in Problem 5.21, optimal linear predictions and the pertaining errors are calculated for the months following January 1996, the last month for which an observation is available. As a second example, in Problem 5.22, assuming $\breve{\sigma}_W^2 = 0.489$ and $\breve{b} = 0.881$ as estimated in the remarks to Fig. 5.5 to be the true parameters in an MA[1] model for the differences of the yearly means in the Basel temperature series, these differences are predicted for the years 1958 and 1959. The predicted differences are then added, to obtain the predictions for the yearly temperature values.

These examples demonstrate that calculations of optimal linear predictions using (5.65) and (5.66) can be lengthy. Less time is needed for a calculation using the difference equation of the ARMA$[p,q]$ model in (5.36).

A derivation of a prediction using the definition of the ARMA$[p,q]$ model is motived by the hope of finding a prediction that can be calculated speedily for use in time-critical applications, e.g., for controlling a manufacturing process. Using the results obtained in Sect. 2.8.4, the k-step prediction of an ARMA$[p,q]$ model is calculated, in Sect. 5.6.4, as a conditional expectation from the difference equation of the model to obtain the optimal one-step prediction proposed in (5.69):

$$\begin{aligned}
&\textit{If the time series } (x_t, x_{t-1}, \ldots, x_{t-N+1} = \boldsymbol{x}_{N,t}) \textit{ is a time slice}\\
&\textit{from a realisation of } (X_t),\ (X_t) \textit{ being an ARMA}[p,q] \textit{ model}\\
&\textit{as defined in (5.36), then}\\
&\widehat{X}_{t+1}(\boldsymbol{X}_{-\infty,t}) = \mathrm{E}\big(X_{t+1}|\boldsymbol{X}_{-\infty,t}\big)\\
&\quad = \sum_{i=1}^{p} a_i x_{t+1-i} \; - \sum_{j=1}^{q} b_i \big(X_{t+1-j} - \widehat{X}_{t+1-j}(\boldsymbol{X}_{-\infty,t})\big)\\
&\textit{is the optimal one-step step prediction of } (X_t),
\end{aligned} \tag{5.69}$$

and the optimal k-step prediction in (5.70):

$$\begin{aligned}
&\textit{Under the assumptions in (5.69),}\\
&\widehat{X}_{t+k}(\boldsymbol{X}_{-\infty,t}) = \mathrm{E}\big(X_{t+k}|\boldsymbol{X}_{-\infty,t}\big)\\
&\quad = \sum_{i=1}^{p} a_i \mathrm{E}\big(X_{t+k-i}|\boldsymbol{X}_{-\infty,t}\big)\\
&\qquad\qquad + \mathrm{E}\big(W_{t+k}|\boldsymbol{X}_{-\infty,t}\big) - \sum_{j=1}^{q} b_j \mathrm{E}\big(W_{t+k-j}|\boldsymbol{X}_{-\infty,t}\big)\\
&\textit{is the optimal } k\textit{-step prediction of } (X_t), \textit{ with, due to}\\
&\textit{the finite past using } \boldsymbol{x}_{N,t}\textit{:}\\
&1.\ \mathrm{E}\big(X_u|\mathbf{X}_{-\infty,t}\big) = \begin{cases} x_u & \textit{for } 1 \le u \le t\\ \widehat{X}_u(\mathbf{X}_{-\infty,t}) & \textit{for } u > t \end{cases}\\
&2.\ \mathrm{E}\big(W_u|\boldsymbol{X}_{-\infty,t}\big) = \begin{cases} \widehat{W}_u = X_u - \widehat{X}_u(\boldsymbol{X}_{-\infty,t}) & \textit{for } 1 \le u \le t\\ \mu_W = 0, & \textit{for } u > t \end{cases}\\
&3.\ \widehat{W}_u = r_u^* \textit{ as defined in (5.62).}
\end{aligned} \tag{5.70}$$

An optimal k-step prediction of an ARMA$[p,q]$ model is calculated using recursion (5.70) provided that the infinite past of the process is known. Although this is not the case in applications, predictions can be calculated using the past of the model, as far back as it is known, by substituting the observed x_u for $\mathrm{E}\big(X_u|\boldsymbol{X}_{-\infty,t}\big)$, and the empirical residuals r_u^* for $\widehat{W}_u$, because the optimal estimation of a random variable is an observation as argued in the remarks to (4.37) and (5.97). Therefore, by calculating as required in

Table 5.5. Predictions for the de-meaned SO index obtained from an ARMA[1,1] model as estimated in Table 5.3 under the assumption that the estimated $\breve{\sigma}_W^2 = 0.5997$, $\breve{\mu}_X = -0.1411$, $\breve{a} = 0.8829$ and $\breve{b} = 0.4649$ are the true model parameters.

	...	$t-4$	$t-3$	$t-2$	$t-1$	t	$t+1$	$t+2$	$t+3$
	...				1995	1996			
	...	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	March	April
SO index $-\breve{\mu}_X$	...	.4411	−0.0588	.1411	−.7588	1.2411			
π weights	...	.0195	.0420	.0903	.1943	.4180			
$\widehat{X}$ (5.65)							.3913		
π weights	...	.0091	.0195	.0420	.0903	.1943	.4180		
$\widehat{X}$(5.66)								.3455	
π weights	...	.0042	.0091	.0195	.0420	.0903	.1943	.4180	
$\widehat{X}$ (5.66)									.3051
r_u^* (5.62)	...	.3543	−.3011	.0373	−.8827	1.4870			
$\widehat{X}$ (5.70)							0.3914	0.3456	0.3051
PVar (5.65)							.5997		
PVar (5.68)								.7045	.7861

(5.69) and (5.70), often good predictions are obtained even without using all observations $(x_t, x_{t-1}, \ldots, x_{t-N+1})$.

The SO index, for example, can be predicted, using an ARMA[1,1] model as estimated in Table 5.3, from January 1996 in forward direction for February, March and April 1996, under the assumption that the estimates are the true coefficients. The results in Table 5.5 are obtained using (5.70) as well as (5.65) and (5.66) assuming the following pasts:

1. when predicting with (5.70): the empirical residuals calculated, in Problem 5.23 forward from time point January 1995 as required in (5.62), act as substitutes for the past innovations;
2. when predicting with (5.65) and (5.66): the predictions are calculated with 19 observations using the π-weights 0.4180, 0.1943, 0.0903, 0.0420, 0.0195, ... obtained from the convolution $(1, -0.4649)^{-1} * (1, -0.8829)$ (Problem 5.21).

Assuming these pasts, both predictions obtained in Table 5.5 are very close in their values. The mean square prediction errors are calculated as required in (5.65) and (5.68) using the ψ-weights −0.4211, −0.3730, −0.3304, −0.2927, ..., which are obtained from the convolution $(1, -0.8829)^{-1} * (1, -0.4649)$.

The predictions arrived at in Table 5.5, applying (5.65) and (5.66), become, after compensating for $\breve{\mu}_X = -0.1411$ and calculating the roots of the

mean square prediction errors, identical with those obtained from R function `predict()`

```
predict(so.arma11fit, n.ahead=3)
$pred       Feb       Mar       Apr    #predictions
1996 0.2502408 0.2043970 0.1639230
$se         Feb       Mar       Apr    #errors
1996 0.7744265 0.8393552 0.8866731
```

that predicts under the assumption that the estimated coefficients are the true coefficients.

Confidence intervals for optimal predictions (prediction intervals) of an ARMA$[p, q]$ model can be calculated if the probability distribution of the prediction error $\hat{e}_{t+k}(\infty) = \sum_{u=0}^{k-1} \psi_u W_{t+k-u}$ obtained in (5.67) is known. If the innovations (W_t) are normally distributed, then $\hat{e}_{t+k}(\infty)$, being a linear combination of normally distributed random variables, is also normally distributed (the very same argument is used to derive (5.56)) and, therefore, the prediction interval (5.71) follows.

Let (X_t) be an ARMA$[p, q]$ model as defined in (5.36) with normally distributed innovations (W_t) and $\widehat{X}_{t+k}(\infty)$ be a prediction calculated using either (5.65) and (5.66) or (5.70) with mean square error $\mathrm{PVar}\widehat{X}_{t+k}(\infty)$ *as required in (5.68). Then*

$$\widehat{X}_{t+k}(\infty) \pm 1.96\sqrt{\mathrm{PVar}\widehat{X}_{t+k}(\infty)} \text{ is a .95 prediction interval for } \widehat{X}_{t+k}(\infty). \tag{5.71}$$

Prediction intervals as defined in (5.71) apply for each prediction individually.

5.5.2 Predictions Using an Estimated ARMA$[p, q]$ Model

Predictions are calculated above, in Sect. 5.5.1, under the assumption that an estimated ARMA$[p, q]$ model is the true model. Do these predictions change when the estimates are not assumed to be the true model coefficients?

Uncertain estimates for the standard deviation σ_W of the innovations cascade linearly to the prediction intervals calculated using (5.71). The distribution of an ML estimate for σ_W is approximated using (5.26,2), provided that the innovations are normally distributed.

Uncertainties in estimated autoregressive and moving-average coefficients can also be accounted for. If the innovations are normally distributed or, asymptotically, if the number of observations is large, then, since (5.26,1) also applies for ARMA$[p, q]$ models, ML estimates for ARMA$[p, q]$ coefficients are multivariate normally distributed. Applying these approximations together with the variances and covariances of the estimates, diagnostics can be constructed that allow to assess the uncertainties introduced when a prediction is calculated using an estimated model.

Table 5.6. Predicting the de-meaned SO index using the ARMA[1,1] model estimated in Table 5.3 taking into account the uncertainties of the estimated $\breve{a} = 0.8829$ and $\breve{b} = 0.4649$, and assuming that (i) $\breve{\sigma}_W^2 = 0.5997$ is the true value and (ii) $\breve{a}$ and $\breve{b}$ are normally distributed with $\breve{\sigma}_a = 0.0233$ and $\breve{\sigma}_b = 0.0432$.

predictions using (5.70)		1996		
.95 prediction intervals (5.71)		Feb.	March	April
.95 interval "expected"		1.9092	1.9908	2.0430
"very high"	$\breve{a} + 1.96\breve{\sigma}_a$, $\breve{b} - 1.96\breve{\sigma}_b$	.5763	.5353	.4971
"high"	$\breve{a} - 1.96\breve{\sigma}_a$, $\breve{b} - 1.96\breve{\sigma}_b$	.4330	.3877	.3245
"expected"	$\breve{a}$, $\breve{b}$	.3914	.3456	.3051
$\sqrt{\text{PVar}}$		.7744	.8394	.8867
"low"	$\breve{a} + 1.96\breve{\sigma}_a$, $\breve{b} + 1.96\breve{\sigma}_b$	.3120	.2971	.2759
"very low"	$\breve{a} - 1.96\breve{\sigma}_a$, $\breve{b} + 1.96\breve{\sigma}_b$	.2065	.1729	.1448
.95 interval "expected"		-1.1264	-1.2996	-1.4328

For instance, the properties discussed in the remarks to (5.26) apply to the covariance matrix of the ARMA[1,1] model estimates in Table 5.3, obtained using `arima()` in Sect. 5.4.3. From the variances in the diagonal of this covariance matrix, the standard deviations $\breve{\sigma}_a = 0.0233$ ($\breve{}$ denotes an ML estimate) and $\breve{\sigma}_b = 0.0432$ pertaining to the estimates $\breve{a} = 0.8829$ and $\breve{b} = 0.4649$ (in Table 5.3 and Sect. 5.4.3) together with the empirical correlation of $\breve{a}$ and $\breve{b}$, $\breve{r}_{ab} = \breve{c}_{ab}/(\breve{\sigma}_a\breve{\sigma}_b) = 0.006914/(0.0233 \times 0.0432) = 0.6867$ are obtained, and these results are then used to construct .95 confidence intervals $\breve{a} \pm 1.96\breve{\sigma}_a = (0.8372, 0.9286)$ and $\breve{b} \pm 1.96\breve{\sigma}_b = (0.3802, 0.5495)$.

For values in these .95 confidence intervals, the predictions given in Table 5.6 are calculated using (5.70) with empirical residuals, as obtained for the predictions in Table 5.5. The predictions, under the assumption that the estimated parameters are the true ones, stem from Table 5.5 and are shown in the line "expected" in Table 5.6. The coefficients being linear in a sequence of predictions calculated using (5.70) for $t+1$, $t+2$, ..., larger predictions (in lines "very high" in Table 5.6) are arrived at for values for a closer to the upper bound and values for b closer to the lower bound of the pertaining .95 confidence intervals than for any other values from these intervals. Vice-versa, the smallest predictions (in lines "very low" in Table 5.6) are obtained for values close to the lower bound for a and close to the upper bound for b.

It is not very likely to obtain the values in lines "very high" and "very low" since the correlation of $\breve{a}$ and $\breve{b}$ is positive ($\breve{r}_{ab} = 0.6867$). Due to this positive correlation, an underestimation of a (i.e., $\breve{a} < a$) is often arrived

at together with an underestimation of b, and vice-versa. In these cases, the effects compensate (lines "low" and "high" in Table 5.6).

From Table 5.6 it is concluded that the predictions for the de-meaned SO index, as given in Table 5.5, can change slightly when uncertainties due to predicting with an estimated model are accounted for. These changes however are small compared to the .95 confidence interval (lines ".95 interval predicted" in Table 5.6) pertaining to the predictions obtained under the assumption that the estimated coefficients ($\breve{a}$ and $\breve{b}$) are the true ones (a and b). Hence, when using the ARMA[1,1] model estimates in Table 5.3 for predicting the SO index, the uncertainties due to the assumption that the estimated coefficients are the true ones can be neglected.

This approximation is feasible because the variances of the estimated coefficients (their standard deviations are calculated above in the remarks to Table 5.6) are small. It is therefore recommended to bear in mind that an AR[p], MA[q] or ARMA[p, q] model will be used for predicting, when the diagnostics introduced in Sect. 5.4.4 are applied to select one of the candidate models identified and estimated as demonstrated in Sects. 5.4.2 and 5.4.3: the estimates for the coefficients in a parsimonious model usually (not always, counter-examples being the ARMA[2,1] and AR[4] models in Table 5.3) have smaller variances than the estimates for the coefficients in a non-parsimonious model.

5.5.3 Summary

AR[p], MA[q] and ARMA[p, q] models as defined in (5.2), (5.28) and (5.36) are linear processes as defined in (2.25). The ARIMA[p, d, q] model as defined in (5.55) becomes a linear process when its differences of order d are calculated using definition (2.40). An AR[p], MA[q] or ARMA[p, q] model is fitted to a stationary time series (or a time series having become stationary after calculating its differences) by performing the steps in (5.1) which require a de-meaning or de-trending of the time series and then the identification, estimation, diagnostic and selection of a model as described in Sect. 5.4.

Then, using the estimated coefficients in recursion (5.70), predictions for time points with no observations can be calculated under the assumption that the predicted values belong to the very same realisation as the observed ones. The prediction is obtained using an explicit or recursive procedure. This is a massive advantage as compared to the calculation of an optimal linear prediction using (2.63), since (2.63,1) requires the inversion of the empirical covariance matrix of the process. This inversion is considered to be time-consuming and possibly delicate in the remarks concluding Sect. 2.7.

5.6 Supplements

The Levinson-Durbin recursion, the partial correlation function of a stationary process and the predictions using an ARMA$[p,q]$ model are obtained, in Sects. 5.6.1, 5.6.2, 5.6.3 and 5.6.4, using longer derivations. The arguments involved are quite straightforward if considered individually; in the chain of the derivations, however, they become powerful enough to prove, e.g., in Sect. 5.6.1, the Levinson-Durbin recursion proposed in (5.14).

5.6.1 Levinson-Durbin Recursion

To arrive at the Levinson-Durbin recursion, we start with the properties proposed in (2.9), i.e., that the covariance function of a stationary stochastic process (X_t) depends on the lag $t-j$ only and is symmetric in its origin: $c_X(t,t-j) = c_X(t-j)$ and $c_X(t,t-j) = c_X(t,t+j)$, $j = 1,2,\ldots,k$. Additionally, $\mu_X = 0$ is assumed. Hence, in the backward prediction $\widehat{X}_t^{(r)}$ of X_t as defined in (5.72):

$$\widehat{X}_t^{(r)} = a_{1k}X_{t+1} + a_{2k}X_{t+2} + \ldots + a_{kk}X_{t+k} \tag{5.72}$$

$$\hat{e}_t^{(r)}(k) = X_t - \widehat{X}_t^{(r)}(k) \tag{5.73}$$

$$\mathrm{PVar}\widehat{X}_t^{(r)}(k) = \mathrm{E}\Big(\big(\hat{e}_t^{(r)}(k)\big)^2\Big) = \mathrm{E}\Big(\big(X_t - \widehat{X}_t^{(r)}(k)\big)^2\Big) \quad \text{minimal} \tag{5.74}$$

the random variables X_t are multiplied with weights $a_{1k},\ldots,a_{kk}$ being identical with the weights used in the forward prediction (5.8), since, by evaluating the optimality conditions (2.74), a system of equations with identical covariances is obtained in both cases. In applications, (5.72) allows for estimating the variable immediately prior to the first observation, whereas (5.8) allows for estimating the variable immediately after the last observation, in both cases with k values following or preceding the variable to be estimated.

Both predictions (5.8) and (5.72) are evaluated, as demonstrated in (2.75) and (2.76), to arrive at the Yule-Walker equations until lag k. These have one and only one solution, and thus (5.75), (5.76) and (5.77) hold.

$$\mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k)\big)X_{t-j}\Big) = \mathrm{E}\Big(\big(\hat{e}_t^{(r)}(k)\big)X_{t+j}\Big) = 0, \qquad j = 1,\ldots,k \tag{5.75}$$

$$\begin{aligned}\mathrm{PVar}\widehat{X}_t^{(v)}(k) &= \mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k)\big)\big(\hat{e}_t^{(v)}(k)\big)\Big) = \mathrm{E}\Big(\big(\hat{e}_t^{(r)}(k)\big)\big(\hat{e}_t^{(r)}(k)\big)\Big)\\ &= \mathrm{PVar}\widehat{X}_t^{(r)}(k) = \mathrm{PVar}(k)\end{aligned} \tag{5.76}$$

$$\begin{aligned}\mathrm{PVar}\widehat{X}_t^{(v)}(k) &= \mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k)\big)X_t\Big) = \mathrm{E}\Big(\big(\hat{e}_t^{(r)}(k)\big)X_t\Big)\\ &= \mathrm{PVar}\widehat{X}_t^{(r)}(k) = \mathrm{PVar}(k)\end{aligned} \tag{5.77}$$

(5.75), (5.76) and (5.77), however, also follow from (5.13,2) where it is stipulated that the Yule-Walker equations connect the covariances of a stationary process with the weights of its optimal linear forward prediction.

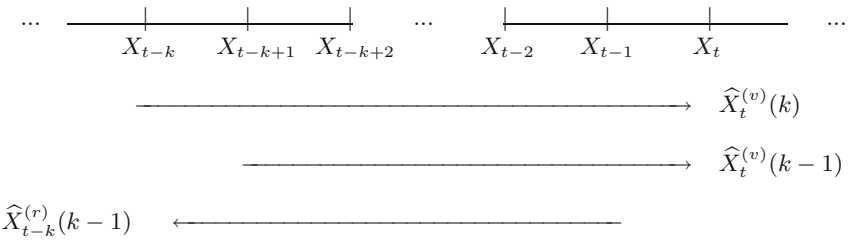


Fig. 5.19. Forward predictions $\widehat{X}_t^{(v)}(k)$ and $\widehat{X}_t^{(v)}(k-1)$ for time X_t using k and $k-1$ variables preceding t; backward prediction $\widehat{X}_{t-k}^{(r)}(k-1)$ for X_{t-k} using $k-1$ following variables. The errors of these predictions are related by equation (5.78).

If the set of the stationary (X_t) is restricted to the set of AR[p] models defined in (5.2), then the Yule-Walker equations connect the covariances of the model with its coefficients as proposed in (5.13,1), and, because in a version of (5.75) applying for causal (as defined in (5.3)) AR[p] models, the white noise variable for time t replaces the prediction error, and we can argue that the covariances $\mathrm{E}(W_t X_{t-j})$ and $\mathrm{E}(W_t X_{t+j})$, $j = 1, \ldots, k$, become identically zero.

It has been shown above, that (5.75), (5.76) and (5.77) are implied by the properties of the covariance function of a stationary stochastic process as proposed in (2.9). From these properties, also (5.78) with θ_k as required in (5.79) is obtained in the remarks to (5.82) and (5.83). Then, from (5.78), the Levinson-Durbin recursion as proposed in (5.14) is obtained in Problems 5.4 and 5.5.

$$\hat{e}_t^{(v)}(k) \;=\; \hat{e}_t^{(v)}(k-1) - \theta_k\big(\hat{e}_{t-k}^{(r)}(k-1)\big) \tag{5.78}$$

The error $\hat{e}_t^{(v)}(k)$ of the forward prediction for X_t using k preceding variables is calculated, in (5.78), by correcting the error $\hat{e}_t^{(v)}(k-1)$ of the forward prediction for X_t using only $k-1$ preceding X_t. The correction applied to $\hat{e}_t^{(v)}(k-1)$ is obtained by multiplying the error of the backward prediction $\hat{e}_{t-k}^{(r)}(k-1)$ for X_{t-k} with the weight θ_k. X_{t-k} is the variable contributing to $\widehat{X}_t^{(v)}(k)$ but not to $\widehat{X}_t^{(v)}(k-1)$. The intervals containing the variables used in these predictions, i.e., the predictions to which the errors in (5.78) pertain, are drawn as arrows in Fig. 5.19.

Fig. 5.19 shows that the predictions $\widehat{X}_t^{(v)}(k-1)$ and $\widehat{X}_{t-k}^{(r)}(k-1)$ are calculated using an identical set of variables X_{t-j}, $j = 1, \ldots, k-1$, being between X_{t-k} and X_t. This implies, using the orthogonality conditions (2.79), that both prediction errors $\hat{e}_t^{(v)}(k-1)$ and $\hat{e}_{t-k}^{(r)}(k-1)$ are not correlated with X_{t-j}, $j = 1, \ldots, k-1$, and therefore, both (5.82) as well as (5.83) follow for $j = 1, \ldots, k-1$ and arbitrary real θ_k. Does a θ_k exist such that (5.82) and (5.83) also hold for $j = k$? Can we calculate this θ_k?

Assume that this θ_k can be calculated from the left side in (5.79). In (5.79), $\mathrm{E}\big(\big(\hat{e}_t^{(v)}(k-1)\big)X_{t-k}\big)$ then becomes, using (5.8,1,2), $\mathrm{E}\big(\big((X_t - \sum_{j=1}^{k-1} a_{j,k-1}X_{t-j}\big)X_{t-k}\big)$, and $\mathrm{E}\big(\big(\big(\hat{e}_{t-k}^{(r)}(k-1)\big)X_{t-k}\big)$, becomes, with (5.77), $\mathrm{PVar}\widehat{X}_{t-k}^{(r)}(k-1)$. (5.79) and (5.80) then follow with $\mu_X = 0$. Like a correlation, $-1 \le \theta_k \le 1$, seeing that (i) identical coefficients and covariances $c_X(1), \ldots, c_X(k-1)$ occur in $\sum_{j=1}^{k-1} a_{j,k-1}c_X(k-j) = \sum_{j=1}^{k-1} a_{j,k-1}c_X(j)$ and (ii) $-c_X(0) \le c_X(k) \le c_X(0)$, as proposed in (2.9,3).

$$\frac{\mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k-1)\big)X_{t-k}\Big)}{\mathrm{E}\Big(\big(\hat{e}_{t-k}^{(r)}(k-1)\big)X_{t-k}\Big)} = \frac{\mathrm{E}\Big(\big(X_t - \sum_{j=1}^{k-1} a_{j,k-1}X_{t-j}\big)X_{t-k}\Big)}{\mathrm{PVar}\widehat{X}_{t-k}^{(r)}(k-1)} \tag{5.79}$$

$$= \frac{\mathrm{E}\Big(X_tX_{t-k} - \sum_{j=1}^{k-1} a_{j,k-1}X_{t-j}X_{t-k}\Big)}{\mathrm{PVar}(k-1)} \tag{5.80}$$

$$= \frac{c_X(k) - \sum_{j=1}^{k-1} a_{j,k-1}c_X(k-j)}{c_X(0) - \sum_{j=1}^{k-1} a_{j,k-1}c_X(j)} = \theta_k \tag{5.81}$$

When θ_k as obtained in (5.81) is substituted in (5.78), then (5.82) and (5.83) follow for $j = 1, \ldots, k$. For $j = 1, \ldots, k-1$, however, (5.82) and (5.83) hold for arbitrary θ_k, as shown above.

$$\mathrm{E}\left(\big(\hat{e}_t^{(v)}(k-1)\big)X_{t-j}\right) = 0 \quad \text{and} \quad \mathrm{E}\Big(\big(\hat{e}_{t-k}^{(r)}(k-1)\big)X_{t-j}\Big) = 0 \tag{5.82}$$

$$\mathrm{E}\left(\Big(\hat{e}_t^{(v)}(k-1) - \theta_k\hat{e}_{t-k}^{(r)}(k-1)\Big)X_{t-j}\right) = \mathrm{E}\Big(\big(\Delta_t(k)\big)X_{t-j}\Big) = 0 \tag{5.83}$$

Both, the differences $\Delta_t(k)$ in (5.83) and the prediction errors $\hat{e}_t^{(v)}(k)$ in (5.75), are not correlated with the X_{t-j}, $j = 1, \ldots, k$. This implies, using (2.78) and (2.79), that linear predictions for X_t with the errors $\hat{e}_t^{(v)}(k)$ (on the left side in (5.78)) and $\Delta_t(k)$ (on the right sides in (5.83) and in (5.78)) are calculated by evaluating identical optimality conditions. Therefore, the systems of equations, and thus the predictions, together with their errors, are identical. Thus (5.78), with θ_k being calculated using predictions and their errors with argument $k-1$ as required in (5.81), allows for calculating $\hat{e}_t^{(v)}(k)$ recursively.

From the recursion (5.78), the Levinson-Durbin recursion as proposed in (5.14) is obtained in Problems 5.4 and 5.5.

5.6.2 Partial Correlation Function

By substituting definitions (5.8) and (5.72) in the covariance in the numerator in (5.15),

$$
\begin{aligned}
&\mathrm{Cov}\big(X_t - \widehat{X}_t^{(v)}(k-1), X_{t-k} - \widehat{X}_{t-k}^{(r)}(k-1)\big)\\
&= \mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k-1)\big)\big(\hat{e}_{t-k}^{(r)}(k-1)\big)\Big)\\
&= \mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k-1)\big)\big(X_{t-k} - a_{1,k-1}X_{t-k+1} - \ldots - a_{k-1,k-1}X_{t-1}\big)\Big)\\
&= \mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k-1)\big)X_{t-k}\Big) - a_{1,k-1}\mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k-1)\big)X_{t-k+1}\Big) - \ldots\\
&\qquad -a_{k-1,k-1}\mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k-1)\big)X_{t-1}\Big)\\
&= \mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k-1)\big)X_{t-k}\Big) \qquad (5.84)
\end{aligned}
$$

the numerator in (5.79), (5.80) and (5.81) as well as in (5.14,1) is obtained. Then, substituting definitions (5.8) and (5.72) in the denominator in (5.15),

$$
\begin{aligned}
&\Big(\mathrm{Var}\big(X_t - X_t^{(v)}(k-1)\big)\mathrm{Var}\big(X_{t-k} - X_{t-k}^{(r)}(k-1)\big)\Big)^{1/2}\\
&= \left(\mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k-1)\big)\big(\hat{e}_t^{(v)}(k-1)\big)\Big)\mathrm{E}\Big(\big(\hat{e}_{t-k}^{(r)}(k-1)\big)\big(\hat{e}_{t-k}^{(r)}(k-1)\big)\Big)\right)^{1/2}\\
&= \Big(\mathrm{PVar}(k-1)\mathrm{PVar}(k-1)\Big)^{1/2} = \mathrm{PVar}(k-1) \qquad (5.85)
\end{aligned}
$$

the denominator in (5.79), (5.80) and (5.81) as well as in (5.14,1) is obtained. Consequently, the sequence (a_{kk}) generated when the Yule-Walker equations are solved using, for example the Levinson-Durbin recursion, is the sequence of partial correlations $a_X(kk)$.

The partial correlation function of the MA[1] model as proposed in (5.34) and (5.35) is obtained from the correlation function of the model as proposed in (2.23). Applying definition (5.15), the Levinson-Durbin recursion is used to solve the Yule-Walker equations:

$$
\begin{aligned}
a_Y(1,1) &= c_Y(1)/c_Y(0) = \frac{-b}{1+b^2}\\
\mathrm{PVar}(1) &= \frac{1+b^2+b^4}{1+b^2} = \frac{(1+b^2+b^4)(1-b^2)}{(1+b^2)(1-b^2)} = \frac{1-b^6}{(1+b^2)(1-b^2)}\\
a_Y(2,2) &= \frac{-b^2}{1+b^2+b^4} = \frac{-b^2(1-b^2)}{1-b^6}\\
\mathrm{PVar}(2) &= \frac{1+b^2+b^4}{1+b^2}\left(1-\left(\frac{-b^2}{1+b^2+b^4}\right)^2\right) = \ldots\\
&= \frac{1+2b^2+2b^4+2b^6+b^8}{(1+b^2)(1+b^2+b^4)} = \frac{1+b^2-b^8-b^{10}}{(1-b^2)(1+b^2)(1+b^2+b^4)}\\
&= \frac{1-b^8}{(1-b^2)(1+b^2+b^4)} = \frac{(1-b^8)(1-b^2)}{(1-b^2)(1-b^6)} = \frac{1-b^8}{1-b^6}\\
a_Y(3,3) &= \frac{-b^3(1-b^2)}{1-b^6} \times \frac{1-b^6}{1-b^8} = \frac{-b^3(1-b^2)}{1-b^8}
\end{aligned}
$$

$$\begin{aligned} \mathrm{PVar}(3) &= \dots \\ &\dots \\ a_Y(k,k) &= \frac{-b^k(1-b^2)}{1-b^{2(k+1)}}, \qquad k>1 \end{aligned} \tag{5.86}$$

5.6.3 Predictions Derived from the AR[∞] and MA[∞] representations of an ARMA[p, q] Model

The one-step prediction of an ARMA$[p,q]$ model as proposed in (5.65) is derived from the AR$[\infty]$ representation (5.47) in the remarks to (5.88) and (5.89). The k-step prediction together with its error are obtained in (5.93), (5.94) and (5.95).

Assuming stationarity of the stochastic process (X_t), a linear prediction is formulated in (5.87), using (2.62) and (5.8) as templates. The prediction is calculated for time point t from all past variables X_{t-1}, X_{t-2}, ... since an infinite number of weights π_t (unlike in (5.8), the weights have only one index) is assumed to be in the absolutely convergent sequence of weights, and therefore $\widehat{X}_t(\infty)$ is written for the prediction.

Let (X_t) be stationary with $\mu_X = 0$ and the sequence of weights (π_t) absolutely convergent. Then:

1. $\widehat{X}_t(\infty) = \pi_1 X_{t-1} + \pi_2 X_{t-2} + \pi_3 X_{t-3} + \dots$ *is the linear forward prediction for X_t, given X_{t-1}, X_{t-2}, ...*
2. $X_t - \widehat{X}_t(\infty) = \hat{e}_t(\infty)$ *is the prediction error*
3. $\mathrm{PVar}\widehat{X}_t(\infty) = \mathrm{E}\Big(\big(\hat{e}_t(\infty)\big)^2\Big) = \mathrm{E}\Big(\big(X_t - \widehat{X}_t(\infty)\big)^2\Big)$ *is required to be minimal to arrive at an optimal estimation.*

(5.87)

From (5.87), infinitely many optimality conditions with infinite linear combinations $\pi_1 X_{t-1} + \pi_2 X_{t-2} + \pi_3 X_{t-3} + \dots$ are obtained. These sums converge since (π_t) absolutely converges. Thus, as (5.10) is calculated from (5.9), the equations in (5.88) are calculated from the optimality conditions pertaining to (5.87).

$$\begin{aligned} c_X(1) &= \pi_1 c_X(0) &&+ \pi_2 c_X(1) &&+ \pi_3 c_X(2) &&+ \dots \\ c_X(2) &= \pi_1 c_X(1) &&+ \pi_2 c_X(0) &&+ \pi_3 c_X(1) &&+ \dots \\ c_X(3) &= \pi_1 c_X(2) &&+ \pi_2 c_X(1) &&+ \pi_3 c_X(0) &&+ \dots \\ &\dots \\ c_X(k) &= \pi_1 c_X(k-1) &&+ \pi_2 c_X(k-2) &&+ \pi_3 c_X(k-3) &&+ \dots \\ c_X(k+1) &= \pi_1 c_X(k) &&+ \pi_2 c_X(k-1) &&+ \pi_3 c_X(k-2) &&+ \dots \\ &\dots \end{aligned} \tag{5.88}$$

The system (5.88) contains infinitely many equations with infinite linear combinations of covariances on the right sides. The linear combinations converge since the sequence (π_t) is assumed to converge absolutely in (5.87).

Proceeding as demonstrated in (5.4) and (5.5), but by substituting π-weights for a-weights, the Yule-Walker equations for lag k pertaining to an AR[∞] model are obtained in (5.89).

$$\begin{array}{rcl} c_X(1) = & \pi_{1k}c_X(0) & + \pi_{2k}c_X(1) \quad + \ldots + \pi_{kk}c_X(k-1) \\ c_X(2) = & \pi_{1k}c_X(1) & + \pi_{2k}c_X(0) \quad + \ldots + \pi_{kk}c_X(k-2) \\ \quad \ldots & & \\ c_X(k) = & \pi_{1k}c_X(k-1) + \pi_{2k}c_X(k-2) + \ldots + \pi_{kk}c_X(0) & \end{array} \tag{5.89}$$

The equations in (5.89) above are identical with those equations in the portion of the system (5.88), being confined by row k and column k on the right side and assuming that the weights are written with only one index.

Continuing to derive the Yule-Walker equations pertaining to an AR[∞] model for lags $k+1$, $k+2$, $k+3$, ..., the system (5.88) is reconstructed, and it is thus shown that the weights $\pi_1, \pi_2, \ldots$, in the AR[∞] representation of an ARMA[p,q] model are also the weights of an optimal linear prediction $\widehat{X}_t(\infty)$ for X_t calculated from X_{t-1}, X_{t-2}, ... The prediction error $\hat{e}_t(\infty)$ is not correlated with the random variables in the prediction, i.e., $\mathrm{E}\big(\hat{e}_t(\infty)X_{t-j}\big) = 0$ for $j = 1, 2, 3, \ldots$. This result follows from the infinite linear combinations in the optimality conditions and (2.79).

When comparing $\mathrm{E}\big(\hat{e}_t(\infty)X_{t-j}\big) = 0$, $j = 1, 2, 3, \ldots$, obtained above with $\mathrm{Cov}(W_t, X_{t-j}) = \mathrm{E}(W_t X_{t-j}) = 0$, $j = 1, 2, 3, \ldots$ as implied by the MA[∞] representation (5.38) of the ARMA[p, q] model, it is concluded that the prediction error $\hat{e}_t(\infty)$ is reconcilable with the causality of the ARMA[p, q] model which requires that the innovation at time point t is not correlated with the past variables in the process. A further comparison of the AR[∞] representation of the ARMA[p,q] model, $X_t - \pi_1 X_{t-1} - \pi_2 X_{t-2} - \pi_3 X_{t-3} - \ldots = W_t$, with the prediction (5.87), shows that the prediction error is the innovation at time point t, $\hat{e}_t(\infty) = W_t$, and the mean square prediction error is the variance of the innovations, $\mathrm{PVar}\widehat{X}_t(\infty) = \sigma_W^2$.

(5.87) defines the prediction for time point t starting at $t-1$. Now, the start time of the prediction is displaced to time point t and the process is predicted for $t+1$ to obtain a one-step prediction for $t+1$. In this one-step prediction, (i) the weights $\pi_1, \pi_2, \ldots$, in the AR[∞] representation of an ARMA[p, q] model are the weights of an optimal linear prediction $\widehat{X}_{t+1}(\infty)$ for X_{t+1}, calculated from X_t, $X_{t-1}, X_{t-2}, \ldots$, (ii) the prediction error is the innovation at time point $t+1$, and (iii) the mean square prediction error is the variance of the innovations. In a one-step prediction for $t+2$, $\widehat{X}_{t+2}(\infty) = \pi_1 X_{t+1} + \pi_2 X_t + \pi_3 X_{t-1} + \ldots$, $\hat{e}_{t+2}(\infty) = W_{t+2}$ and $\mathrm{PVar}\widehat{X}_{t+2}(\infty) = \sigma_W^2$ are arrived at. Thus, the errors $\big(\hat{e}_{t+1+j}(\infty)\big)$ in a sequence of one-step predictions $\big(\widehat{X}_{t+1+j}(\infty)\big)$ for $\big(X_{t+1+j}\big)$, $j = 0, 1, 2, 3, \ldots$, are a white noise process.

The results obtained above are summarised in (5.65).

The error and the mean square error of a k-step prediction in (5.67) and (5.68) are obtained from the MA[∞] representation (5.38) of the ARMA[p, q] model variable at time point $t+k$ (remember: $\psi_0 = 1$) in (5.90).

$$X_{t+k} = \sum_{u=0}^{\infty} \psi_u W_{t+k-u} \tag{5.90}$$

$$\widehat{X}_{t+k}(\infty) = \sum_{u=0}^{k-1} \psi_u W_{t+k-u} + \sum_{u=k}^{\infty} \psi_u W_{t+k-u} \tag{5.91}$$

$$= \sum_{u=0}^{k-1} \psi_u \mathrm{E} W_{t+k-u} + \sum_{u=k}^{\infty} \psi_u W_{t+k-u} \tag{5.92}$$

$$= \sum_{u=k}^{\infty} \psi_u W_{t+k-u} \tag{5.93}$$

$$\hat{e}_{t+k}(\infty) = X_{t+k} - \widehat{X}_{t+k}(\infty) = \sum_{u=0}^{k-1} \psi_u W_{t+k-u} \tag{5.94}$$

$$\mathrm{PVar}\widehat{X}_{t+k}(\infty) = \sigma_W^2 \sum_{u=0}^{k-1} \psi_u^2 \tag{5.95}$$

In (5.91), the prediction for a future time point $t + k$, starting from the present time point t, $\widehat{X}_{t+k}(\infty)$, is the sum of two sums: (i) the sum of the weighted future innovations at $t+k, \ldots, t+1$, and (ii) the sum of the present innovation at t and the past innovations at $t-1, t-2, \ldots$. This linear combination of the ψ-weights and the innovations is shown to be, in Problem 5.20, an optimal linear prediction for X_{t+k}. Substituting the expectations for the future innovations $W_{t+1}, \ldots, W_{t+k}$, (5.92) is arrived at. How can this step be justified, the innovations not being observable either in the past or present nor in the future? In an application, i.e., when calculating $\widehat{X}_{t+k}(\infty)$ with an estimated model,

1. the present innovation W_t and past innovations $W_{t-1}, \ldots$, can be replaced by estimates for the residuals of the estimated model as defined in (5.62) and
2. the innovations in the future, $W_{t+1}, \ldots, W_{t+k}$, cannot be estimated since the innovations are non-correlated: $W_{t+1}, \ldots, W_{t+k}$, are replaced by their expectations, being optimal predictions as shown in (2.83).

Since the expectations of the innovations are, as required in (5.36) identically zero, (5.93) is obtained. The error and the mean square error of the k-step prediction in (5.94) and (5.95) immediately follow from (5.93).

5.6.4 Predictions Derived from the Difference Equation of an ARMA$[p, q]$ Model

An optimal prediction is calculated, using (2.82), as a conditional expectation from the conditional probability density of the random variable to be predicted, given the observations. The conditional expectation can be used as

optimal prediction only for special cases because the conditional probability density is calculated from the joint probability density of the random variables $X_1, \ldots, X_N$ which is usually very difficult to obtain. When these densities are not available, however, a conditional expectation can be calculated (for special cases) using the same formulae for calculating the expectations of random variables as proposed in (1.15) and (1.16), as these formulae also apply for conditional expectations. Using these formulae, the optimal one-step prediction (5.96) of an ARMA$[p, q]$ model (X_t) is derived below.

$$\widehat{X}_{t+1}(\boldsymbol{X}_{N,t}) = \mathrm{E}\big(X_{t+1}|\boldsymbol{X}_{N,t}\big) \tag{5.96}$$

$$= \sum_{i=1}^{p} a_i \mathrm{E}\big(X_{t+1-i}|\boldsymbol{X}_{N,t}\big) \tag{5.97}$$

$$+\mathrm{E}\big(W_{t+1}|\boldsymbol{X}_{N,t}\big) \tag{5.98}$$

$$-\sum_{j=1}^{q} b_j \mathrm{E}\big(W_{t+1-j}|\boldsymbol{X}_{N,t}\big) \tag{5.99}$$

To arrive at (5.97), (5.98) and (5.99), we start by writing the definition (i.e., the difference equation) of the ARMA$[p, q]$ model in (5.36) for time point $t+1$: $X_{t+1} = \sum_{i=1}^{p} a_i X_{t+1-i} + W_{t+1} - \sum_{j=1}^{q} b_j W_{t+1-j}$, assume that the prediction is calculated from the N observed values $(x_t, x_{t-1}, \ldots, x_{t-N+1}) = \boldsymbol{x}_{N,t}$ of the random variables $(X_t, X_{t-1}, \ldots, X_{t-N+1}) = \boldsymbol{X}_{N,t}$, and then write the one-step prediction $\widehat{X}_{t+1}(X_{N,t})$ as conditional expectation in (5.96).

Under the assumption that the coefficients in the ARMA$[p, q]$ model are known, $\widehat{X}_{t+1}(X_{N,t})$ becomes the sum of three conditional expections that are calculated as follows:

1. $\sum_{i=1}^{p} a_i \mathrm{E}\big(X_{t+1-i}|\boldsymbol{X}_{N,t}\big)$ in (5.97) is obtained using the optimality of the conditional expectation as proposed in (2.82) because the conditional expectation of an observed random variable is the observed value. For this prediction, the prediction error and thus the mean square prediction error are identically zero, as argued in the remarks to (4.37).
2. $\mathrm{E}\big(W_{t+1}|\boldsymbol{X}_{N,t}\big)$ in (5.98) follows since W_{t+1} is not correlated with X_t, X_{t-1}, …, the model being causal: $\mathrm{E}\big(W_{t+1}|\boldsymbol{X}_{N,t}\big) = \mathrm{E}W_{t+1} = \mu_W = 0$.
3. $\sum_{j=1}^{q} b_j \mathrm{E}\big(W_{t+1-j}|\mathbf{X}_{N,t}\big)$ in (5.99) contains the conditional expectations in the first column of the table below which can be calculated using the random variables in the second column (and additional random variables further back in the past):

$$\begin{array}{ll} b_1 \mathrm{E}\big(W_t|(X_t, X_{t-1}, \ldots, X_{t-N+1})\big) & W_t \text{ and } X_t \\ b_1 \mathrm{E}\big(W_t|(X_t, X_{t-1}, \ldots, X_{t-N+1})\big) & W_{t-1} \text{ and } X_t,\, X_{t-1} \\ b_q \mathrm{E}\big(W_{t+1-q}|(X_t, X_{t-1}, \ldots, X_{t-N+1})\big) & W_{t+1-q} \text{ and } X_t, \ldots, X_{t+1-q} \end{array}$$

The random variables in the pairs (W_t, X_t), (W_{t-1}, X_t), (W_{t-1}, X_{t-1}), … are dependent, X_t being, in (5.38), a function of the W_t, W_{t-1}, W_{t-2}, …,

and thus a calculation of the conditional expectations in (5.99) becomes difficult.

This obstacle can, however, be circumvented by substituting observed innovations w_t, w_{t-1}, ..., w_{t+1-q} for the conditional expectations. The w_t, w_{t-1}, ..., w_{t+1-q} are available in simulation experiments; in applications, however, they are not available. This second obstacle is circumvented by substituting estimates of the innovations $\widehat{W}_t(\boldsymbol{X}_{N,t}) = X_t - \widehat{X}_t(\boldsymbol{X}_{N,t})$, $\widehat{W}_{t-1}(\mathbf{X}_{N,t}) = X_{t-1} - \widehat{X}_{t-1}(\boldsymbol{X}_{N,t})$, ... in the prediction. Hence, the prediction becomes recursive, and a non-finite past $(X_t, X_{t-1}, X_{t-2}, \ldots) = \boldsymbol{X}_{-\infty,t}$ becomes a prerequisite.

Consequently, the one-step prediction in (5.69) and k-step prediction in (5.70) become recursions substituting, as required in (5.97), (5.98) and (5.99), (i) the future conditional expectations of the model variables with their predictions, (ii) the future conditional expectations of the innovations with $\mu_W = 0$ (the expectations in the model (5.36)), (iii) the past conditional expectations of the model variables with the observed values and, (iv) the past conditional expectations of the innovations with the empirical residuals of the model calculated using (5.62).

5.7 Problems

5.1. Select an AR[1] model for the measurements of the wind speeds in the turbulent atmospheric flow in Fig. 2.1 with their empirical correlation function in Fig. 2.3. Then calculate, using (5.5) and (5.6), the Yule-Walker estimate for the coefficient in this model. Plot a simulation obtained from the estimated model together with its empirical correlation function and compare with the plots in Fig. 2.11.

5.2. Calculate the mean square prediction error $\mathrm{PVar}\widehat{X}_t^{(v)}(k)$ in (5.11).

$$
\begin{aligned}
&\mathrm{PVar}\widehat{X}_t^{(v)}(k) = \mathrm{E}\Big(\big(X_t - \widehat{X}_t^{(v)}(k)\big)^2\Big) \\
&= \mathrm{E}\Big(\big(X_t - (a_{1k}X_{t-1} + a_{2k}X_{t-2} + \ldots + a_{kk}X_{t-k})\big)^2\Big) \\
&= \mathrm{E}\Big(\big(\hat{e}_t^{(v)}(k)\big(X_t - (a_{1k}X_{t-1} + a_{2k}X_{t-2} + \ldots + a_{kk}X_{t-k})\big)\Big) \\
&= \mathrm{E}\big(\hat{e}_t^{(v)}(k)X_t\big) - a_{1k}\mathrm{E}\big(\hat{e}_t^{(v)}(k)X_{t-1}\big) - \ldots - a_{kk}\mathrm{E}\big(\hat{e}_t^{(v)}(k)X_{t-k}\big) \\
&= \mathrm{E}\big(\hat{e}_t^{(v)}(k)X_t\big) \qquad (5.100) \\
&= \ldots
\end{aligned}
$$

(5.100) follows from the optimality conditions (5.9). Compare with the optimality conditions (2.74).

5.3. From (5.8), (5.11) and (5.12)

$$\begin{aligned}\mathrm{PVar}X_t^{(v)}(k) &= \mathrm{E}\Big(\big(X_t - \widehat{X}_t^{(v)}(k)\big)^2\Big) \\ &= c_X(0) - \big(a_{1k}c_X(1) + a_{2k}c_X(2) + \ldots + a_{kk}c_X(k)\big)\end{aligned}$$

is obtained, from which the optimality conditions (5.9) can be derived.

5.4. Prove (5.14,1,2). In a first step, substitute definitions (5.72) and (5.73) for $\widehat{X}_t^{(r)}(k)$ and $\hat{e}_t^{(r)}(k)$, as well as definition (5.8,1) for $\widehat{X}_t^{(v)}(k)$ in recursion (5.78), to derive

$$\sum_{j=1}^{k} a_{jk}X_{t-j} = \sum_{j=1}^{k-1} a_{j,k-1}X_{t-j} + \theta_k\Big(X_{t-k} - \sum_{j=1}^{k} a_{j,k-1}X_{t-k+j}\Big)$$

Evaluating this sum in the second step,

$$\begin{array}{ccc}
a_{1k}X_{t-1} & a_{1,k-1}X_{t-1} & -\theta_k a_{1,k-1}X_{t-k+1} \\
+ & + & + \\
a_{2k}X_{t-2} & a_{2,k-1}X_{t-2} & -\theta_k a_{2,k-1}X_{t-k+2} \\
+ & + & + \\
\vdots & \vdots & \vdots \\
+ & + & + \\
a_{k-1,k}X_{t-k+1} & a_{k-1,k-1}X_{t-k+1} & -\theta_k a_{k-1,k-1}X_{t-1} \\
+ & & \\
a_{kk}X_{t-k} & \theta_k X_{t-k} &
\end{array}$$

$$\textstyle\sum_{j=1}^{k} a_{jk}X_{t-j} = \sum_{j=1}^{k-1} a_{j,k-1}X_{t-j} + \theta_k X_{t-k} - \theta_k \sum_{j=1}^{k} a_{j,k-1}X_{t-k+j}$$

is arrived at. Derive (5.14,1,2) by evaluating the equations for X_{t-j}, $j = 1, \ldots, k$.

5.5. From (5.72), (5.73) and (5.74) as well as (5.11) and (5.12), the first = in (5.14,3) is obtained. Prove the second = by multiplying (5.78) with X_t,

$$\hat{e}_t^{(v)}(k)X_t = \hat{e}_t^{(v)}(k-1)X_t - \theta_k \hat{e}_{t-k}^{(r)}(k-1)X_t$$

taking the expectations, and, using (5.77) to arrive at

$$\mathrm{PVar}(k) = \mathrm{PVar}(k-1) - \theta_k \mathrm{E}\Big(\hat{e}_{t-k}^{(r)}(k-1)X_t\Big).$$

Compare this result with definition (5.72).

5.6. Show that the Yule-Walker estimates $\hat{a}_{12}$ and $\hat{a}_{22}$ as obtained in the remarks to (5.14) are identical with an empirical version of (2.53).

5.7. Calculate the differences of the yearly values of shortwave incoming radiation at Hohenpeissenberg station in Fig. 3.1, and then plot these differences together with their empirical correlation and partial correlation functions. Compare them with the plots in Figs. 5.1 and 5.3.

5.8. Show that $\tilde{a}_{kk}$ as proposed in (5.21,2) minimises the sum in (5.21,1). Substitute the forward and backward predictions in (5.21,1) with the expressions obtained in (5.19) and (5.20)

$$\sum_{t=k+1}^{N} \left(\left(\tilde{e}_t^{(v)}(k) \right)^2 + \left(\tilde{e}_{t-k}^{(r)}(k) \right)^2 \right)$$

$$\sum_{t=k+1}^{N} \left(\left(\tilde{e}_t^{(v)}(k-1) - \tilde{a}_{kk} \tilde{e}_{t-k}^{(r)}(k-1) \right)^2 + \left(\tilde{e}_{t-k}^{(r)}(k-1) - \tilde{a}_{kk} \tilde{e}_t^{(v)}(k-1) \right)^2 \right.$$

$$\sum_{t=k+1}^{N} \left(\left(e_1 - \tilde{a}_{kk} e_2 \right)^2 + \left(e_2 - \tilde{a}_{kk} e_1 \right)^2 \right)$$

$$\ldots$$

and then evaluate

$$\frac{\mathrm{d}}{\mathrm{d}a_{kk}} \left(\sum_{t=k+1}^{N} \left(e_1^2 + e_2^2 \right) - 4\tilde{a}_{kk} \sum_{t=k+1}^{N} \left(e_1 e_2 \right) + \tilde{a}_{kk}^2 \sum_{t=k+1}^{N} \left(e_1^2 + e_2^2 \right) \right) = 0.$$

5.9. $\left(\boldsymbol{X}_t, \boldsymbol{X}_p^{(\uparrow)} \right)$ is the matrix in (5.23). What is the difference between $\left(\left(\boldsymbol{X}_t, \boldsymbol{X}_p^{(\uparrow)} \right) \right)^T \left(\left(\boldsymbol{X}_t, \boldsymbol{X}_p^{(\uparrow)} \right) \right) = \boldsymbol{C}_X^{(\uparrow)}$ and $\hat{\boldsymbol{C}}_X(p)$ in the empirical Yule-Walker equations (5.7)? $(\boldsymbol{X}_t, \boldsymbol{X}_p)$ is the matrix in the remarks to (5.23). What is the difference between $(\boldsymbol{X}_t, \boldsymbol{X}_p)$ and $\hat{\boldsymbol{C}}_X(p)$?

5.10. Estimate an AR[2] model for the de-meaned horizontal wind speeds as plotted in Fig. 5.1 using regression techniques and forward displacements of the time series. Compare your results with the regression estimates in Table 5.1 obtained with backward displacements of the time series. Repeat this investigation for the de-meaned time series of the amount of a product formed during a chemical reaction as plotted in Fig. 5.3.

5.11. Simulate a few realisations using

```
#white noise with 0 expectation and unit variance
a1 <- 2.7607        #AR[4] coefficients
a2 <- -3.8200
a3 <- 2.6535
a4 <- -0.9238
arsim <- arima.sim(model=list(ar=c(a1,a2,a3,a4)),n=20000,sd=1.0)
#
```

```
b1 <- 2.7607        #MA[4] model
b2 <- -3.8200       #change signs of moving
b3 <- 2.6535        #average coefficients as
b4 <- -0.9238       #demonstrated in Sect. 5.4.3
masim <- arima.sim(model=list(ma=-c(b1,b2,b3,b4)),n=40000,sd=1.0)
```

and plot the empirical correlation and partial correlation functions calculated from the simulated realisations.

Then simulate an invertible MA[2] process. Choose the weights (b_1, b_2), after substituting b_1 and b_2 for a_1 and a_2 in Fig. 2.17, from different regions in the (b_1, b_2)-plane. Calculate the empirical correlation and partial correlation functions from the simulated realisations and compare with the results obtained in Problem 2.21.

5.12. Please multiply the left side of the following equation

$$\begin{aligned}(\psi_0 + \psi_1 z + \psi_2 z^2 + \ldots)(1 - a_1 z - \ldots - a_p z^p) \\ = (1 - b_1 z - \ldots - b_q z^q - 0 \times z^{q+1} - \ldots)\end{aligned} \quad (5.101)$$

as follows

$$\begin{array}{llllll} \psi_0 & +\psi_1 z & +\psi_2 z^2 & +\psi_3 z^3 + \ldots & +\psi_p z^p & +\psi_{p+1} z^{p+1} + \ldots \\ & -\psi_0 a_1 z & -\psi_1 a_1 z^2 & -\psi_2 a_1 z^3 + \ldots & -\psi_{p-1} a_1 z^p & -\psi_p a_1 z^{p+1} - \ldots \\ & & & & \ldots & \\ & & & & -\psi_0 a_p z^p & -\psi_1 a_p z^{p+1} - \ldots \end{array}$$

and compare, for identical powers of z, the sums on the left side with the coefficients on the right side of (5.101). Performing this, (5.41) and (5.42) are arrived at.

5.13. Apply `arima.sim()` to simulate a stationary and invertible ARMA[1,1] model. Choose the weights a and b such that the pairs (a, b) are in the quadrants of the region $-1 < a < 1$ and $-1 < b < 1$ for each pair $(a - b) > 0$ or $(a - b) < 0$. From the simulated realisations, now calculate and plot the empirical correlation and partial correlation functions.

5.14. Estimate an ARIMA$[p, d, q]$ model for the residuals, as plotted in Fig. 5.7, of a first order linear model for the trend in Basel yearly temperature values plotted in Fig. 2.13.

5.15. Estimate a seasonal ARIMA$[0,1,1]_{12}$ model for the logarithms of the MLO atmospheric CO_2 monthly values, plotted below in Fig. 5.9.

5.16. Estimate candidate models for the SO index given in Table 5.3, using all available data (plot (a) in Fig. 5.14). Compare your results with the estimates in Table 5.3 obtained from the contiguous SO index monthly values (plot (b) in Fig. 5.14).

5.17. Calculate the empirical residuals of the ARMA[1,1] model in Table 5.3 fitted to the SO index, using a calculator and the recursion (5.62). (The result is given in Table 5.4). For time points later than July 1933, your results come closer and closer to the residuals calculated by `arima()`. Why?

5.18. Is there a model in the set of candidate models given in Table 5.3 that is completely in agreement with (5.63)?

5.19. Using `arima()`, estimate

1. AR[1] and AR[2] models for the wind speeds plotted in Fig. 2.1
2. an AR[2] model for the product of a chemical reaction plotted in Fig. 5.3
3. an ARIMA[0,1,1] model for the yearly values in the Basel temperature series plotted in Fig. 2.13, and
4. an ARIMA[2,0,0]$\times$[0,1,1]$_{12}$ for the logarithms of the monthly CO_2 concentrations plotted in Fig. 5.9 (b).

Then calculate and plot, for each estimated model, the diagnostics as proposed in (5.60), (5.61) and (5.63).

5.20. Show that linear combinations constructed from the innovations of an ARMA$[p, q]$ process and the ψ-weights in its MA$[\infty]$ representation (5.38), $\widehat{X}_{t+1} = W_{t+1} - \psi_1 W_t - \psi_2 W_{t-1} - \ldots$, are optimal linear predictions for X_{t+1}, given the innovations. Obtain the optimality conditions by calculating the partial derivatives for ψ_j, $j = 1, 2, 3, \ldots$:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\psi_j}\mathrm{E}\Big(\big(X_{t+1} - (W_{t+1} - \psi_1 W_t - \psi_2 W_{t-1} - \ldots)\big)^2\Big) &= 0 \\ &\ldots \\ 2\psi_j \sigma_W^2 &= 0 \end{aligned}$$

5.21. Using (5.65) and (5.66), predict the SO index for three months following the month with the last observation, i.e., for February, March and April 1996, under the assumption that the ARMA[1,1] model estimated in Sect. 5.4.3 is the true model, i.e., that $\sigma_W^2 = 0.5997$, $a = 0.8829$ and $b = 0.4649$ as well as $\mu_X = -0.1411$.

Calculate the π-weights 0.4180, 0.1943, 0.0903, 0.0420, 0.0195, 0.0091, 0.0042, 0.0020, ... in the AR$[\infty]$ representation (5.47) of the model by convolving $(1, -0.4649)^{-1} * (1, -0.8829) = (1, -\pi_1, -\pi_2, \ldots)$, and then the ψ weights -0.4180, -0.3690, -0.3258, -0.2876, ... in the MA$[\infty]$ representation (5.38) of the model by convolving $(1, -0.8829)^{-1} * (1, -0.4649) = (1, -\psi_1, -\psi_2, \ldots)$. The SO index observations in the years 1995 and 1996 follow:

year	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
1995	-0.6	-0.5	0.2	-1.1	-0.7	-0.2	0.3	-0.1	0.3	-0.2	0.0	-0.9
1996	1.1											

Multiply four and then eight π-weights with the de-meaned observations to obtain the de-meaned predictions. Compare with the de-meaned predictions as calculated in Table 5.5.

5.22. Using (5.65) and (5.66), predict the differences $X_t = Y_t - Y_{t-1}$ of the Basel temperature yearly means. Assume that the estimates obtained in the remarks to (5.57) and Fig. 5.5 are the true parameters $\sigma_W^2 = 0.469$ and $b = 0.881$ of an MA[1] model for (X_t), and estimate X_{1958} and X_{1959}. From these, calculate estimates for Y_{1958} and Y_{1959}. The Basel yearly temperature values for the years 1947 through to 1957 follow:

year	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1957
(Y_t), ^{0}C	10.57	9.98	10.48	10.02	9.70	9.78	9.63	9.01	9.17	7.92	9.51
(X_t), ^{0}C	0.98	−0.59	0.5	−0.46	−0.32	0.08	−0.15	0.62	0.16	−1.25	1.59

5.23. As Problem 5.21, but using (5.70).

5.24. As Problem 5.22, but using (5.70).

5.25. Predict the SO index for February, March and April 1996 using one of the autoregressive models of higher order in Table 5.3 and assuming that the estimated model is the true model. If the residuals of the model are not correlated then, as argued in Sect. 5.4.5, the model is an alternative to a very parsimonious ARMA[1,1] or AR[4] model. Use `arima()` and `predict()`. Compare your predictions with those in Table 5.5 calculated with an ARMA[1,1] model.

5.26. The predictions in Problem 5.25 are calculated with estimated parameters and thus are afflicted with additional uncertainties. Try to assess these uncertainties qualitatively and compare with the additional uncertainties obtained in Table 5.6 for the predictions calculated using the ARMA[1, 1] model.

5.27. Calculate, using `predict()` and the model estimated in Problem 5.15, one-, two-, three- and four-step predictions for the logarithms of the monthly CO_2 concentrations in Fig. 5.9 (b).

5.28. The definitions of the AR[p], MA[q] and ARMA[p, q] models in (5.2), (5.28) and (5.36) as well as the definition of the ARIMA[p, d, q] model in (5.55) consist of two parts: a formula with the model coefficients (the "deterministic" part) and the properties of the innovations (the "stochastic" part). These two parts usually occur in statistical models. Give examples of such models and compare their "stochastic" parts with the properties of the innovations.

5.29. The duration of extreme El Niño events is usually longer than one month. For example, the last events (August 1982 through to March 1983, and December 1991 through to March 1992) shown in Fig. 5.14 are captured by the AR[4] and ARMA[1,1] models in Table 5.3 and thus are no longer visible in the residuals of these models as plotted above, in Figs. 5.15, 5.16 and 5.17. In these plots, however, three extremely small values are visible. Could these be outliers due to gross measurement errors? Compare the SO index values with these residuals (the SO index time series downloaded from the address in [86] is used in Sects. 5.4.3 and 5.4.4) with the values in the SO index time series which can be obtained from the addresses given in [115].

5.30. Apply (5.64) to the estimated AR[4] candidate model in Table 5.3.

5.31. The very parsimonious ARIMA[2,0,0]$\times$[0,1,1]$_{12}$ model fits, as can be concluded from Fig. 5.13, the logarithms of the monthly CO_2 concentrations in Fig. 5.9 (b). These concentrations are the result of the transport of carbon, in its many forms, between many reservoirs (atmosphere, surface water of the oceans, deep water of the oceans, biomass, fossil oil and coal reservoirs, as well as other rocks containing carbon) [71]. One property of the ARIMA$[2,0,0]\times[0,1,1]_{12}$ model pertains to one of these carbon cycles.

6 Fourier Transforms of Deterministic Functions

A variable (e.g., the temperature at a point on the earth's surface) is deterministic on condition that (i) it can be incorporated in a deterministic model (e.g., a general circulation model), and (ii) no probabilistic model exists (examples are given in Chaps. 1 through to 5, and in Chaps. 7, 9 and 10), that better fits the data or allows for better predictions. Once a deterministic variable is observed, a deterministic function is obtained.

Can a deterministic function $f(t)$ be represented using a linear combination of trigonometric functions? A linear combination of trigonometric functions is a sum of trigonometric oscillations, defined by their periods or frequencies and by their coefficients. If such a linear combination exists then it is called a Fourier representation or a Fourier transform of $f(t)$, and the coefficients of the trigonometric oscillations are called Fourier coefficients. The Fourier coefficients and also the squares of their absolute values can be plotted against the frequencies of the oscillations. The squared absolute values of the Fourier coefficients as function of the frequencies are called spectrum. A spectrum is usually the key to an in-depth analysis of the observed variable.

In Sect. 6.1, a function is reconstructed using a linear combination of trigonometric values with coefficients that are easily obtained provided that the trigonometric functions are orthogonal, i.e., are elements of the sets defined in Sect. 6.2. Fourier transforms are introduced in Sect. 6.3 and the spectrum in Sect. 6.4.

Usually, in applications, a real-valued function $f(t)$ with real argument t is observed at discrete time points to obtain a sequence $(g_{t\Delta t})$, Δt being the sampling interval, in a finite interval for N time points. These restrictions imposed by the observation can produce distortions in the Fourier transform calculated from $(g_{t\Delta t})$ as compared with the Fourier transform of $f(t)$, i.e., the Fourier transform that could be obtained assuming that $f(t)$ is known for all real t. These distortions are dealt with in Sects. 6.5, 6.6 and 6.7.

Convolution, as introduced in Sect. 2.4, is revisited in Sect. 6.8.

The last two sections contain, as usual, the supplements and the problems.

Fourier transforms are dealt with in [18] and [32]. In [18], however, the examples stem from Electrical Engineering.

6.1 Adding Trigonometric Functions

In this section, the basic idea behind Fourier theory is demonstrated, i.e., that a real- or complex-valued deterministic function $f(t)$, or sequence (f_t), can be represented as a linear combination of a set of trigonometric oscillations, called a *Fourier representation.* The trigonometric oscillations together with their *Fourier coefficients* result from a *Fourier analysis,* and the Fourier representation is obtained in a *Fourier synthesis.* The pertaining propositions will be given in Sects. 6.2 and 6.3. As a first example, a periodic sequence is represented as a linear combination of trigonometric oscillations with complex-valued coefficients that are easily obtained using R function `fft()`. For each oscillation thus obtained, the squared absolute value of its coefficient is plotted against its frequency to arrive at the *spectrum* of the sequence. In the spectrum, periodicities become visible that cannot be easily spotted in a plot of the sequence.

6.1.1 Representation of a Sequence using Trigonometric Functions

A periodic function $g(x)$ as defined in (6.1)

A real-valued function $g(x)$ is called periodic with period p, on condition that $g(x) = g(x + np)$, x, p real, n integer. (6.1)

is known for all x provided that it is known in an arbitrary interval of length p, p the period of $g(x)$. Prominent examples are the trigonometric functions: $\sin(x)$ is periodic with period 2π since $\ldots = \sin(x - 4\pi) = \sin(x - 2\pi) = \sin(x) = \sin(x + 2\pi) = \sin(x + 4\pi) = \ldots$, whereas $\sin(2\pi x)$ is periodic with period 1 since $\sin(2\pi x) = \sin(2\pi(x)) = \sin(2\pi(x + 1)) = \sin(2\pi(x - 1)) = \sin(2\pi(x + 2)) = \ldots$.

As a second example, the sawtooth-shaped sequence $(f_t) = (g_t) + (h_t)$, as shown in Fig. 6.1 (a), is generated using the following R expressions

```
Vlu <- (4:0)/5          #lu: left below
Vru <- (0:4)/5          #ru: right below
Vlo <- (9:1)            #lo: left above
Vro <- (1:9)            #ro: right above
V <- c(Vlu,Vru)         #as in the remarks
V <- c(V,Vro)           #to Fig. 2.16
V <- c(V,Vlo)           #V has now 28 values
V5 <- rep(V,5)          #sequence g with period 28
v <- c(1.0,0.5,0.0,0.5,1.0)     #v has 5 values
v28 <- rep(v,28)                #sequence h with period 5
saw <- V5 + v28                 #sequence f with periods 28 and 5
t <- 0:(length(saw)-1)          #time from 0 through to 139
```

which do not contain a random number generator, and, consequently, sequence (f_t) is deterministic and periodic with period 140. R vector `t` contains time points chosen from 0 through to 139 due to the reasons given in

Sect. 6.2.3. (f_t) is the sum of two periodic sequences (g_t) and (h_t) with periods $p_g = 28$ and $p_h = 5$ time units. p_g is readily seen in Fig. 6.1 (a), whereas p_h can be detected only when the plot is analysed in detail.

Can the sawtooth-shaped sequence (f_t) be represented using trigonometric functions? If a Fourier representation of (f_t) exists then, hopefully, the periodic properties of (f_t) become more clearly visible in a plot constructed from the Fourier coefficients than in a plot of (f_t).

In R, Fourier coefficients are calculated using a time-saving algorithm which is implemented in R function `fft()`. `fft()` calculates a discrete Fourier transform as defined in (6.22,4). Using `fft()`, the Fourier coefficients are obtained in a complex-valued R vector. Hence, some elementary properties of complex numbers are refreshed in (6.2).

Let $z = a + \mathrm{i}b$ be in the plain of complex numbers. Then:
1. $a = \mathrm{re}(z)$, *is called the real part* z
2. $b = \mathrm{im}(z)$, *the imaginary part of* z
3. $\overline{z} = a - \mathrm{i}b$, *the complex-conjugate to* z,
4. $|z| = \mathrm{mod}(z) = \sqrt{a^2 + b^2}$, *the absolute value or modulus of* z,
5. $\varphi = \arg(z)$, $\tan(\varphi) = b/a$, *the phase of* z,
6. $z = |z|\big(\cos(\varphi) + \mathrm{i}\sin(\varphi)\big) = |z|\mathrm{e}^{\mathrm{i}\varphi}$, $a = |z|\cos(\varphi)$, $b = |z|\sin(\varphi)$, *where* $\mathrm{mod}(z)$, $\arg(z)$, $\mathrm{re}(z)$, $\mathrm{im}(z)$ *are real numbers.*

(6.2)

The sawtooth-shaped sequence in Fig. 6.1 (a) is Fourier analysed using the following R expressions. (In a first reading, please ignore the references to Sect. 6.3).

```
x <- saw          #sawtooth-shaped sequence Fig. 6.1 (a)
N <- length(x)
#discrete Fourier transform as in (6.22,4), xfft being
xfft <- fft(x, inverse=T)               #a complex R vector
xtscaled <- (xfft/N)*2                  #scale as in (6.21),
xtscaled[1] <- xtscaled[1]/2            #(6.22) and (6.25)
xtscaled[N/2+1] <- xtscaled[N/2+1]/2    #for even N only
#the complex coefficients of (N/2+1) trigonometric
#oscillations are in the first half of xtscaled
xt <- xtscaled[1:(N/2+1)]
sk <- (0:(N/2))/N  #the frequencies sk of the oscillations
A <- Re(xt) #cos()-coefficients Ak of the oscillations
B <- Im(xt) #sin()-coefficients Bk of the oscillations
absval <- Mod(xt)  #moduli of the complex xt[1], ...
phase <- Arg(xt)   #phase of the complex xt[1], ...
#A, B, absval qnd phase are now real-valued R vectors
```

This Fourier analysis results in the Fourier coefficients A_k and B_k of $N/2+1 = 71$ trigonometric oscillations with frequencies $s_k = k/N$, $k = 0, 1, \ldots, 70$ (in `sk` with indices from 1 through to 71) since there are $N = 140$ values in the sequence. The number of the trigonometric oscillations is related to N because a trigonometric oscillation of a given frequency is determined by two

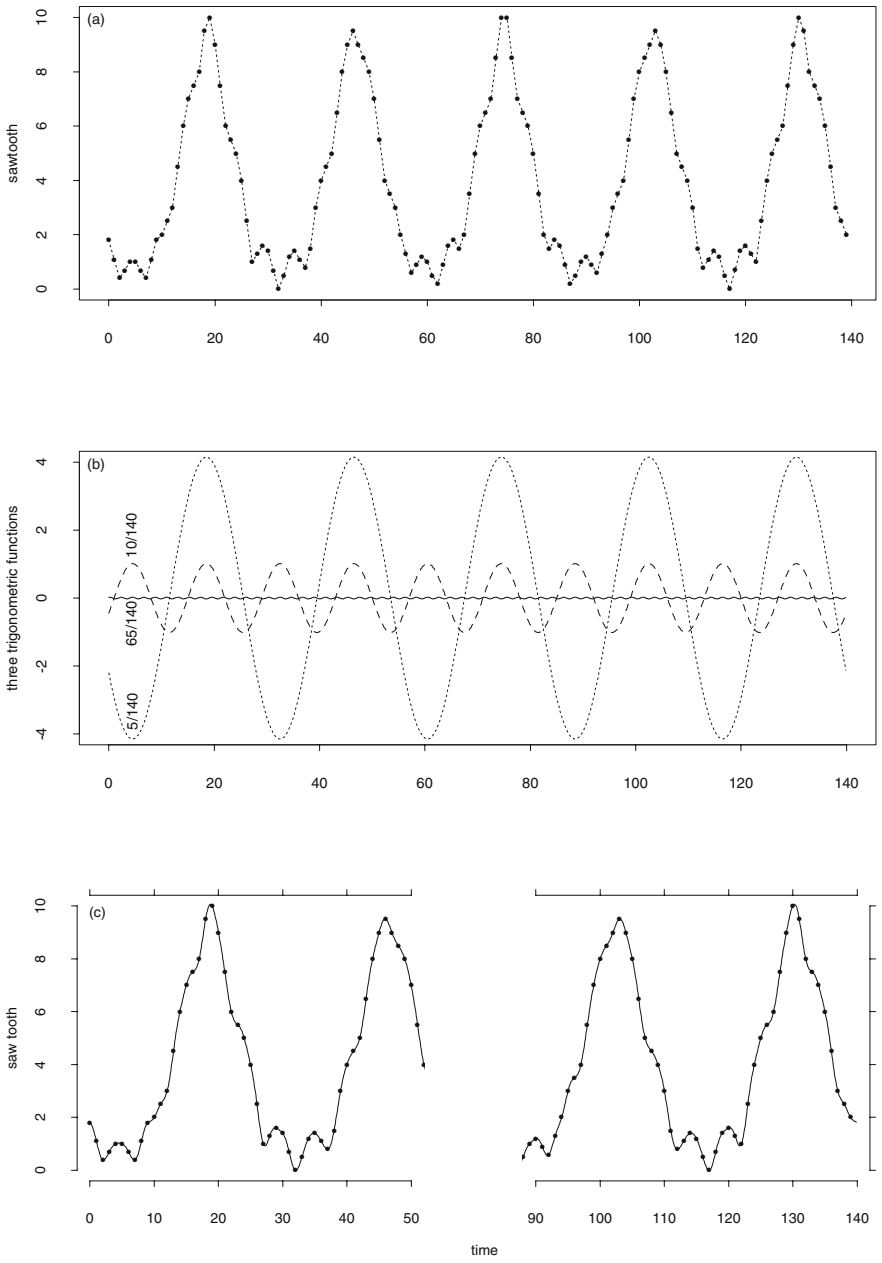


Fig. 6.1. A sawtooth-shaped sequence (plot a) is reconstructed by means of adding trigonometric functions (those with frequencies 5/140, 10/140 and 65/140 are shown in plot (b)) to obtain its Fourier representation in plot (c).

points. The first oscillation has frequency $s_0 = 0$ and is thus a constant A_0, A_0 being the mean of the sawtooth-shaped sequence.

Using A_k and B_k in the following R expressions

```
res <- 10      #resolution of time
hrlen <- res*N #time 0 at index 1, time 1 at index 11, etc.
timehr <- (0:(hrlen-1))/res  #vector for the time
xhr <- 1:hrlen               #vector for an oscillation
#here the oscillation with frequency 5/140 = 0.0375 in Fig. 6.1 (b)
k <- 6
xhr <- (rep(A[k],hrlen))*cos(2*pi*sk[k]*timehr) +
         (rep(B[k],hrlen))*sin(2*pi*sk[k]*timehr)
```

trigonometric oscillations are obtained, three of which are plotted in Fig. 6.1 (b), i.e., those with frequencies $s_5 = 5/140 = 0.0357$, $s_{10} = 10/140 = 0.0714$ and $s_{65} = 65/140 = 0.4643$ and therefore with periods of 28, 14 and 2.1538 time units. These oscillations are calculated and plotted for the time points $0.0, 0.1, 0.2, \ldots, 0.9, 1.0, 1.1, \ldots, 139.0, 139.1, \ldots, 139.9$.

The Fourier representation of the sawtooth-shaped sequence is arrived at by adding the oscillations obtained above

```
#oscillation no. 1 has frequency zero and thus is constant A_0
xhracc <- rep(A[1], hrlen)   #vector for summation (accumulator)
for(k in (2:(N/2)+1) ) {     #add oscillations
   xhr <- (rep(A[k],hrlen))*cos(2*pi*sk[k]*timehr) +
             (rep(B[k],hrlen))*sin(2*pi*sk[k]*timehr)
   xhracc <- xhracc + xhr                              }
```

and then plotting the sum as a solid line in Fig. 6.1 (c), where the sawtooth-shaped sequence is plotted with symbol •, as above, in plot (a). From plot (c) in Fig. 6.1 it is obvious that the sequence is reconstructed, for $t = 0, \ldots, 139$, from the trigonometric oscillations, as required in (6.22,3).

When all trigonometric oscillations are plotted (only those with frequencies 5/140, 10/140 and 65/140 are shown in Fig. 6.1 (b), as more than three oscillations cannot be distinguished clearly when plotted), it becomes obvious that (i) most oscillations are negligibly small in their absolute values, i.e., in $|F_k| = (A_k^2 + B_k^2)^{-1/2}$, (ii) a few have small $|F_k|$ and (iii) two or three have large $|F_k|$.

On the one hand, an oscillation is found to be not negligibly small in its absolute value on condition that its period is "somehow reconcilable" with the period of a fluctuation in the sawtooth-shaped sequence (f_t). For example, the absolute value of the oscillation with frequency 5/140 is quite large as can be seen in Fig. 6.1 (b). The period of this oscillation is 28 time units, and thus identical with period p_g of sequence (g_t) which is added to sequence (h_t) to obtain (f_t) as plotted in Fig. 6.1 (a). Like this oscillation, those with frequencies 10/140 (with period 14 = 28/2) and 65/140 (with period 2.1538 = 28/13) have absolute values that are not negligible, even though there are no fluctuations with periods 14 or 2.1538 in (f_t). On the other hand,

oscillations with periods not being "somehow reconcilable" with the period of a fluctuation in (f_t) are found to be negligibly small or identically zero in their absolute values. For example, the oscillation with frequency 1/140 (not plotted in Fig. 6.1 (a)) has a zero amplitude.

This behaviour is plausible since the oscillations belong to the Fourier representation of (f_t). The meaning of "somehow reconcilable" will be more precisely described in the remarks to Fig. 6.2.

Since the periods of the fluctuations in the sawtooth-shaped sequence (f_t) are found by comparing the absolute values of the trigonometric oscillations in the Fourier representation of (f_t), the idea arises to associate the absolute value $|F_k| = (A_k^2 + B_k^2)^{-1/2}$ with the frequency s_k of each oscillation in a table, as is demonstrated in Table 6.1. Usually, however, the squared absolute values $|F_k|^2 = A_k^2 + B_k^2$ are plotted against the frequencies s_k as shown in Fig. 6.2 (b). $|F_k|^2$ as function of the s_k is called the spectrum of (f_t). The reasons for plotting the squared absolute values of the oscillations and also the definition of the spectrum are given in Sect. 6.4. In Fig. 6.2 (b), the ratios of the spectral values $|F_k|^2$ cannot be seen since most $|F_k|^2$ are small. This obstacle is circumvented by plotting the logarithms of $|F_k|^2$ (if larger than -4 to avoid an overloaded plot) against the frequencies as shown in Fig. 6.2 (a). When the logarithms are plotted, ratios become differences and therefore, in Fig. 6.2, the logarithmic plot (a) is preferred to the non-logarithmic plot (b).

It is obvious from Fig. 6.2 and Table 6.1 that the oscillation with frequency 0.0357 (and thus a period of 28 time units) has a large absolute value $|F_k|$, and that there are oscillations with spectral values $|F_k|^2 \geq 0.05$ ($\log_{10}(0.05^2) \approx -2.60$) pertaining to frequencies 0.0714, 0.1074, 0.1438, 0.1785 and 0.2857 (all multiples of 0.0357, and thus having periods of 14, 9.33, 7, 5.6 and 3.5 time units). These oscillations are marked with $\circ$ in Table 6.1 and their sum as obtained in Problem 6.1 approximates sequence (g_t). (g_t) is used above to construct the sawtooth-shaped sequence $(f_t) = (g_t) + (h_t)$ and has a period of 28 time units. In Fig. 6.2 and Table 6.1, also oscillations with frequencies 0.2 and its twofold 0.4 (and thus with periods of 5 and 2.5 time units) have large spectral values. These are marked with $*$ in Table 6.1 and their sum approximates sequence (h_t) with period 5 time units used to construct (f_t) above. Remember that the Fourier representation of (f_t) in Fig. 6.1 (c) is the sum of all oscillations.

Hence, when Fig. 6.1 (a) is compared with Fig. 6.2 (a), it becomes obvious that the periods of the fluctuations in the sawtooth-shaped sequence are more easily detected in the plot of the spectrum than in the plot of the sequence.

The first trigonometric oscillation in the Fourier representation of a sequence with N values (e.g., there are $N = 140$ values in the sawtooth-shaped sequence (f_t) as plotted in Fig. 6.1 (a)) has frequency $s_1 = 1/N$ and period $p_1 = N$. This oscillation is called *fundamental oscillation* pertaining to the sequence and its frequency is called *fundamental frequency* (pertaining to the sequence). All further oscillations in the Fourier representation have frequen-

Table 6.1. Frequencies s_k, periods and absolute values $|F_k| = (A_k^2 + B_k^2)^{-1/2}$ of the trigonometric oscillations in the Fourier representation of sequence $(f_t) = (g_t) + (h_t)$ as plotted in Fig. 6.1 (c). Only the values pertaining to oscillations with $|F_k|^2 \geq 0.05$ are given. They belong to two families with frequencies being multiples of 0.0357 ($\circ$) and 0.2 ($*$) (time units)$^{-1}$.

frequency	0	0.0357	0.0714	0.1074	0.1438	0.1785	0.2000	0.2857	0.4000
period	∞	28	14	9.33	7	5.6	5	3.5	2.5
absolute value	3.95	4.15	1.01	0.15	0.10	0.23	0.53	0.07	0.08
		$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$*$	$\circ$	$*$

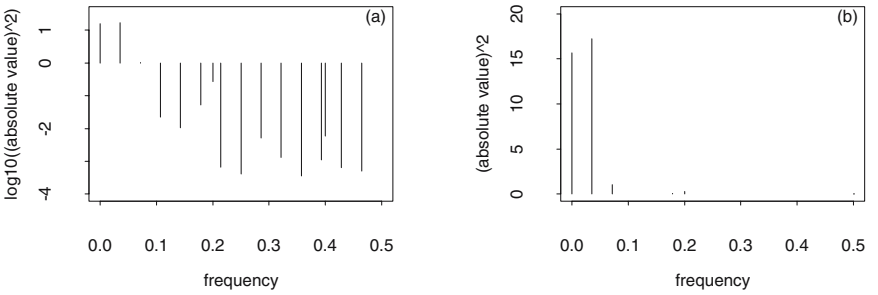


Fig. 6.2. Spectrum of (f_t) in Fig. 6.1 (a), i.e., squares of the absolute values of the trigonometric oscillations in the Fourier representation of (f_t) in Fig. 6.1 (c) (on the right, plot (b)), and their logarithms, if larger than -4 (on the left, plot (a)), plotted against the frequencies.

cies $s_k = k/N$, $k = 2, \ldots, m$, as defined in (6.21,1), that are multiples of the fundamental frequency, i.e., $s_k = ks_1$, and thus have periods p_k such that $kp_k = N$. They are therefore also periodic with period N. The frequency s_{k+1} is called the kth *harmonic* of the fundamental frequency s_1, i.e., s_2 is the first harmonic of s_1. For example, the Fourier representation $f_p(t)$ of the sawtooth-shaped sequence (f_t), as plotted in Fig. 6.1 (c), is a periodic function with period $N = 140$, i.e., $\ldots = f_p(0) = f_p(-140) = f_p(140) \ldots = f_p(0 + n \times 140)$, $n = \ldots, -1, 0, 1, \ldots$. The periodicity of f_p is demonstrated in Fig. 6.3.

In a Fourier representation $g_p(t)$ of a sequence (g_t), to any trigonometric oscillation with frequency s_k pertain oscillations with harmonic frequencies ls_k, $l = 2, 3, \ldots$, that are multiples of s_k; however, only those with $ls_k \leq 1/2$ can be calculated as shown in Sect. 6.5. These harmonic oscillations are periodic with period $1/s_k$. This is the reason why a fluctuation of non-trigonometric shape that is repeated with period $1/s_k$ in the sequence (g_t) under analysis can be represented by the sum of the oscillation with fre-

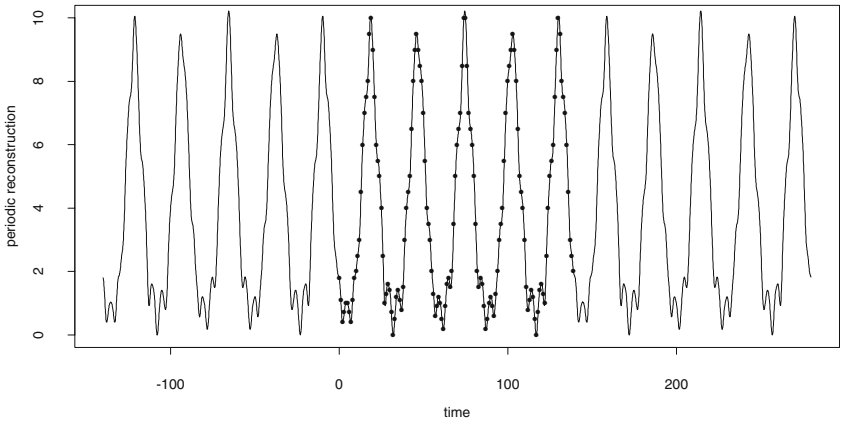


Fig. 6.3. The Fourier representation $f_p(t)$ of the sawtooth-shaped sequence (f_t) as plotted in Fig. 6.1 (c) is periodic with period of 140 time units. (f_t) is plotted for $t = 0, 1, \ldots, 139$, $f_p(t)$ for $-140.0, -139.9, \ldots$ through to $279.8, 279.9$

quency $1/s_k$ and the pertaining harmonics, as demonstrated in Problem 6.1 using the example sawtooth-shaped sequence (f_t) as plotted in Fig. 6.1 (a). Obviously, on the one hand, a strong (with large amplitudes) fluctuation in (g_t) with period $1/s_k$ induces, in $g_p(t)$, a trigonometric oscillation with frequency s_k and the pertaining harmonics such that the sum of the squared absolute values of their coefficients is large. These are the oscillations in $g_p(t)$ having periods being "somehow reconcilable" as argued above, in the remarks to Fig. 6.1, with periodic fluctuations in (g_t), $g_p(t)$ being the Fourier representation of (g_t). On the other hand, a weak (with small amplitudes) fluctuation in (g_t) with period $1/s_k$ induces in $g_p(t)$ a family of trigonometric oscillations with frequency s_k and its multiples such that the sum of the squared absolute values of their coefficients is small.

It is demonstrated, above in this section, that the periods of the fluctuations in the sawtooth-shaped sequence (f_t) are more easily detected in a plot of its spectrum $|F_k|^2$ than in a plot of (f_t) (the sequence is plotted in Fig. 6.1 (a), its spectrum in Fig. 6.1 (a)). This is an example of a successful application of a Fourier analysis for detecting periodic fluctuations "hidden" in sequence (f_t).

Fourier analysis is also able to detect "hidden periodicities" in real data when the following recipe is applied. (i) Given a sequence of observations (g_t), calculate the coefficients A_k and B_k of the trigonometric oscillations in the Fourier representation $g_p(t)$ of (g_t) using R expressions as in the remarks to Fig. 6.1. (ii) Plot the spectrum of (g_t), i.e., the squared absolute values of the oscillations, $|G_k|^2 = A_k^2 + B_k^2$, against their frequencies s_k in logarithmic scale. (iii) In the spectrum, try to identify sets of large values having frequen-

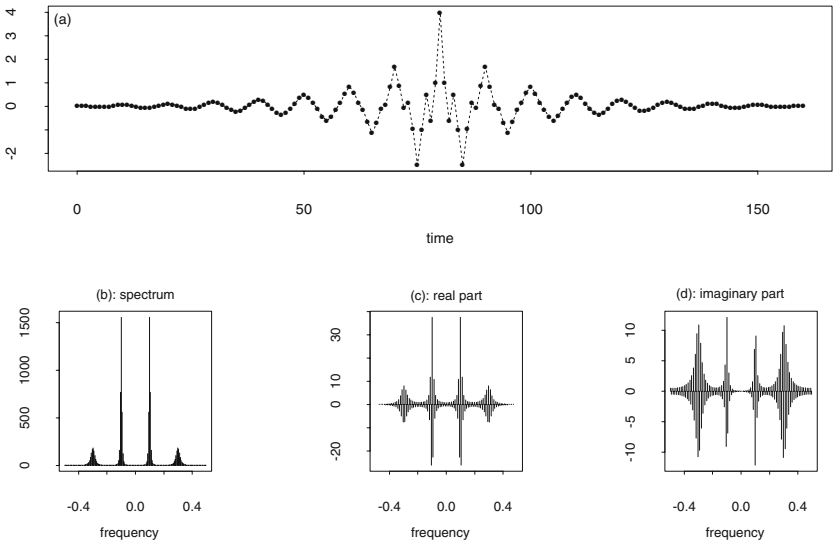


Fig. 6.4. Fourier transform (in plots (c) and (d)) of a deterministic sequence (g_t) as shown in plot (a), and its spectrum in plot (b).

cies s_k and multiples thereof. Such a set of oscillations can be induced by a fluctuation with period $1/s_k$ in (g_t). However, when this recipe is applied to observations of a deterministic or a random function, some pitfalls have to be circumvented using the theory and procedures introduced in this and the following chapters.

6.1.2 Fouriertransforms: a Preview to Chap. 6

The Fourier representation of the saw-tooth shaped sequence that is successfully computed in Sect. 6.1.1 (in the remarks to Fig. 6.1) suggests that a real-valued deterministic sequence of finite length N given for $t = 0, 1, \ldots, N-1$ can be represented by adding the mean of the sequence and $N/2$ trigonometric oscillations. The $N/2$ complex coefficients of the oscillations are the result of a discrete Fourier transform of the sequence. The discrete Fourier transform is defined in (6.22). As a second example, the deterministic real-valued sequence available in `/path/fig64.dat` and plotted in Fig. 6.4 (a) is Fourier transformed to arrive at the complex-valued sequence as plotted in Fig. 6.4 (c) and (d) for positive and negative frequencies. Negative frequencies are introduced in the remarks to (6.24). Fig. 6.4 also contains the spectrum of the sequence in plot (b). Obviously, the spectrum is symmetric since it is a quadratic function, and therefore, it is often plotted only against the positive frequencies, as shown in Fig. 6.2.

Why are trigonometric oscillations of the frequencies $1/N$, $2/N$, ... used in a Fourier representation? Do these allow for a representation with the smallest number of oscillations? How are the coefficients of the oscillations computed? Do also functions $f(t)$ defined for real t have a Fourier representation? Answers to these questions are given in Sects. 6.2 and 6.3.

6.1.3 Spectra of Time Series: a Preview to Chaps. 7, 9 and 10

Using R expressions as proposed in Sect. 6.1.1 (in the remarks to Fig. 6.1), a preliminary empirical spectrum of a time series can be calculated and plotted. In this case, plotting the logarithms of the squared absolute values of the trigonometric oscillations is desirable because (i), as in the deterministic case, ratios become differences, and (ii), there are statistical reasons (becoming obvious in Figs. 9.3 and 9.4, as well as in Fig. 9.23) for plotting a spectrum in logarithmic scale. To give examples, four preliminary empirical spectra of four time series are computed as demonstrated in Sect. 6.1.1 and thereafter plotted in Fig. 6.5, i.e., those of the

1. NAO index as plotted in Fig. 2.10, in plot (a)
2. amount of a product formed during a chemical reaction as plotted in Fig. 5.3, in plot (b)
3. horizontal wind speeds measured in a turbulent atmospheric flow as plotted in Fig. 2.1, in plot (c)
4. SO index from January 1933 through to January 1996 as plotted in Fig. 5.14, in plot (d).

It is obvious from these preliminary empirical spectra that, in a local (local properties are defined in (2.55,1)) mean, the absolute values of the trigonometric oscillations with low frequencies are larger than those with high frequencies in plot (c), whereas, in plot (b), the absolute values of the trigonometric oscillations increase with increasing frequency. Contrary to the above preliminary empirical spectra that increase or decrease in a local mean, the spectrum in plot (a) remains constant, and the one in plot (d) has a peak a for frequencies ≈ 0.02 month^{-1}, both in their local means.

The trigonometric oscillations in a Fourier representation of a time series have frequencies that depend on the sampling interval used for the measurements of the time series, i.e., one year in the case of the NAO index, one second in the case of the wind speed series, and one month in the case of the SO index monthly values. In the case of the amount of a product formed during a chemical reaction there is no time unit.

The preliminary empirical spectra in Fig. 6.5 are calculated under the assumption that the time series are deterministic sequences as those, for example, plotted in Figs. 6.1 (a) and 6.4 (a). However, the time series used for the calculation of the empirical spectra in Fig. 6.5 are realisations of linear processes (2.25), as shown in Chap. 2 in the remarks to Figs. 2.10 and 2.18 (for the NAO index) as well as in Chap. 5 in Sects. 5.1.3 (for the wind speed

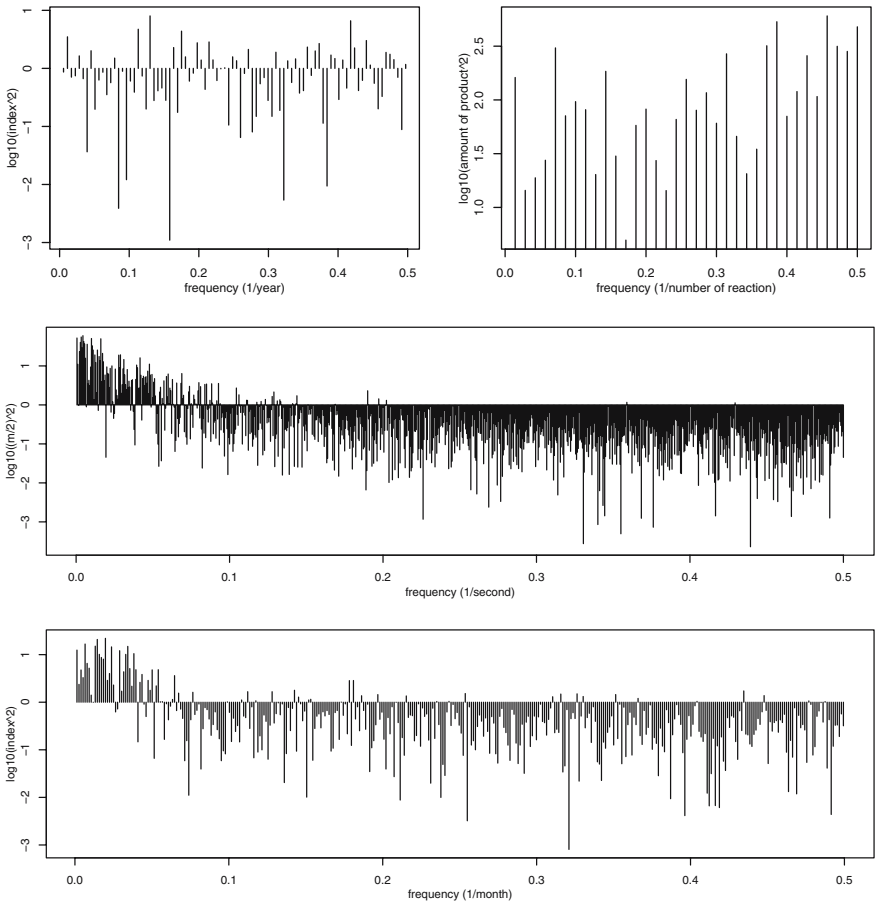


Fig. 6.5. Preliminary empirical spectra of following time series: (a) the NAO index (Jones) as plotted in Fig. 2.10, (b) the amount of a product formed during a chemical reaction as plotted in Fig. 5.3, (c) the horizontal wind speeds as plotted in Fig. 2.1, (d) SO index from January 1933 through to January 1996 as plotted in Fig. 5.14.

series and the amount of a product formed during a chemical reaction) and 5.4.4 (for the SO index).

Does this assumption, clearly not being reconcilable with the probabilistic properties of the observed variables, induce the extremely strong fluctuations in the preliminary empirical spectra in Fig. 6.5? This question arises because these preliminary empirical spectra neither resemble plot (a) in Fig. 6.2, showing the spectrum of the sequence in Fig. 6.1 (a) with a few oscillations having large absolute values, nor plot (b) in Fig. 6.4 which shows the relatively smooth spectrum of the sequence in Fig. 6.4 (a).

Since a linear process has a smooth spectrum, as shown in Sect. 7.4, it is supposed that the probabilistic properties of the variable observed have to be accounted for when the spectrum of the process is estimated from a realisation. Such estimators are introduced in Chap. 9.

6.2 Linear Vector Spaces

This section introduces us to the properties of linear combinations of trigonometric functions used in a Fourier representation (an example is plotted in Fig. 6.1 (c)). The trigonometric functions in a Fourier representation are orthogonal, as summarised in Sect. 6.2.3 and derived in Sect. 6.9.1. The orthogonality of sine and cosine functions with Fourier frequencies is shown, in Sect. 6.2.2, to favour the relatively straightforward calculation of their coefficients. The orthogonality of functions or sequences is a concept borrowed from geometry and is meaningful on condition that functions or sequences are considered to be vectors (in Sect. 6.2.1).

6.2.1 Linear Vector Spaces

When constructing linear combinations of functions, initial problems have to be solved which are similar to those encountered when constructing linear combinations of vectors in three-dimensional Euclidean space: a point with coordinates x, y, z is a three-tuple (x, y, z), and the position (radius) vector from the origin to this point can be written as linear combination $v = xi + yj + zk$ of the basis vectors i, j, k. (In this section, vectors are not written bold). This is geometrically clear and physically meaningful as the basis is defined by unit distances on the three axes being perpendicular to each other.

Despite being geometrically no longer obvious and physically no longer immediately meaningful, the generalisation of the above linear combinations of functions to higher dimensional vector spaces R^N is known from algebra. In the higher dimensional case, the points are N-tuples $\phi = (\varphi_1, \ldots, \varphi_N)^T$, φ_i real numbers, and the pertaining "position" vectors are linear combinations of N mutually independent basis vectors $b_1, b_2, \ldots, b_N$: $v = \varphi_1 b_1 + \ldots + \varphi_N b_N$, N being the dimension of the vector space. Operations are then defined on the vectors in R^N; from these operations, the addition, the multiplication with a real number (i.e., with a scalar), the scalar product and the norm are used in this section.

The geometric concepts originating from Euclidean space are now applied to functions $\phi(t)$, $a \leq t \leq b$, in a N- or ∞-dimensional vector space: the $\phi(t)$ can be added and/or multiplied with a scalar and thus linear combinations of the $\phi(t)$ can be obtained. For example, the Fourier representation of the sawtooth-shaped sequence (f_t), $t = 0, 1, \ldots, 139$, as plotted in Fig. 6.1 (c), is a linear combination of (trigonometric) functions in a vector space with $N = 140$.

If the argument t of the functions $\phi(t)$, $a \le t \le b$, is a real number, then the $\phi(t)$ have an infinite number of components, i.e., values of $\phi(t)$, $a \le t \le b$. In this case, the $\phi(t)$ are infinite-dimensional vectors in an infinite-dimensional vector space. If the argument t of the functions $\phi(t)$, $a \le t \le b$, is an integer number, then the $\phi(t)$ have a finite number of components, one for each t in $a \le t \le b$. In Chaps. 1 through to 5 and also in Sect. 6.1, such functions are called sequences (x_t), (f_t), etc., usually with argument $t = 1, \ldots, N$. In this section, and further on in this chapter, the term function is also used for sequences $\phi(t)$, $\psi(t)$, $\phi_n(t)$, $y(x)$, $z(x)$, $f(t)$, $x(t)$, etc., the argument t being a real or an integer number (if integer, then usually $t = 0, 1, \ldots, N-1$). For example, using this terminology and assuming that the origin in the time domain is arbitrary, a time series (x_t), $t = 1, 2, \ldots, N$, is a vector $x(t)$ with components $x(0), x(1), \ldots, x(N-1)$.

For each t, there is a vector in the basis of the vector space. If t is integer then there are N basis vectors, if t is real then there are an infinite number of basis vectors. Functions in a linear vector space can be multiplied with each other to obtain the scalar product of a pair of functions, and, when a function is multiplied with itself, its "length" (borrowing from the original geometric concepts), i.e., its norm is obtained.

These properties are summarised in (6.3), where the set of functions with finite norm in the interval $a \le t \le b$ is defined.

Let $\phi(t)$, $\phi_k(t)$, $\psi(t)$, $\psi_k(t)$, $k = 1, 2, \ldots$ be real-valued functions of t, $a \le t \le b$, t integer or real, and a, b, d_k real constants. Then:

1. $\psi(t) = d_1\phi_1(t) + d_2\phi_2(t)$ *is a linear combination of $\phi_1(t)$ and $\phi_2(t)$.*
2. $\langle \phi\psi \rangle = \begin{cases} \sum_{a \le t \le b} \phi(t)\psi(t) & \text{for integer } t \\ \int_a^b \phi(t)\psi(t)\mathrm{d}t & \text{for real } t \end{cases}$
   is the scalar (or dot) product of $\phi(t)$ and $\psi(t)$. $\langle\ \rangle$ usually denote the inner product in a vector space. (6.3)
3. $\|\phi(t)\|^2 = \begin{cases} \sum_{a \le t \le b} \big(\phi(t)\big)^2 & \text{for integer } t \\ \int_a^b \big(\phi(t)\big)^2 \mathrm{d}t & \text{for real } t \end{cases}$
   is the squared norm of $\phi(t)$.
4. *The set of functions $\big(\phi_k(t)\big)$ (sets of functions are enclosed in parentheses ( ) and not braces { } in this chapter) with $\|\phi_k(t)\|^2 < \infty$ is called $L^2(a, b)$. $L^2(a, b)$ is the linear vector space of functions with finite norm on the real $a \le t \le b$.*

For example, the sawtooth-shaped sequence (f_t) as plotted in Fig. 6.1 (a) is in $L^2(0, 139)$, and the sequence (g_t) as plotted in Fig. 6.4 (a) is in $L^2(0, 160)$, the argument t being integer in both cases. The result of Problem 1.12 implies that the Gaussian function $y(x) = \mathrm{e}^{-bx^2}$, x real, is in $L^2(-\infty, \infty)$ for $b > 0$. Function $y(x) = d \times x$, x real and d a real constant, is in $L^2(a, b)$, yet neither in $L^2(-\infty, \infty)$ nor in $L^2(a, \infty)$ nor in $L^2(-\infty, b)$.

In $L^2(a,b)$, the Cauchy-Schwarz inequality as proposed in (6.4) holds

$$\left(\int_a^b \phi(t)\psi(t)\mathrm{d}t\right)^2 \leq \int_a^b \big(\phi(t)\big)^2\mathrm{d}t \int_a^b \big(\psi(t)\big)^2\mathrm{d}t \tag{6.4}$$

which is used, for example, to show (1.10,4). (6.4) also holds for the absolute values of complex-valued functions as shown in [18]. Applying the Cauchy-Schwarz inequality, the triangle inequality $\|\phi(t)+\psi(t)\| \leq \|\phi(t)\| + \|\psi(t)\|$ is easily obtained in Problem 6.2.

Using the "vector properties" of functions as defined in (6.3), an answer to the question can be given, whether a function can be approximated by a sequence of functions. In a (linear) vector space, usually the norm is applied to find out how accurately a vector has been approximated by means of a sequence of vectors, as defined in (6.5).

In $L^2(a,b)$, a sequence of functions $\phi_n(t)$, $n=1,2,\ldots$, converges in mean square to a function $\phi(t) \in L^2(a,b)$, on condition that $\lim_{n\to\infty}\int_a^b \big(\phi_n(t)-\phi(t)\big)^2\mathrm{d}t = 0$. To denote convergence in mean square, $\phi_n(t) \overset{\text{ms}}{\to} \phi(t)$ is used or, alternatively, (i) $\phi_n(t)$ converges in mean square to $\phi(t)$, (ii) $\phi_n(t)$ converges in $L^2(a,b)$ to $\phi(t)$ or (iii) $\phi_n(t)$ converges in the norm to $\phi(t)$. (6.5)

In Fig. 6.1 (c), for example, the sawtooth-shaped sequence (f_t) is the function $\phi(t)$, $t=0,\ldots,139$, to be approximated, and $\phi_n(t)$ is the sum of the trigonometric oscillations subsequent to the addition of oscillation $k=n$: convergence in the norm is arrived at with $k=n=70$. In this example, $(f_t)=\phi(t)$ and all $\phi_n(t)$, $n=1,\ldots,70$, are in $L^2(0,139)$, t being integer.

Examples for sequences of functions in $L^2(-\infty,\infty)$ and with real argument t are plotted in Fig. 6.6. These plots are generated in Problem 6.3. The Gaussian functions $y_k(x) = \mathrm{e}^{-(1/10)^k x^2}$ in plot (a) are in $L^2(-\infty,\infty)$ for finite k; for $k\to\infty$, however, functions $y_k(x)$ degenerate to a constant (plotted with a solid line) not being in $L^2(-\infty,\infty)$. The functions $z_k(x) = \mathrm{e}^{-10^k x^2}$ in plot (b) are in $L^2(-\infty,\infty)$ for $k=1,2,3,\ldots$ and the sequence of functions $z_k(x)$ converges in $L^2(-\infty,\infty)$ to the constant 0, i.e., $z_k(x) \overset{\text{ms}}{\to} 0$ for $k\to\infty$, although $z_k(0)=1$ for all k. In these examples, the limiting functions (solid lines in Fig. 6.6) can be deduced directly from the definitions of $y_k(x)$ and $z_k(x)$.

The convergence in mean square of a sequence of functions $\phi_n(t)$ to a limiting function $\phi(t)$ can be shown even when $\phi(t)$ is not known. In this case, $\phi_n(t)$ can be proved to converge using a Cauchy sequence. In general, a Cauchy sequence allows for proving a mean-square convergence in arbitrary vector spaces (an example for a Cauchy sequence in the vector space of random variables $L^2(\Omega,F)$ defined in (7.10) is given in (7.12); in (6.6), however, it is defined for the case of a convergence in $L^2(a,b)$. (6.6,1) is the

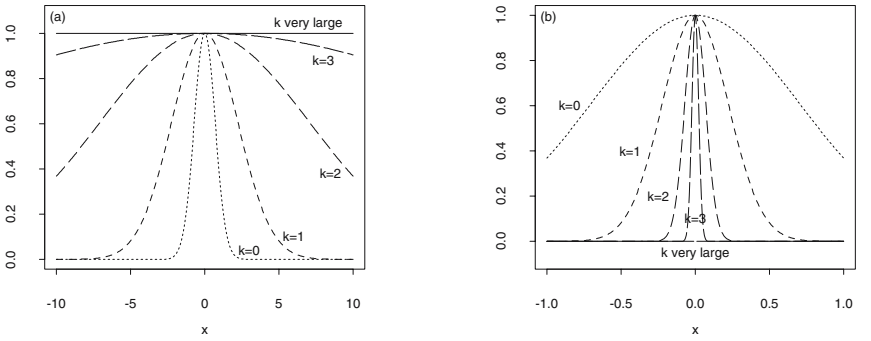


Fig. 6.6. Sequences of functions: (a) $y_k(x) = \mathrm{e}^{-(1/10)^k x^2}$, (b): $z_k(x) = \mathrm{e}^{-10^k x^2}$, $k = 0, 1, 2, \ldots$. The functions for $k \to \infty$ are plotted with solid lines.

definition of the Cauchy sequence in $L^2(a, b)$, (6.6,2) is a property of $L^2(a, b)$ (see, e.g., [126]) and (6.6,3) is the definition of the complete vector space.

Let $\phi_n(t)$, $n = 1, 2, \ldots$, be a sequence of functions in $L^2(a, b)$, and ϵ a real number, positive and arbitrarily small. Then:

1. *$\phi_n(t)$ is called a Cauchy sequence in $L^2(a, b)$ on condition that, given an ϵ, a positive integer number $N(\epsilon)$ exists such that $\int_a^b \big(\phi_n(t) - \phi_m(t)\big)^2 \mathrm{d}t < \epsilon$ for $n, m > N(\epsilon)$.* (6.6)
2. *Every Cauchy sequence $\phi_n(t)$ in $L^2(a, b)$ is associated with a function $\phi(t)$ in $L^2(a, b)$ such that $\phi_n(t) \overset{\text{ms}}{\to} \phi(t)$.*
3. *A vector space with property (6.6,2) is called complete or a Hilbert space, e.g., $L^2(a, b)$ is a complete vector space.*

For example, $z_k(x) = \mathrm{e}^{-10^k x^2}$, $k = 1, 2, 3, \ldots$, in Fig. 6.6 (b) is a Cauchy sequence in $L^2(-\infty, \infty)$, whereas $y_k(x) = \mathrm{e}^{-(1/10)^k x^2}$, $k = 1, 2, 3, \ldots$, in Fig. 6.6 (a) is not a Cauchy sequence in $L^2(-\infty, \infty)$, as shown in Problem 6.3. Since $L^2(-\infty, \infty)$ is complete, $z_k(x)$ converges in $L^2(-\infty, \infty)$, whereas $y_k(x)$ does not converge. In both cases, the limiting functions are known (plotted with solid lines in Fig. 6.6) and therefore, in Problem 6.3, the Cauchy sequences are calculated for illustrating definition (6.6,1).

6.2.2 Orthogonal Functions

Applying definitions (6.3,2,3), the angle θ between functions $\phi(t)$ and $\psi(t)$ in $L^2(a, b)$ is obtained as $\cos\theta = \langle\phi\psi\rangle / (\|\phi(t)\| \times \|\psi(t)\|)$. If $\langle\phi\psi\rangle = 0$ then $\theta = \pi/2$, and the functions are said to be orthogonal, as defined in (6.7,1).

In (6.7,3.1), $\delta_{kl} = 1$ for $k = l$ and $\delta_{kl} = 0$ for $k \neq l$, $k, l = 1, 2, \ldots$, is called Kronecker δ.

In (6.7,3.2), $\Phi(t) \stackrel{\text{ms}}{=} \sum_{k=1}^{\infty} d_k \phi_k(t)$ means that a linear combination of the $\phi_k(t)$ converges in mean square as defined in (6.5) to $\Phi(t)$: $\Phi(t) \stackrel{\text{ms}}{=} \sum_{k=1}^{\infty} d_k \phi_k(t)$ is written for $\sum_{k=1}^{n} d_k \phi_k(t) \stackrel{\text{ms}}{\to} \Phi(t)$, $n = 1, 2, \ldots$.

1. *Two functions $\phi_1(t)$ and $\phi_2(t)$ in $L^2(a,b)$ are said to be orthogonal on condition that their scalar product is identically zero, i.e., $\langle \phi_1 \phi_2 \rangle = 0$.*
2. *A function $\phi(t)$ in $L^2(a,b)$ is called normalised on condition that its norm amounts to unity, i.e., $\|\phi(t)\| = 1$.*
3. *A set of functions $\big(\phi_k(t) \in L^2(a,b)\big)$, $k = 1, 2, ...,$ is called:* (6.7)

3.1 *orthonormal on condition that $\langle \phi_k \phi_l \rangle = \delta_{kl}$, $k, l = 1, 2, \ldots,$ δ_{kl} being the Kronecker δ as in the above remarks;*

3.2 *an orthonormal basis in $L^2(a,b)$ on condition that, besides being orthonormal, $\Phi(t) \stackrel{\text{ms}}{=} \sum_{k=1}^{\infty} d_k \phi_k(t)$ holds for any function $\Phi(t)$ in $L^2(a,b)$, $\stackrel{\text{ms}}{=}$ as in the above remarks.*

For example, definitions (6.7) are applied to the Fourier representation in Fig. 6.1 (c) of the sawtooth-shaped sequence (f_t). (f_t), being in $L^2(0, 139)$, plays the role of $\Phi(t)$ in $L^2(a,b)$ in (6.7); the trigonometric functions, pairs of which are amalgamated in the trigonometric oscillations, play the role of the $\phi_k(t)$, with the Fourier coefficients A_k and B_k, obtained with the R expressions in the remarks to Fig. 6.1 (c), acting as the d_k. Obviously, in Fig. 6.1 (c), the Fourier representation converges in $L^2(0, 139)$ to (f_t). Trigonometric functions with identical frequencies, i.e., the Fourier frequencies, but having distinct Fourier coefficients, are used in the Fourier representations when the experiment is repeated with any other sequence in $L^2(0, 139)$. This suggests that trigonometric functions with Fourier frequencies, properly normalised, are an orthonormal basis in $L^2(0, 139)$, as can be deduced from the sums and orthogonality relations below in (6.10), (6.11), (6.12) and (6.13).

Prior to introducing orthogonal trigonometric functions in Sect. 6.2.3, an answer to the following question has to be found: given a function $\Phi(t)$ and an orthonormal basis, both in $L^2(a,b)$, how can the coefficients d_k in (6.7,3.2) be calculated such that the linear combination of functions $\phi_k(t)$ in the basis converges in mean square to $\Phi(t)$? In the case of the sawtooth-shaped sequence (f_t) in Fig. 6.1: how can the Fourier coefficients A_k and B_k be obtained?

In general, a coefficient d_l ($l = 1, 2, \ldots$ substituted in (6.7,3) for $k = 1, 2, \ldots$) is calculated by multiplying both sides of $\stackrel{\text{ms}}{=}$ in (6.7,3.2) with $\phi_l(t)$, integrating from a to b and taking advantage of the properties of the orthonormal basis

$$\int_a^b \Phi(t)\phi_l(t)\mathrm{d}t \stackrel{\text{ms}}{=} \int_a^b \phi_l(t) \sum_{\ell=1}^{\infty} d_\ell \phi_\ell(t)\mathrm{d}t = \sum_{\ell=1}^{\infty} d_\ell \int_a^b \phi_l(t)\phi_\ell(t)\mathrm{d}t = d_l \quad (6.8)$$

since $\int_a^b \phi_l(t)\phi_\ell(t)\mathrm{d}t = 0$, for $l \neq \ell$ and $\int_a^b \phi_l(t)\phi_\ell(t)\mathrm{d}t = 1$, for $l = \ell$. In (6.8), the order of the sum and integral can be changed due to $d_k = \langle \Phi(t)\phi_k(t) \rangle$,

provided that any representation $\Phi(t) \overset{\text{ms}}{=} \sum_{k=1}^{\infty} d_k \phi_k(t)$ does exist at all. If such a representation exists, then $\sum_{k=1}^{n} \phi_k(t)$, $n = 1, 2, \ldots$, being a special case of a linear combination, converges in mean square and the order of the sum and the integral can be reversed.

The conditions for a convergence depend on function $\Phi(t)$ and the orthonormal basis $\big(\phi_k(t)\big)$. For a proof of a possible convergence, usually coefficients d_l are calculated as required in (6.8), and then the linear combination of the functions in the basis is shown to converge in mean square. If it converges then the coefficients have been determined.

6.2.3 Orthogonal Trigonometric Functions

Are trigonometric functions with Fourier frequencies, possibly subsequent to having been normalised, an orthonormal basis in $L^2(a, b)$ as defined in (6.7)? To give an answer, two cases are distinguished: either the argument is an integer or a real number.

In the first case, an orthonormal basis can be obtained, subsequent to an adequate normalisation, from the set of trigonometric functions as defined in (6.9).

$$\left(\cos\big(2\pi\frac{k}{N}t\big), \sin\big(2\pi\frac{k}{N}t\big)\right) \qquad \begin{array}{l} t, k = \ldots, -1, 0, 1, \ldots \text{ and} \\ 0 \le t \le N-1,\ 0 \le k \le (N/2) \end{array} \tag{6.9}$$

For $k = 0$, the cosine function becomes identically 1 and the sine function identically 0. For $k = N/2$ and even N, the argument of the functions becomes πt and, consequently, the cosine has alternating values 1 or -1 whereas the sine becomes identically 0, as for $k = 0$. The set (6.9) therefore contains $2(m + 1)$ trigonometric functions, m the largest integer number such that $m \le (N/2)$. When the sine functions being identically 0 are excluded, the number of functions in (6.9) is reduced by 2 functions for even N and by 1 function for odd N. Due to this reduction, the set defined in (6.9) contains N trigonometric functions with properties (6.10), (6.11), (6.12) and (6.13).

$$\sum_{t=0}^{N-1} \cos\big(2\pi\frac{k}{N}t\big) = \begin{cases} 0 & \text{for } k \neq 0 \\ N & \text{for } k = 0 \end{cases} \qquad \text{and} \qquad \sum_{t=0}^{N-1} \sin\big(2\pi\frac{k}{N}t\big) = 0 \tag{6.10}$$

$$\sum_{t=0}^{N-1} \sin\big(2\pi\frac{k}{N}t\big)\cos\big(2\pi\frac{l}{N}t\big) = 0 \qquad \text{for } k, l = 0, 1, ..., m \tag{6.11}$$

$$\sum_{t=0}^{N-1} \cos\big(2\pi\frac{k}{N}t\big)\cos\big(2\pi\frac{l}{N}t\big) = \begin{cases} 0 & \text{for } k \neq l \\ N/2 & \text{for } k = l \neq 0,\ N/2 \\ N & \text{for } k = l = 0,\ N/2 \end{cases} \tag{6.12}$$

$$\sum_{t=0}^{N-1} \sin\big(2\pi\frac{k}{N}t\big)\sin\big(2\pi\frac{l}{N}t\big) = \begin{cases} 0 & \text{for } k \neq l \\ N/2 & \text{for } k = l \neq 0,\ N/2 \end{cases} \tag{6.13}$$

In (6.10), the sums of the functions in (6.9) are identically zero for $k \neq 0$ as derived in (6.128) and (6.129). In (6.130), the cosine functions are shown to be orthogonal as required in (6.12); orthogonality relations (6.11) and (6.13) are obtained with similar derivations. In these sums and orthogonality relations, $k, l = 0, 1, ...m$, m the largest integer number such that $m \leq (N/2)$. Thus, although the trigonometric functions in the set (6.9) are orthogonal, they are, however, not an orthonormal basis.

The trigonometric functions in the Fourier representation in Fig. 6.1 (c) of the sawtooth-shaped sequence (f_t), for example, are in the set defined in (6.9) with $N = 140$, $m = 70$, $t = 0, 1, \ldots, 139$ and $k = 0, 1, \ldots, 70$.

In the second case, the argument of the functions is real and an orthonormal basis can be obtained, subsequent to adequately normalising, from the set as defined in (6.14)

$$\left(\cos\big(2\pi\frac{k}{T}t\big), \sin\big(2\pi\frac{k}{T}t\big)\right) \qquad t \text{ real}, \quad k = 0, 1, 2, ... \tag{6.14}$$

which contains an infinite number of trigonometric functions being periodic in an interval of length T. Their integrals in (6.15) and (6.16) are arrived at in (6.131). Using these integrals, the orthogonality relations in (6.17), (6.18) and (6.19) can be derived. Prior to these derivations, the products are written as sums applying the elementary trigonometric identities.

$$\int_{-T/2}^{T/2} \cos\big(2\pi\frac{k}{T}t\big)\mathrm{d}t = \begin{cases} 0 & \text{for } k \neq 0 \\ T & \text{for } k = 0 \end{cases} \tag{6.15}$$

$$\int_{-T/2}^{T/2} \sin\big(2\pi\frac{k}{T}t\big)\mathrm{d}t = 0 \qquad \text{for } k = 1, 2, 3, ... \tag{6.16}$$

$$\int_{-T/2}^{T/2} \sin\big(2\pi\frac{k}{T}t\big)\cos\big(2\pi\frac{l}{T}t\big)\mathrm{d}t = 0 \qquad \text{for } k, l = 0, 1, ... \tag{6.17}$$

$$\int_{-T/2}^{T/2} \cos\big(2\pi\frac{k}{T}t\big)\cos\big(2\pi\frac{l}{T}t\big)\mathrm{d}t = \begin{cases} 0 & \text{for } k \neq l \\ T/2 & \text{for } k = l \neq 0 \\ T & \text{for } k = l = 0 \end{cases} \tag{6.18}$$

$$\int_{-T/2}^{T/2} \sin\big(2\pi\frac{k}{T}t\big)\sin\big(2\pi\frac{l}{T}t\big)\mathrm{d}t = \begin{cases} 0 & \text{for } k \neq l \text{ oder } k = l = 0 \\ T/2 & \text{for } k = l \neq 0 \end{cases} \tag{6.19}$$

Using three trigonometric functions from the set (6.14), for example, function $f(t) = 2t$, $-1/2 \leq t \leq 1/2$, is approximated in Fig. 6.8.

In the following section, a function $f(t)$ being in $L^2(a, b)$ is approximated in the sense that the norm of differences becomes minimal, as required in (6.5), using trigonometric functions. The trigonometric functions are in the set (6.14) for real t, and in the set (6.9) for $t = \ldots, -1, 0, 1, \ldots$. However, the norms of the functions in both sets are not unity, as required in (6.7) and (6.8). This has to be accounted for in the approximations.

6.3 Fourier Transforms

A Fourier transform generates, using a formula $\mathcal{F}_{-\mathrm{i}}(\ )$, from a function $f(t)$ (t usually being the time) a function $F(s)$ (s being the frequency of trigonometric oscillations): $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$. $f(t)$ and $F(s)$ are called a *Fourier transform pair:* usually a small letter is used for the function with argument t, and a capital letter for the function with argument s. In this section, four cases of Fourier transform pairs are introduced:

1. *t integer, (f_t) in $L^2(0, N-1)$, $F(s_k)$ for a finite number of s_k,*
2. *t real, $f(t)$ in $L^2(-T/2, T/2)$, $F(s_k)$ for an infinite number of s_k,*
3. *t real, $f(t)$ in $L^2(-\infty, \infty)$, s real, $F(s)$ in $L^2(-\infty, \infty)$, and*
4. *t integer, (f_t) in $L^2(-\infty, \infty)$, s real, $F(s)$ in $L^2(-1/2, 1/2)$.*

(6.20)

6.3.1 Case 1: Discrete Fourier Transform

The calculation of the Fourier coefficients A_k and B_k in the Fourier representation of the sawtooth-shaped sequence in Fig. 6.1 (c) demonstrates the first case in (6.20,1).

Let (f_t), $t = 0, 1, 2, \ldots, N-1$, be a sequence in $L^2(0, N-1)$. Then:

1. *The frequencies $s_k = k/N$, with $k = 0, 1, \ldots, m$, m the largest integer $\leq N/2$, are called Fourier frequencies of (f_t), and $\Delta s_k = s_{k+1} - s_k = (k+1)/N - k/N = 1/N$ is the distance of the s_k.*
2. *In case of even N:*

$$A_k = \begin{cases} (2/N)\sum_{t=0}^{N-1} f_t \cos(2\pi s_k t) & \text{for } k = 1, 2, \ldots, m-1 \\ (1/N)\sum_{t=0}^{N-1} f_t \cos(2\pi s_k t) & \text{for } k = 0, m \end{cases}$$

$$B_k = \begin{cases} (2/N)\sum_{t=0}^{N-1} f_t \sin(2\pi s_k t) & \text{for } k = 1, 2, \ldots, m-1 \\ 0 & \text{for } k = 0, m \end{cases}$$

and, in case of odd N:

$$A_k = \begin{cases} (2/N)\sum_{t=0}^{N-1} f_t \cos(2\pi s_k t) & \text{for } k = 1, 2, \ldots, m \\ (1/N)\sum_{t=0}^{N-1} f_t \cos(2\pi s_k t) & \text{for } k = 0 \end{cases}$$

$$B_k = \begin{cases} (2/N)\sum_{t=0}^{N-1} f_t \sin(2\pi s_k t) & \text{for } k = 1, 2, \ldots, m \\ 0 & \text{for } k = 0 \end{cases}$$

are called Fourier coefficients of the sequence (f_t), $t = 0, 1, \ldots, N-1$.

3. *If (f_t), k, m, N, s_k, A_k, and B_k are defined as above, then $f_m(t) = \sum_{k=0}^{m}\big(A_k \cos(2\pi s_k t) + B_k \sin(2\pi s_k t)\big)$ is called finite Fourier series and $f_m(t) = (f_t)$, for $t = 0, 1, ..., N-1$.*

(6.21)

Since the sine and cosine functions with argument $2\pi s_k t$ are in the set of trigonometric functions (6.9) and thus are orthogonal in $L^2(0, N-1)$ as proposed in (6.11), (6.12) and (6.13), and since (f_t) is also in $L^2(0, N-1)$, the Fourier coefficients A_k and B_k are obtained applying (6.8) and thus (6.21,2) is arrived at, provided it is taken into account that (i) t is an integer number and

(ii) the trigonometric functions in (6.9) are orthogonal but not normalised. Using the A_k and B_k as coefficients in a linear combination of the sines and cosines, the finite Fourier series in (6.21,3) is arrived at, which converges in mean square to (f_t). In (6.21,3), $=$ substitutes $\overset{\text{ms}}{=}$ as, obviously, $f_m(t)$ converges in $L^2(0, N-1)$ to (f_t). When it is obvious that a sequence of functions converges in $L^2(\ldots,\ldots)$ to a limiting function, then often $=$ is written for $\overset{\text{ms}}{=}$ and even for $\overset{\text{ms}}{\to}$.

The convergence in mean square (6.21,3) is implied by (6.22,3) (with due substitutions applying (6.25)). (6.22,3) is derived in (6.135) and (6.136). Consequently, any (f_t), $t = 0, 1, \ldots, N-1$, has a Fourier representation as linear combination of the orthogonal trigonometric functions (6.9). The existence of a Fourier representation is assumed in Sect. 6.1.2.

For example, the Fourier representation, i.e., the finite Fourier series $f_m(t)$ (6.21,3), t an integer number, converges in $L^2(0, 139)$, i.e., for $t = 0, 1, \ldots, 139$, to the sawtooth-shaped sequence (f_t), as demonstrated in Fig. 6.1 (c). In Fig. 6.1 (c), $f_m(t)$ is plotted with a solid line, it is calculated for $t = 0.0, 0.1, 0.2, \ldots, 139.8, 139.9$, and, being a linear combination of the trigonometric functions (6.9), is a smooth function. As demonstrated in Fig. 6.1 (c), $f_m(t)$ can be calculated for real t; convergence to the sawtooth-shaped sequence (f_t), however, is arrived at for $t = 0, 1, \ldots, 138, 139$.

The sequences of the Fourier coefficients (A_k) and (B_k) pertaining to the sequence (f_t) are not arrived at, in practical applications, from computations as required in (6.21,2), but by way of normalising the Fourier transform (F_k) of (f_t) as defined in (6.22):

1. *If* (f_t) *is a sequence in* $L^2(0, N-1)$ *then*
   $(F_k) = \mathcal{F}_{-\mathrm{i}}\big(f_t\big) = \frac{1}{N}\sum_{t=0}^{N-1} f_t \mathrm{e}^{-\mathrm{i}2\pi(k/N)t}$, *for* $k = \ldots, -1, 0, 1, \ldots$
   $= \frac{1}{N}\sum_{t=0}^{N-1} f_t \cos\big(2\pi(k/N)t\big) - \frac{\mathrm{i}}{N}\sum_{t=0}^{N-1} f_t \sin\big(2\pi(k/N)t\big)$
   *is called discrete Fourier–*i*-transform of* (f_t).
2. *If* (F_k) *is a sequence in* $L^2(0, N-1)$ *then*
   $(f_t) = \mathcal{F}_{+\mathrm{i}}\big(F_k\big) = \sum_{k=0}^{N-1} F_k \mathrm{e}^{+\mathrm{i}2\pi(t/N)k}$, *for* $t = \ldots, -1, 0, 1, \ldots$
   *is called discrete Fourier+*i*-transform of* (F_k).
3. (f_t) *and* (F_k) *are a Fourier transform pair,*
   since $(f_t) = \mathcal{F}_{+\mathrm{i}}\big(F_k\big)$, *where* $(F_k) = \mathcal{F}_{-\mathrm{i}}\big(f_t\big)$.
4. *Often, the* +i*-transform is normalised, not the* −i*-transform:*
   $(F_k) = \mathcal{F}_{-\mathrm{i}}\big(f_t\big) = \sum_{t=0}^{N-1} f_t \mathrm{e}^{-\mathrm{i}2\pi(k/N)t}$, *for* $k = \ldots, -1, 0, 1, \ldots$
   $(f_t) \;= \mathcal{F}_{+\mathrm{i}}\big(F_k\big) = \frac{1}{N}\sum_{k=0}^{N-1} F_k \mathrm{e}^{+\mathrm{i}2\pi(t/N)k}$, *for* $t = \ldots, -1, 0, 1, \ldots$

(6.22)

$\mathcal{F}_{-\mathrm{i}}(\ )$ and $\mathcal{F}_{+\mathrm{i}}(\ )$, used for Fourier transforms, are operators, similar to the derivative $(\mathrm{d}/\mathrm{d}t)(\ )$. Since the trigonometric functions (6.9) are orthogonal, but not normalised, either the −i- or the +i-transform are normalised such that (6.22,3) holds. (6.22,3) is derived in (6.135) and (6.136).

R function `fft(...,inverse=T)` transforms as required in the first line in (6.22,4), i.e., without normalising in the −i-transform. Consequently, the complex R vector resulting from `fft(...,inverse=T)` is normalised as demonstrated in the remarks to Fig. 6.1 to obtain the Fourier coefficients. The Fourier coefficients are defined in (6.21,2) for Fourier frequencies $s_k \leq 1/2$ because $0 \leq k \leq m$; whereas, in (6.22,1), the Fourier −i-transform is defined for all integer k, i.e., also for $k < 0$ and therefore also for negative frequencies. This allows for a definition that is symmetric to those of the Fourier +i-transform in (6.22,2). Symmetric definitions are desirable and also possible because the transforms are periodic (6.1) with period N.

The Fourier −i-transform (F_k) as in (6.22,1) is shown to be periodic with period N in Problem 6.4: $(F_k) = (F_{k+nN})$, for integer k, n, N. Since also $\mathrm{e}^{\mathrm{i}2\pi(t/N)k}$ is periodic with period N for $k = \ldots, -1, 0, 1, \ldots$, the product $F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k}$ is periodic with period N, and therefore, being a sum of such products, $\mathcal{F}_{+\mathrm{i}}(F_k)$ is periodic with period N, $t, k = \ldots, -1, 0, 1, \ldots$. Consequently, at the exterior of the interval $t = 0, 1, \ldots, N-1$, $\mathcal{F}_{+\mathrm{i}}(F_k)$ becomes a periodic version of (f_t), since $\mathcal{F}_{+\mathrm{i}}(F_k) = f_t$ for $t = 0, 1, \ldots, N-1$ as required in (6.22,3).

For example, the sawtooth-shaped sequence (f_t) in Fig. 6.1 (a) is defined for $t = 0, 1, \ldots, 139$. From its Fourier transform (F_k), calculated as required in (6.22,1), the Fourier representation $(f_t) = \mathcal{F}_{+\mathrm{i}}(F_k)$, $t = 0, 1, \ldots, 139$, as shown in Fig. 6.1 (c), is obtained by applying (6.22,2) and (6.22,3). $\mathcal{F}_{+\mathrm{i}}(F_k)$ is periodic with period 140, three periods being plotted in Fig. 6.3. Since the sawtooth-shaped sequence (f_t) is constructed, in the remarks to Fig. 6.1 (a), as a periodic function with period 140, $(f_t) = \mathcal{F}_{+\mathrm{i}}(F_k)$ for $t = \ldots, -1, 0, 1, \ldots$.

The Fourier representation of a non-periodic sequence, however, converges for $t = 0, 1, \ldots, N-1$, although not for all t. For example, $\mathcal{F}_{+\mathrm{i}}(G_k)$ of the non-periodic sequence (g_t) in Fig. 6.4 (a) converges to (g_t) for $t = 0, 1, \ldots, 160$, for $t < 0$ and $t > 160$; however, $(g_t) \neq \mathcal{F}_{+\mathrm{i}}(G_k)$, since (g_t) is not periodic.

Since $F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k}$, with $F_k = \mathcal{F}_{-\mathrm{i}}(f_t)$, $k = \ldots, -1, 0, 1, \ldots$, is periodic with period N, (i) the Fourier representation of (f_t), $\mathcal{F}_{+\mathrm{i}}(F_k)$, is periodic with period N, as shown above, and (ii), using the remark to (6.1), the sum in (6.23) can be evaluated in arbitrary intervals of length N

$$f_t = \sum_{k=n}^{N+n-1} F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k} \qquad \text{with } n, t = \ldots, -1, 0, 1, \ldots \tag{6.23}$$

$$= \begin{cases} \sum_{k=-(N-1)/2}^{(N-1)/2} F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k} & \text{for odd } N \\ \sum_{k=-(N/2-1)}^{N/2} F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k} & \text{for even } N \end{cases} \tag{6.24}$$

which implies that the sequence $F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k}$ in (6.22,2) can be summed within arbitrary intervals of length N, e.g., in (6.24), within the interval $k = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$, with $k = 0$ in its centre, by assuming that N is odd and then substituting $n = -(N-1)/2$ in (6.23). If the

sums are calculated for arbitrary intervals then negative frequencies are introduced (the Fourier frequencies s_k as defined in (6.21,1) are non-negative): $\ldots, -s_2 = -2/N, -s_1 = -1/N, s_0 = 0, s_1 = 1/N, s_2 = 2/N, \ldots$. For symmetry reasons, a Fourier transform is often plotted against the negative and positive frequencies in the sums (6.24), as demonstrated in Fig. 6.4.

The periodic Fourier transform $(F_k) = (F_{k+nN})$ of (f_t) implies $F_k = F_{k+N} = F_{k-N}$ and, therefore, directly from (6.22,1): $F_k = \overline{F_{-k}} = \overline{F_{N-k}}$ ($\overline{z}$ being, as defined in (6.2,3), the complex-conjugate to z) is obtained. Multiplying both sides of (6.22,1) with the factor 2, the first line in (6.25) is arrived at. Comparing this line with (6.21), the second and third lines are derived.

$$
\begin{aligned}
&2F_k = \frac{2}{N}\sum_{t=0}^{N-1} f_t \cos\bigl(2\pi(k/N)t\bigr) - \frac{2\mathrm{i}}{N}\sum_{t=0}^{N-1} f_t \sin\bigl(2\pi(k/N)t\bigr)\\
&2F_k = F_k + \overline{F_{N-k}} = F_k + \overline{F_{-k}} = A_k - \mathrm{i}B_k\\
&2F_k = \begin{cases} A_k - \mathrm{i}B_k & \textit{for } k \geq 1\\ 2A_0 & \textit{for } k = 0\\ A_{|k|} + \mathrm{i}B_{|k|} & \textit{for } k = \leq -1 \end{cases}\\
&A_k = 2\mathrm{re}(F_k),\ \textit{for } k \neq 0;\ A_m = F_m,\ \textit{for even } N;\ A_0 = F_0\\
&B_k = 2\mathrm{im}(F_k),\ \textit{for } k \neq 0, m;\ B_m = 0,\ \textit{for even } N;\ B_0 = 0
\end{aligned}
\tag{6.25}
$$

A_k in the fourth line is obtained using $A_k = ((1/2)A_k - (1/2)\mathrm{i}B_k) + ((1/2)A_k + (1/2)\mathrm{i}B_k) = F_k + \overline{F_k} = F_k + F_{-k} = 2\mathrm{re}(F_k)$, $\mathrm{re}(z)$ being the real part of a complex number z, as defined in (6.2,1). B_k in the fifth line follows from a similar argument. Applying (6.25), Fourier coefficients A_k and B_k (6.21,2) can be computed from a discrete Fourier transform F_k (6.22,1), and vice-versa. If, however, the F_k are calculated as required in (6.22,4), they then have to be normalised before applying (6.25), as demonstrated in the remarks to Fig. 6.1.

If a sequence $(f_t) = \bigl(f_0, f_1, \ldots, f_{N-2}, f_{N-1}\bigr)$ is concatenated, on its right side, with a sequence $(0, 0, \ldots, 0)$ containing $N' - N$ values which are identically zero, then a sequence $(g_t) = \bigl(g_0 = f_0,\ g_1 = f_1,\ \ldots,\ g_{N-2} = f_{N-2},\ g_{N-1} = f_{N-1},\ g_N = 0, \ldots, g_{N'-1} = 0\bigr)$ with N' values is obtained, i.e., (g_t) is in $L^2(0, N'-1)$. It is said that (f_t) is *zero-padded* to obtain (g_t). From a discrete Fourier transform of a zero-padded sequence as defined in (6.22,4) a Fourier transform pair with frequencies s'_k as in (6.26) and (6.27) is obtained. If $s'_k = k/N'$ in (6.26) and (6.27) is identical with $s_k = k/N$ in (6.22,4), then both sums $F_k = \sum_{t=0}^{N-1} f_t \mathrm{e}^{-\mathrm{i}2\pi(k/N)t}$ and $G_k = \sum_{t=0}^{N'-1} g_t \mathrm{e}^{-\mathrm{i}2\pi(k/N')t}$, contain identical non-zero terms, and, consequently, $G_k = F_k$.

$$
(G_k) = \mathcal{F}_{-\mathrm{i}}\bigl(g_t\bigr) = \sum_{t=0}^{N'-1} g_t \mathrm{e}^{-\mathrm{i}2\pi(k/N')t},\ \text{for } k = \ldots, -1, 0, 1, \ldots \tag{6.26}
$$

$$
(g_t) = \mathcal{F}_{+\mathrm{i}}\bigl(G_k\bigr) = \frac{1}{N'}\sum_{k=0}^{N'-1} G_k \mathrm{e}^{+\mathrm{i}2\pi(t/N')k} \text{ for } t = \ldots, -1, 0, 1, \ldots \tag{6.27}
$$

Consequently, by zero-padding a sequence, its Fourier transform could be calculated using a finer grid of frequencies than would be the case without zero-padding. With, e.g., $N' = 2N$, the distance of the frequencies reduces to $\Delta s'_k = 1/N' = 1/(2N)$, i.e., to half the distance of the Fourier frequencies pertaining to the original sequence, i.e., $\Delta s_k = 1/N$ as defined in (6.21,1). Examples are given in Problem 6.5.

Both pairs of sequences, (f_t) and $(F_k) = F(s_k)$ in (6.22) and (6.24) as well as (g_t) and $(G_k) = G(s_k)$ in (6.26) and (6.27), are Fourier transform pairs. If no distinction is made between $\mathcal{F}_{+\mathrm{i}}(\ )$ and $\mathcal{F}_{-\mathrm{i}}(\ )$, then the type, i.e., $+\mathrm{i}$ or $-\mathrm{i}$, of the Fourier transform will become obvious from the context. In (6.22) and (6.21), as a first convention, t is used for the time and $s_k = t/N$ for the frequencies, and, as a second convention, a Fourier transform pair is written using the same letter: with a small letter when the argument is the time, with a capital letter when the argument is the frequency.

In (6.22), each complex F_k is the sum of N complex multiplications. Since (F_k) is periodic with period N, as proposed in the remarks to (6.22), a Fourier transform is arrived at by means of N^2 complex multiplications, and, therefore, the time required for the calculations increases in relation to N^2, N the number of observations. The increase with the square of N can be reduced without loss in the accuracy of the result by using the fast Fourier transform (FFT) algorithm as introduced in [39]. When this sophisticated algorithm is applied to compute a discrete Fourier transform, the number of multiplications is reduced, depending on how N can be factorised. For example, $N = 140$ (the number of values in the sawtooth-shaped sequence in Fig. 6.1) is the product of the factors $2 \times 2 \times 5 \times 7$. If $N = 2^r$, i.e., the number of values in the sequence to be transformed is a power of two, then the Fourier transform is computed with only $2rN$ complex multiplications, on condition that the FFT algorithm is applied. For example, the discrete Fourier transform of a sequence with $N = 512 = 2^9$ is obtained by way of $2 \times 9 \times 512 = 9216$ complex multiplications, whereas, when computing using (6.22,1) without modifications, $512 \times 512 = 262144$ complex multiplications have to be performed to arrive at the same result. The FFT is described, e.g., in [18]. The FFT algorithm is made available in R as function `fft()`, and applied as demonstrated in the following examples. Experiments with `fft()` are to be performed in Problem 6.6. Favourable factorisations are obtained with `nextn()`.

As a first example, the sequence (g_t) with an odd number of values ($N = 161$), as plotted in Fig. 6.4 (a), is Fourier transformed using `fft()`. `xt <- fft(x,inverse=T)`, `x` a real-valued vector containing (g_t), computes, using the FFT algorithm, the discrete Fourier transform $(G_k) = G(s_k) = \mathcal{F}_{-\mathrm{i}}(g_t)$ as required in (6.22,4) with the result in `xt`.

```
#x is a real R vector with the sequence in Fig. 6.4 (a)
#as available in file /path/fig64.dat
N <- length(x)
```

```
m <- floor(N/2)          #as required in (6.21)
xt <- fft(x, inverse=T) #discrete -i-transform (6.22,4)
```

`xt` is a complex-valued R vector having the same length as `x`. The first value in `xt` is G_0, the discrete Fourier transform for $s_0 = 0$. The next values are $G_1, \ldots, G_m$, pertaining to frequencies $s_1 = 1/N, \ldots, s_m = m/N$, followed by $G_{m+1}, \ldots, G_N$, pertaining to frequencies $-s_m = -m/N, \ldots, -s_1 = -1/N$ being negative. In the above example, the complex-valued vector `xt` resulting from `fft()` contains values

```
...
[79]    0.0239256- 0.49006583i   -0.0142419+ 0.48643952i
[81]    0.0047249- 0.48426303i    0.0047249+ 0.48426303i
[83]   -0.0142419- 0.48643952i    0.0239256+ 0.49006583i
...
```

with indices from 1 through to $N = 161$, whereas, in (6.21,4), the sums are calculated for $k = 0, 1, \ldots, N-1$. From these values, it is concluded that `xt[81]` $= \overline{\texttt{xt[82]}} = \overline{\texttt{xt[161−81+2]}}$, `xt[79]` $= \overline{\texttt{xt[84]}} = \overline{\texttt{xt[161−79+2]}}$, as required in the third line in (6.25).

The same order of values has to be adhered to in vector `x` used as argument for `fft(x,...)`, on condition that the sequence to be transformed is given for negative time points, as demonstrated in Sect. 6.9.8, Figs. 6.33 and 6.34.

The order of the values in `xt` is inherited by its real and imaginary parts as well as by its absolute value and its phase as defined in (6.2), all of which are obtained as real R vectors using `Re(xt)`, `Im(xt)`, `Mod(xt)`, `Arg(xt)`. This is accounted for when, for example using the following R expressions,

```
#Fourier frequencies as defined in (6.21,1) in the same order
#as the values in xt; c() concatenates R vectors,
sk <- c( (0:m)/N,-(m:1)/N )              #odd N
#sk <- c( (0:m)/N,-((m-1):1)/N )         #even N
plot(sk, (Mod(xt)*Mod(xt)), type="n", xlab=" ", ylab=" ")
for(k in 1:N ) { #plot vertical lines in Fig. 6.4
  lines( c(sk[k],sk[k]), c(0.0,(Mod(xt)*Mod(xt))[k]) ) }
#another possible plot
  lines(sk[1:(m+1)],(Mod(xt)*Mod(xt))[1:(m+1)]) #sk >= 0
  lines(sk[(m+2):N],(Mod(xt)*Mod(xt))[(m+2):N]) #sk < 0
  lines(c(sk[1],sk[N]),
        c((Mod(xt)*Mod(xt))[1],(Mod(xt)*Mod(xt))[N])) #gap
#end another possible plot
```

the squared absolute values in the Fourier transform of the sequence (g_t) in Fig. 6.4 (a) are plotted against the frequencies $-s_m, \ldots, -s_1, 0, s_1, \ldots, s_m$ to obtain the spectrum in Fig. 6.4 (b).

As a second example, the sawtooth-shaped sequence in Fig. 6.1 (a), with an even ($N = 140$) number of values, is Fourier transformed as required in (6.21,4) using the R expressions in the remarks to Fig. 6.1 (c). Here, a part of the complex-valued result is given.

```
...
[57] -1.652476e+00+5.085798e+00i -2.289524e-15+6.637337e-15i
[59]  1.384425e-16-1.736015e-16i -3.884569e-15+8.655689e-15i
[61]  1.603875e+00-7.723857e-01i  1.063581e-14-3.017723e-15i
...
[81]  1.603875e+00+7.723857e-01i -2.612836e-15-8.103348e-15i
[83] -2.154693e-17+3.691674e-16i  4.060969e-15-4.772066e-15i
[85] -1.652476e+00-5.085798e+00i -2.582466e-01-2.292002e+00i
...
```

Using the indices of this R vector which run from 1 through to $N = 140$ (and not from 0 through to 139), one obtains, for example, `xfft[57]` $= \overline{\texttt{xfft[85]}} =$ $\overline{\texttt{xfft[140−57+2]}}$, and it is clearly seen that the F_k are identically zero for practical purposes, with two exceptions: (i) $F_{n\times 5}$, $n = 1, 2, \ldots, 13$, the coefficients of the trigonometric oscillation with frequency 5/140 and the pertaining harmonics, (ii) F_{28} and F_{56}, the coefficients for the oscillation with frequency 28/140 and its first harmonic. This finding is in agreement with the spectrum plotted in Fig. 6.2.

Sect. 6.3.1 is summarised as follows. A sequence (f_t) being in $L^2(0, N-1)$ is represented by its finite Fourier series $f_m(t)$ (6.21,3) which converges in mean square to (f_t). $f_m(t)$ is a linear combination of the trigonometric functions (6.9) with Fourier coefficients A_k and B_k. The A_k and B_k are calculated, using (6.25), either from a discrete Fourier transform (F_k) of (f_t) as defined in (6.22,1), or, when (F_k) is calculated as required in (6.22,4), subsequent to normalising (F_k). From $(F_k) = \mathcal{F}_{-\mathrm{i}}(f_t)$ and $(f_t) = \mathcal{F}_{+\mathrm{i}}(F_k)$ in (6.22,3) it is concluded that (f_t) and (F_k) are a Fourier transform pair as required in (6.20,1), since (F_k) can be written as $F(s_k)$, with $s_k = k/N$, the Fourier frequencies as defined in (6.21,1).

In this first case (6.20,1) of a Fourier transform pair, the time is discrete, and, when the sampling interval Δt is set to unity and the first observation is taken at zero time, the observations become a sequence (f_t), $t = 0, 1, 2, \ldots, N-1$. $F(s_k)$, the Fourier transform of (f_t), is a sequence defined for a finite number of discrete frequencies s_k, i.e., the Fourier frequencies as in (6.21,1).

6.3.2 Case 2: Fourier Series

In the second case (6.20,2) of a Fourier transform pair, a real-valued function $f(t)$ with real argument t, being in $L^2(-T/2, T/2)$, is approximated using a linear combination of the orthogonal trigonometric functions (6.14). The frequencies k/T, $k = 1, 2, \ldots$, of these functions are discrete and not bounded, as there is an infinite number of functions in this set.

The functions in (6.14) are orthogonal, but not normalised as concluded from (6.15) and (6.16). This has to be accounted for (as in (6.21) and (6.22) when approximating sequence (f_t) using the functions (6.9)) when $f(t)$ is approximated in $L^2(-T/2, T/2)$ using the functions (6.14) together with Fourier

coefficients A_k and B_k which are obtained, in (6.28,2), by applying (6.8). A_0, for example, is normalised with $(2/T)$ in (6.28,2) though $(1/T)$ would be correct as deduced from (6.15). This error is compensated in (6.28,4).

Let $f(t)$, t real, be a real-valued function in $L^2(-T/2, T/2)$. Then:

1. *$s_k = k/T$, with $k = 0, 1, \ldots$, are the Fourier frequencies of $f(t)$.*
2. *$A_k = (2/T)\int_{-T/2}^{T/2} f(t)\cos(2\pi s_k t)\mathrm{d}t$, for $k = 0, 1, \ldots$, and $B_k = (2/T)\int_{-T/2}^{T/2} f(t)\sin(2\pi s_k t)\mathrm{d}t$, for $k = 1, 2, \ldots$, are the Fourier coefficients of $f(t)$.*
3. *$f_M(t) = A_0/2 + \sum_{k=1}^{M}\big(A_k\cos(2\pi s_k t) + B_k\sin(2\pi s_k t)\big)$ is the Fourier partial sum pertaining to $f(t)$, $0 < M < \infty$ and integer.*
4. *$f_{M\to\infty}(t) = A_0/2 + \sum_{k=1}^{\infty}\big(A_k\cos(2\pi s_k t) + B_k\sin(2\pi s_k t)\big)$ is the Fourier series pertaining to $f(t)$ with the property $f_{M\to\infty}(t) = f(t)$. If $f(t)$ jumps in t_i then $f_{M\to\infty}(t)$ converges to $\big(f(t_i^-) + f(t_i^+)\big)/2$, $f(t_i^-)$, $f(t_i^+)$ the left and right limits of $f(t)$ in t_i.*
5. *If (f_t) is periodic with period T then $f_{M\to\infty}(t) = f(t)$ also at the exterior of $L^2(-T/2, T/2)$, i.e., for all t.*

(6.28)

The Fourier series (6.28,4) converges in $L^2(-T/2, T/2)$ to $f(t)$, a result obtained in the basic lectures in calculus.

Most functions $f(t)$ have a finite norm in $-T/2 \le t \le T/2$ and are, therefore, in $L^2(-T/2, T/2)$. For example, a function is in $L^2(-T/2, T/2)$, even though it has a finite number of jumps (i.e., a finite number of discontinuities) and/or is not smooth in a finite number of points (i.e., has a finite number of discontinuities in its first derivative) in $-T/2 \le t \le T/2$. Examples are plotted in Fig. 6.7. For applications, it is important to know that most functions are well-behaved in the sense that their Fourier series converges.

The Fourier series in (6.28,4) can be written in complex notation. Applying identities $\mathrm{e}^{\mathrm{i}x} = \cos x + \mathrm{i}\times\sin x$, $\cos x = (1/2)(\mathrm{e}^{\mathrm{i}x} + \mathrm{e}^{-\mathrm{i}x})$ and $\sin x = (1/2\mathrm{i})(\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x})$ to sequences A_k and B_k in (6.28,2), the Fourier transform pair $f(t)$ and (F_k) in (6.29) and (6.30) is obtained.

$$f(t) = \mathcal{F}_{+\mathrm{i}}(F_k) = \mathcal{F}_{+\mathrm{i}}\big(F(s_k)\big) = \sum_{k=-\infty}^{\infty} F_k \mathrm{e}^{\mathrm{i}2\pi s_k t} \quad \frac{-T}{2} \le t \le \frac{T}{2} \tag{6.29}$$

$$(F_k) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = \frac{1}{T}\int_{-T/2}^{T/2} f(t)\mathrm{e}^{-\mathrm{i}2\pi s_k t}\mathrm{d}t \tag{6.30}$$

$$= \begin{cases} (1/2)\big(A_k - \mathrm{i}B_k\big) & k \ge 1 \\ (1/2)A_0 & k = 0 \\ (1/2)\big(A_{|k|} + \mathrm{i}B_{|k|}\big) & k \le -1 \end{cases} \tag{6.31}$$

The Fourier transform pair in (6.29) and (6.30) is a more symmetric formulation of the convergence of the Fourier series in (6.28,4), similar to the Fourier transform pair in (6.22,3) which is a more symmetric formulation of

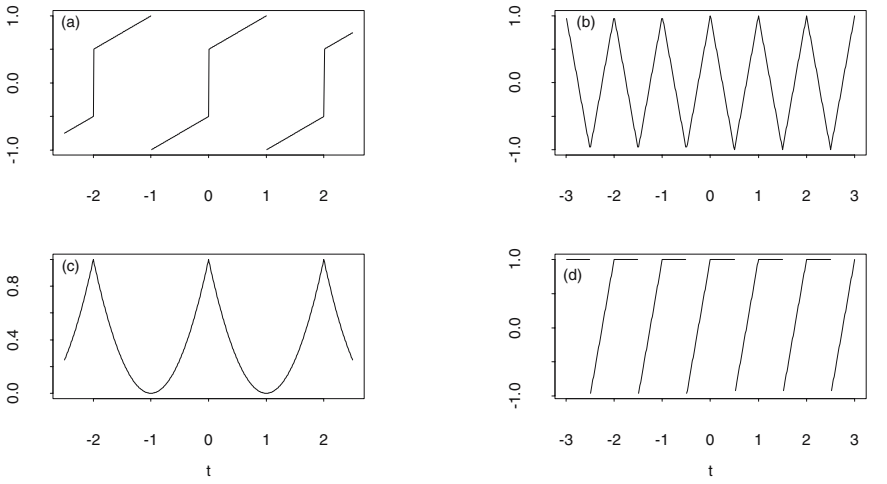


Fig. 6.7. Periodic functions of the real argument t which are reconcilable with (6.28), i.e., which are well-behaved in the sense that their Fourier series exists.

the convergence of the finite Fourier series in (6.21,3). The mathematically wishful symmetry in the formulation is, however, counterbalanced with the introduction of negative frequencies.

A periodic function with period T can always be defined in the interval $-T/2 \le t \le T/2$, as concluded from the remarks to (6.1). This interval is required in definitions (6.28) as well as (6.29) and (6.30). For examples, the functions in Fig. 6.7 plots (a) and (c) are defined in $-1 \le t \le 1$, whereas $-1/2 \le t \le 1/2$ lends itself for the definition of the functions in Fig. 6.7, plots (b) and (d).

When a function $f(t)$ to be transformed is symmetric, the calculations of the sums and integrals in (6.28) as well as (6.29) and (6.30) becomes easier. A catalogue of symmetries favouring the evaluation of sums and integrals is given in (6.32).

Let $f(x)$ be a real-valued function. Then:
1. *$f(x)$ is an even function on condition that $f(x) = f(-x)$.*
2. *$f(x)$ is an odd function on condition that $f(x) = -f(-x)$.*
3. *$f(x)$ is a mixed function if $f(x)$ is neither even nor odd.*
4. *A product of either two even or two odd functions is even.*
5. *A product of an even and an odd function is odd.*
6. *Every $f(x)$ can be split into an even part* $\mathrm{even}(f(x))$ *and an odd part* $\mathrm{odd}(f(x))$.

(6.32)

(6.32,4,5) directly follow from definitions (6.32,1,2). (6.32,6) is also obtained from (6.32,1,2): $f(x) = \mathrm{even}(f(x)) + \mathrm{odd}(f(x))$ and $f(-x) = \mathrm{even}(f(-x)) + \mathrm{odd}(f(-x)) = \mathrm{even}(f(-x)) - \mathrm{odd}(f(x))$ imply $\mathrm{even}(f(x)) = \big(f(x)+f(-x)\big)/2$

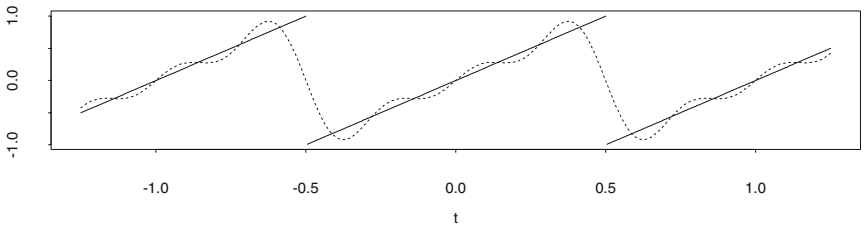


Fig. 6.8. Function $f(t) = 2t$, $-1/2 \le t \le 1/2$, and periodically repeated at the exterior of this interval, i.e., $f(t) = f(t+n)$, $n = \ldots, -1, 0, 1, \ldots$ (solid line) together with the pertaining Fourier series for $M = 3$.

and $\mathrm{odd}(f(x)) = \big(f(x) - f(-x)\big)/2$. In the case of complex-valued functions, both the real and the imaginary part can be even, odd or mixed, and therefore more cases are distinguished to describe the symmetry of a product of complex-valued functions [18].

For example, the Fourier series pertaining to function $f(t) = 2t$, $-1/2 \le t \le 1/2$, $f(t) = f(t+n)$, $n = \ldots, -1, 0, 1, \ldots$, is computed using its periodic and symmetric properties. $f(t)$ is plotted with solid lines in Fig. 6.8.

The A_k and B_k in the Fourier series of $f(t)$ are calculated as required in (6.28,2) by taking advantage of $f(t)$ being odd (6.32,2). The integrals for A_k

$$A_k = \frac{2}{1}\int_{-1/2}^{0} 2t\cos(2\pi\frac{k}{1}t)\mathrm{d}t + \frac{2}{1}\int_{0}^{1/2} 2t\cos(2\pi\frac{k}{1}t)\mathrm{d}t = 0$$

become identically zero for $k = 0, 1, \ldots$, since $f(x)\cos(x) = -f(-x)\cos(-x)$ as implied by (6.32,5). No separate calculation for A_0 is required, as the error in (6.28,2) is compensated in (6.28,4). The second $=$ in the following calculation for B_k

$$B_k = \frac{2}{1}\int_{-1/2}^{1/2} 2t\sin(2\pi\frac{k}{1}t)\mathrm{d}t = 8\int_{0}^{1/2} t\sin(2\pi\frac{k}{1}t)\mathrm{d}t = -\frac{2}{\pi k}\cos(\pi k)$$

is obtained since $f(t)\sin(t) = f(-t)\sin(-t)$, and the third $=$ from an evaluation of the expression $\big[(-t/(2\pi k))\cos(2\pi kt) + (1/(4\pi^2k^2))\sin(2\pi kt)\big]_0^{1/2}$, resulting from integration by parts. Thereafter, the Fourier series is arrived at by substituting B_k in (6.28,4).

$$\begin{aligned} f_M(t) &= \sum_{k=1}^{M} -\frac{2}{\pi k}\cos(\pi k)\sin(2\pi kt) \qquad M \to \infty \\ &= \frac{2}{\pi}\Big(\sin(2\pi t) - (1/2)\sin(4\pi t) + (1/3)\sin(6\pi t) - \ldots \end{aligned}$$

The approximation as obtained above is plotted, for $M = 3$, with a broken line in Fig. 6.8. This approximation comes closer to $f(t)$ with increasing

M, and convergence in mean square as required in (6.28) is arrived at for $M \to \infty$. Further examples are given in Problems 6.7 and 6.8.

Sect. 6.3.2 is summarised as follows. In both cases of Fourier transforms (6.20,1,2), a sequence (f_t) or function $f(t)$, both defined in a closed interval and possibly periodic, are represented as a linear combination of trigonometric functions. If $t = \ldots, -1, 0, 1, \ldots$ and (f_t) is in $L^2(0, N-1)$, then the finite Fourier series (6.21,3), being a linear combination of N (i.e., a finite number) of trigonometric functions in basis (6.9), converges in $L^2(0, N-1)$ to (f_t); whereas, if t is real and $f(t)$ in $L^2(-T/2, T/2)$ then the Fourier series (6.28,4), being a linear combination of an infinite number of trigonometric functions in basis (6.14), converges in $L^2(-T/2, T/2)$ to $f(t)$. These results are summarised as Fourier transform pairs proposed in (6.22), as well as in (6.29) and (6.30), both having discrete Fourier frequencies s_k.

However, the following distinction should be made. In the case of the transform pair as defined in (6.22,3), (F_k) is periodic with period N and is therefore completely determined, i.e., known for $k = \ldots, -1, 0, 1, \ldots$, on condition that N complex-valued F_k are calculated as Fourier transform of (f_t), $t = 0, 1, \ldots, N-1$. Consequently, (f_t) results from the +i-transform as a sum of N terms, i.e., $\mathcal{F}_{+\mathrm{i}}(F_k)$ converges in mean square for $t = 0, 1, \ldots, N-1$ to (f_t). In the case of the transform pair as defined in (6.29) and (6.30), in contrast, (F_k) will be completely determined on condition that an infinite number of complex F_k are calculated as Fourier transform of $f(t)$. Consequently, in the +i-transform, $f(t)$ is obtained as the sum of an infinite number of terms, i.e., $\mathcal{F}_{+\mathrm{i}}(F_k)$ converges in mean square for $-T/2 \le t \le T/2$ to $f(t)$. In both cases however, $\mathcal{F}_{+\mathrm{i}}(F_k)$ is periodic because the trigonometric functions in either (6.9) or (6.14) are periodic with periods N or T. For that reason, convergence is obtained even at the exterior of the intervals $0, 1, \ldots, N-1$ or $-T/2 \le t \le T/2$ provided that (f_t) or $f(t)$ are periodic with periods N or T, as shown in Figs. 6.3 and 6.8.

6.3.3 Case 3: Fourier Integrals

Does a real-valued function $f(t)$ in $L^2(-\infty, \infty)$, t real, have a Fourier transform, as postulated in (6.20,3)? A plausible answer to this question is given by deducing case (6.20,3) from (6.20,2) and then letting $T \to \infty$. In this manoeuvre, $f(t)$ in (6.28) as well as in (6.29) and (6.30) is replaced with a function $f_T(t) = f(t)$ in $-T/2 \le t \le T/2$ and being periodic with period T, $f_T(t + nT)$, at the exterior of this interval. For $T \to \infty$, $f_T(t)$ has a Fourier transform as defined in (6.30) on condition that, in the limit for $T \to \infty$, the integrals for the A_k and B_k in the Fourier series (6.28,4) do exist, and, therefore, the Fourier series converges in $L^2(-T/2, T/2)$ to $f_T(t)$. When T increases, the distance $\Delta s_k = s_{k+1} - s_k$ decreases, s_k being the Fourier frequencies defined in (6.28,1) and used in (6.29) and (6.30). This suggests that, for $T \to \infty$, the set of the discrete s_k becomes the set of all points s in the real frequency domain.

Hence, the Fourier integrals in (6.35) can be made plausible in the following steps. The integral in (6.31) is substituted for F_k in the sum in (6.29), and, since $f(t) = f_T(t)$ in $-T/2 \le t \le T/2$, (6.33) is obtained. (6.34) follows since $1/T = s_k - s_{k-1} = \Delta s_k$. If $T \to \infty$ then $\Delta s_k \to 0$ and, believably, the series in (6.33) becomes the first integral in (6.35).

$$f(t) = f_T(t) \;=\; \sum_{k=-\infty}^{\infty} \left(\frac{1}{T} \int_{-T/2}^{T/2} f(t) \mathrm{e}^{-\mathrm{i}2\pi s_k t} \mathrm{d}t \right) \mathrm{e}^{\mathrm{i}2\pi s_k t} \tag{6.33}$$

$$f(t) = f_T(t) \;=\; \sum_{k=-\infty}^{\infty} \left(\int_{-T/2}^{T/2} f(t) \mathrm{e}^{-\mathrm{i}2\pi s_k t} \mathrm{d}t \right) \mathrm{e}^{\mathrm{i}2\pi s_k t} \Delta s_k \tag{6.34}$$

$$f(t) = \int_{-\infty}^{\infty} F(s) \mathrm{e}^{\mathrm{i}2\pi s t} \mathrm{d}s, \text{ with } F(s) = \int_{-\infty}^{\infty} f(t) \mathrm{e}^{-\mathrm{i}2\pi s t} \mathrm{d}t \tag{6.35}$$

In (6.35), $F(s) = \mathcal{F}_{-\mathrm{i}}(f(t))$ and $f(t) = \mathcal{F}_{+\mathrm{i}}(F(s))$ are a Fourier transform pair. (6.35) is derived, as is (6.28,4), in the basic lectures in calculus, provided that $f(t)$ has the properties required in (6.36).

$f(t)$ is a function defined for all real t. If
$\| f(t) \|^2 = \int_{-\infty}^{\infty} \big(f(t)\big)^2 \mathrm{d}t < \infty$, *i.e., if $f(t)$ is in $L^2(-\infty, \infty)$,* (6.36)
then the integrals in (6.35) converge in $L^2(-\infty, \infty)$.

Many functions dealt with in applications are in $L^2(-\infty, \infty)$ as required in (6.36): the functions in Fig. 6.9, for example, decay rapidly enough to 0 for $-\infty \leftarrow t$ and $t \to \infty$ such that they are in $L^2(-\infty, \infty)$.

Prior to demonstrating (6.35) with the Fourier transform pairs in Figs. 6.9 and 6.10, it is helpful to derive, applying symmetries (6.32) and definitions (6.37,1,2), the properties as proposed in (6.37,3,4,5,6).

Let $\mathrm{even} f(t)$ *and* $\mathrm{odd} f(t)$ *be the even and odd parts as defined in (6.32,6) of a function $f(t)$ in $L^2(-\infty, \infty)$. Then:*

1. $\int_{t=0}^{\infty} \mathrm{even}\big(f(t)\big) \cos(2\pi s t) \mathrm{d}t = \mathcal{F}_{\cos}\big(f(t)\big)$ *is the Fourier cosine-transform,*
2. $\int_{t=0}^{\infty} \mathrm{odd}\big(f(t)\big) \sin(2\pi s t) \mathrm{d}t = \mathcal{F}_{\sin}\big(f(t)\big)$ *the Fourier sine-transform of $f(t)$, and the following hold:*
3. $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = \mathrm{re}\big(F(s)\big) - \mathrm{im}\big(F(s)\big)\mathrm{i} = 2\mathcal{F}_{\cos}\big(f(t)\big) - 2\mathcal{F}_{\sin}\big(f(t)\big)\mathrm{i}$
4. $\overline{F(s)} = \mathcal{F}_{+\mathrm{i}}\big(f(t)\big) = \mathrm{re}\big(F(s)\big) + \mathrm{im}\big(F(s)\big)\mathrm{i} = 2\mathcal{F}_{\cos}\big(f(t)\big) + 2\mathcal{F}_{\sin}\big(f(t)\big)\mathrm{i}$
5. $f(t) = \mathcal{F}_{+\mathrm{i}}\big(F(s)\big) = \mathrm{re}\big(f(t)\big) + \mathrm{im}\big(f(t)\big)\mathrm{i} = 2\mathcal{F}_{\cos}\big(F(s)\big) + 2\mathcal{F}_{\sin}\big(F(s)\big)\mathrm{i}$
6. $\overline{f(t)} = \mathcal{F}_{-\mathrm{i}}\big(F(s)\big) = \mathrm{re}\big(f(t)\big) - \mathrm{im}\big(f(t)\big)\mathrm{i} = 2\mathcal{F}_{\cos}\big(F(s)\big) - 2\mathcal{F}_{\sin}\big(F(s)\big)\mathrm{i}$

(6.37)

The derivation of (6.37,3,4,5,6) is straightforward. For example, the second = in decomposition (6.37,4) is arrived at in Problem 6.9 and, thereafter, the result follows from the definitions of the Fourier cosine- and sine-transforms. Definitions and properties (6.37) are formulated for Fourier integrals (6.35),

i.e., for case (6.20,3), they however also apply to Fourier transform pairs (6.20,1), (6.20,2), and (6.20,4).

(6.37) implies that even functions have a real Fourier transform, odd functions a pure imaginary one and mixed functions a complex transform. Hence, if e.g., $f(t)$ is an even function, $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = \mathcal{F}_{+\mathrm{i}}\big(f(t)\big) = \mathcal{F}\big(f(t)\big) = 2\mathcal{F}_{\cos}\big(f(t)\big)$ is obtained for its Fourier transform $F(s)$.

For instance, the real-valued functions $f(t)$ in the left column in Fig. 6.9 (plots a,d,g) are Fourier transformed to the complex-valued functions $F(s)$, whose real and imaginary parts are plotted in the centre (plots b,h) and right columns (plots f,i) in this figure. All $f(t)$ as well as $\mathrm{re}\big(F(s)\big)$ and $\mathrm{im}\big(F(s)\big)$ are in $L^2(-\infty,\infty)$, as required in (6.36), and a Fourier $-$i-transform is performed: from $f(t)$ in the time domain the pertaining $F(s)$ in the frequency domain is obtained as required in (6.35).

As a first example, the *rectangle function* $h(t)$ with height a and width $2b$ as defined in (6.38) and plotted in Fig. 6.9 (a) (for $a = 1$ and $b = 1/2$ and therefore with unit integral), is Fourier transformed to $H(s)$ in (6.39).

$$h(t) = \begin{cases} a & \text{for } -b \le t \le b \\ 0 & \text{for } |t| > b \end{cases} \tag{6.38}$$

$$\begin{aligned} H(s) = \mathcal{F}_{-\mathrm{i}}\big(h(t)\big) &= \int_{-\infty}^{\infty} a\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t = \int_{-b}^{b} a\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\ &= 2a\int_0^b \cos(2\pi st)\mathrm{d}t = 2a\left[\frac{1}{2\pi s}\sin(2\pi st)\right]_{t=0}^{t=b} = 2ab\frac{\sin(2\pi bs)}{2\pi bs} \end{aligned} \tag{6.39}$$

If x is substituted for the argument s in $H(s)$ and the factor $2ab$ is omitted, then the $(\sin x)/x$ *function* is arrived at. The $(\sin x)/x$ function has its properties enumerated in (6.40)

1. $(2\pi bx)^{-1}\sin(2\pi bx) = 0$ *for* $bx = \ldots, -2, -1, 1, 2, \ldots$
2. $\lim_{x\to 0}(2\pi bx)^{-1}\sin(2\pi bx) = 1$
3. $(2\pi bx)^{-1}\sin(2\pi bx)$ *is not the derivative of a simple function*
4. $\int_{-\infty}^{\infty}\big((2\pi bx)^{-1}\sin(2\pi bx)\big)\mathrm{d}x = 1/|2b|$,

(6.40)

which are derived in Problem 6.10. (6.40,2) asserts the existence of the $(\sin x)/x$ function for $x = 0$, difficulties arising from (6.40,3) are circumvented by calculating definite integrals using a series expansion, and, due to (6.40,4), the $(\sin x)/x$ function is in $L^2(-\infty,\infty)$ which implies that also $H(s)$ in (6.39) is in $L^2(-\infty,\infty)$. Consequently, (6.36) and (6.35) are applied to obtain $h(t)$ as Fourier $+$i-transform from $H(s)$:

$$\begin{aligned} h(t) = \mathcal{F}_{+\mathrm{i}}\big(H(s)\big) &= \int_{-\infty}^{\infty} H(s)\mathrm{e}^{+\mathrm{i}2\pi ts}\mathrm{d}s = \int_{-\infty}^{\infty} 2ab\frac{\sin(2\pi bs)}{2\pi bs}\mathrm{e}^{+\mathrm{i}2\pi ts}\mathrm{d}s \\ &= 2ab\int_{-\infty}^{\infty}\frac{\sin(2\pi bs)\cos(2\pi st)}{2\pi bs}\mathrm{d}s \qquad \text{with (6.32,4)} \end{aligned}$$

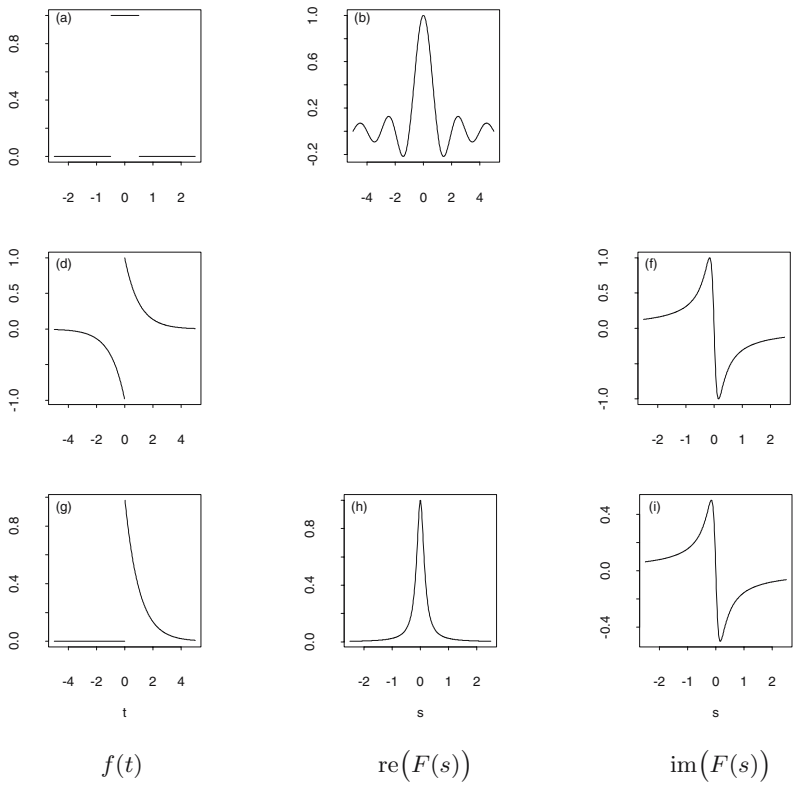


Fig. 6.9. Real and imaginary parts of Fourier transforms $F(s)$ of functions: (i) $f(t) = 1$ for $-1/2 \le t \le 1/2$, $f(t) = 0$ at the exterior of this interval (plots a,b); (ii) $f(t) = \mathrm{e}^{-t}$ for $t \ge 0$, $f(t) = -\mathrm{e}^{t}$ for $t < 0$ (plots d,f); (iii) $f(t) = \mathrm{e}^{-t}$ for $t \ge 0$, $f(t) = 0$ for $t < 0$ (plots g,h,i)

.

$$
\begin{aligned}
&= 2ab \int_{-\infty}^{\infty} \left(\frac{\sin(2\pi bs - 2\pi st)}{4\pi bs} + \frac{\sin(2\pi bs + 2\pi st)}{4\pi bs} \right) \mathrm{d}s \\
&= 2ab \int_{-\infty}^{\infty} \left(\frac{b-t}{2b} \frac{\sin(2\pi(b-t)s)}{2\pi(b-t)s} + \frac{b+t}{2b} \frac{\sin(2\pi(b+t)s}{2\pi(b+t)s)} \right) \mathrm{d}s \\
&= \frac{a(b-t)}{2|b-t|} + \frac{a(b+t)}{2|b+t|} = \begin{cases} a & \text{for } -b \le t \le b \\ 0 & \text{for } |t| > b \end{cases} \qquad \text{with (6.40,4)} \qquad (6.41)
\end{aligned}
$$

and thus, $h(t)$ and $H(s)$ in (6.38) and (6.39) are a Fourier transform pair as required in (6.35). This example demonstrates that a harmless function with two jump discontinuities can have a Fourier transform that does not jump and is smooth but has the remarkable properties of the $(\sin x)/x$ function, enumerated in (6.40).

As a second example, the Fourier transform $F(s)$ of the negative exponential function $f(t)$ in (6.42) is calculated. $f(t)$ is an odd function and, therefore, in its transform $F(s) = \mathcal{F}_{-\mathrm{i}}(f(t))$ remains only the imaginary part, as concluded from (6.37). $F(s)$ is calculated as Fourier sine-transform applying (6.37,2), or, with less effort, as exponential transform. In this case, $F(s)$ in (6.43) is the imaginary part of two times the integral for $t > 0$.

$$f(t) = \begin{cases} \mathrm{e}^{-at} & \text{for } t \geq 0 \\ -\mathrm{e}^{at} & \text{for } t < 0 \end{cases} \tag{6.42}$$

$$\begin{aligned} F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big) &= \int_{-\infty}^{0} -\mathrm{e}^{at}\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t + \int_{0}^{\infty} \mathrm{e}^{-at}\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\ &= \mathrm{im}\left(2\int_{0}^{\infty} \mathrm{e}^{-at}\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t\right) = 2\mathrm{im}\int_{0}^{\infty} \mathrm{e}^{-t(a+\mathrm{i}2\pi s)}\mathrm{d}t \qquad (6.43) \\ &= 2\mathrm{im}\left[\frac{1}{a+\mathrm{i}2\pi s}(-1)\mathrm{e}^{-t(a+\mathrm{i}2\pi s)}\right]_{t=0}^{t=\infty} \\ &= 2\mathrm{im}\left(\frac{1}{a+\mathrm{i}2\pi s}\right) = 2\mathrm{im}\left(\frac{(a-\mathrm{i}2\pi s)}{a^2+(2\pi s)^2}\right) = \frac{-\mathrm{i}4\pi s}{a^2+(2\pi s)^2} \qquad (6.44) \end{aligned}$$

$f(t)$ and $F(s)$ are plotted, for $a = 1$, in Fig. 6.9, (d) and (f).

As a third example, the Fourier transform $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ of the function $f(t)$ in (6.45) is calculated. $f(t)$ is mixed, which implies that $F(s)$ has a real and an imaginary part. The integration is performed as above, for the second example, however, now $F(s)$ is the full complex integral, not only its imaginary part.

$$f(t) = \begin{cases} \mathrm{e}^{-at} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases} \tag{6.45}$$

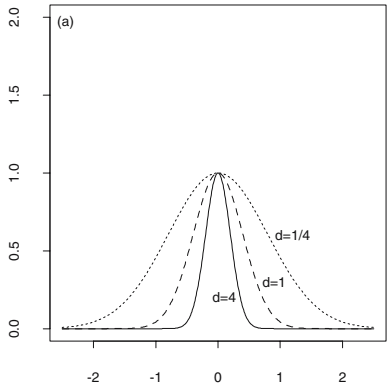
$$F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = \int_{0}^{\infty} \mathrm{e}^{-at}\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t = \frac{(a-\mathrm{i}2\pi s)}{a^2+(2\pi s)^2} \tag{6.46}$$

$f(t)$ and $F(s)$ are plotted for $a = 1$ in Fig. 6.9, (g), (h) and (i).

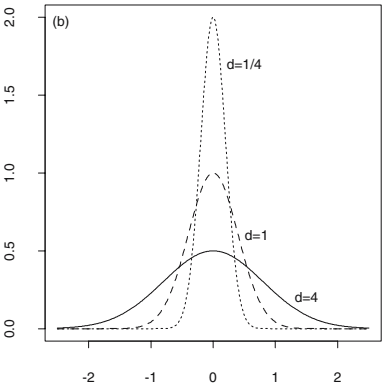
As a last example, the Gaussian function $f(t) = \mathrm{e}^{-d\pi t^2}$, $d > 0$, is Fourier transformed. $f(t)$ is even and its Fourier transform $F(s)$ is obtained as function $G(u, r)$ of arguments u and r,

$$\begin{aligned} \frac{1}{2}F(s) = \frac{1}{2}\mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = \frac{1}{2}\mathcal{F}_{-\mathrm{i}}\big(\mathrm{e}^{-\pi dt^2}\big) &= \int_{0}^{\infty} \mathrm{e}^{-\pi dt^2}\cos(2\pi st)\mathrm{d}t \\ = \int_{0}^{\infty} \mathrm{e}^{-ut^2}\cos(rt)\mathrm{d}t = G(u,r) \qquad & \begin{matrix} u = d\pi \\ r = 2\pi s \end{matrix} \end{aligned}$$

subsequent to substituting $u = d\pi$ and $r = 2\pi s$. The derivative of $G(u, r)$ with respect to r is integrated by parts to obtain a first order linear differential equation, which is thereafter solved using c_2 as obtained in Problem 1.12.



$f(t) = \mathrm{e}^{-d\pi t^2}$



$F(s) = \frac{1}{\sqrt{d}}\mathrm{e}^{-\pi\frac{1}{d}s^2}$

Fig. 6.10. There are functions which are their own Fourier transforms. An example is the Gaussian function $f(t) = \mathrm{e}^{-\pi dt^2}$ with $d = 1$. Plots of the transform pairs $f(t)$ and $F(s)$ for $d = 1$, $d = 1/4$ and $d = 4$.

$$\frac{\mathrm{d}}{\mathrm{d}r}G(u,r) = \int_0^\infty \mathrm{e}^{-ut^2}(-t)\sin(rt)\mathrm{d}t$$

$$= \left(\frac{1}{2u}\mathrm{e}^{-ut^2}\sin(rt)\Big|_0^\infty - \frac{r}{2u}\int_0^\infty \mathrm{e}^{-ut^2}\cos(rt)\mathrm{d}t\right) = -\frac{r}{2u}G(u,r)$$

$$\frac{1}{G(u,r)}\mathrm{d}G = -\frac{r}{2u}\mathrm{d}r \qquad \ln G(u,r) = -\frac{r^2}{4u} + c_1 \qquad G(u,r) = c_2\mathrm{e}^{-r^2/(4u)}$$

$$c_2 = G(u,0) = \int_0^\infty \mathrm{e}^{-ut^2}\mathrm{d}t = \frac{\sqrt{\pi}}{2\sqrt{u}}$$

$$F(s) = 2G(u,r) = \frac{\sqrt{\pi}}{\sqrt{u}}\mathrm{e}^{-r^2/(4u)} = \frac{1}{\sqrt{d}}\mathrm{e}^{-\pi d^{-1}s^2} \tag{6.47}$$

Parameter d in $f(t)$ and $F(s)$ (in $F(s)$ its inverse $1/d$ is multiplied with s^2) scales either the time axis t with $\sqrt{d}$ or the frequency axis s with $1/\sqrt{d}$. If d increases then $f(t)$ becomes narrower and the area under $f(t)$ decreases; if, however, d decreases then $F(s)$ becomes narrower and higher, with the area under $F(s)$ remaining constant. For $d = 1$, $f(t) = \mathrm{e}^{-\pi t^2}$ and $F(s) = \mathrm{e}^{-\pi s^2}$ are obtained. Consequently, $f(t)$ for $d = 1$ is a function which is its own Fourier transform, as demonstrated in Fig. 6.10.

Fourier transform pairs as defined in (6.35) (examples are plotted in Figs. 6.9 and 6.10) have the properties proposed in (6.48) which are derived in Sect. 6.9.4. However, properties (6.48,1,. . .,7) apply to Fourier transform pairs for all cases in (6.20). Derivations are straightforward, e.g., (6.48,7) is obtained in (6.145) for case (6.20,4).

$$
\begin{aligned}
&1.\ F(0) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}2\pi 0t}\mathrm{d}t = \int_{-\infty}^{\infty} f(t)\mathrm{d}t \\
&2.\ F(-s) = \overline{F(s)} \\
&3.\ \mathcal{F}_{-\mathrm{i}}\big(f(at)\big) = (1/|a|)F(s/a),\ \textit{for } a \neq 0 \\
&4.\ \mathcal{F}_{-\mathrm{i}}\big(f(t-b)\big) = F(s)\mathrm{e}^{-\mathrm{i}2\pi bs} \\
&5.\ h(t) = af(t) + bg(t)\ \textit{implies}\ H(s) = aF(s) + bG(s) \\
&6.\ h(t) = f(t) * g(t)\ \textit{implies}\ H(s) = F(s)G(s) \\
&7.\ h(t) = f(t)g(t)\ \textit{implies}\ H(s) = F(s) * G(s)
\end{aligned}
\tag{6.48}
$$

In (6.48,3), a stretching of the time dimension is related to a shrinking of the frequency dimension, and vice-versa, as demonstrated using the Gaussian function in Fig. 6.10. In this example, $\sqrt{d}$ ($\sqrt{d}\times t\times\sqrt{d}\times t = d\times t^2$) substitutes a in (6.48,3). A more symmetric version of (6.48,3) is given in (6.98).

In (6.48,4), a translation of $f(t)$ produces changes in the phase (6.2,5) of its transform $F(s)$, whereas the absolute value (6.2,4) of $F(s)$ remains unchanged. When, for example, (6.48,4) is applied to the rectangle function (6.38), a result is obtained that is identical to the one arrived at in Problem 6.11. In Problem 6.11, the Fourier transform of the displaced rectangle function $f(t) = a$, for $c - b \leq t \leq c + b$, and $f(t) = 0$, for t at the exterior of this interval, is arrived at by direct integration.

Both convolution theorems (6.48,6,7) are often used, as demonstrated in Problem 6.12 as well as in (6.57), (6.75), (6.94) and (6.113).

6.3.4 Case 4: Fourier Series with t and s Reversed

In Sects. 6.3.1, 6.3.2, and 6.3.3, Fourier transform pairs are introduced for the cases (6.20,1,2,3). The last case remains: $t = \ldots, -1, 0, 1, \ldots$, (f_t) is a sequence in $L^2(-\infty, \infty)$, s is real, $F(s)$ is in $L^2(-1/2, 1/2)$ and periodic with period 1, with the Fourier transform pair as proposed in (6.49) and (6.50):

$$
(f_t) = \mathcal{F}_{+\mathrm{i}}\big(\mathcal{F}_{-\mathrm{i}}(f_t)\big) = \int_{-1/2}^{1/2}\Big(\sum_{u=-\infty}^{\infty} f_u\mathrm{e}^{-\mathrm{i}2\pi su}\Big)\mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}s \tag{6.49}
$$

$$
F(s) = \mathcal{F}_{-\mathrm{i}}\Big(\mathcal{F}_{+\mathrm{i}}\big(F(s)\big)\Big) = \sum_{t=-\infty}^{\infty}\Big(\int_{-1/2}^{1/2} F(r)\mathrm{e}^{\mathrm{i}2\pi rt}\mathrm{d}r\Big)\mathrm{e}^{-\mathrm{i}2\pi st} \tag{6.50}
$$

with t, u integer and s, r real

(6.49) is derived by evaluating the integral of the sum as a sum of integrals. This is possible for the reason that (f_t) is required to be in $L^2(-\infty, \infty)$,

$$
\begin{aligned}
(f_t) = \mathcal{F}_{+\mathrm{i}}\big(\mathcal{F}_{-\mathrm{i}}(f_t)\big) &= \int_{-1/2}^{1/2}\Big(\sum_{u=-\infty}^{\infty} f_u\mathrm{e}^{-\mathrm{i}2\pi su}\Big)\mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}s \\
&= \sum_{u=-\infty}^{\infty} f_u\int_{-1/2}^{1/2}\mathrm{e}^{-\mathrm{i}2\pi su}\mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}s = \sum_{u=-\infty}^{\infty} f_u\int_{-1/2}^{1/2}\mathrm{e}^{-\mathrm{i}2\pi(u-t)s}\mathrm{d}s
\end{aligned}
$$

$$= \sum_{u=-\infty}^{\infty} f_u \frac{\sin\big(\pi(u-t)\big)}{\pi(u-t)} \qquad t, u = \ldots, -1, 0, 1, \ldots$$

$$= \sum_{u=-\infty}^{\infty} f_u \begin{cases} 1, & \text{for } u = t \\ 0, & \text{for } u \neq t \end{cases} \qquad t, u = \ldots, -1, 0, 1, \ldots \tag{6.51}$$

and (6.51) is arrived at using the properties (6.40) of the $(\sin x)/x$ function.

(6.50) is obtained assuming that a real-valued function $G(s)$, s real, is in $L^2(-1/2, 1/2)$ and periodic with period 1. $G(s)$ shares these properties with (i) function $F(s)$ as required above as a precondition for (6.49) and (6.50) and (ii) function $f(t)$ as required in (6.28) (with $T = 1$). Consequently, $G(s)$ has a Fourier series that converges as proposed in (6.28,4), implying that also $F(s)$ has a convergent Fourier series. The Fourier series pertaining to $F(s)$ is the infinite sum for $t = \ldots, -1, 0, 1, \ldots$ over the integrals $\int_{-1/2}^{1/2} F(r)\mathrm{e}^{\mathrm{i}2\pi rt}\mathrm{d}r$ in (6.50). This is obtained with $T = 1$ from (6.29) and (6.30).

Above, $F(s)$ in (6.50) is associated with $f(t)$ in (6.28) although the roles that time and frequency play in the Fourier transform pair (6.49) and (6.50) are interchanged when compared with the Fourier series (6.28) or the more symmetric Fourier transform pair (6.29) and (6.30): in (6.49) and (6.50), the time $t = \ldots, -1, 0, 1, \ldots$ is discrete and the frequency s is real, whereas, in (6.29) and (6.30), the time t is real and the frequency $s_k = k/T$ is discrete (with $s_k = k$, $k = \ldots, -1, 0, 1, \ldots$ on condition that $T = 1$).

For example, the Fourier transforms of the rectangle and triangle sequences are derived in the following paragraphs. These sequences are, together with their Fourier transforms, the focal point of the discussions in Sects. 6.5, 8.3 and 8.5.

$$\begin{aligned} &1.\ \Pi_N(t) = \begin{cases} 1 & \textit{for } t = 0, 1, \ldots, N-1 \\ 0 & \textit{for } t < 0 \textit{ and } N-1 < t \end{cases} \\ &\quad \textit{is called rectangle sequence,} \\ &2.\ \Pi_N^{(e)}(t) = \begin{cases} 1 & \textit{for } t = -\frac{N-1}{2}, \ldots, -1, 0, 1, \ldots, \frac{N-1}{2} \\ 0 & \textit{for } t < -\frac{N-1}{2} \textit{ and } \frac{N-1}{2} < t \end{cases} \\ &\quad \textit{is called even rectangle sequence.} \end{aligned} \tag{6.52}$$

$\Pi_N(t)$ in (6.52,1) has N elements with value 1, the first one in time point $t = 0$, the last one in $t = N-1$. $\Pi_N^{(e)}$ in (6.52,2) also has N elements with value 1, however with centre in $t = 0$ which implies that N is odd. For example, $\Pi_7^{(e)}(t) = (\ldots, 0, 1, 1, 1, 1, 1, 1, 1, 0, \ldots)$ for $t = (\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots)$ is the even rectangle sequence with 7 non-zero values. $\Pi_N^{(e)}(t)$ is even and $\Pi_N(t)$ is mixed, as defined in (6.32).

The Fourier transforms of the rectangle sequences in (6.52) are calculated as required in (6.49): $F(s) = \sum_{t=-\infty}^{\infty} \Pi_N(t)\mathrm{e}^{-\mathrm{i}2\pi st} = \sum_{t=0}^{N-1} 1 \times \mathrm{e}^{-\mathrm{i}2\pi st}$, for $(f_t) = \Pi_N(t)$, and $F(s) = \sum_{t=-\infty}^{\infty} \Pi_N^{(e)}\mathrm{e}^{\mathrm{i}2\pi st} = \sum_{t=-(N-1)/2}^{(N-1)/2} 1 \times \mathrm{e}^{-\mathrm{i}2\pi st}$, for $(f_t) = \Pi_N^{(e)}(t)$, are obtained; both transforms are given in (6.53).

$$\mathcal{F}_{-\mathrm{i}}\big(\Pi_N(t)\big) = \sum_{t=0}^{N-1} 1 \times \mathrm{e}^{-\mathrm{i}2\pi st} \qquad \mathcal{F}_{-\mathrm{i}}\big(\Pi_N^{(e)}(t)\big) = \sum_{t=-(N-1)/2}^{(N-1)/2} 1 \times \mathrm{e}^{-\mathrm{i}2\pi st} \tag{6.53}$$

The sum on the left in (6.53) is evaluated and the Fourier transform of the rectangle sequence

$$\mathcal{F}_{-\mathrm{i}}\big(\Pi_N(t)\big) = \begin{cases} \frac{\sin(\pi Ns)}{\sin(\pi s)} \mathrm{e}^{-\mathrm{i}\pi(N-1)s} & \text{for } -1/2 \le s \le 1/2 \text{ and } s \ne 0 \\ N & \text{for } s = 0 \end{cases} \tag{6.54}$$

is arrived at in Problem 6.15. From (6.54), the Fourier transform of the even rectangle sequence is obtained using $\mathrm{e}^{-\mathrm{i}2\pi st} = \mathrm{e}^{-\mathrm{i}2\pi s(t+(N-1)/2)} \times \mathrm{e}^{\mathrm{i}2\pi s(N-1)/2}$. This Fourier transform is called *Dirichlet kernel* $D_N(s)$.

$$\begin{aligned} \mathcal{F}_{-\mathrm{i}}\big(\Pi_N^{(e)}(t)\big) = D_N(s) &= \sum_{t=-(N-1)/2}^{(N-1)/2} 1 \times \mathrm{e}^{-\mathrm{i}2\pi st} \\ &= \mathrm{e}^{\mathrm{i}2\pi s(N-1)/2} \sum_{t=-(N-1)/2}^{(N-1)/2} 1 \times \mathrm{e}^{-\mathrm{i}2\pi s(t+(N-1)/2)} \\ &= \mathrm{e}^{\mathrm{i}\pi s(N-1)} \sum_{t=0}^{N-1} 1 \times \mathrm{e}^{-\mathrm{i}2\pi st} \\ &= \begin{cases} \frac{\sin(\pi Ns)}{\sin(\pi s)} & \text{for } -1/2 \le s \le 1/2 \text{ and } s \ne 0 \\ N & \text{for } s = 0 \end{cases} \end{aligned} \tag{6.55}$$

$D_N(s)$ in (6.55) is a real-valued function since it is the Fourier transform $\mathcal{F}_{-\mathrm{i}}(\Pi_N^{(e)}(t))$ of an even sequence, whereas $\mathcal{F}_{-\mathrm{i}}(\Pi_N(t))$ in (6.54) is a complex-valued function.

The even triangle sequence $\Lambda_{2N+1}^{(e)}$ is defined in (6.56). $\Lambda_{2N+1}^{(e)}$ has $2N-1$ (i.e., an odd number) non-zero values with centre $t = 0$. For example, $\Lambda_7^{(e)}(t) = (\ldots, 0, 0, 1/3, 2/3, 1, 2/3, 1/3, 0, 0, \ldots)$ for $t = (\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots)$ is the even triangle sequence with $N = 3$.

$$\Lambda_{2N+1}^{(e)}(t) = \begin{cases} 1 - \frac{|t|}{N} & \textit{for } t = -N, \ldots, -1, 0, 1, \ldots, N \\ 0 & \textit{for } t \textit{ at the exterior of this interval} \end{cases} \tag{6.56}$$

is called an even triangle sequence.

In (6.56), N is the number of non-zero values in one half of $\Lambda_{2N+1}^{(e)}$, whereas, in (6.52), N is the total number of non-zero values.

The Fourier transform of the even triangle sequence is called *Fejer kernel* $F_N(s)$. $F_N(s) = \mathcal{F}_{-\mathrm{i}}\big(\Lambda_{2N+1}^{(e)}(t)\big)$ is obtained in (6.57) by convolving the even rectangle sequence with itself using the paper strip device (2.38), applying convolution theorem (6.48,6), and then substituting $D_N(s)$ as defined in (6.55).

$$\frac{1}{N}\Big(\big(\Pi_N^{(e)}(t)\big) * \big(\Pi_N^{(e)}(t)\big)\Big) = \Lambda_{2N+1}^{(e)}(t)$$

$$\frac{1}{N}\Big(\mathcal{F}_{-\mathrm{i}}\big(\Pi_N^{(e)}(t)\big)\Big)\Big(\mathcal{F}_{-\mathrm{i}}\big(\Pi_N^{(e)}(t)\big)\Big) = \mathcal{F}_{-\mathrm{i}}\big(\Lambda_{2N+1}^{(e)}(t)\big)$$

$$\frac{1}{N}\left(\frac{\sin(\pi N s)}{\sin(\pi s)}\right)\left(\frac{\sin(\pi N s)}{\sin(\pi s)}\right) = \frac{1}{N}\left(\frac{\sin(\pi N s)}{\sin(\pi s)}\right)^2$$

$$F_N(s) = \left\{ \begin{array}{ll} \frac{1}{N}\left(\frac{\sin(\pi N s)}{\sin(\pi s)}\right)^2 & -\frac{1}{2} \le s \le \frac{1}{2},\ s \neq 0 \\ N & s = 0 \end{array} \right\} = \mathcal{F}_{-\mathrm{i}}\big(\Lambda_{2N+1}^{(e)}(t)\big) \qquad (6.57)$$

Both kernels $D_N(s)$ and $F_N(s)$ have similar properties which can be deduced directly from their definitions and which are summarised in (6.58) and demonstrated in Fig. 6.11.

1. $D_N(s)$ *and* $F_N(s)$ *are periodic with period 1*
2. $D_N(s)$ *and* $F_N(s)$ *are even*
3. $F_N(s) \ge 0$ *(not true for* $D_N(s)$*)* (6.58)
4. $D_N(s) = 0$ *and* $F_N(s) = 0$ *for* $s = \pm 1/N, \pm 2/N, \pm 3/N, \ldots$
5. $D_N(s) = N$ *and* $F_N(s) = N$ *for* $s \to 0$ *and* $0 \leftarrow s$.

Both, $D_N(s)$ and $F_N(s)$, are quotients of odd (sine) functions and therefore are even, as concluded from (6.32,4), and both are periodic with period 1. Consequently, they are known for all s provided that they are known in the interval $0 \le s \le 1/2$. For $s \to 0$, $D_N(s) = F_N(s) = N$, as obtained in (6.55) and (6.57). Since $\sin(\pi N s) = 0$ for $s = \pm 1/N, \pm 2/N, \pm 3/N, \ldots$, also $D_N(s)$ and $F_N(s)$ become identically zero for these s. However, there are two differences between $D_N(s)$ and $F_N(s)$: (i) $F_N(s) \ge 0$ for all s whereas $D_N(s) < 0$ for, e.g., $1/N < s < 2/N$, and (ii) in $D_N(s)$, N is the number of non-zero values in the even rectangle sequence, whereas, in $F_N(s)$, N is the number of non-zero values in half the even triangle sequence.

Further properties of $D_N(s)$ and $F_N(s)$ as proposed in (6.59) do not follow directly from the definitions but are still easily derived:

1. $\int_{-1/2}^{1/2} D_N(s)\mathrm{d}s = 1$ *and* $\int_{-1/2}^{1/2} F_N(s)\mathrm{d}s = 1$*, and*
2. $F_N(s) \to 0$ *for* $-1/2 \le s \le 1/2$ *and* $s \neq 0$*,* (6.59)
   provided that $N \to \infty$.

(6.59,1) results from (6.48,1) since $\Pi_N^{(e)}(t)$ and $D_N(s)$ as well as $\Lambda_{2N+1}^{(e)}(t)$ and $F_N(s)$ are Fourier transform pairs with $\Pi_N^{(e)}(t) = \Lambda_{2N+1}^{(e)}(t) = 1$ for $t = 0$. (6.59,2) holds since $(1/N)\left((\sin(\pi N s))^2/(\sin(\pi s))^2\right) \le (1/N)\big(1/(\sin(\pi s))^2\big)$ and $(1/N)\big(1/(\sin(\pi s))^2\big) \to 0$ for $-1/2 \le s \le 1/2$ and $s \neq 0$, on condition that $N \to \infty$.

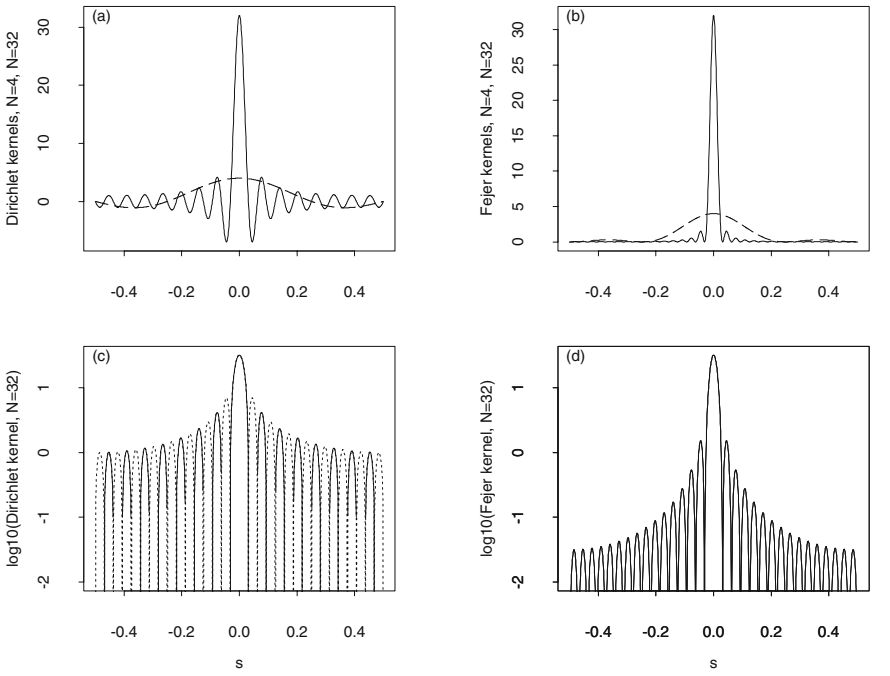


Fig. 6.11. Dirichlet kernels $D_N(s)$ (a) and Fejer kernels (b) for $N = 4$ (broken lines) and $N = 32$ (solid lines). Below, plots (c) and (d), base-10 logarithms of $D_{32}(s)$ and $F_{32}(s)$ (negative side lobes of $D_N(s)$ using at dotted line).

6.3.5 Delta Function

The Fourier transform of a function as required in (6.20) is obtained using the appropriate definition, as introduced in Sects. 6.3.1, 6.3.2, 6.3.3 and 6.3.4. There are, however, functions which are not in $L^2(-\infty, \infty)$ and thus have no Fourier transform. Examples are a real constant $f(t) = a$, or the trigonometric function $f(t) = \cos(2\pi s_1 t)$, s_1 a constant frequency. Fourier transforms of functions not being in $L^2(-\infty, \infty)$ can be defined provided that the notion of function is expanded by introducing the delta or impulse function $\delta(x)$.

$\delta(x)$ is called delta function, impulse function or Dirac δ, on condition that:

$$\begin{aligned} &1.\ \delta(x) = \lim_{n\to\infty} \delta_n(x) = \begin{cases} 0 & \textit{for } x \neq 0 \\ \infty & \textit{for } x = 0 \end{cases} \\ &2.\ \int_{-\infty}^{\infty} \delta_n(x)\mathrm{d}x = 1, \textit{ for } n = 1, 2, \ldots \end{aligned} \tag{6.60}$$

$\delta(x)$ in (6.60) is defined as the limiting case of $\delta_n(x)$, with $\delta_n(x)$ being a sequence of functions having the properties required in (6.60): $\delta(x)$ is not a function in the usual sense as, for example, $f(t) = \cos(2\pi s_1 t)$, s_1 a real

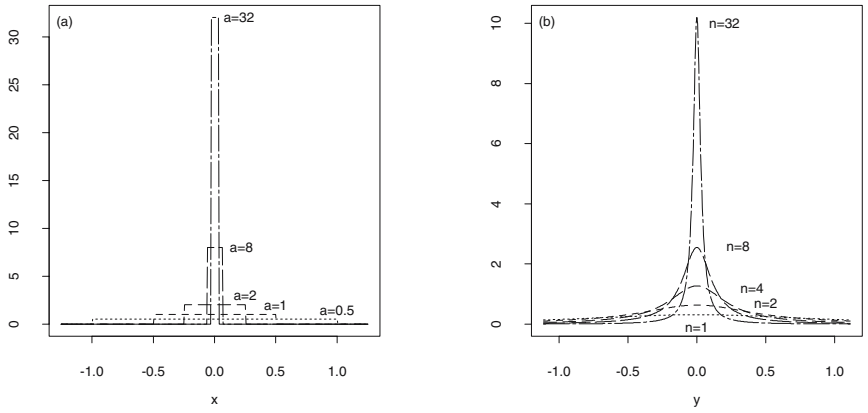


Fig. 6.12. On the left, in plot (a), functions for $a = 0.5, 1, 2, 8, 32$ and thus $b_a = 1, 0.5, 0.25, 0.0625, 0.015625$ in the sequence $g_a(x)$ in (6.61); on the right, plot (b), functions for $n = 1, 2, 4, 8, 32$ in the sequence $f_n(y)$ in (6.62).

constant, or $f(x) = (1/(\sigma\sqrt{2\pi}))\mathrm{e}^{-(1/2)((x-\mu)/\sigma)^2}$, μ, σ real constants with $\sigma > 0$.

Quite often, the sequences of functions (6.61) and (6.62) are used as defining sequences for $\delta(x)$. If, in (6.61), $a = 1, 2, \ldots$ is assumed then $b_a = 1/2, 1/4, \ldots$ is implied since $2ab = 1$, and, therefore, $g_a(x)$ becomes a sequence of rectangles with decreasing widths and increasing heights, yet all having unit area, as plotted in Fig. 6.12 (a). In the limit, $g_a(x)$ becomes a "rectangle" with infinite height and zero width, still with unit area. Thus, $g_a(x)$ is a defining sequence for $\delta(x)$.

$$\delta_a(x) = g_a(x) = \begin{cases} a & \text{for } |x| \le b_a \\ 0 & \text{else} \end{cases} \quad \text{with } 2ab_a = 1 \text{ for } a = 1, 2, 3, \ldots \tag{6.61}$$

$$\delta_n(y) = f_n(y) = \frac{(1/n)}{\pi(y^2 + (1/n)^2)} \qquad n = 1, 2, 3, \ldots \tag{6.62}$$

If, in (6.62), $n = 1, 2, 3, 4, 5, \ldots$ is assumed then a sequence of functions with increasing height n/π for $y = 0$ is obtained, as plotted in Fig. 6.12 (b). Since, as derived in Problem 6.18, (i) $\int_{-\infty}^{\infty} f_n(y)\mathrm{d}y = 1$ for all n and (ii) $f_n(y) \to 0$, for $n \to \infty$ and $y \neq 0$, $f_n(y)$ has the properties required in (6.60). Hence, $f_n(y)$ is a defining sequence for $\delta(x)$.

Other sequences of functions with properties as required in (6.60) are feasible as a defining sequence for $\delta(x)$, e.g., a sequence of Gaussian functions $F(s)$ as proposed in (6.47), $d = 1, 2, \ldots$. The functions in a defining sequence for $\delta(x)$ resemble, in the limit, a spike possessing unit area but "infinite" height and "zero" diameter, independent from the form of the functions, i.e., Gaussian, rectangle, etc., in the non-limiting cases.

A periodic version of the delta function $\delta(s + n)$, $n = \ldots, -1, 0, 1, \ldots$ can be defined when the functions in the defining sequence (i) are periodic

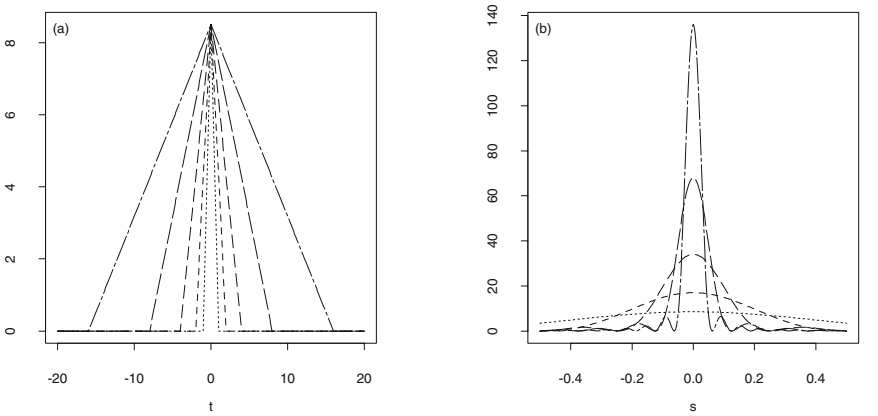


Fig. 6.13. A constant and its Fourier transform defined by the sequence $aG_b(s)$ in (6.63): sequences of triangle functions (plot a) and their Fourier transforms (plot b) for $a = 8.5$, $b = 1, 2, 4, 8, 16$.

with period 1, and (ii) have, in definition interval $-1/2 \le s \le 1/2$, the properties required in (6.60). For example, a sequence of Fejer kernels $F_N(s)$, $N = 1, 2, \ldots$, having the properties (6.59) is a defining sequence for the periodic delta function.

With the definition of the delta function in (6.60) you are now prepared to calculate Fourier transforms of functions that are not in $L^2(-\infty, \infty)$. For example, $f(t) = a$ is arrived at as the limiting case of the sequence of triangle functions $ag_b(t)$ in (6.63) which increase in width for increasing b, as plotted in Fig. 6.13 (a). $ag_b(t)$ is Fourier transformed, in Problem 6.12, to $aG_b(s)$,

$$ag_b(t) = a \begin{cases} 1 - \frac{1}{b}|t| & |t| \le b \\ 0 & |t| > b \end{cases} \qquad aG_b(s) = a \begin{cases} \frac{1}{b}\left(\frac{\sin(\pi bs)}{\pi s}\right)^2 & s \ne 0 \\ b & s = 0 \end{cases} \tag{6.63}$$

also in (6.63). When b increases, the oscillations in $aG_b(s)$ increase in height and number, as plotted in Fig. 6.13 (b). In the limit, i.e., if $b \to \infty$ then $G_b(s) \to \infty$ for $s = 0$ and $G_b(s) \to 0$ for $s \ne 0$, since $(1/b)\big((\sin(\pi bs))/(\pi s)\big)^2 \le (1/b)\big(1/(\pi s)\big)^2$ and $(1/b)\big(1/(\pi s)\big)^2 \to 0$, $s \ne 0$, on condition that $b \to \infty$. Further, (6.48,1) implies that $\int_{-\infty}^{\infty} G_b(s)\mathrm{d}s = g_b(0) = 1$ for $b = 1, 2, \ldots$, i.e., also in the limiting case when $b \to \infty$. Hence, $G_b(s)$, $b = 1, 2, 3, \ldots$, is a defining sequence for $\delta(s)$ as required in (6.60). Consequently, in (6.63), $ag_b(t) = a$ and $aG_b(s) = a\delta(s)$, on condition that $b \to \infty$, implying that $a\delta(s)$ is the Fourier transform of the constant $f(t) = a$.

This example suggests the following procedure to obtain the Fourier transform of a function $f(t)$ that is not in $L^2(-\infty, \infty)$. Approximate $f(t)$ using a sequence of functions $f_n(t)$ which are (i) in $L^2(-\infty, \infty)$ and, (ii) converge to $f(t)$. $f_n(t)$ is chosen such that the sequence of Fourier transforms $F_n(s) =$

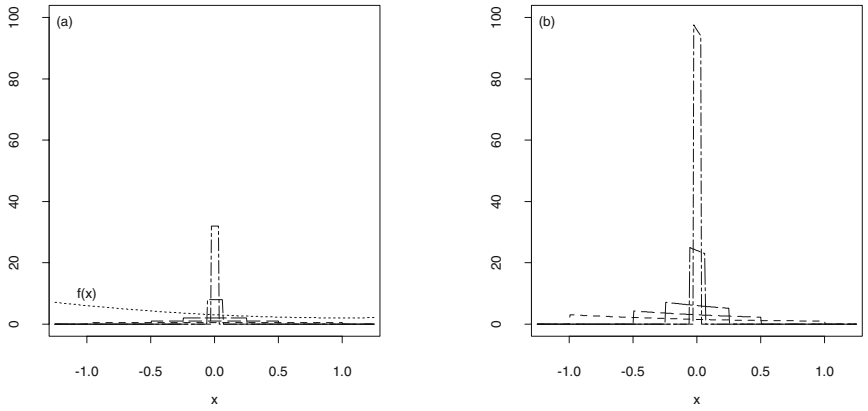


Fig. 6.14. On the left, in plot (a), rectangle functions with height a and width $2b_a$ in the sequence $\delta_a(x)$ as defined in (6.61) for $a = 0.5, 1.0, 2.0, 8.0, 32.0$ together with function $f(x)$. On the right, in plot (b), products $f(x)\delta_a(x)$ for $a = 0.5, 1.0, 2.0, 8.0, 32.0$.

$\mathcal{F}_{-\mathrm{i}}\big(f_n(t)\big)$ is a defining sequence for $\delta_n(s)$ as required in (6.60), which implies that, for all n including the limiting case, $F_n(s)$ is in $L^2(-\infty, \infty)$. Using this procedure, the constant a is Fourier transformed, above, to $a\delta(s)$, and, in Problem 6.19, the trigonometric functions $a_1 \cos(2\pi s_1 t)$ and $a_2 \sin(2\pi s_2 t)$, a_1, a_2, s_1, s_2 real constants, are shown to have Fourier transforms $(a_1/2)\delta(s - s_1) + (a_1/2)\delta(s + s_1)$ and $(a_2/2)\delta(s - s_2) - (a_2/2)\delta(s + s_2)$.

$a\delta(s)$ and $(a_2/2)\delta(s - s_2)$, and also $\delta(c_1 x + c_2)$, $f(x)\delta(x)$, $\int_{-\infty}^{\infty} f(x)\delta(x)\mathrm{d}x$ and $G(y) * \delta(y)$ are examples given to illustrate terms incorporating a Dirac δ. To evaluate such a term, substitute an appropriate $\delta_n(x)$ for $\delta(x)$, add, multiply, integrate, etc. and thereafter obtain the result in the limiting case, as demonstrated by the following examples.

As a first example, the evaluation of the product $f(x)\delta(x)$ is demonstrated in Fig. 6.14. Plot (a) shows functions in sequence $\delta_a(x)$ as defined in (6.61) for $a = 0.5, 1.0, 2.0, 8.0, 32.0$, and function $f(x)$. $f(x)$ is continuous in $x = 0$ with $f(0) > 0$. If the delta function in the product $f(x)\delta(x)$ is substituted by the sequence $\delta_a(x)$ (6.61), then the sequence of functions $\Pi_a(x) = f(x)\delta_a(x) = af(x)$, for $|x| < b_a$, and $\Pi_a(x) = f(x)\delta_a(x) = 0$, for x at the exterior of this interval, is obtained. Functions $\Pi_a(x)$ for $a = 0.5, 1.0, 2.0, 8.0, 32.0$ are plotted in Fig. 6.14 (b). Like $\delta_a(x)$, $\Pi_a(x)$ becomes identically zero for $x < -b_a$ and $b_a < x$; for $-b_a \le x \le b_a$, however, $\Pi_a(x)$ is an approximate rectangle having width $2b_a$ and height $af(-b_a) \approx af(0) \approx af(b_a)$. For its integral, $\int_{-\infty}^{\infty} \Pi_a(x)\mathrm{d}x \approx af(0) \times 2b_a \approx f(0)$ is obtained, since $2ab_a = 1$ as required in (6.61). Consequently, $\int_{-\infty}^{\infty} \Pi_a(x)\mathrm{d}x = \int_{-\infty}^{\infty} f(x)\delta_a(x)\mathrm{d}x \approx f(0)$ results.

The approximations of the height of $\Pi_a(x)$, and therefore also of its integral, improve when its width $2b_a$ decreases as a increases. In the limiting case

for $a \to \infty$, $-b_a \to 0$ and $0 \leftarrow b_a$ are arrived at and, consequently, $2b_a \to 0$ as well as $af(-b_a) = af(0) = af(b_a)$, implying $\amalg_a(x) = af(0)$ for $x = 0$ and $\amalg_a(x) = 0$ for $x \neq 0$ on condition that $a \to \infty$. $\int_{-\infty}^{\infty} \amalg_a(x)\mathrm{d}x = f(0)$ also holds in this case, since, in (6.61), $2ab_a = 1$ in the limiting case as well.

The same argument is used when assumption $f(0) > 0$ is omitted. If, for example, $f(x) < 0$ in a small neighbourhood about 0 where $f(x)$ is intersected by the rectangle functions $\delta_a(x)$, negative products $f(x)\delta_a(x)$ are obtained (imagine the positive products in Fig. 6.14 (b) as being symmetric to the x-axis). This completes the derivation of (6.64).

$$\begin{array}{l} f(x)\delta(x) = f(0)\delta(x) \text{ and } \int_{-\infty}^{\infty} \delta(x)f(x)\mathrm{d}x = f(0) \\ \text{on condition that } f(x) \text{ is continuous in } x = 0. \end{array} \tag{6.64}$$

(6.64) implies (6.65), (6.66), (6.67) and (6.68) provided that $f(x)$ is continuous for all x.

$$\int_{-\infty}^{\infty} \delta(a-x)f(x)\mathrm{d}x = f(a) = \int_{-\infty}^{\infty} \delta(x)f(a-x)\mathrm{d}x \tag{6.65}$$

$$\int_{-\infty}^{\infty} \delta(a+x)f(x)\mathrm{d}x = f(-a) = \int_{-\infty}^{\infty} \delta(x)f(x+a)\mathrm{d}x \tag{6.66}$$

$$\int_{-\infty}^{\infty} \delta(x-y)f(y)\mathrm{d}y = f(x) = \delta(x) * f(x) \tag{6.67}$$

$$\int_{-\infty}^{\infty} f(x-y)\delta(y)\mathrm{d}y = f(x) = f(x) * \delta(x) \tag{6.68}$$

In propositions (6.67) and (6.68), the convolution of $f(x)$ with $\delta(x)$ results in $f(x)$. If, in convolution integral (2.28,1), function $g(x)$ is substituted with sequence $\delta_n(x)$ defining $\delta(x)$, then the integral exists since one of its functions stems from a defining sequence $\delta_n(x)$ for $\delta(x)$ and is thus in $L^2(-\infty, \infty)$ by definition. Hence, $\delta(x)$ is the identity element under convolution (2.28,1) for functions with a real argument. In contrast, (e_i) in (2.41) is the identity element under convolution (2.28,2) for sequences.

As a second example, the evaluation of $\delta(c_1x + c_2)$, i.e., a delta function with an argument being a linear combination of variable x and constants c_1 and c_2, is demonstrated in the following derivation of (6.73). By substituting $\delta_a(x)$ as defined in (6.61) for $\delta(x)$ in $\delta(c_1x + c_2)$, (6.69) is obtained. In (6.70), $g_a(x)$ is a rectangle with height a and width $2b_a/c_1$, $a, b_a, c_1 > 0$, with an area identical to those of the rectangle with height a/c_1 and width $2b_a$, $a, b_a, c_1 > 0$, as given by $g_{ac}(x)$ in (6.71).

$$g_a(x) = \begin{cases} a & \text{for } |c_1x + c_2| \leq b_a \\ 0 & \text{for } x \text{ at the exterior of this interval} \end{cases} \tag{6.69}$$

$$= \begin{cases} a & \text{for } (-b_a/c_1 - c_2/c_1) \leq x \leq (b_a/c_1 - c_2/c_1) \\ 0 & \text{for } x \text{ at the exterior of this interval} \end{cases} \tag{6.70}$$

$$g_{ac}(x) = \begin{cases} a/|c_1| & \text{for } (-b_a - c_2/c_1) \leq x \leq (b_a - c_2/c_1) \\ 0 & \text{for } x \text{ at the exterior of this interval} \end{cases} \tag{6.71}$$

$$= \begin{cases} a/|c_1| & \text{for } |x + c_2/c_1| \le b_a \\ 0 & \text{for } x \text{ at the exterior of this interval} \end{cases} \tag{6.72}$$

$$\delta(c_1 x + c_2) = \frac{1}{|c_1|}\delta(x + c_2/c_1) \tag{6.73}$$

$g_a(x)$ in (6.70) can be substituted by $g_{ac}(x)$ in (6.71), because in a defining sequence $\delta_n(x)$ for $\delta(x)$, the form of the functions is not important in the limiting case, as argued in the remarks to (6.62). Clearly, both $g_a(x)$ and $g_{ac}(x)$, resemble a spike having unit area but "infinite" height and "zero" diameter in the limiting case, and thus have the properties required in (6.60). Consequently, both sequences define the delta function, and (6.73) follows easily by comparing (6.72) with (6.69).

(6.73) can also be obtained with the following plausible argument. The integrals of the functions in a defining sequence for the delta functions are the c_1-fold of the unit area when the x-axis expands with factor c_1. The too large integrals are compensated by multiplying the functions with factor $1/c_1$, e.g., in the case of rectangle functions, reducing the height of the rectangles by the same factor that expands their widths. The translation c_2 of the functions in the defining sequence does not, however, influence their form.

As a third example, the derivative $\dot{\delta}(x)$ of $\delta(x)$ is obtained in Problem 6.20. Obviously, to arrive at the derivative of the delta function, a sequence of functions is substituted whose first derivatives can be easily calculated.

How can I envisage a mental picture of the delta function, you may ask at the end of this section introducing the delta function. Imagine a sequence of unit-area impulses (as used in Mechanics) which are measured with an instrument having a limited resolution. When the impulses become more and more concentrated, although their form can no longer be resolved, their integral can still be measured. When the impulses become infinitely concentrated and infinitely strong, you will have conceptualized a sequence of delta functions. Important is the integral of the impulses; their form is unimportant. In Mathematics and Physics, similar idealised notions such as point observations (examples are given in Sect. 4.6.2), point masses, point charges, etc., are frequently used; clearly, however, these objects do not exist.

6.4 Spectra

In Sect. 6.1, the squared absolute values of the Fourier transform of a deterministic sequence are called spectrum (plural: spectra) when plotted against their frequencies. In such a plot, the frequencies of the oscillations in a sequence become visible, as demonstrated for the sawtooth-shaped sequence in Figs. 6.1 and 6.2. "Hidden periodicities" in a sequence or function can be revealed by calculating a Fourier transform and then a spectrum, not only when dealing with sequences as required in (6.21,1) but in all cases in (6.20). In this section, however, spectra are defined only for the cases (6.20,2,3).

When, in case (6.20,2), a function is approximated in $L^2(-\infty,\infty)$ by its Fourier series, the pertaining mean square error is analysed and the result in (6.74) is arrived at.

1. *If $f(t)$ and $f_M(t)$ are as required in (6.28) then from the mean square approximation error* $\int_{-T/2}^{T/2}\big(f(t)-f_M(t)\big)^2\mathrm{d}t \geq 0$,
2. $(2/T)\int_{-T/2}^{T/2}(f(t))^2\mathrm{d}t \geq A_0^2/2+\sum_{k=1}^{M}(A_k^2+B_k^2)$ *is obtained, which is called Parseval's identity (=) or Bessel's inequality (>).*
3. *If (i) $f(t)$ and $f_M(t)$ are as required in (6.28) and (ii) $M\to\infty$, then the series of squared Fourier coefficients* $A_0^2/2+\sum_{k=1}^{\infty}(A_k^2+B_k^2)$ *converges to* $(2/T)\int_{-T/2}^{T/2}(f(t))^2\mathrm{d}t$.

(6.74)

(6.74,3) follows from the remarks to (6.149) where Bessel's inequality is obtained. Parseval's identity is shown by multiplying both sides of $f(t)=\lim_{M\to\infty} f_M(t)=A_0/2+\sum_{k=1}^{\infty}(A_k\cos(2\pi s_k t)+B_k\sin(2\pi s_k t))$ in (6.28) with $f(t)$, and thereafter integrating from $-T/2$ to $T/2$. Due to the convergence in $L^2(-T/2,T/2)$, the integral of the series becomes the series of the integrals of the terms and the result is obtained as in the derivation of (6.147) where the term $2\int_{-T/2}^{T/2} f(t)f_M(t)\mathrm{d}t$ is integrated for finite M.

In case (6.20,3), function $f(t)$ is $L^2(\infty,\infty)$ as required in (6.36), and therefore in (6.35), its Fourier transform $F(s)$ and thus $|F(s)|^2$ are functions of the real frequencies s. Fourier transform pairs as in (6.35) have properties (6.48). Hence, applying (6.48,7), from the first line in (6.75), the convolution integral in the second line is obtained, and Parseval's identity follows with (6.48,2).

$$\begin{aligned}\int_{-\infty}^{\infty}\big(f(t)\big)^2\mathrm{d}t &= \int_{-\infty}^{\infty}\Big(\big(f(t)\big)\big(\big(f(t)\big)\Big)\mathrm{e}^{-\mathrm{i}2\pi rt}\mathrm{d}t \quad \text{for } r=0\\ &= \int_{-\infty}^{\infty}F(s)F(r-s)\mathrm{d}r = \int_{-\infty}^{\infty}|F(s)|^2\mathrm{d}s \quad \text{for } r=0\end{aligned} \tag{6.75}$$

$|F(s)|^2$ is an even function of s provided that $f(t)$ is a real-valued function of t.

The left-side integrals in (6.75) and (6.74,2, Parseval's identity) are often proportional to the total energy of a physical system, e.g., when $f(t)$ is a speed or a voltage. In (6.75), the total energy is dissipated in $-\infty<t<\infty$, and in (6.74,2) in $-T/2\leq t\leq T/2$. Consequently, the power of the system is integrated on both sides in (6.75) and (6.74,2, Parseval's identity) to obtain the total energy: on the left sides over the time, and on the right sides over the frequencies. In both cases (6.20,2,3), therefore, Parseval's identity distributes the energy dissipated in $-T/2\leq t\leq T/2$ or $-\infty<t<\infty$ over the frequencies. Even when $\big(f(t)\big)^2$ is not the power of a physical system, the

power of a function $f(t)$ with real argument t is said to be distributed over the frequencies in (6.74,2) and (6.75).

This distribution cannot be separated from the Fourier representation $F(s_k)$ or $F(s)$ of $f(t)$, i.e., from the representation of $f(t)$ as a linear combination of trigonometric functions. Also sequences (f_t) (i.e., functions with integer argument t) have Fourier representations $F(s_k)$ or $F(s)$, as in (6.22) or (6.49) and, consequently, also the *power of sequences* can be distributed over the frequencies.

Let $f(t)$ be a function with Fourier transform $\mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ or (f_t) be a sequence with Fourier transform $\mathcal{F}_{-\mathrm{i}}(f_t)$ as defined in (6.22), (6.30), (6.35) or (6.49). Then:

1. $\big|\mathcal{F}_{-\mathrm{i}}\big(f(t)\big)\big|^2$ *or* $\big|\mathcal{F}_{-\mathrm{i}}(f_t)\big|^2$ *is called spectrum or power spectrum of $f(t)$. Less frequently used are:* (6.76)
2. *discrete power spectrum for $|F(s_k)|^2$ as function of the Fourier frequencies s_k in case (6.20,2) with $\mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = F(s_k)$,*
3. *energy density spectrum for $|F(s)|^2$ as function of frequency s, s real, in case (6.20,3) with $\mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = F(s)$.*

Definition (6.76,3) reflects the following considerations: $|F(s)|^2 \mathrm{d}s$ is, in (6.75), the part of the total energy of $f(t)$ that pertains to oscillations in the Fourier representation of $f(t)$ having frequencies between s and $s + \mathrm{d}s$. Thus, $|F(s)|^2$ is an energy density, with properties similar to the probability density of a random variable: $|F(s)|^2$ is the energy density in a small neighbourhood about s, and not the energy in point s.

In applications, a spectrum is computed in two steps as demonstrated in Sect. 6.1. In a first step, the observations (g_t) are discretely Fourier transformed using (6.22) or (6.26) to arrive at $G(s_k)$ or $G(s'_k)$. In the second step, $|G(s_k)|^2$ or $|G(s'_k)|^2$ are plotted against the frequencies, preferably, as demonstrated in Fig. 6.2 and recommended in the introductory remarks to Sect. 6.1.3, after taking the logarithm.

6.5 Aliasing and Leakage

Time is a real dimension. In Geosciences therefore, variables that depend on time are functions $f(t)$ of the real argument t, on condition that they are, as is the case in this chapter, assumed to be deterministic. Furthermore, $f(t)$ is assumed to be in $L^2(-\infty, \infty)$. This second assumption is not unrealistic for the following reason. Usually, $f(t)$ is observed for a finite interval (i.e., during a finite period) and, therefore, can be modified at the exterior of the observational interval such that it is in $L^2(-\infty, \infty)$. If $f(t)$ is in $L^2(-\infty, \infty)$ then it has a Fourier transform $F(s)$ as in (6.35), $F(s)$ being a complex-valued function of the real argument s.

$f(t)$ is measured to obtain the sequence of observations (g_t) by discretising the time and confining the measurements to a finite interval. The Fourier

transform of (g_t), i.e., $G(s_k)$ as required in (6.28), is a complex-valued function of the Fourier frequencies s_k, thus being a function of a discrete argument, whereas $F(s)$ is a complex-valued function of the real argument s. $G(s_k) = F(s)$ for $s_k = s$ would be a wishful result. This remains however, unless adequate precautions are taken, a wish, as shown in this section.

6.5.1 Measuring a Function $f(t)$ with Real Argument t

For example, a Fourier transform pair $f(t)$ and $F(s)$, with both $f(t)$ and $F(s)$ in $L^2(-\infty, \infty)$, is plotted in Fig. 6.15 (a) and (b) in the intervals $-20.0 \le t \le 20.0$ and $-2.0 \le s \le 2.0$. The damped real-valued oscillation $f(t)$ is chosen to be even (6.32,1), and consequently, $F(s)$ is also real-valued and even as concluded from (6.37,1) and (6.32,4). $f(t)$ and $F(s)$ in Fig. 6.15 result with $s_1 = 0.1$, $s_2 = 0.7$, $a_1 = 0.05$ and $a_2 = 0.15$ in (6.77) and (6.78).

$$f(t) = \sum_{l=1}^{2} f_l(t) \text{ with } f_l(t) = 2\mathrm{e}^{-|a_l t|}\cos(2\pi s_l t) \tag{6.77}$$

$$F(s) = \sum_{l=1}^{2} F_l(s) \text{ with } F_l(s) = \frac{2a_l}{a_l^2 + \big(2\pi(s-s_l)\big)} + \frac{2a_l}{a_l^2 + \big(2\pi(s+s_l)\big)} \tag{6.78}$$

$$(g_t) = f(t) \text{ for } t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2 \tag{6.79}$$

$$G(s_k) = 2 \sum_{t=0}^{(N-1)/2} g_t \cos(2\pi s_k t) \quad k = 0, 1, \ldots, (N-1)/2 \tag{6.80}$$

$f(t)$ is the sum of two cosine functions with frequencies s_l which are damped using negative exponential functions with exponent a_l, $l = 1, 2$. $F(s)$ has peaks at frequencies s_l and decreases for $|s| > \max(s_l)$, $\max(s_l)$ the maximum frequency s_l. $f_l(t)$ is Fourier transformed, in Problem 6.23, to $F_l(s)$ in (6.78). $F(s)$, the Fourier transform of $f(t)$, is obtained using (6.48,5).

Geophysical variables are often measured with constant sampling intervals Δt during a finite period T. In this chapter as well as in Chaps. 7, 8, 9 and 10, $\Delta t = 1$ is assumed. This does not imply a basic restriction, as discussed in the remarks to (6.92) in Sect. 6.5.2. From observations of $f(t)$, beginning at an integer time point, a sequence (g_t) as defined in (6.79) is obtained. (g_t) is even, a property inherited from $f(t)$, and, therefore, its Fourier transform $G(s_k)$ is computed as required in (6.80) for frequencies $0 \le s_k \le 1/2$. $G(s_k)$ is even and periodic with period 1.

In Fig. 6.15, for example, $N = 41$ and the measurement of $f(t)$ as in plot (a) produces sequence (g_t), $t = -20, \ldots, -1, 0, 1, \ldots, 20$, as in plot (c). By Fourier transforming (g_t) as required in (6.80), $G(s_k)$ in plot (d) is obtained. $G(s_k)$ is defined for discrete $-1/2 \le s_k \le 1/2$ and is periodic with period 1. In contrast, $F(s)$ is defined for real $-\infty \le s \le \infty$. Comparing plots (b) and (d) in Fig. 6.15, it is concluded that $G(s_k) \ne F(s_k)$, $-1/2 \le s_k \le 1/2$, for the following two reasons:

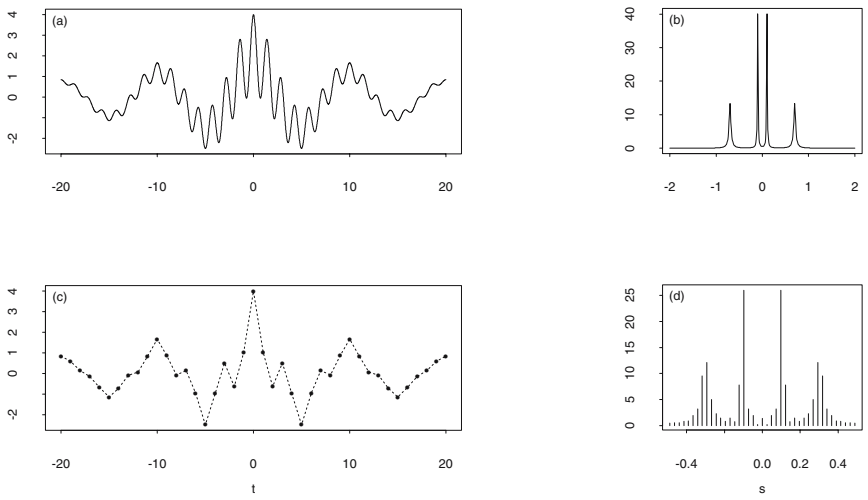


Fig. 6.15. Pairs of fourier transforms as defined in (6.77) and (6.78) (above, plots (a) and (b)) as well as in (6.79) and (6.80) (below, plots (c) and (d)) in the intervals $-20.0 \leq t \leq 20.0$ as well as $2.0 \leq s \leq 2.0$ and $0.5 \leq s \leq 0.5$, for $s_1 = 0.1$, $s_2 = 0.7$, $a_1 = 0.05$ and $a_2 = 0.15$ as well as $N = 41$.

1. the peak in $F(s)$ at frequency 0.1 appears in $G(s_k)$ at frequencies $s_k \approx 0.1$, however with reduced height;
2. the peak in $F(s)$ at frequency 0.7 appears in $G(s_k)$ at frequencies $s \approx 0.3$ with approximately the same height as in $F(s)$.

In this example, the first distortion in $F(s)$ is judged to be a less serious distortion than the second one. Consequently, prior to giving a second demonstration of the first distortion in Fig. 6.17, the second one, i.e., the displacement of a peak, is made plausible.

To $f(t)$, as plotted in Fig. 6.15 (a), pertains the real-valued Fourier transform $F(s)$ as plotted in Fig. 6.15 (b). $F(s)$ is, because real-valued and positive for all s, the absolute value of the trigonometric oscillations in the Fourier representation (6.35) of $f(t)$. The oscillations with frequencies $s \approx 0.7$ in $F(s)$ are clearly apparent in $G(s_k)$, as plotted in Fig. 6.15 (d), with discrete frequencies $s_k \approx -0.3$ or $s_k \approx 0.3$, both $F(s)$ and $G(s_k)$ being even. The reason for this phenomenon is made plausible in Fig. 6.16. If, for example, the trigonometric oscillations as plotted in Fig. 6.16 are observed, using a sampling interval $\Delta t = 1$ at integer time points, oscillations with frequencies s_2, s_3 and s_4 appear as if they had frequency s_1. s_1 is called *alias frequency* of s_2, s_3 and s_4.

In general, *aliasing* occurs when oscillations in $F(s)$ with $|s| > 1/2$ appear in $G(s_k)$, $-1/2 \leq s_k \leq 1/2$, with ($f(t)$ and $F(s)$ as well as (g_t) and $G(s_k)$

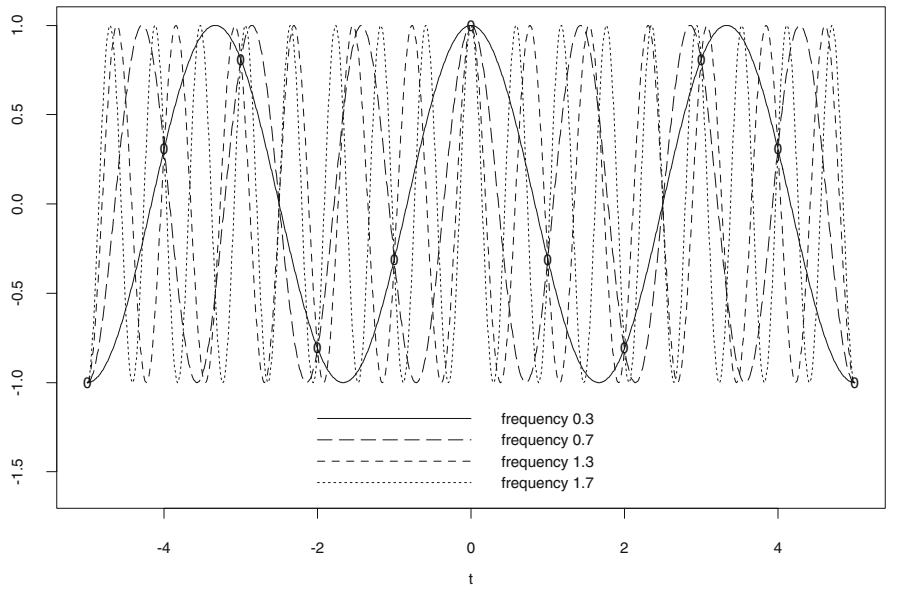


Fig. 6.16. Functions $\cos(2\pi s_i t)$, $s_1 = 0.3$, $s_2 = 0.7$, $s_3 = 1.3$, $s_4 = 1.7$, cannot be distinguished from each other when measuring at time points $\ldots, -4, -3, -2, -1, 0, 1, 2, 3, 4, \ldots$ (plotted with symbol 0).

being Fourier transform pairs as demonstrated in Fig. 6.15. The aliasing is thoroughly dealt with in Sect. 6.5.2. In applications, a possible distortion of a Fourier transform or spectrum due to aliasing can be avoided or substantially reduced on condition that the sampling interval Δt is chosen to be small enough, as proposed in the remarks to (6.92). This is feasible when an instrument is available that allows for measuring with a sufficiently small Δt. Examples are given in Sect. 9.4.

When a function $f(t)$ in $L^2(-\infty, \infty)$ is observed in order to obtain its Fourier transform $F(s)$, distortions which are not due to aliasing can also occur. These distortions arise from observing $f(t)$, which is defined for all real $-\infty < t < \infty$, in a finite interval. An observation interval is always finite because resources available for a measurement are limited. For example, $f(t)$ is assumed as in (6.77) for $-\infty < t < \infty$ with $s_1 = 0.01$, $s_2 = 0.02$, $a_1 = 0.002$ and $a_2 = 0.0001$. $f(t)$ is shown in Fig. 6.17, plot (a), for $-100.0 \le t \le 100.0$, and the logarithm of its Fourier transform $F(s)$ below, in plots (c) and (e), for $-0.06 \le s \le 0.06$ and $0.20 \le s \le 0.32$. $F(s)$ has two sharp and narrow peaks close to frequency 0, and $f(t)$ is only slightly damped.

$f(t)$ is observed at time points $t = -100, \ldots, -1, 0, 1, \ldots, 100$ to obtain sequence (g_t) as defined in (6.79). (g_t) is shown in Fig. 6.17 (b). Thereafter, (g_t) is zero-padded symmetrically on both sides until $N' = 401$ to obtain

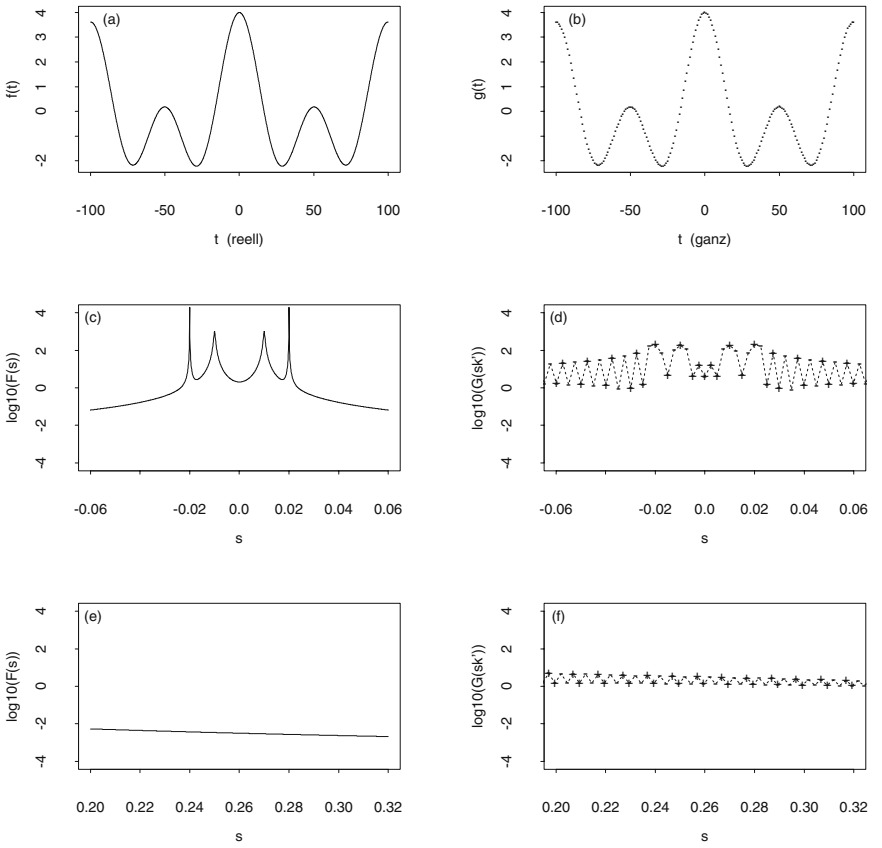


Fig. 6.17. Fourier transform pairs $f(t)$ and $F(s)$ as defined in (6.77) (6.78) (on the left) as well as (g_t) and $G(s'_k)$ as defined in (6.79) (6.80)(on the right) for $s_1 = 0.01$, $s_2 = 0.02$, $a_1 = 0.002$, $a_2 = 0.0001$, $N = 201$ and $N' = 401$; N' and $G(s'_k)$ as required in (6.26) and (6.26). $\log_{10}\big(|G(s'_k)|\big)$ for $G(s'_k) > 0$ is plotted using symbol $+$, for $G(s'_k) < 0$ using $-$.

its discrete Fourier transform $G(s'_k)$ at frequencies s'_k in distance $\Delta s'_k = 1/N' = 1/401 \approx 0.0025$, as concluded from (6.26), i.e., a reduction to half of the distance $\Delta s_k = 1/N = 1/201 \approx 0.005$ of the Fourier frequencies as in (6.21,1). Thus, using a finer grid of frequencies, hopefully, the sharp peaks in $F(s)$ become better visible in $G(s'_k)$. $\log_{10}|G(s'_k)|$ is shown in Fig. 6.17, plots (d) and (f), for $-0.06 \le s_k \le 0.06$ and $0.20 \le s_k \le 0.32$, with both intervals containing approximately 50 discrete frequencies s'_k.

In both intervals, $G(s'_k)$ is a very poor approximation of $F(s)$. This (not very encouraging) result is obtained when plots (d) and (f) are compared with plots (c) and (e), in Fig. 6.17. It particularly irritates that:

1. $G(s'_k)$ takes negative values whereas $F(s) > 0$ for all s, and
2. $|G(s'_k)| \ll F(s)$ at the peaks of $F(s)$, but $|G(s'_k|)| \gg F(s)$ between and even in larger distance from the peaks ($\ll$ ($\gg$) smaller (larger) by orders of magnitude).

These distortions are said to be due to *leakage*. Clearly, they cannot be neglected and, therefore, they are analysed in Sects. 6.5.3 and 6.7.

Examples such as those given in Figs. 6.15 and 6.17 demonstrate that there are Fourier transform pairs $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ as defined in (6.35) having the following property: $F(s)$ cannot be obtained without distortions from observations of $f(t)$. These distortions occur when $f(t)$ is observed at N discrete time points to obtain sequence (g_t), after which the Fourier transform $G(s_k)$ of (g_t) is calculated. In both examples, $G(s_k) \neq F(s_k)$, $-1/2 \leq s_k \leq 1/2$, due to aliasing and leakage, both are described in Sects. 6.5.2 and 6.5.3.

As a prerequisite for the derivations in Sects. 6.5.2 and 6.5.3, the discretisation of time, when observing with sampling interval Δt, is modelled in (6.81), (6.82) and (6.84), assuming a non-finite duration of the observations. In (6.81), $f(t)$ is assumed to be (i) a real-valued function with the real argument t, and (ii) in $L^2(-\infty, \infty)$. Its Fourier transform $F(s)$, as required in (6.35), is also in $L^2(-\infty, \infty)$ and is defined for all real s.

$$\mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = F(s) = \int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \quad s,\, t \text{ real} \tag{6.81}$$

$$(g_t) = f(t) \text{ for } t = \ldots, -1, 0, 1, \ldots \text{ with} \tag{6.82}$$

$$(g_t) = \mathcal{F}{+}\mathrm{i}\big(G(s)\big) = \int_{-1/2}^{1/2} G(s)\mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}s \tag{6.83}$$

$$(g_{t\Delta t}) = f(t\Delta t) \text{ for } t = \ldots, -1, 0, 1, \ldots \text{ with} \tag{6.84}$$

$$(g_{t\Delta t}) = \mathcal{F}{+}\mathrm{i}\big(G(s)\big) = \int_{-1/(2\Delta t)}^{1/(2\Delta t)} G(s)\mathrm{e}^{\mathrm{i}2\pi st\Delta t}\mathrm{d}s \tag{6.85}$$

When $f(t)$ is measured at time points $t = \ldots, -1, 0, 1, \ldots$, i.e., using sampling interval $\Delta t = 1$, sequence (g_t) in (6.82) is obtained, and, when using a general sampling interval Δt not required to be 1, $(g_{t\Delta t})$ in (6.84) is arrived at. Since both (g_t) and $(g_{t\Delta t})$ are in $L^2(-\infty, \infty)$, a property inherited from $f(t)$, both have periodic Fourier transforms $G(s)$ as defined in (6.49) and (6.50) (modified for a general Δt in the case of $(g_{t\Delta t})$). Consequently, (6.83) and (6.85) are obtained. In contrast, (g_t) and $G(s_k)$ in (6.79) and (6.80) are a Fourier transform pair as defined in (6.22).

In applications, an observation is restricted to a finite interval in the time domain, as demonstrated in Figs. 6.15 and 6.17. For the descriptions of aliasing and leakage, following in Sects. 6.5.2 and 6.5.3, it is however,

assumed that, in (6.82) and (6.84), (g_t) and $(g_{t\Delta t})$ are available for all t and $t\Delta t$, $t = \ldots, -1, 0, 1, \ldots$.

6.5.2 Aliasing

Experiments such as those in Figs. 6.15 and 6.16 suggest that aliasing occurs once the time is discretised when a variable is observed. A discretisation of time solely (i.e., without confining the measurements to a finite period) is modelled in (6.82) or (6.84). Assuming $f(t)$, $(g_{t\Delta t})$ and $F(s)$ as defined in (6.81) and (6.84), $(g_{t\Delta t})$ is represented, in (6.86), as a Fourier transform of $F(s)$. Clearly, the derivation starts with this Fourier representation because $(g_t\Delta t)$ is contructed from $f(t)$.

$$(g_{t\Delta t}) = \int_{-\infty}^{\infty} F(s)\mathrm{e}^{\mathrm{i}2\pi st\Delta t}\mathrm{d}s \qquad t,j \text{ integer, } s \text{ real} \tag{6.86}$$

$$= \sum_{j=-\infty}^{\infty} \left(\int_{s=-(2j-1)/(2\Delta t)}^{(2j+1)/(2\Delta t)} F(s)\mathrm{e}^{\mathrm{i}2\pi st\Delta t}\mathrm{d}s \right) \tag{6.87}$$

$$= \sum_{j=-\infty}^{\infty} \left(\int_{r=-1/(2\Delta t)}^{+1/(2\Delta t)} F(r + j/\Delta t)\mathrm{e}^{\mathrm{i}2\pi(r+j/\Delta t)t\Delta t}\mathrm{d}r \right) \tag{6.88}$$

$$= \sum_{j=-\infty}^{\infty} \left(\int_{r=-1/(2\Delta t)}^{+1/(2\Delta t)} F(r + j/\Delta t)\mathrm{e}^{\mathrm{i}2\pi rt\Delta t}\mathrm{d}r \right) \tag{6.89}$$

$$= \int_{r=-1/(2\Delta t)}^{+1/(2\Delta t)} \left(\sum_{j=-\infty}^{\infty} F(r + j/\Delta t) \right) \mathrm{e}^{\mathrm{i}2\pi rt\Delta t}\mathrm{d}r \tag{6.90}$$

$$\mathcal{F}_{-\mathrm{i}}(g_{t\Delta t}) = G(s) = \sum_{j=-\infty}^{\infty} F(s + j/\Delta t) \quad \text{for } |s| \le \frac{1}{2\Delta t} \tag{6.91}$$

(6.87) holds, since an integral can be evaluated as a sum of integrals. (6.88) is obtained by substituting $s = r + j/\Delta t$ for s, to obtain an inner integral with bounds that no longer depend on j. This advantage is paid for by having $j/\Delta t$ in the exponent of the e function. For $kt = \ldots, -1, 0, 1, \ldots$, $\mathrm{e}^{\mathrm{i}2\pi kt} = \cos(2\pi kt) + \mathrm{i} \times \sin(2\pi kt) = 1$, and thus $\mathrm{e}^{\mathrm{i}2\pi(s+j/\Delta t)t\Delta t} = \mathrm{e}^{\mathrm{i}2\pi s\Delta t}\mathrm{e}^{\mathrm{i}2\pi jt} = \mathrm{e}^{\mathrm{i}2\pi s\Delta t}$ is arrived at. Using this result, (6.89) follows with $\mathrm{re}\big(F(s)\big)$ and $\mathrm{im}\big(F(s)\big)$ being in $L^2(-\infty, \infty)$. Hence, the summation and integration can be performed in reversed order, and (6.90) is arrived at.

When comparing (6.90) with (6.85), it follows that $\sum_{j=-\infty}^{\infty} F(r + j/\Delta t)$ is the Fourier transform $G(r)$ of $(g_{t\Delta t})$, r substituted for s, as stipulated in (6.91). In the above derivation of (6.91), $F(s)$ is the Fourier transform of $f(t)$, the function being observed at time points $t\Delta t$, $t = \ldots, -1, 0, 1, \ldots$ to obtain sequence $(g_{t\Delta t})$. Consequently, $G(s)$ in (6.91) is a sum of $F(s)$ and its duplicates which are displaced by $j/\Delta t$, $j = \ldots, -1, 0, 1, \ldots$. For every frequency s in $-1/(2\Delta t) \le s \le 1/(2\Delta t)$, therefore, $G(s)$ is the sum

of the values of $F(s)$ for frequencies $s \pm 1/\Delta t, s \pm 2/\Delta t, s \pm 3/\Delta t, \ldots$. Thus, trigonometric oscillations in $F(s)$ with frequencies $s \pm 1/\Delta t, s \pm 2/\Delta t, s \pm 3/\Delta t, \ldots$, appear with frequency s in $G(s)$.

Let $f(t)$, $(g_{t\Delta t})$, $F(s)$ and $G(s)$ be as in (6.81), (6.84) and (6.91). Then:

1. *s is called alias frequency of $s \pm j/\Delta t$, $j \neq 0$.* (6.92)
2. *The highest frequency that is not an alias of a lower one is called Nyquist frequency: $s^{(Ny)} = 1/(2\Delta t)$.*

Using the Nyquist frequency (6.92,2), two cases can be distinguished. In the first case, $F(s) = 0$ for $|s| > 1/(2\Delta t)$. This implies that $F(s)$ in (6.91) remains invariant once its displaced copies are added, resulting in $G(s) = F(s)$ for $-1/(2\Delta t) \leq s \leq 1/(2\Delta t)$. In this case, $G(s)$ will be a periodic version of $F(s)$ and no information on the oscillations inherent in $f(t)$ will be lost when t is discretised to obtain $(g_{t\Delta t})$. Hence, $f(t)$ can be interpolated from $(g_{t\Delta t})$. An interpolation function is obtained in the following calculation, assuming $\Delta t = 1$ and, therefore, observations (g_u), $u = \ldots, -1, 0, 1, \ldots$. Then, $G(s) = F(s)$ for $-1/2 \leq s \leq 1/2$ implies $f(t) = \int_{-1/2}^{1/2} F(s) \mathrm{e}^{\mathrm{i}2\pi st} \mathrm{d}s = \int_{-1/2}^{1/2} G(s) \mathrm{e}^{\mathrm{i}2\pi st} \mathrm{d}s$, t and s real. Substituting $G(s) = \mathcal{F}_{-\mathrm{i}}(g_u) = \sum_{u=-\infty}^{\infty} g_u \mathrm{e}^{-\mathrm{i}2\pi su}$, u integer, s real, $f(t) = \int_{-1/2}^{1/2} \big(\sum_{u=-\infty}^{\infty} g_u \mathrm{e}^{-\mathrm{i}2\pi su}\big) \mathrm{e}^{\mathrm{i}2\pi st} \mathrm{d}s$ is arrived at. The order of integration and summation is reversed, as in the derivation of (6.51), and $f(t) = \sum_{u=-\infty}^{\infty} g_u \big(\sin(\pi(u-t))/(\pi(u-t))\big)$ is obtained, u integer, s real, which can be used for interpolating $f(t)$ from (g_u).

In the second case, $F(s) \neq 0$ for $s < -1/(2\Delta t)$ and $s > 1/(2\Delta t)$, implying $F(s) \neq 0$ for frequencies $s \pm 1/\Delta t, s \pm 2/\Delta t, s \pm 3/\Delta t, \ldots$. Since, in (6.91), $G(s)$ is the sum of $F(s)$ and its displaced duplicates, $G(s)$ is not a periodic version of $F(s)$ in $-1/(2\Delta t) \leq s \leq 1/(2\Delta t)$ due to aliasing. This case (with $\Delta t = 1$) is demonstrated using the Fourier transform pair $f(t)$ and $F(s)$ in Fig. 6.15, (a) and (b). $F(s)$ is re-plotted with higher resolution of the time domain in Fig. 6.18 (a), and $G(s)$, obtained as required in (6.91) from $F(s)$, is plotted in Fig. 6.18 (b). $F(s)$ has a peak at frequency 0.7 which appears in $G(s)$ at its alias frequency -0.3 or 0.3, $F(s)$ and $G(s)$ being even. Aliasing is also seen in the spectra $|G(s)|^2$ and $|G(s_k)|^2$. $|G(s)|^2$ in Fig. 6.18 (c) is calculated using $G(s)$, and $|G(s_k)|^2$ in plot (d) is calculated using a discrete Fourier transform, as in (6.80), from (g_t) in the time slice $t = -80, \ldots, -1, 0, 1, \ldots, 80$. Consequently, with $N = 161$, this time slice is four times longer than the one used in Fig. 6.15 (c) and (d), where $G(s_k)$ has been calculated from (g_t) for $t = -20, \ldots, -1, 0, 1, \ldots, 20$.

Let $F(s)$ be a function in the frequency domain that cannot be measured but which is known to be approximately identically zero for $|s| > s_0$, i.e., $F(s) \approx 0$ for $|s| > s_0$, and let the Fourier transform of $F(s)$, i.e., $f(t) = \mathcal{F}_{+\mathrm{i}}\big(F(s)\big)$, be a function such that it can be measured in the time domain. Then, using the sampling interval $\Delta t = 1/(2s_0)$, $F(s) \approx 0$ at the exterior of $-1/(2\Delta t) \leq s \leq 1/(2\Delta t)$ and thus almost all oscillations in $f(t)$ are captured

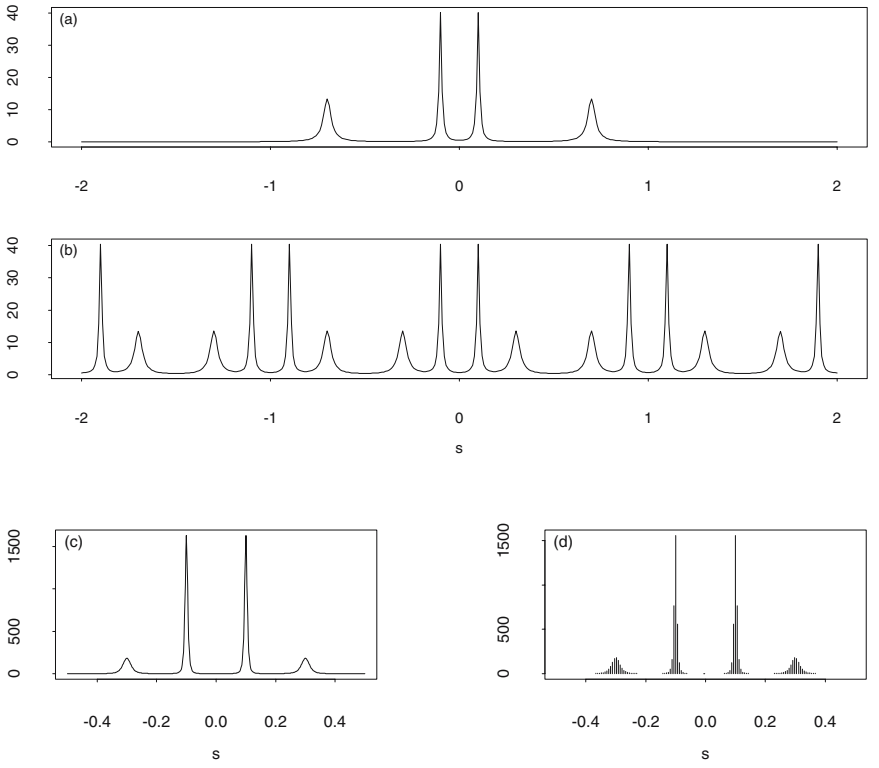


Fig. 6.18. Fourier transform $F(s)$ (plot a) of the function $f(t)$ in Fig. 6.15 (a), $G(s)$ calculated from $F(s)$ using (6.91) (plot b), spectrum $|G(s)|^2$ calculated from $G(s)$ (plot c), and $|G(s_k)|^2$ calculated from a time slice of the observations (g_t) (plot d).

by the measurement. In applications, Δt is chosen such that $F(s)$ is small at the exterior of $-1/(2\Delta t) \leq s \leq 1/(2\Delta t)$. This usually is feasible since a function $F(s)$ being in $L^2(-\infty, \infty)$ becomes approximately identically zero for $-\infty \leftarrow s$ and $s \rightarrow \infty$. Examples for choosing the sampling interval with respect to the Nyquist frequency are given in Sect. 9.4, in the remarks to Figs. 9.28 and 9.34.

Subsequent to the measurement, the observations are dealt with as sequence (g_t) with argument $t = \ldots, -1, 0, 1, \ldots$. This is not a principal restriction, because a sampling interval $\Delta t \neq 1$ can be adapted, as implemented in `spec.univariate()` introduced in Sect. 9.5.6.

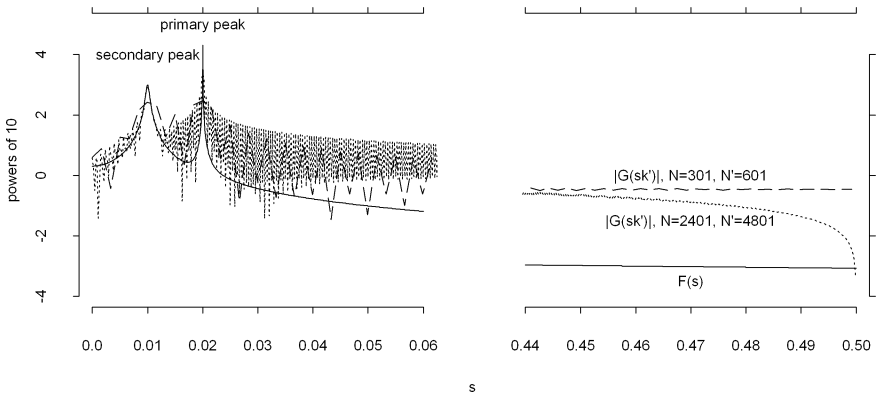


Fig. 6.19. Function $F(s)$ (solid line) and discrete Fourier transforms $G(s'_k)$ (broken lines) as in Fig. 6.17, for $N = 301$ and $N' = 601$ as well as $N = 2401$ and $N' = 4801$.

6.5.3 Leakage

A discrete Fourier transform of (g_t), (g_t) being observations of the damped oscillations $f(t)$ as required in (6.79), generates $G(s_k) = \mathcal{F}_{-\mathrm{i}}(g_t)$, and, as demonstrated in Figs. 6.15 and 6.17, $G(s_k) \neq F(s_k)$, $-1/2 \leq s_k \leq 1/2$, due to aliasing and leakage. Distortions in $G(s_k)$ due to leakage are assumed to be caused, in the remarks to Fig. 6.17, by an observational period confined to $N = 201$. Does a further experiment with $N > 201$ confirm this speculation?

$F(s)$ as in Fig. 6.17 (c) and (e) is drawn again, using a solid line, in Figs. 6.19 and 6.20. At the exterior of $-1/2 \leq s \leq 1/2$, $F(s) < 10^{-3}$ and the function then decreases with increasing $|s|$. Hence, $F(s) < 10^{-3}$ for $s \pm 1, s \pm 2, \ldots$, $-1/2 \leq s \leq 1/2$, and, applying (6.91), it is concluded that $G(s) \approx F(s)$ in $-1/2 \leq s \leq 1/2$, $G(s)$ being the Fourier transform of (g_t). Consequently, distortions in $G(s)$ due to aliasing can be neglected. Remember that: (g_t) is obtained from $f(t)$ for $t = \ldots, -1, 0, 1, \ldots$ as defined in (6.83), i.e., that (g_t) is not confined to an observational period.

From (g_t), three time slices $(g_t^{(s)})$, $(g_t^{(m)})$ and $(g_t^{(l)})$ as defined in (6.79) are selected, with $N^{(s)} = 301$, $N^{(m)} = 2401$ and $N^{(l)} = 65535$, all three with $\Delta t = 1$ since the sampling interval is inherited from (g_t). $(g_t^{(s)})$ and $(g_t^{(m)})$ are symmetrically zero-padded (a sequence is zero-padded, in (6.26) and (6.27), only on the right side) to obtain sequences with $N^{(s)\prime} = 601$ and $N^{(m)\prime} = 4801$ values, whereas, following the recommendations in the remarks to (6.27), $(g_t^{(l)})$ is padded with only one zero on the right side, to arrive at a sequence with $N^{(l)\prime} = 65536 = 2^{16}$ values. Thereafter, the discrete Fourier transforms $G^{(s)}(s'_k)$, $G^{(m)}(s'_k)$ and $G^{(l)}(s'_k)$ are computed from $(g_t^{(s)})$, $(g_t^{(m)})$ and $(g_t^{(l)})$ using `fft()`, as demonstrated in Sects. 6.3.1 and 6.9.8.

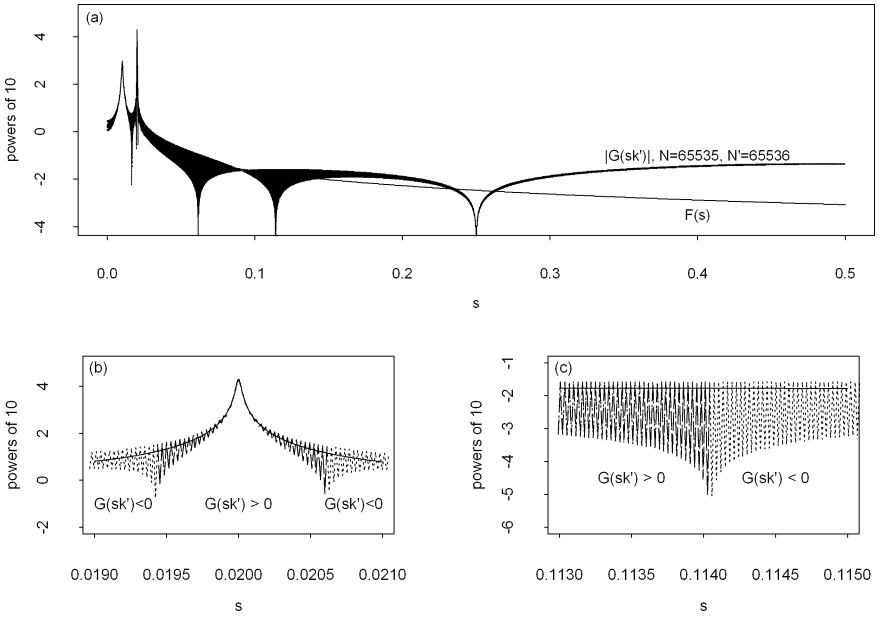


Fig. 6.20. Function $F(s)$ (solid line) and discrete Fourier transforms $G(s'_k)$ (broken lines) as in Fig. 6.17, for $N = 65535$ and $N' = 65536$ (b), (c) and (d).

Since these Fourier transforms take positive and negative values, the logarithms of their absolute values are plotted in Figs. 6.19 and 6.20, using a solid line when positive, a broken line when negative (not visible in plot Fig. 6.20 (a)). Obviously, $\Delta s'_k$ decreases with increasing length of the sequence, and $|G(s'_k)|$ comes closer to $F(s) \approx G(s)$ at least in a small neighbourhood of the peaks in $F(s)$: $G^{(m)}(s'_k)$ traces the secondary peak, $G^{(l)}(s'_k)$ the primary peak. At large distances away from those peaks (i.e., for higher frequencies), however, the absolute values of all three discrete Fourier transforms, $G^{(s)}(s'_k)$, $G^{(m)}(s'_k)$ and $G^{(l)}(s'_k)$, are much larger than $F(s)$.

The experiments shown in Figs. 6.17, 6.19 and 6.20 are repeated, in Problem 6.25, using functions $F(s)$ as defined in (6.78). The $F(s)$ used in Problem 6.25 are required to (i) be still small for $s > 0.5$ such that aliasing can be neglected, and (ii) have peaks of moderate heights. From these experiments, leakage is concluded to have the following appearance:

1. $G(s_k)$ or $G(s'_k)$ deviate from $F(s)$ due to distortions that cannot be neglected on condition that the peaks in $F(s)$ are rather high. Higher peaks generate larger distortions than smaller ones, thereby suggesting that the distortions depend on the differences of the maxima and minima in $F(s)$.
2. Peaks in $F(s)$ are traced by $G(s_k)$ with negligible distortions provided that enough observations are available. To resolve a narrow peak, more

observations are needed than for resolving a wide peak, since $\Delta s_k = 1/N$ in (6.21).

3. Distortions are not confined to a small neighbourhood of the peaks, rather $|G(s'_k)| \gg F(s)$ at larger distances away from the peaks. These distortions even occur when N is large enough to trace the narrowest peak.

Since leakage depends (at least partially) on N, i.e., on the number of observations in applications, restricting observations to a finite time slice, e.g., $-(N-1)/2 \le t \le (N-1)/2$, of sequence (g_t) is modelled in (6.94).

Sequence (g_t), as used below in (6.93), (6.94) and (6.95), is defined in (6.82) for $t = \ldots, -1, 0, 1, \ldots$, and consequently, its Fourier transform $G(s)$ is obtained using (6.49). (g_t) originates from $f(t)$ as required in (6.82), i.e., $f(t)$ is a function with real argument t, being in $L^2(-\infty, \infty)$, and having Fourier transform $F(s)$ as defined in (6.81). Please bear in mind that $f(t)$ and $F(s)$ are a Fourier transform pair as defined in (6.35), whereas (g_t) and $G(s)$ are a Fourier transform pair as defined in (6.49) and (6.50). In $-1/2 \le s \le 1/2$, $F(s) \approx G(s)$, as concluded from (6.91), on condition that a possible aliasing can be neglected, as demonstrated in Figs. 6.17, 6.19 and 6.20. The mathematical model for limiting the observational period to a finite time slice of (g_t), e.g., $-(N-1)/2 \le t \le (N-1)/2$ as proposed in (6.94), applies, however, to every (g_t) as obtained in (6.82) from $f(t)$; in particular, it does not depend on $F(s)$ and a possible aliasing. These assumptions are not restrictive.

$$G(s) = \mathcal{F}_{-\mathrm{i}}(g_t) = \sum_{t=-\infty}^{\infty} g_t \mathrm{e}^{-\mathrm{i}2\pi st} \quad (g_t) \text{ as in (6.82)} \tag{6.93}$$

$$G(s) * D_N(s) = \sum_{t=-\infty}^{\infty} \Big((g_t) \big(\Pi_N^{(e)}(t) \big) \Big) \mathrm{e}^{-\mathrm{i}2\pi st} \tag{6.94}$$

$$\int_{-1/2}^{1/2} G(r) D_N(s-r) \mathrm{d}r = \sum_{t=-(N-1)/2}^{(N-1)/2} g_t \mathrm{e}^{-\mathrm{i}2\pi st} \quad \begin{array}{l}\text{convolved as} \\ \text{in (6.144)}\end{array} \tag{6.95}$$

The finite duration of a measurement is modelled in (6.94) by multiplying (g_t) with the even rectangle sequence $\Pi_N^{(e)}(t)$ as defined in (6.52): the observed (g_t), $t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$, are weighted with $\Pi_N^{(e)}(t) = 1$, the non-observed (g_t) with $\Pi_N^{(e)}(t) = 0$. Consequently, only the observed (g_t) contribute to the Fourier sum on the right side in (6.94).

The left side in (6.94) is obtained with a version of the convolution theorem (6.48,7) as derived in (6.145) for the case of a Fourier transform pair as in (6.49) and (6.50): convolving $G(s)$ and $D_N(s)$ in the frequency domain is the counterpart of multiplying (g_t) and $\Pi_N^{(e)}(t)$ in the time domain, since $G(s)$ and $D_N(s)$ are the Fourier transforms of (g_t) and $\Pi_N^{(e)}(t)$. Thereafter, (6.95) follows. From (6.95) it is concluded that $G(s_k) = \int_{-1/2}^{1/2} G(r) D_N(s_k - r) \mathrm{d}r$,

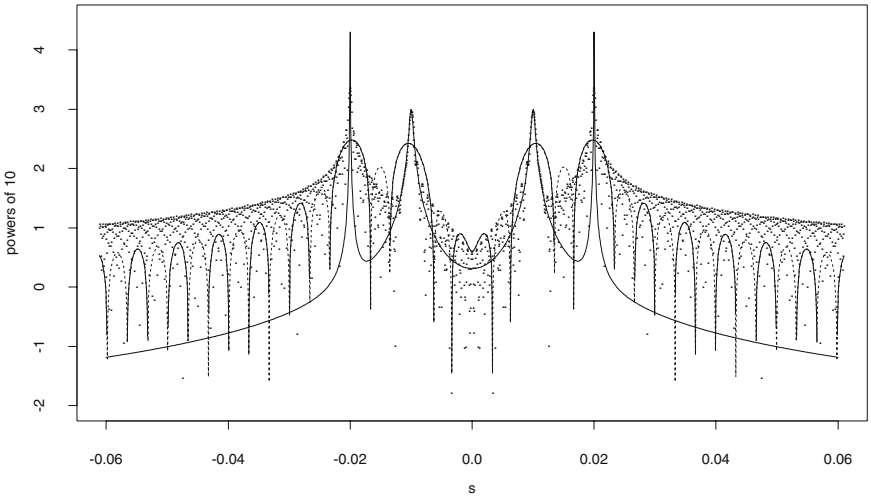


Fig. 6.21. Effects of the main lobe in $D_N(s)$ in (6.95). Function $F(s) \approx G(s)$ as in Figs. 6.17 as well as in Figs. 6.19 and 6.20 together with the absolute value of convolutions $G(s) * D_N(s)$ as required in (6.95) for $N = 301$ (solid when positive, broken when negative) and $N = 2401$ (with dots).

with $G(s_k)$ being a discrete Fourier transform of (g_t) in a finite time interval $t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N+1)/2$, i.e., that $G(s_k)$ is a distortion of $G(s)$ generated by its convolution with the Dirichlet kernel $D_N(s)$ (6.55), having properties (6.58) and (6.59). This distortion is called *leakage*.

For example, the sums in (6.95) (in the exponential version of the discrete Fourier transform) and (6.80) (in the cosine version) become identical for frequencies s_k provided that (g_t) as defined in (6.79) is substituted in (6.95).

The main lobe and the side lobes of $D_N(s)$ have distinct effects on the convolution integral $\int_{-1/2}^{1/2} G(r) D_N(s-r) \mathrm{d}r$ in (6.95). When convolved with $D_N(s)$, peaks in $G(s)$ that are narrower than the main lobe in $D_N(s)$ are blurred, i.e., they become less high and wider. In Fig. 6.21, for example, $F(s)$ as in Figs. 6.19 and 6.20 is plotted together with convolutions $G(s) * D_N(s)$ in (6.95) for $N = 301$ and $N = 2401$. Remember that $F(s) \approx G(s)$ since aliasing can be neglected, as argued in the remarks to Fig. 6.19. The width of the secondary peak in $G(s)$ is larger than the width of the main lobe in $D_N(s)$ for $N = 2401$, and thus the secondary peak persists under the convolution with $D_N(s)$: $G(s) * D_N(s) \approx G(s) \approx F(s)$ in Fig. 6.21 and $G(s'_k) \approx F(s'_k)$ in Fig. 6.19 locally within a small neighbourhood of the secondary peak of $F(s)$. For $N = 301$ however, the secondary peak is blurred by the convolution, as seen in Figs. 6.19 and 6.21. The primary peak of $F(s)$, being narrow, is blurred when convolved with $D_N(s)$ for both $N = 301$ and $N = 2401$. For

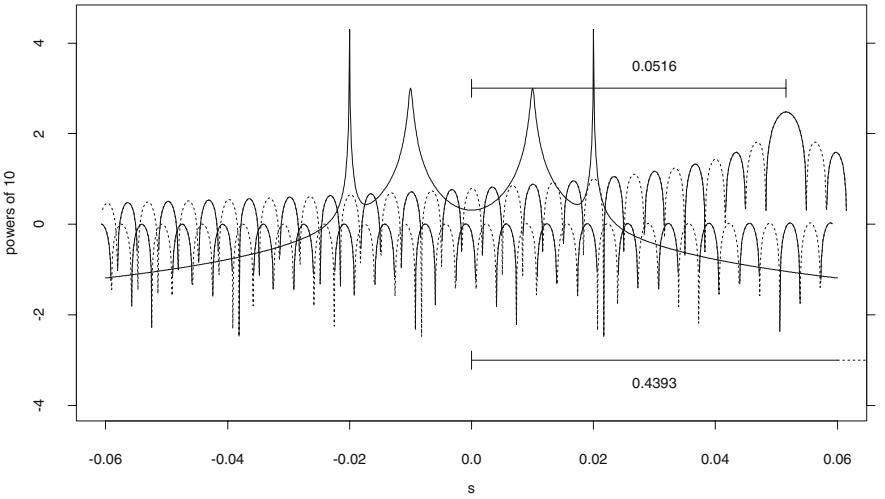


Fig. 6.22. Effects of the side lobes in $D_N(s)$ in (6.95). Function $F(s) \approx G(s)$ as in Figs. 6.17 as well as Figs. 6.19 and 6.20 together with $D_N(s - s_1)$ and $D_N(s - s_2)$ for $N = 301$ and displacements $s_1 = 31/601 \approx 0.01516$ and $s_2 = 264/601 \approx 0.4393$.

N larger by an order of magnitude, however, even the primary peak persists when convolved with $D_N(s)$, as shown for $N = 65535$ in Fig. 6.20 (b).

A side lobe in $D_N(s)$ becomes coincident with a peak in $G(s)$ when the displacement s in convolution integral $\int_{-1/2}^{1/2} G(r) D_N(s-r) \mathrm{d}r$ (6.95) becomes larger than approximately the widths of the main lobe in $D_N(s)$ and the peak in $G(s)$. For these displacements, this convolution integral oscillates and, possibly, $|\int_{-1/2}^{1/2} G(r) D_N(s-r) \mathrm{d}r| \gg |F(s)|$ for frequencies in a large distance from a peak in $F(s)$, since (i) the convolution integral is dominated by the primary peak in $G(s)$, and (ii) the side lobes in $D_N(s)$ alternate in their sign with very slowly decaying absolute values.

In Fig. 6.22, for example, $F(s) \approx G(s)$ (as in Fig. 6.21) is plotted together with duplicates of $D_N(s)$, $N = 301$, for displacements $s_1 = 31/601 \approx 0.0516$ and $s_2 = 264/601 \approx 0.4393$. Displacements being multiples of $1/601 = 1/N' = \Delta s'_k$ are chosen because s'_k is the resolution of the discrete Fourier transform $G^{(s)}(s'_k)$ of the time slice $(g_t^{(s)})$ for $N = 301$ (and zero-padded until $N' = 601$) as plotted in Fig. 6.19. Shifting $D_N(s)$ by s_1, a negative side lobe becomes coincident with the primary peak in $F(s)$ and, therefore, convolution integral $\int_{-1/2}^{1/2} G(r) D_N(s-r) \mathrm{d}r$ (6.95) becomes negative, as is apparent in Fig. 6.21. For a slightly larger displacement of $D_N(s)$, however, a positive side lobe becomes coincident with the primary peak in $F(s)$ and the convolution integral becomes positive. $\int_{-1/2}^{1/2} G(r) D_N(s-r) \mathrm{d}r$ fluctuates from positive to negative values and vice-versa because it is dominated by

the primary peak located at frequency $s = 0.02$ in $F(s)$ as $F(0.02) \gg F(s)$ for s not in a very small neighbourhood of $s = 0.02$. The primary peak in $F(s)$ dominates, however, the convolution integral not only for moderate but also for large displacements of $D_N(s)$, e.g., for $s_2 = 264/601 \approx 0.4393$, a side lobe, still being relatively high, aligns with the primary peak in $F(s)$, as is obvious from Fig. 6.22. For this reason, $|G(s'_k)| \gg F(s)$ for $N = 301$, $N = 2401$ and $N = 65535$ in Figs. 6.19 and 6.20.

Convolution $G(s) * D_N(s)$ in (6.94) and (6.95) is derived for functions $f(t)$, $F(s)$ and $G(s)$ as well as for sequence (g_t) as in (6.81), (6.82) and (6.83). Possible consequences of $G(s) * D_N(s)$ are: (i) a blurring of high and sharp peaks in $G(s)$, (ii) oscillations in $G(s) * D_N(s)$, and (iii) $|G(s) * D_N(s)| \gg |G(s)|$ for frequencies in large distances from a peak in $F(s)$. These distortions are due to leakage. Extending the observational period by increasing N, $G(s) * D_N(s) \approx G(s)$ at the peaks in $F(s)$ is arrived at; for large N, however, still $|G(s) * D_N(s)| \gg |G(s)|$ results for frequencies in a larger distance from the peaks. This leakage can be reduced as demonstrated in Figs. 6.27 and 6.28 on condition that $D_N(s)$ in $G(s) * D_N(s)$ is substituted with an even function $H_N^{(e)}(s)$ which has smaller side lobes that $D_N(s)$.

Convolution $G(s) * D_N(s)$ in (6.94) and (6.95) not only produces distortions when $F(s)$ in (6.81) and (6.82) has sharp peaks and large differences in its maximum and minimum but also when $F(s)$ has a jump discontinuity. For example, sequence $(g_t) = \sin(0.1\pi t)/(0.1\pi)$, $t = \ldots, -1, 0, 1, \ldots$ in Fig. 6.23 (a) and function $G(s)$ in Fig. 6.23 (b) are a Fourier transform pair as proposed in (6.49) and (6.50). $G(s)$ is the rectangle function (6.38) with $b = 0.05$ and $a = 10$ defined in $-1/2 \le s \le 1/2$ and periodic with period 1. The Fourier sum of (g_t) for $t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$, $\sum_{t=-(N-1)/2}^{(N-1)/2} g_t \mathrm{e}^{-\mathrm{i}2\pi st}$ is identical with the Fourier sum $\sum_{t=-\infty}^{\infty} (g_t^{\vdash t \dashv}) \mathrm{e}^{-\mathrm{i}2\pi st}$ on condition that $(g_t^{\vdash t \dashv}) = (g_t)\big(\Pi_N^{(e)}(t)\big)$, i.e., $(g_t^{\vdash t \dashv})$ being the product of (g_t) with the even rectangle sequence $\Pi_N^{(e)}(t)$, as plotted for $N = 127$ in Fig. 6.23 (c). The Fourier transform of $(g_t^{\vdash t \dashv})$ is $G^{\vdash t \dashv}(s)$. For $N = 127$, $\log_{10}|G^{\vdash t \dashv}(s)|$ is plotted in $0 \le s \le 1/2$ in Fig. 6.23 (d).

Bearing in mind that $G(s)$ in Fig. 6.23 (b) is plotted in usual scale whereas the logarithms of the absolute values of $G^{\vdash t \dashv}(s)$ are plotted in Fig. 6.23 (d), the effect of the convolution of $G(s)$ with $D_N(s)$, $N = 127$, to arrive at $G^{\vdash t \dashv}(s)$, becomes apparent when these plots are compared. The main lobe in $D_N(s)$ blurs the jump of $G(s)$ in $s = 0.05$ and, due to the side lobes in $D_N(s)$, $|G^{\vdash t \dashv}(s)| \approx 10^{-1}$ for $s \ge 0.05$ despite $G(s) = 0$ for $s \ge 0.05$. In addition, the side lobes in $D_N(s)$ generate small oscillations in $G^{\vdash t \dashv}(s)$ for s smaller than (but close to) the frequency $s = 0.05$ with the discontinuity in $G(s)$. These small oscillations which are often visible as an overshoot in $G^{\vdash t \dashv}(s)$ are called *Gibbs phenomenon*.

Aliasing and leakage have been introduced in this section using mathematical models describing (i) the discretisation of time when observing a function $f(t)$ with real argument t and being in $L^2(-\infty, \infty)$ as defined in

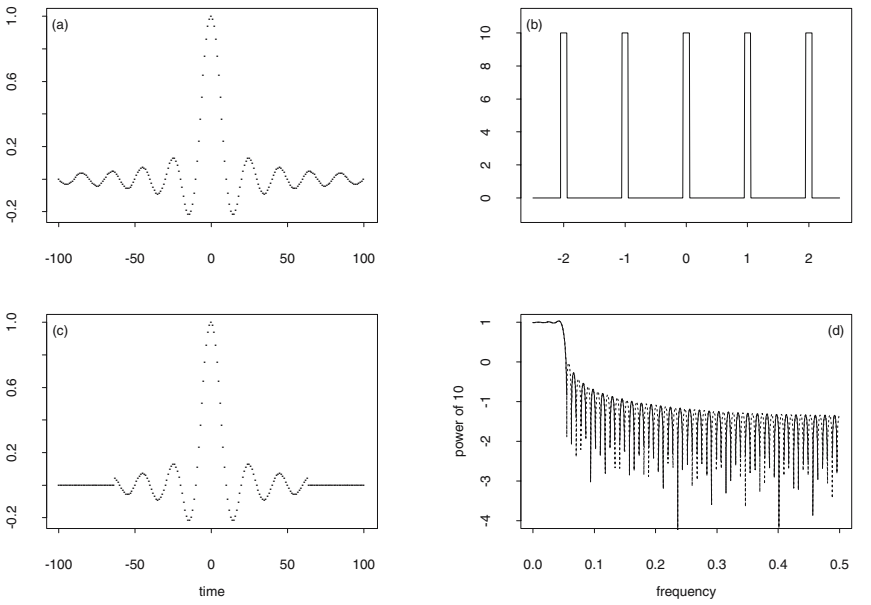


Fig. 6.23. Fourier transform pairs as in (6.49) and (6.50). On the left in plots (a) and (c), sequences of the argument $t = \ldots, -1, 0, 1, \ldots$; on the right in plots (b) and (d), functions defined in $-1/2 \leq s \leq 1/2$ and periodically continued on the exterior of this interval: (g_t) and $G(s)$ (a,b) as well as $(g_t^{\vdash t \dashv}) = (g_t)\big(\Pi_{127}^{(e)}(t)\big)$ and $G^{\vdash t \dashv}(s)$ (c,d). Plot (d) shows $\log_{10}|G^{\vdash t \dashv}(s)|$ for $0 \leq s \leq 1/2$, using broken lines when $G^{\vdash t \dashv}(s) < 0$.

(6.82) and (6.84) and (ii) the finite duration of the period with observations in (6.94). Both, aliasing and leakage, depend on the form of $F(s)$, i.e., the Fourier transform of the function being observed, as demonstrated by the examples in Figs. 6.15, 6.17, 6.19, 6.20 and 6.23. How can the form of a function be described?

6.6 Width of Functions in a Fourier Transform Pair

The form of a function or sequence can be described using the characteristics to be defined in this section. For functions with at least one peak, the dynamic range and the width at half height as defined in Sect. 6.6.1 can be calculated. Other widths (equivalent, σ-, and autocorrelation) are defined in Sects. 6.6.2 and 6.6.3 for functions in $L^2(-\infty, \infty)$ (with further restrictions in the event of the equivalent width and σ-width). Using these definitions, more light is shed on the observation that "wide" functions have a "narrow" Fourier transform, and vice-versa. This reciprocity relationship is captured

by, e.g., the similarity theorem (6.98) or the fundamental uncertainty relationship (6.107). For sequences, a related formulation can be obtained using the definitions of band- and time-limited sequences in Sect. 6.6.4. Both terms, wide and narrow, although being somewhat vague, will be used without quotation marks to describe roughly the form of a function in the following parts of the book.

6.6.1 Dynamic Range and Width at Half Height

The dynamic range and the width at half height are used to describe the form of a real-valued function $f(x)$ with real argument x having at least one peak. The dynamic range is defined in (6.96), the width at half height in (6.97).

A function having either a high peak or a deep valley with steep sides has a large dynamic range, whereas a function having only small peaks and valleys has a small dynamic range.

Let $f(x)$ be a real-valued function with $f(x) > 0$ in $x_1 \le x \le x_2$.
Then $\mathrm{dyn}\big(f(x)\big) = 10\log_{10}\left(\dfrac{\max_{x_1\le x\le x_2} f(x)}{\min_{x_1\le x\le x_2} f(x)}\right)$ (6.96)
is called dynamic range of $f(x)$ in $x_1 \le x \le x_2$.

$\mathrm{dyn}\big(f(x)\big)$ is measured in decibel (dB). The dB is a logarithmic unit used to describe a ratio, e.g., 10 dB are an order of magnitude.

For example, $F(s)$ as plotted in Figs. 6.19 and 6.20 has a large dynamic range: approximately $(10(4-(-1)) = 50\ dB$ for $0 \le s \le 0.06$, and approximately $(10(4-(-3)) = 70\ dB$ for $0 \le s \le 0.5$ as is obvious from the plots in logarithmic scale. In contrast, the Gaussian function $f(t) = \mathrm{e}^{-d\pi t^2}$ with $d = 1/4$ as plotted in Fig. 6.10 has a small dynamic range as it increases, in $-1 \le t \le 1$, from 0.5 to 1.0 and thereafter decreases to 0.5: the dynamic range of $f(t)$ in this interval is approximately 3 dB.

Using the dynamic range, i.e., the logarithm of the ratio of the maxima and minima of a function, details in the form of a function can no longer be captured. For example, the twin peak of $F(s)$ as plotted in Figs. 6.19 and 6.20 or the jump discontinuity of $G(s)$ in Fig. 6.23 (b) cannot be inferred from the large dynamic range of both functions. Consequently, a further definition is required for the description of functions with peaks and valleys.

Let $F(s)$ be a real-valued function with a maximum (a peak) in s_i.
Then $\mathcal{B}_{\triangle}\big(F(s)\big) = s_\jmath - s_\imath$ is called width at half height of $F(s)$
on condition that $s_\imath < s_i$ and $s_i < s_\jmath$ do exist such that (6.97)
$F(s_\imath) = F(s_\jmath) = F(s_i)/2$. The width at half height of
a minimum (a valley) in $F(s)$ is calculated accordingly.

The width at half height of a function with more than one peak and/or valley is the width at half height of the narrowest peak or valley. If a function $H(s)$

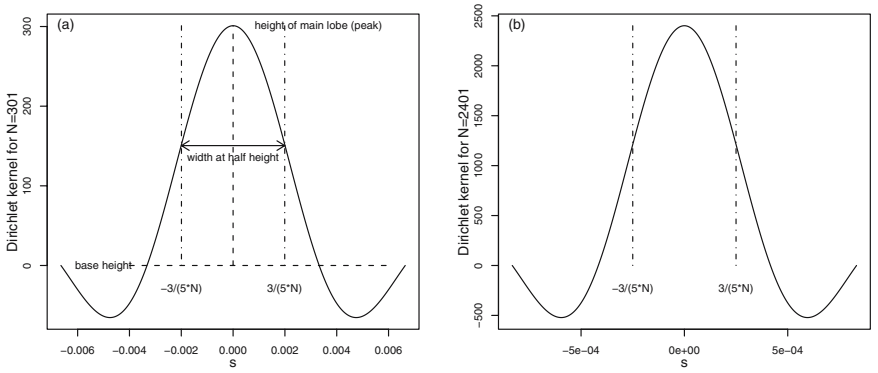


Fig. 6.24. Dirichlet kernels for $N = 301$ (a) and $N = 2401$ (b) together with their widths at half height.

has main and side lobes as do the Dirichlet and the Fejer kernels in (6.55) and (6.57), then $\mathcal{B}_{\triangle}^{\cap}\big(H(s)\big)$ is the width at half height of the main lobe of $H(s)$.

As a first example, $\mathcal{B}_{\triangle}^{\cap}\big(D_N(s)\big)$, the width at half height of the main lobe of the Dirichlet kernel (6.55), is obtained by plotting $D_N(s)$ in non-logarithmic scale (as required in (6.97)) in the interval $-2/N \le s \le 2/N$ for $N = 301$ and $N = 2401$ in Fig. 6.24. The main lobe of $D_N(s)$ with height $D_N(s) = N$ for $s = 0$ as in (6.58,5) is confined symmetrically on both sides by defining $D_N(s) = 0$ as base height, and, given this base height, (6.58,4) implies the interval $-1/N \le s \le 1/N$. In this interval, $D_N(s) = N/2$ for $s_i \approx -3/(5N)$ and $s_j \approx 3/(5N)$ and, therefore, $\mathcal{B}_{\triangle}^{\cap}\big(D_N(s)\big) \approx 6/(5N)$. In Fig. 6.24, $\mathcal{B}_{\triangle}^{\cap}\big(D_{301}(s)\big) \approx 0.004$ and $\mathcal{B}_{\triangle}^{\cap}\big(D_{2401}(s)\big) \approx 0.0005$.

An alternative base height could be used to confine the main lobe of $D_N(s)$. For example, using the minima of the valleys on both sides of the main lobe, another value for $\mathcal{B}_{\triangle}^{\cap}\big(D_N(s)\big)$ would be obtained. This example demonstrates that the width at half height of a function $f(x)$ depends on how the peaks in $f(x)$ are confined. For the Fejer kernel $F_N(s)$ in (6.57), $\mathcal{B}_{\triangle}^{\cap}\big(F_N(s)\big) \approx 7/(8N)$ is obtained also using base height $F_N(s) = 0$.

As a second example, the width at half height of $F(s)$ in (6.78) with $s_1 = 0.01$, $s_2 = 0.02$, $a_1 = 0.002$ and $a_2 = 0.0001$ is determined. $F(s)$ is plotted in Figs. 6.17, 6.19 and 6.20, and, because it has two peaks, its width at half height is the minima of those widths obtained for the two peaks and the valley in-between, as required in the remark to (6.97). To calculate these widths, $F(s)$ is plotted with a sufficiently high resolution, e.g., as in Fig. 6.25, base height 0 is chosen (although $F(s) \approx 0$ in $0.01 \le s \le 0.02$ as seen in Fig. 6.21), and

1. for the left peak with $F(s) \approx 1000$ for $s_1 = 0.01$, $F(s) \approx 500$ for $s_{1l} = 0.0097$ and $s_{1r} = 0.0103$, implying $s_{1r} - s_{1l} = 6 \times 10^{-4}$,

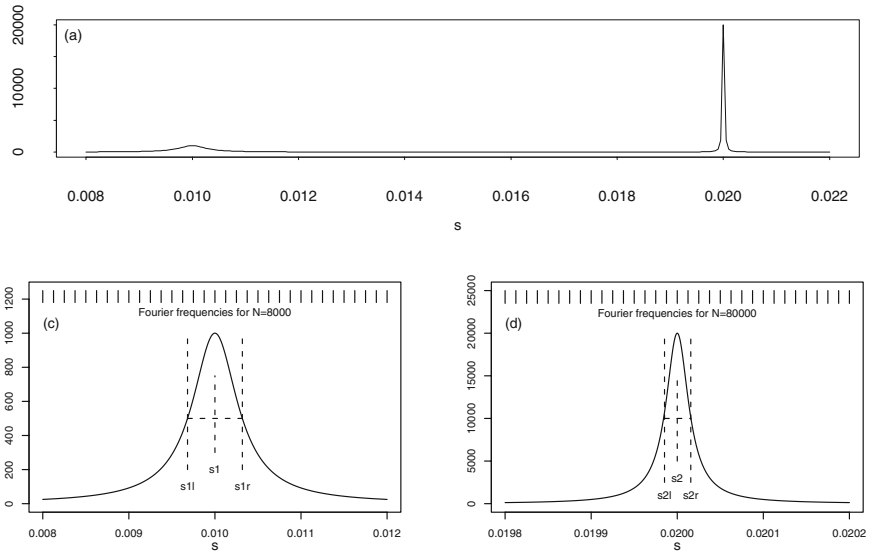


Fig. 6.25. Function $F(s)$ in Figs. 6.17, 6.19 and 6.20 (there plotted in logarithmic scale) for the interval $0.08 \leq s \leq 0.022$ (a); below peaks of $F(s)$ together with their widths at half height and Fourier frequencies for $N = 8000$ (c) and $N = 80000$ (d). The rationale for plotting the Fourier frequencies is diagnostic (6.116).

2. for the right peak with $F(s) \approx 20000$ for $s_2 = 0.02$, $F(s) \approx 10000$ for $s_{2l} = 0.019985$ and $s_{2r} = 0.020015$, implying $s_{2r} - s_{2l} = 3 \times 10^{-5}$, and
3. for the valley in-between as difference of the frequencies at which $F(s)$ is at half heights of the left and the right peak, $s_{2l} - s_{1r} \approx 1 \times 10^{-2}$,

are arrived at. Frequencies $s_{1l}, s_1, s_{1r}, s_{2l}, s_2$ and s_{2r} are shown in plots (c) and (d) in Fig. 6.25. Consequently, $\mathcal{B}_{\triangle}\big(F(s)\big) = 3 \times 10^{-5}$, as required in the remark to (6.97).

6.6.2 Equivalent Width and σ-Width

Let $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ be a Fourier transform pair as defined in (6.35). Then $f(t)$ is narrow on condition that $F(s)$ is wide, and vice-versa, as demonstrated by (i) the Gaussian functions in Fig. 6.10, (ii) the damped oscillations $f(t)$ and their Fourier transforms $F(s)$ in Figs. 6.15 and 6.17, and (iii), for a Fourier transform pair as defined in (6.49) and (6.50), the even rectangle sequence $\Pi_N^{(e)}(t)$ (6.52) and the Dirichlet kernel $D_N(s)$ (6.55). This reciprocity is described by propositions (6.48,3), (6.99,2), (6.107) and (6.109). The introduction of these reciprocity relationships for Fourier transform pairs commences in this section with (6.98) (a version of (6.48,3)) and (6.99,2), whereas (6.107) and (6.109) are proposed in Sects. 6.6.3 and 6.6.4.

(6.48,3) is derived in (6.139), where, in the remarks, also a more symmetric version of (6.48,3) is given, which is used to reformulate (6.48,3) in (6.98).

$$\begin{aligned}&\textit{Similarity theorem: If } F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big) \textit{ as defined}\\ &\textit{in (6.35) then } \mathcal{F}_{-\mathrm{i}}\big(|a|^{1/2} f(at)\big) = (1/|a|^{1/2}) F(s/a).\end{aligned} \tag{6.98}$$

From (6.98) it is concluded that $|a|^{1/2} f(at)$ and $(1/|a|^{1/2})F(s/a)$ are a Fourier transform pair on condition that $f(t)$ and $F(s)$ are a Fourier transform pair, i.e., that $f(t)$ shrinks when $F(s)$ expands, and vice-versa, as demonstrated in Fig. 6.10 using the Gaussian functions. (6.98) applies to all cases of Fourier transform pairs as defined in (6.20), not only for pairs as defined in (6.35).

In (6.99), the reciprocity in a pair of Fourier transforms as defined in (6.35) is described, using the equivalent widths of $f(t)$ and $F(s)$. The equivalent width $\mathcal{B}_\sqcap\big(f(t)\big)$ of function $f(t)$ is defined, in (6.99,1), as the width of a rectangle with height $f(0)$ and area $\int_{-\infty}^{\infty} f(t)\mathrm{d}t$, i.e., a rectangle function as defined in (6.38), having height $f(0)$, width $\mathcal{B}_\sqcap\big(f(t)\big)$ and area $\int_{-\infty}^{\infty} f(t)\mathrm{d}t$, acts as a substitute for $f(t)$. Hence, a geometrically meaningful equivalent width can be calculated only for functions with properties as required in (6.99,1), provided that $f(0)$ and the integral exist.

1. *Let $f(t)$ be in $L^2(-\infty,\infty)$, real-valued, continuous and positive for all t, and have its maximum in $t = 0$, then $\mathcal{B}_\sqcap\big(f(t)\big) = \big(\int_{-\infty}^{\infty} f(t)\mathrm{d}t\big)/f(0)$ is called equivalent width of $f(t)$.*
2. *If, in addition, $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ then (6.48,1) implies*

$$\mathcal{B}_\sqcap\big(f(t)\big) = \frac{\int_{-\infty}^{\infty} f(t)\mathrm{d}t}{f(0)} = \frac{F(0)}{\int_{-\infty}^{\infty} F(s)\mathrm{d}s} = \frac{1}{\mathcal{B}_\sqcap\big(F(s)\big)} \tag{6.99}$$

(6.99,2), which can be also be written $\mathcal{B}_\sqcap\big(f(t)\big) \times \mathcal{B}_\sqcap\big(F(s)\big) = 1$, also holds for Fourier transform pairs as defined in (6.49) and (6.50) having a sum of $f(t)$ substituted for the integral, since (6.48,1) also applies in this case. For example, the equivalent widths of the even triangle sequence $\Lambda^{(e)}_{2N+1}(t)$ as defined in (6.56) and of its Fourier transform $F_N(s)$ as proposed in (6.57) are calculated, resulting in $\mathcal{B}_\sqcap\big(\Lambda^{(e)}_{2N+1}(t)\big) = N/1 = N$ and $\mathcal{B}_\sqcap\big(F_N(s)\big) = 1/N$, owing to (i) $\sum_{t=-\infty}^{\infty} \Lambda^{(e)}_{2N+1}(t) = N$ as well as $F_N(s) = N$ for $s = 0$, and (ii) $\Lambda^{(e)}_{2N+1}(t)\big) = 1$ for $t = 0$ as well as $\int_{-\infty}^{\infty} F_N(s)\mathrm{d}s = 1$ as proposed in (6.59). Consequently, $\mathcal{B}_\sqcap\big(\Lambda^{(e)}_{2N+1}(t)\big) \times \mathcal{B}_\sqcap\big(F_N(s)\big) = 1$ as required in (6.99,2). Please show, as a straightforward exercise, that (6.99,2) holds for $\Pi^{(e)}_N$ (6.52,2) and its Fourier transform $D_N(s)$ (6.55).

In the case of a Fourier transform pair as in (6.35), (6.99,2) is demonstrated by calculating the equivalent widths of Gaussian functions (6.47) being plotted for some d in Fig. 6.10. Using (6.99,1), $\mathcal{B}_\sqcap\big(f(t)\big) = 1/\sqrt{d}$ and $\mathcal{B}_\sqcap\big(F(s)\big) = \sqrt{d}$ are obtained, in accordance with (6.99,2).

Gaussian function $F(s)$ in Fig. 6.10 (b) becomes, subsequent to substituting $\sigma\sqrt{2\pi}$ for $\sqrt{d}$, f for F and x for s, $f(x) = (\sigma\sqrt{2\pi})^{-1}\mathrm{e}^{-(1/2)(x/\sigma)^2}$, the probability density of the normal distribution with expectation $\mu = 0$ and standard deviation σ. Consequently, $\mathcal{B}_{\sqcap}\big(f(x)\big) = \sigma\sqrt{2\pi}$ is obtained as the equivalent width of this normal density, and clearly, $\mathcal{B}_{\sqcap}\big(f(x)\big) \neq \sigma^2$, i.e., the equivalent width is not identical with the second moment of $f(x)$, even though both are connected to the concentration of this distribution, as shown in the remarks to (1.17).

An equivalent width can be calculated for every probability density with expectation being identically zero, because a probability density has, provided that $\mu = 0$, the properties required in (6.99,1). In general, however, the equivalent width is not identical with the second moment of a function, as is obvious from a comparison of definitions (6.99,1) and (6.100).

Let $f(t)$ be in $L^2(-\infty,\infty)$, real-valued and positive for all t. Then:
1. $\mu_f = \int_{-\infty}^{\infty} tf(t)\mathrm{d}t$ *is called the first moment, and*
2. $\sigma_f^2 = \int_{-\infty}^{\infty}(t-\mu_f)^2 f(t)\mathrm{d}t$ *the second moment of $f(t)$.* (6.100)

Using the second moment of a function with a first moment being identically zero, the σ-width of a function is defined in (6.101).

Let $f(t)$ be in $L^2(-\infty,\infty)$, real-valued and positive for all t with $\mu_f = 0$ and $\int_{-\infty}^{\infty} f(t)\mathrm{d}t = 1$. Then
$\mathcal{B}_\sigma\big(f(t)\big) = \big(12\int_{-\infty}^{\infty} t^2 f(t)\mathrm{d}t\big)^{1/2}$ *is called σ-width of $f(t)$.* (6.101)

When definitions (6.101) and (6.100),2) are compared, it is concluded that $\mathcal{B}_\sigma\big(f(t)\big)$ is proportional with $2\sqrt{3}$ to σ_f, the root of the second moment of $f(t)$. The factor $2\sqrt{3}$ makes the σ-width comparable to the width of a rectangle function with unit area. This is easily shown by integrating a rectangle function $f(t)$ as defined in (6.38) with $a = 1/(2b)$ to obtain its second moment $\sigma_f^2 = b^2/3$ from which $\sqrt{12}\sigma_f = 2\sqrt{3}\sigma_f = 2b$ is obtained as its σ-width.

6.6.3 Autocorrelation Width and Uncertainty Relationship

Since $\mu_f = 0$ is assumed in (6.101), the σ-width is geometrically meaningful only when computed for a function $f(x)$ having its maximum in or close to $x = 0$, a property shared with the equivalent width as defined in (6.99,1). To describe a function $f(x)$ having a peak not close to $x = 0$, either the width at half height (6.97) is calculated, or, as an alternative, the autocorrelation width as defined in (6.104), subsequent to the preliminaries in (6.102) and (6.103).

When convolving sequences (a_i) and (b_i) as required in (2.28,2), the order of the values in either (a_i) or (b_i) is reversed, as demonstrated when convolving using the paper strip device (2.38). When convolving functions $f(t)$ and

$g(t)$ with real argument t as required in (2.28,1), the direction of the t-axis in either $f(t)$ or $(g(t)$ is reversed, as demonstrated in Problem 2.11. In a similar manner, however without changing the sign of the argument in one of the sequences or functions, the $\star$-transformation is defined in (6.102).

Let $f(t)$ and $g(t)$ be real-valued functions as well as (f_t) and (g_t) be real-valued sequences in $L^2(-\infty,\infty)$. Then:

1. $h(t) = \big(f(t)\big) \star \big(g(t)\big) = \int_{-\infty}^{\infty} f(u)g(t+u)\mathrm{d}u$
2. $(h_t) = (f_t) \star (g_t) = \sum_{u=-\infty}^{\infty} f_u g_{t+u}$ (6.102)

The $\star$-transformation has properties similar to those of the convolution, with the following differences.

Convolution theorem (6.48,7) is derived twice: (i) in (6.143) for Fourier transform pairs as in (6.35), and (ii) in (6.145) for Fourier transform pairs as in (6.49) and (6.50). A corresponding theorem is derived in (6.150) for the $\star$-transformation: if $(h_t) = (f_t)\star(g_t)$ then $H(s) = F(s)\overline{G(s)}$ on condition that $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$, $G(s) = \mathcal{F}_{-\mathrm{i}}\big(g(t)\big)$ and $H(s) = \mathcal{F}_{-\mathrm{i}}\big(h(t)\big)$. This theorem also applies to sequences being $\star$-transformed.

The convolution is commutative whereas the $\star$-transformation is not commutative, as shown in the remarks to (2.35).

Using the $\star$-transformation, the autocorrelation of a deterministic function is defined in (6.103).

Let $f(t)$ be a real-valued function in $L^2(-\infty,\infty)$. Then

$$f\star f(t) = \big(f(t)\big)\star\big(f(t)\big) = \begin{cases} \int_{-\infty}^{\infty} f(u)f(u+t)\mathrm{d}u & \text{for real } t \\ \sum_{u=-\infty}^{\infty} f_u f_{u+t} & \text{for integer } t \end{cases} \tag{6.103}$$

is called autocorrelation of $f(t)$.

For example, $h\star h(t) = 10$ for $t = 0$ in Fig. 6.34 (c) and, as implied by this example, $f \star f(t) > 1$ is not excluded in (6.103) whereas, in Statistics, a correlation is always between -1 and 1.

From definition (6.103) it is concluded that $f\star f(t) = f\star f(-t)$, and therefore, $f\star f(t)$ is an even function as defined in (6.32). Consequently, applying (6.37,1), the Fourier transform of $f\star f(t)$ is real-valued, with $\mathcal{F}_{-\mathrm{i}}\big(f\star f(t)\big) = F(s)\overline{F(s)} = |F(s)|^2$, with $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$, as shown in the remarks to (6.150). Further, $f\star f(t)$ is maximal in $t = 0$ as derived in (6.151).

As an aside, an autocorrelation shares the property of being an even function with (i) the covariance and correlation functions of a stochastic process since (2.4,3) implies (2.7,1), and (ii) the empirical covariance and correlation functions of a time series as a direct consequence of (2.1,3). In contrast, the self-convolution $f*f(t)$ of a function $f(t)$ is, in general, a mixed function as defined in (6.32), and $\mathcal{F}_{-\mathrm{i}}\big(f*f(t)\big) = F(s)F(s)$ in (6.48,6) is complex-valued. Examples are plotted in Figs. 6.33, 6.34 and 6.35.

Owing to these properties, i.e., $\int_{-\infty}^{\infty} f\star f(t)\mathrm{d}t < \infty$, $f\star f(0) > 0$ and $f\star f(0) \geq f\star f(t)$, the equivalent width of an autocorrelation as defined in

(6.104) is geometrically meaningful in most cases. The third = in (6.104) is obtained applying (6.48,1) and $\mathcal{F}_{-\mathrm{i}}\big(f\star f(t)\big) = F(s)\overline{F(s)} = |F(s)|^2$ as shown in the remarks to (6.150).

Let $f(t)$ be a real-valued function in $L^2(-\infty,\infty)$. Then

$$\mathcal{B}_\star\big(f(t)\big) = \mathcal{B}_\sqcap\big(f\star f(t)\big) = \frac{\int_{-\infty}^{\infty} f\star f(t)\mathrm{d}t}{f\star f(0)} = \frac{F(0)\overline{F(0)}}{\int_{-\infty}^{\infty}\big(F(s)\overline{F(s)}\big)\mathrm{d}s} \tag{6.104}$$

is called autocorrelation width of $f(t)$.

For example, the equivalent width, the σ-width and the autocorrelation width of the sequences on the left in Fig. 6.36 are computed in Problem 6.27.

The functions and/or sequences in a Fourier transform pair are subject to a reciprocity: the function or sequence in the time domain is narrow on condition that its Fourier transform in the frequency domain is wide, and vice-versa, as demonstrated by the examples listed in the introduction to Sect. 6.6.2. This reciprocity is described by the similarity theorem in (6.98), applying the equivalent widths in (6.99,2) or, as a third possibility, by the fundamental uncertainty relationship in (6.107). In (6.107), the somewhat special second moments (written as in [18]) are arrived at under the assumption that $f(t)$ and $F(s)$ are a Fourier transform pair as defined in (6.35) using the following arguments. Since, in (6.35), $F(s)$ can be complex-valued when, e.g., $f(t)$ is real-valued, the second moment of both functions cannot be calculated using (6.100), a definition applying to real-valued functions only. However, the squares of the absolute values of both functions, $|f(t)|^2$ and $|F(s)|^2$, are (i) even as required in (6.32) and (ii) real-valued owing to (6.37). Additionally, $\int_{-\infty}^{\infty}|f(t)|^2\mathrm{d}t = \int_{-\infty}^{\infty}|F(s)|^2\mathrm{d}s$ is implied by (6.75).

In addition, $\int_{-\infty}^{\infty}|f(t)|^2\mathrm{d}t = \int_{-\infty}^{\infty}|F(s)|^2\mathrm{d}s = 1$ is assumed. This property is arrived at by standardising $f(t) = f(t)/\Sigma(f^2)$, with $\Sigma(f^2) = \int_{-\infty}^{\infty}|f(t)|^2\mathrm{d}t$. Then $|f(t)|^2$ as well as $|F(s)|^2$ have the properties of a probability density function, and the second moments of both $|f(t)|^2$ as well as $|F(s)|^2$ are obtained in (6.105), provided that $\mu^2_{|f^2|} = \mu^2_{|F^2|} = 0$, as required in (6.101). These moments are then multiplied in (6.106).

$$\sigma^2_{|f^2|} = \int_{-\infty}^{\infty} t^2|f(t)|^2\mathrm{d}t \qquad \sigma^2_{|F^2|} = \int_{-\infty}^{\infty} s^2|F(s)|^2\mathrm{d}s \tag{6.105}$$

$$\sigma^2_{|f^2|}\sigma^2_{|F^2|} = \left(\int_{-\infty}^{\infty} t^2|f(t)|^2\mathrm{d}t\right)\left(\int_{-\infty}^{\infty} s^2|F(s)|^2\mathrm{d}s\right) \tag{6.106}$$

$$\sigma^2_{|f^2|}\sigma^2_{|F^2|} \geq \frac{1}{16\pi^2} \tag{6.107}$$

From (6.106), inequality (6.107) is obtained, e.g., in [18]. This inequality is called *uncertainty relationship*, fundamental uncertainty relationship or Heisenberg's uncertainty relationship.

Assume that $\{f(t)\}$ is the set of all real-valued $f(t)$ in $L^2(-\infty,\infty)$ and $\{F(s)\}$ the set of all possibly complex-valued $F(s)$ with $\mathrm{re}(F(s))$ and

$\mathrm{im}(F(s))$ in $L^2(-\infty,\infty)$. Then the uncertainty relationship restricts the pairs of functions $\big(f(t),F(s)\big)$ in the cross-product of $\{f(t)\}$ and $\{F(s)\}$ to those with $\sigma^2_{|f^2|}\sigma^2_{|F^2|} \geq 1/16\pi^2$ as being feasible in a Fourier transform pair. Thus, the uncertainty relationship sets a lower bound to $\sigma^2_{|f^2|}\sigma^2_{|F^2|}$, $f(t)$ and $F(s)$ being a Fourier transform pair (6.35). The uncertainty relationship has multi-faceted consequences [40], when the lower bound is interpreted as physical limit in Quantum Mechanics by Heisenberg, cf. [65] or [64].

When applied to $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ as defined in (6.35), the uncertainty relationship implies that $f(t)$ is narrow on condition that $F(s)$ is wide, and vice-versa, since, obviously, $|f(t)|^2$ is wide provided that $f(t)$ is also wide and $|f(t)|^2$ is narrow provided that $f(t)$ is narrow as well. Thus, the uncertainty relationship is the third reciprocity relationship for Fourier transform pairs, the first one being (6.98) and the second one (6.99,2). The fourth reciprocity relationship (6.109) is proposed using definitions (6.108) in Sect. 6.6.4.

6.6.4 Time-limited and Band-limited Sequences

With $a = 1$, the rectangle function (6.38) becomes $\sqcap_b(t) = 1$, for $|t| \leq b$, and $\sqcap_b(t) = 0$, for $|t| > b$. The Fourier transform of $\sqcap_b(t)$ is $\mathcal{F}_{-\mathrm{i}}\big(\sqcap_b(t)\big) = \big(2b\sin(2\pi bs)\big)/(2\pi bs)$. $\sqcap_b(t)$ has two discontinuities: it jumps from 0 to 1 in $t = -b$ and then back to 0 in $t = b$. In contrast, $\big(2b\sin(2\pi bs)\big)/(2\pi bs)$ oscillates on both sides of its primary peak taking positive and negative values in $-\infty < s < \infty$. Consequently, no frequency $s_1 > 0$ exists such that $\big(2b\sin(2\pi bs)\big)/(2\pi bs) = 0$ for $|s| > s_1$.

Similarly, the triangle function $g_b(t)$ in (6.63) becomes identically zero at the exterior of $-b \leq t \leq b$ and its Fourier transform $G_b(s)$, being positive for $-\infty < s < \infty$, has an infinite number of secondary peaks separated by $G_b(s) = 0$ on both sides of the main peak, as demonstrated in Fig. 6.13. This behaviour is shared by Fourier transform pairs as defined in (6.49) and (6.50), as demonstrated by the even rectangle and triangle sequences $\mathit{\Pi}^{(e)}_N(t)$ and $\mathit{\Lambda}^{(e)}_N(t)$ as defined in (6.52) and (6.56) together with their Fourier transforms, i.e., the Dirichlet and the Fejer kernels $D_N(s)$ and $F_N(s)$ in (6.55) and (6.57).

However, there are Fourier transform pairs that do not behave in this manner. In both domains, the Gaussian functions $f(t) = \mathrm{e}^{-d\pi t^2}$ and $F(s) = (1/\sqrt{d})\mathrm{e}^{-\pi\frac{1}{d}s^2}$ in Fig. 6.10, for example, are continuous and smooth and do not become identically zero for intervals $|t| > t_1$ or $|s| > s_1$. The difference in their behaviour suggests that the oscillations in the Fourier transforms of the rectangle and triangle functions and sequences are generated by the property that these functions become abruptly identically zero at the exterior of a finite interval in the time domain.

This suggestion is accepted because, hopefully, a closer look at the behaviour of $\mathit{\Pi}^{(e)}_N(t)$ and $D_N(s)$ and similar Fourier transform pairs could result in a means to reduce the height of the side lobes of $D_N(s)$ and, consequently,

any leakage produced by the side lobes of $D_N(s)$ in (6.93) and (6.94). The following investigation begins with definitions (6.108):

1. *A real-valued sequence* $(g_t^{\vdash t \dashv})$, $t = \ldots, -1, 0, 1, \ldots$, *is called time- or index-limited, on condition that* $(g_t^{\vdash t \dashv}) = 0$ *for* $|t| > (N-1)/2$, N *a positive and odd integer number.*
2. *A real-valued sequence* $(g_t^{\vdash s \dashv})$, $t = \ldots, -1, 0, 1, \ldots$, *is called frequency- or band-limited, on condition that (i)* $G^{\vdash s \dashv}(s) = 0$ *for* $|s| > s_0$, $0 < s_0 < 1/2$ *and (ii)* $(g_t^{\vdash s \dashv})$ *and* $G^{\vdash s \dashv}(s)$ *are a Fourier transform pair as in (6.49) and (6.50).*

(6.108)

In (6.108), a superscript $^{\vdash t \dashv}$ denotes a finite interval in the time domain and a $^{\vdash s \dashv}$ denotes a finite interval in the frequency domain. A finite interval in the frequency domain is called a frequency band or band. Consequently, a sequence having property (6.108,2) is called band-limited.

The interval $|t| \leq (N-1)/2$ in (6.108,1) with its centre in $t = 0$ does not induce a fundamental restriction, since the origin of the time domain can be shifted, e.g., in the centre of $N_1 \leq t \leq N_2$ in case of a sequence being identically zero at the exterior of $t = N_1, \ldots, N_2$ with $N_1 < N_2$ such that the interval contains an odd number of time points. For example, the rectangle sequence $\Pi_N(t)$ in (6.52) is time-limited as defined in (6.108,1), on condition that the origin of the time domain is shifted in the centre of $\Pi_N(t)$ and N is odd.

Without shifting the origin of the time domain, the even rectangle and triangle sequences as defined in (6.52) and (6.56), as well as the sequence $(g_t^{\vdash t \dashv})$ in Fig. 6.23 (c), are time-limited sequences as defined in (6.108,1). The Fourier transforms of these sequences are the Dirichlet and the Fejer kernels as well as function $(G^{\vdash t \dashv}(s) = \mathcal{F}_{-\mathrm{i}}(g_t^{\vdash t \dashv})$ as plotted in Fig. 6.23 (d), all of which are not identically zero (except their nulls) in $-1/2 \leq s \leq 1/2$. Consequently, these sequences are not band-limited. In contrast, sequence (g_t) in Fig. 6.23 (a) is band-limited because its Fourier transform $G(s)$ in Fig. 6.23 (b), being a rectangle function as defined in (6.38), becomes identically zero for $|s| > b = 0.05$. Sequence (g_t) is, however, not time-limited.

The above examples demonstrate property (6.109) of real-valued sequences.

A real-valued sequence is either time- or band-limited, i.e., it does not have both properties. (6.109)

(6.109) is easily derived using convolution theorems (6.48,6,7) which also apply to Fourier transform pairs as defined in (6.49) and (6.50).

Firstly, it is shown that a time-limited sequence is not band-limited. Let $(g_t^{\vdash t \dashv})$ be a time-limited sequence as required in (6.108,1). Then a sequence (g_t), $t = \ldots, -1, 0, 1, \ldots$, exists such that $g_t^{\vdash t \dashv} = g_t$ for $t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$. At the exterior of this interval, (g_t) is arbitrary, possibly it may even have the property that a frequency s_0 exists

such that $G(s) = 0$ for $|s| > s_0$, $0 < s_0 < 1/2$, i.e., (g_t) may be band-limited. From (g_t), $(g_t^{\vdash t \dashv})$ is obtained as $(g_t^{\vdash t \dashv}) = (g_t)\Pi_N^{(e)}(t)$ in the time domain, and the convolution theorems imply $G^{\vdash t \dashv}(s) = G(s) * D_N(s)$ in the frequency domain. In $-1/2 \le s \le 1/2$, $D_N(s)$ becomes identically zero only for $s = \pm 1/N, \pm 2/N, \pm 3/N, \ldots$ as required in (6.58) which implies that $G(s) * D_N(s) = 0$ only for isolated frequencies s, i.e., at its nulls. Consequently, there is no $0 < s_0 < 1/2$ such that $G^{\vdash t \dashv}(s) = 0$ for $|s| > s_0$ i.e., $(g_t^{\vdash t \dashv})$ is not band limited. This is demonstrated by the sequences (g_t) and $(g_t^{\vdash t \dashv})$ and their Fourier transforms $G(s)$ and $G^{\vdash t \dashv}(s)$ in Fig. 6.23.

Secondly, it is shown, in Problem 6.29, that a band-limited sequence is not time-limited.

The definitions introduced in this section are used to describe the form of a function or sequence. When they are applied to functions and/or sequences pertaining to a Fourier transform pair, four reciprocity relationships can be proposed which all imply, in different ways, that wide functions have a narrow Fourier transform, and vice-versa. The fourth reciprocity relationship (6.109), requiring that a real-valued sequence is either time- or band-limited, is connected with the property that a time-limited sequence has a Fourier transform which oscillates, i.e., has a primary peak flanked by secondary peaks on both sides. Obviously, a time-limited sequence $(g_t^{\vdash t \dashv})$ as defined in (6.108,1) can approach zero, close to the endpoints of interval $|t| \le (N-1)/2$ where $(g_t^{\vdash t \dashv}) \ne 0$, in many possible ways. Does the form of $(g_t^{\vdash t \dashv})$ close to $-(N-1)/2$ and $(N-1)/2$, have an impact on the secondary peaks in its Fourier transform $G^{\vdash t \dashv}(s)$? An answer is given in Sect. 6.7.

6.7 Using a Data Taper to Reduce Leakage

Damped oscillation $f(t)$ and its Fourier transform $F(s)$ as defined in (6.77) and (6.78) are plotted for $s_1 = 0.01$, $s_2 = 0.02$, $a_1 = 0.002$ and $a_2 = 0.0001$ in Figs. 6.17, 6.19 and 6.20. $f(t)$ is an even (6.32) function and, because $f(t)$ is even, $F(s)$ is real-valued and even as concluded from (6.37). In (6.79), sequence (g_t) is constructed from $f(t)$ such that (g_t) is both real-valued and even, and thus has a discrete Fourier transform $G(s_k) = \mathcal{F}_{-\mathrm{i}}(g_t)$ with these properties. These even example functions are special cases selected to demonstrate, in Sect. 6.5, aliasing and leakage using a relatively small number of plots (when dealing with complex-valued functions, their real as well as imaginary parts have to be plotted). Aliasing is reduced, as proposed in the remarks to (6.92), by choosing a sampling interval Δt such that $F(s)$ is small at the exterior of $-1/(2\Delta t) \le s \le 1/(2\Delta t)$, on condition that an instrument and data acquisition system is available that allows for the possibly small Δt required. Leakage due to the main lobe of $D_N(s)$ in (6.94) and (6.95) is reduced by increasing N as demonstrated in Fig. 6.21, whereas leakage due to the side lobes of $D_N(s)$ remains substantial with increasing N as demonstrated in Figs. 6.20 and 6.22.

6.7.1 Data Taper

Leakage due to the side lobes of $D_N(s)$ can be reduced on condition that $D_N(s)$ in the convolution integral $G(s) * D_N(s)$ proposed in (6.94) and (6.95) is substituted with an even function $H_N^{(e)}(s)$, having smaller side lobes than $D_N(s)$. $H_N^{(e)}(s)$ is the Fourier transform of the data taper $h_N^{(e)}(t)$ being introduced in (6.110).

Let (i) $H_N^{(e)}(s) = \mathcal{F}_{-\mathrm{i}}\big(h_N^{(e)}(t)\big)$ *be a Fourier transform pair as defined in (6.49) and (6.50), with* $N = 1, 2, 3, \ldots$, *and (ii)* w *be a smooth (i.e., with continuous derivative) function that maps the real numbers into the closed interval* $[0, 1]$ *such that* $w\big(-(x-1)\big) = w(x-1)$, $w(1) = 1$ *and* $w(x) = 0$ *for* x *at the exterior of* $0 \le x \le 2$. *Then:*

1. $h_N^{(e)}(t) = w\big((2t)/N + 1\big)$ *is an even data taper provided that* N *is odd, and* $H_N^{(e)}(s)$ *is the kernel pertaining to* $h_N^{(e)}(t)$,
2. $h_N(t) = w\big((2t)/N\big)$ *is a data taper with* $H_N(s)$ *being its kernel.*

Other names for $h_N(s)$ *and* $h_N^{(e)}(s)$ *are data window or fader.* (6.110)

Both, $h_N^{(e)}(t)$ and $H_N^{(e)}(s)$, are real-valued and even and thus are special cases. This minor disadvantage is overcompensated, as in Sect. 6.5, by the reduction of the number of plots needed for demonstrating convolutions with kernel $H_N^{(e)}(s)$, e.g., $G(s) * H_N^{(e)}(s)$ as in 6.114. A general real-valued data taper, i.e., a data taper that is not restricted to being even, is obtained in (6.110,2) by shifting $h_N^{(e)}(t)$. $h_N(t)$ can be applied to observations available for time points $t = 0, 1, \ldots, N-1$.

The kernel $g(t, u)$ and the limits a and b, in (2.33), distinguish one particular integral transform from another. The convolution integral as defined in (2.28,1) is shown to be, in (2.34), time-invariant as defined in (2.32). For example, $D_N(s)$ and $F_N(s)$, when used in a convolution integral, are the kernels pertaining to $\Pi_N^{(e)}(t)$ and $\Lambda_N^{(e)}(t)$, i.e., the even rectangle and triangle sequences as defined in (6.52) and (6.56).

An even data taper $h_N^{(e)}(t)$ as defined in (6.110) is a time-limited sequence as required in (6.108). Consequently, $h_N^{(e)}(t)$ is not band-limited, as concluded from (6.109), a property which implies that no frequency $0 < s_0 < 1/2$ does exist such that the pertaining kernel $H_N^{(e)}(s) = 0$ for $|s| > s_0$. As a second property, inferred from the examples in Sect. 6.6.4, $H_N^{(e)}(s)$ oscillates, i.e., has a primary peak flanked by secondary peaks on both sides.

From $h_N^{(e)}(0) = 1$, $\int_{-1/2}^{1/2} H_N^{(e)}(s)\mathrm{d}s = 1$ is obtained applying (6.48,1).

Definition (6.110) is relatively strict. In particular, w is required to be a smooth function: this excludes $\Pi_N^{(e)}(t)$ and $\Lambda_N^{(e)}(t)$ from being an even data taper as required in (6.110). The property that w is a smooth function is "inherited" by $h_N^{(e)}(t)$ (in the sense that $h_N^{(e)}(t)$ appears smooth when plotted)

and thus, hopefully, asserts that the side lobes of $H_N^{(e)}(s)$ are smaller than those of $D_N(s)$. This favourable property is hoped for because the Gaussian functions $f(t) = \mathrm{e}^{-d\pi t^2}$ and $F(s) = d^{-1/2}\mathrm{e}^{-\pi\frac{1}{d}s^2}$ in Fig. 6.10, although not becoming identically zero for intervals $t_1 \le t \le t_2$ or $s_1 \le s \le 2_2$, are smooth in both domains.

As an example for a sequence having the properties required in (6.110), the even cosine taper is defined in (6.111,2).

Let t be an integer, N a positive integer, p real with $0 \le p \le 0.5$, and q the largest integer with $q \le pN$. Then

1. $$\widehat{h_N}(t) = \begin{cases} 0 & \text{for } t < 0 \\ \frac{1}{2}\left(1 - \cos\left(\frac{\pi(2t+1)}{2q}\right)\right) & \text{for } 0 \le t < q \\ 1 & \text{for } q \le t < N - q \\ \frac{1}{2}\left(1 - \cos\left(\frac{\pi(2(N-1-t)+1)}{2q}\right)\right) & \text{for } N - q \le t < N \\ 0 & \text{for } N \le t \end{cases} \qquad (6.111)$$

is called cosine taper (or cosine bell taper) for N and p.

If, in addition, N is odd and $M = (N-1)/2$, then

2. $\widehat{h_N^{(e)}}(t) = \widehat{h_N}(t+M)$ *is called even cosine taper (or even cosine bell taper) for N and p.*

An (even) cosine taper as defined in (6.111) is obtained in R using `spec.taper()`. With, for example,

```
h <- as.vector(spec.taper(rep(1,65),p=0.10)  #N=65 and p=0.10
t1 <- 0:64              #time points for cosine taper
hg <- c(rep(0,10),h)    #for plotting
hg <- c(hg,rep(0,10)    #R vector hg now has 85 values
t2 <- -42:42            #time points for even cosine taper
plot(t2,hg)
```

the even cosine taper for $N = 65$ and $p = 0.10$ is plotted in Fig. 6.26 (c). The pertaining kernel in Fig. 6.26 (d) is obtained, in R, as demonstrated in Sect. 6.9.8. Despite being calculated as discrete Fourier transforms, the even sequences on the left side in Fig. 6.26, and the pertaining kernels on the right side, are Fourier transform pairs as defined in (6.49) and (6.50).

pN in (6.111) is the length of the flattened tails on both ends of the cosine taper $\widehat{h_N}(t)$ or the even cosine taper $\widehat{h_N^{(e)}}(t)$. For p close to 0, the tails in $\widehat{h_N}(t)$ or $\widehat{h_N^{(e)}}(t)$ are shorter than those obtained for larger p, and the side lobes in the pertaining kernels $\widehat{H_N}(s) = \mathcal{F}_{-\mathrm{i}}\big(\widehat{h_N}(t)\big)$ and $\widehat{H_N^{(e)}}(s) = \mathcal{F}_{-\mathrm{i}}\big(\widehat{h_N^{(e)}}(t)\big)$ are higher than those in the kernels pertaining to $\widehat{h_N}(t)$ or $\widehat{h_N^{(e)}}(t)$ with larger p. If $p = 0$, then $\widehat{h_N^{(e)}}(t) = \mathit{\Pi}_N^{(e)}(t)$ implying $\widehat{H_N^{(e)}}(s) = D_N(s)$. With p increasing and finally coming close to 0.5, the tails in $\widehat{h_N^{(e)}}(t)$ increase in length and the side lobes in $\widehat{H_N^{(e)}}(s)$ decrease in height, as demonstrated in Fig. 6.26. The decreasing height of the side lobes in $\widehat{H_N^{(e)}}(s)$ is, however, paid for by the

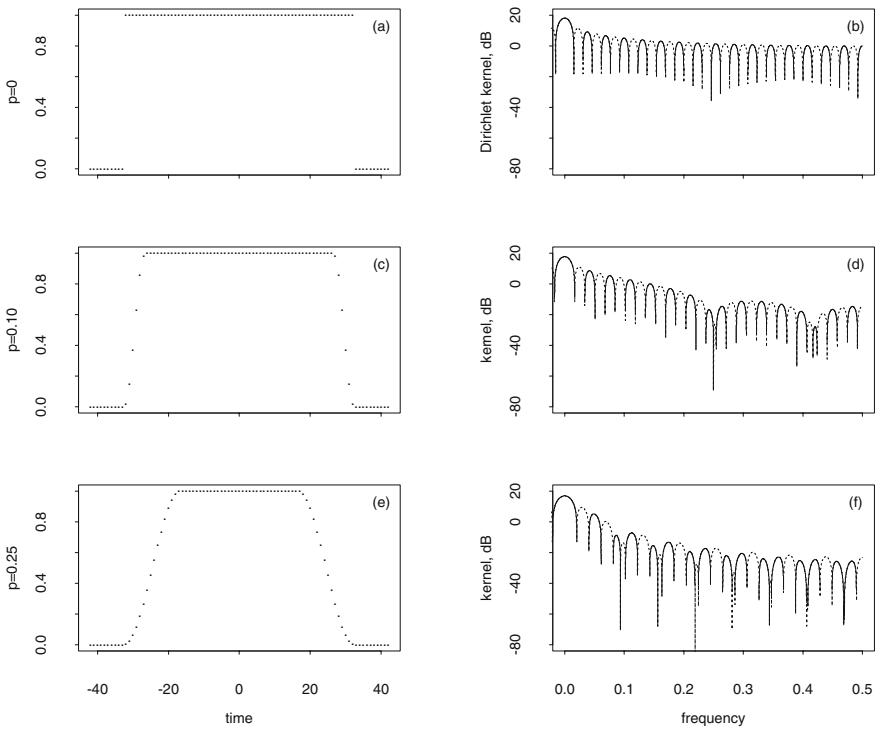


Fig. 6.26. Even cosine tapers $\widehat{h_N^{(e)}}(t)$ as defined in (6.111,2) (on the left) together with their kernels $\widehat{H_N^{(e)}}(s)$ (on the right, in dB, negative secondary peaks with broken lines), for $p = 0.00, 0.10, 0.25$ and $N = 65$ (from top to bottom).

main lobe decreasing in height and increasing in width, as is obvious from Table 6.2.

The even cosine tapers and the pertaining kernels given as examples in Fig. 6.26 and Table 6.2 suggest that the absolute values of first order differences, calculated as required in (2.39,2) for the tails in an even data taper $h_N^{(e)}(t)$, influence the height of the side lobes in the pertaining kernel $H_N^{(e)}(s)$: large side lobes in $H_N^{(e)}(s)$ pertain to steep tails in $h_N^{(e)}(t)$. Further examples for this dependence are given in Problem 6.27.

Subsequent to the above discussion of the even data taper and the pertaining kernel as defined in (6.110), you are now prepared to reconsider the derivation of the convolution integral in (6.95). This convolution produces the distortions due to leakage, which are discussed at length in Sect. 6.5.3.

Assume that (g_t) and $G(s)$, in (6.112), have the same properties as in (6.93), thereafter substitute an even data taper $h_N^{(e)}(t)$ for the even rectangle sequence $\Pi_N^{(e)}(t)$ and, consequently, the pertaining kernel $H_N^{(e)}(s)$ for the

Table 6.2. Height and width at half height $\mathcal{B}^{\cap}_{\triangle}\big(\widehat{H^{(e)}_N}(s)\big)$ of the main peaks of kernels $\widehat{H^{(e)}_N}(s)$ as defined in (6.110) pertaining to even cosine tapers $\widehat{h^{(e)}_N}(t)$ as defined in (6.111,2) for $N = 65$. $\mathcal{B}^{\cap}_{\triangle}\big(\widehat{H^{(e)}_N}(s)\big)$ is computed as required in the remarks to (6.97).

p	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
height $\approx$	65	62	59	55	52	49	46	39	33
$\mathcal{B}^{\cap}_{\triangle}\big(\widehat{H_N}(s)\big) \approx$	.0186	.0196	.0206	.0218	.0230	.0242	.0254	.0281	.0308

Dirichlet kernel $D_N(s)$ in (6.94) and (6.95) to obtain (6.113) and (6.114).

$$G(s) = \mathcal{F}_{-\mathrm{i}}(g_t) = \sum_{t=-\infty}^{\infty} g_t \mathrm{e}^{-\mathrm{i}2\pi st} \quad (g_t) \text{ as in (6.82)} \tag{6.112}$$

$$G(s) * H^{(e)}_N(s) = \sum_{t=-\infty}^{\infty} \Big((g_t)\big(h^{(e)}_N(t)\big)\Big)\mathrm{e}^{-\mathrm{i}2\pi st} \quad \begin{matrix}\text{convolved}\\ \text{as in}\\ (6.144)\end{matrix} \tag{6.113}$$

$$\int_{-1/2}^{1/2} G(r)H^{(e)}_N(s-r)\mathrm{d}r = \sum_{t=-(N-1)/2}^{(N-1)/2} \big(h^{(e)}_N(t)\big)(g_t)\mathrm{e}^{-\mathrm{i}2\pi st} \tag{6.114}$$

A high peak in $G(s)$ dominates the convolution integrals on the left sides in (6.95) and (6.114), as demonstrated for the case of convolving with $D_N(s)$ in Figs. 6.21 and 6.22. Consequently, the distortions of $G(s)$, due to the side lobes of the kernels in the convolution integrals, will be smaller in (6.114) than in (6.95) on condition that $H^{(e)}_N(s)$ has smaller side lobes than $D_N(s)$. This desired reduction of the leakage is arrived at, applying convolution theorem (6.48,7), when, in the time domain, (g_t) is multiplied with $h^{(e)}_N(t)$ (and not with $\Pi^{(e)}_N(t)$). A substitution of $\Pi^{(e)}_N(t)$ with $h^{(e)}_N(t)$ aiming at the reduction of the leakage due to the side lobes of the kernels in convolution integrals (6.95) and (6.114) is called *tapering*.

If $H^{(e)}_N(s)$, $N = 1, 2, \ldots$, is a defining sequence for delta function $\delta(s)$ as required in (6.60), then $\int_{-1/2}^{1/2} G(r)H^{(e)}_N(s-r)\mathrm{d}r = G(s)$ for $N \to \infty$, as concluded from (6.68). For example, $F_N(s)$, the Fourier transform of the triangle sequence (6.56), is a kernel with this property, as is obvious from the remarks to (6.63).

For finite N in (6.114), however, sequence (g_t), $t = \ldots, -1, 0, 1, \ldots$, is seen through the data window $h^{(e)}_N(t)$ in a finite interval $t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$ as product $(g_t)h^{(e)}_N(t)$. "Data window" as alternative to "data taper" originates from this perception of the right side in (6.114). As a consequence of the Fourier sum being calculated from a

limited and weighted part of (g_t), $G(s)$ is distorted by the convolution with $H_N^{(e)}(s)$, resulting in leakage. The leakage depends on $G(s)$, $H_N^{(e)}(s)$ and N since the results obtained in the experiments with $D_N(s)$ in Figs. 6.17, 6.19, 6.20, 6.21, and 6.22, as well as in Problem 6.24, also apply to convolution integrals with kernel $H_N^{(e)}(s)$.

Having assimilated the Nyquist frequency as defined in (6.92), the dynamic range and width at half height as in (6.96) and (6.97), and also the tapering as in the remarks to (6.114), you are now able to successfully circumvent the traps set by aliasing and leakage as demonstrated in Sect. 6.7.2.

6.7.2 Aliasing and Leakage Circumvented

In applications, an even function $F(s)$ is often not known in detail, however, approximations for its dynamic range (6.96) and its width at half height (6.97) are available and it is known that $F(s)$ is in $L^2(-\infty,\infty)$ with $F(s) \approx 0$ for $|s| > s_0$. This knowledge stems from previous experiments and/or is derived from theory. $F(s)$ is obtained as Fourier transform of $f(t)$ on condition that (i) $f(t)$ and $F(s)$ are a Fourier transform pair as defined in (6.35), and (ii) $f(t)$ can be measured in the time domain whereas $F(s)$ is not observable in the frequency domain.

$f(t)$ is observed using a sampling interval $\Delta t < 1/(2s_0)$ as proposed in the remarks to (6.92), to obtain (g_t), $t = \ldots, -1, 0, 1 \ldots$, provided that (i) an instrument and a data acquisition system exist that allow for measurements using the required small Δt, and (ii) it is assumed, admittedly not being realistic, that the period of measurement is not limited. To (g_t), the sequence of observations defined for all integer t, pertains $G(s) = \mathcal{F}_{-\mathrm{i}}(g_t)$, its Fourier transform as defined in (6.49) and (6.50). Since $F(s) \approx 0$ for $|s| > s_0$ and $\Delta t < 1/(2s_0)$, aliasing becomes negligible implying $G(s) \approx F(s)$ and, therefore, $\mathrm{dyn}\big(F(s)\big) \approx \mathrm{dyn}\big(G(s)\big)$ as well as $\mathcal{B}_{\triangle}\big(F(s)\big) \approx \mathcal{B}_{\triangle}\big(G(s)\big)$, with $\mathrm{dyn}\big(F(s)\big)$ being the dynamic range (6.96) and $\mathcal{B}_{\triangle}\big(F(s)\big)$ the width at half height (6.97) of function $F(s)$. In applications, to become realistic, the observational period is limited, in this example to $t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$, and, consequently, the measurement suffers from leakage due to the convolution with kernel $H_N^{(e)}(s)$ on the left side in (6.114) distorting $G(s)$ when, on the right side in (6.114), (g_t) is multiplied with an even data taper $h_N^{(e)}(t)$ as defined in (6.110).

Diagnostics for the leakage attributed to the main lobe in $H_N^{(e)}(s)$ and also for that due to the side lobes in $H_N^{(e)}(s)$ can be obtained from $\mathcal{B}_{\triangle}\big(F(s)\big)$ and $\mathrm{dyn}\big(F(s)\big)$, since $\mathcal{B}_{\triangle}\big(F(s)\big) \approx \mathcal{B}_{\triangle}\big(G(s)\big)$ and $\mathrm{dyn}\big(F(s)\big) \approx \mathrm{dyn}\big(G(s)\big)$, as argued above. However, these diagnostics can only be calculated on condition that both are available: on the one hand the dynamic range $\mathrm{dyn}\big(F(s)\big)$ (6.96) as well as the width at half height $\mathcal{B}_{\triangle}\big(F(s)\big)$ (6.97) of $F(s)$, and on the other hand $\mathcal{B}^{\cap}_{\triangle}\big(H_N^{(e)}(s)\big)$, i.e., the width at half height of the main lobe of $H_N^{(e)}(s)$, as

defined in the remarks to (6.97). In this example, $\mathrm{dyn}\big(F(s)\big)$ and $\mathcal{B}_{\triangle}\big(F(s)\big)$ are, above, assumed to be known, and $\mathcal{B}^{\cap}_{\triangle}\big(H^{(e)}_N(s)\big)$ can be determined as demonstrated in Fig. 6.24 for the case of the Dirichlet kernel.

To obtain a diagnostic for the leakage resulting from the main lobe in $H^{(e)}_N(s)$, Priestley in [113] argues that a sharp peak in $G(s)$, when convolved with $H^{(e)}_N(s)$ on the left side in (6.114), (i) becomes blunt when $\mathcal{B}_{\triangle}\big(G(s)\big) < \mathcal{B}^{\cap}_{\triangle}\big(H^{(e)}_N(s)\big)$ but (ii) suffers less and remains tolerably sharp when $\mathcal{B}_{\triangle}\big(G(s)\big) > \mathcal{B}^{\cap}_{\triangle}\big(H^{(e)}_N(s)\big)$. Consequently, narrow forms in $G(s)$ are not distorted (or only to a negligible extent) when convolved with $H^{(e)}_N(s)$ on condition that

$$\mathcal{B}^{\cap}_{\triangle}\big(H^{(e)}_N(s)\big) \le \mathcal{B}_{\triangle}\big(G(s)\big)/2. \tag{6.115}$$

Diagnostic (6.115) allows for approximating the length of the observational period to obtain a tolerably small resolution in the frequency domain, provided that parameter N in (6.110), i.e., the width of the even data taper $h^{(e)}_N(t)$, can be computed from $\mathcal{B}^{\cap}_{\triangle}\big(H^{(e)}_N(s)\big)$.

For example, $F(s)$, $f(t)$ and (g_t) as defined in (6.77), (6.78) and (6.79) with $s_1 = 0.01$, $s_2 = 0.02$, $a_1 = 0.002$ and $a_2 = 0.0001$, are plotted in Figs. 6.17, 6.19 and 6.20. For this example, aliasing is negligible, as argued in Sect. 6.5.3. The width at half height of $F(s)$ is determined in Fig. 6.25: $\mathcal{B}_{\triangle}\big(F(s)\big) = \mathcal{B}_{\triangle}\big(F(s), \text{sharper peak}\big) = 3 \times 10^{-5}$. How many observations of $f(t)$ have to be made to resolve $F(s)$ without tapering?

When no data taper is applied to the observations, then $G(s)$ is convolved with $D_N(s)$ as in (6.95). Consequently, $\mathcal{B}^{\cap}_{\triangle}\big(D_N(s)\big) \approx 6/(5N)$, as determined in the remarks to Fig. 6.24, and $\mathcal{B}_{\triangle}\big(F(s)\big)$, as obtained above, are substituted in diagnostic (6.115) to obtain $6/(5N) \le 1.5 \times 10^{-5}$, and thus $N \ge 80000$ is arrived at. With $N = 65535$, a smaller N, a quite good resolution of the primary peak of $F(s)$ is obtained in Fig. 6.20 (b).

Neglecting the finer details of the kernel in (6.114) (i.e., the accurate width of its main lobe), the length of the observational period can also be approximated using simply the distance between the Fourier frequencies, i.e., $\Delta s_k = 1/N$ as defined in (6.21,1). For example, the Fourier frequencies for $N = 80000$ drawn in Fig. 6.25 (d) and those for $N = 8000$ in Fig. 6.25 (c), demonstrate that the forms in a function $F(s)$ are approximately captured with an N such that

$$1/N = \Delta s_k < \mathcal{B}_{\triangle}\big(F(s)\big)/2. \tag{6.116}$$

In most applications, N is approximated using diagnostic (6.116) with the result that narrow peaks in $F(s)$ are not unduly flattened when $G(s) \approx F(s)$ is convolved with kernel $H^{(e)}_N(s)$, on the left side in (6.114). Consequently, and because it can be easily calculated, (6.116) is often preferred to (6.115).

A diagnostic for the leakage due to the side lobes in $D_N(s)$ is feasible using $\mathrm{dyn}\big(G(s)\big) \approx \mathrm{dyn}\big(F(s)\big)$ since a high peak or a high jump in $F(s)$ not only

dominates $\int_{-1/2}^{1/2} G(r)D_N(s-r)\mathrm{d}r$ in (6.95) but also generates a large dynamic range in $F(s)$. From extensive experiments with damped oscillations $f(t)$ as defined in (6.77) and (6.78) (e.g., these functions are plotted for $s_1 = 0.01$, $s_2 = 0.02$, $a_1 = 0.002$, $a_2 = 0.0001$ in Fig. 6.17) the threshold in diagnostic (6.117) is obtained.

Let $f(t)$ and $F(s)$ as well as (g_t) and $G(s)$ be as defined in (6.81) and (6.82) and used in (6.93), Δt small enough such that an aliasing can be neglected and thus $F(s) \approx G(s)$, $\mathit{\Pi}_N^{(e)}(t)$ the even rectangle sequence and $D_N(s)$ the Dirichlet kernel as used in (6.94) and (6.95). Then the side lobes of $D_N(s)$ in $\int_{-1/2}^{1/2} G(r)D_N(s-r)\mathrm{d}r$ on the left side in (6.95) generate distortions that are:

1. *negligibly small for* $\mathrm{dyn}\big(G(s)\big) \leq 10$ dB *but*
2. *cannot be neglected for* $\mathrm{dyn}\big(G(s)\big) > 10$ dB.

(6.117)

Often, functions $F(s)$ have a dynamic range larger than 10 dB. Consequently, a leakage due to the side lobes of $D_N(s)$ cannot, in most cases, be neglected.

If diagnostic (6.117) arouses suspicion of leakage, then observations (g_t), $t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$ are multiplied with a sequence of even cosine tapers $\widehat{h_N^{(e)}}(t)$ for $p_0 = 0.00$ (in this case, the observations are multiplied with the even rectangle sequence), $p_1 = 0.05$, $p_2 = 0.10$, ..., $\widehat{h_N^{(e)}}(t)$ as defined in (6.111,2). These products are Fourier transformed using (6.22) or (6.26) to obtain $G^{(p_0)}(s'_k)$ $G^{(p_1)}(s'_k)$, $G^{(p_2)}(s'_k)$, ..., which are then plotted. $\widehat{h_N^{(e)}}(t)$ with increasing p are applied, until $\big|G^{(p_n)}(s'_k)\big| \approx \big|G^{(p_{n-1})}(s'_k)\big|$ is arrived at: $G^{(p_n)}(s'_k)$ is assumed to be the best approximation of $G(s)$ that can be obtained using an even cosine taper.

For example, (g_t) generated as required in (6.82) from the damped oscillations $f(t)$ as defined in (6.77) and (6.78) with $s_1 = 0.01$, $s_2 = 0.02$, $a_1 = 0.002$, $a_2 = 0.0001$ (this is the example in Figs. 6.17 and 6.19) is multiplied with the even cosine taper $\widehat{h_N^{(e)}}(t)$ for $N = 65565$ and $p_1 = 0.05$ which is plotted in Fig. 6.27. The product $(g_t)\big(\widehat{h_N^{(e)}}(t)\big)$ thus obtained for the time slice $t = -(N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$ is zero-padded on its right side to obtain $N' = 65536$ values which then are Fourier transformed to arrive at (finally) $G^{(p_1)}(s'_k)$ as plotted in Fig. 6.28.

From Fig. 6.28 it becomes evident that (i) $G^{(p_1=0.05)}(s'_k)$ is much closer to $F(s)$ than $G^{(p_0=0)}(s'_k)$ as plotted in Figs. 6.17, 6.19 and 6.20, and (ii) $G^{(p_1)}(s'_k) > 0$ for $0 \leq s'_k \leq 1/2$. Hence, $G^{(p_1)}(s'_k)$ performs much better than $G^{(p_0)}(s'_k)$. However, $G^{(p_1)}(s'_k)$ is still subject to leakage: (i) it has small oscillations due to the side lobes in $\widehat{H_N^{(e)}}(s)$, the kernel pertaining to $\widehat{h_N^{(e)}}(t)$ as is demonstrated in Fig. 6.28 (b), and (ii) $G^{(p_1)}(s'_k) > F(s)$ substantially (though not by orders of magnitude) for frequencies not close to the peaks in $F(s)$, as is seen in Fig. 6.28 plots (a) and (c). Trying $\widehat{h_N^{(e)}}(t)$ for $p_2 = 0.10$,

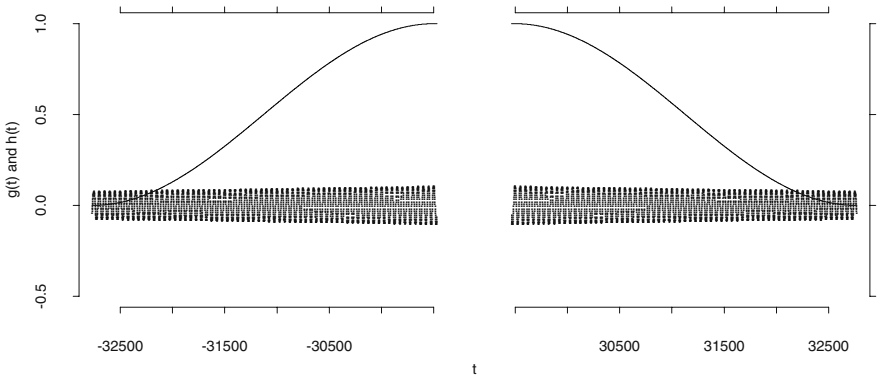


Fig. 6.27. Sequence (g_t) as in Fig. 6.17 for $N = 65535$ (broken line) and even cosine taper $\widehat{h_N^{(e)}}(t)$ (6.111,2) for $p = 0.05$ and $N = 65535$ (solid line).

$p_3 = 0.15$, $p_4 = 0.20$, ..., as recommended above, a $G^{(p_n)}(s'_k)$ close to $F(s)$ for all $0 \le s'_k \le 1/2$ is obtained in Problem 6.28.

Comparing $\left|G^{(p_0)}(s'_k)\right|$ with $\left|G^{(p_n)}(s'_k)\right|$, an approximation for the error due to leakage is obtained that would result were the observations not multiplied with an even data taper as defined in (6.110). In the above example, it is concluded from Figs. 6.19 and 6.20 on the one hand, as well as from Fig. 6.28 and the plots obtained in Problem 6.28 on the other hand, that a leakage in the order of magnitudes has to be taken into account on condition that no suitable data taper has been applied.

This section is summarised as follows. The problem described in Sect. 6.5.1, i.e., to determine $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$, both $f(t)$ and $F(s)$ real-valued and even, from observations $(g_t) = f(t)$ for $t = (N-1)/2, \ldots, -1, 0, 1, \ldots, (N-1)/2$, is tackled in two steps, as shown above, in Sect. 6.7.2.

In the first step, by (i) applying (6.92), a sampling interval $\Delta t < 1/(2s_0)$ small enough to avoid aliasing, and, (ii) applying (6.116), the number $1/N = \Delta s_k < \mathcal{B}_{\triangle}\big(F(s)\big)/2$ of observations to be made to resolve narrow peaks in $F(s)$, are obtained. Δt imposes restrictions on the equipment used for the measurement and from N depend its operating costs, i.e., these diagnostics allow for assessing the resources needed to arrive at $F(s)$.

In the second step, observations (g_t) are multiplied with an even data taper $h_N^{(e)}(t)$ as defined in (6.110) since, in most cases, $\mathrm{dyn}\big(F(s)\big) \approx \mathrm{dyn}\big(G(s)\big) >$ 10 dB and tapering is called for, as concluded from diagnostic (6.117). $h_N^{(e)}(t)$ has a kernel $H_N^{(e)}(s)$ with smaller side lobes than $D_N(s)$. $H_N^{(e)}(s)$ is the kernel in the convolution integral in (6.114), and $D_N(s)$ is the kernel in (6.95).

Even example functions are used to introduce, in Sect. 6.5, the traps set by aliasing and leakage, and to demonstrate, in Sect. 6.7, how these pitfalls

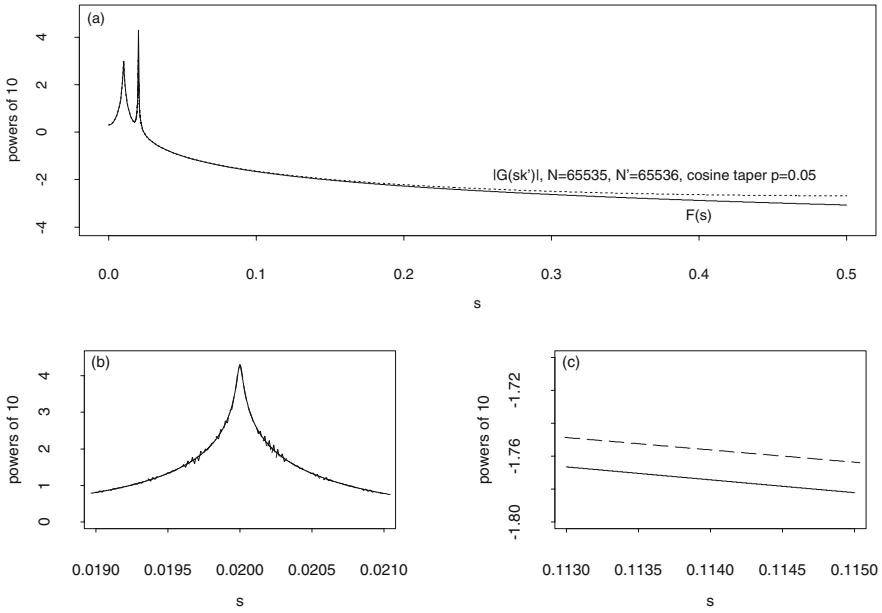


Fig. 6.28. $F(s)$ (solid line) and discrete Fourier transform (broken line) of the product $(g_t)\widehat{h_N^{(e)}}(t)$, (g_t) and $\widehat{h_N^{(e)}}(t)$ as in Fig. 6.27.

can be circumvented. In applications however, a data taper as defined in (6.110,2) is used, i.e., a data taper that is not restricted to being even.

6.8 Convolution II

The stationary solution of the first order linear difference equation derived in the remarks to (2.27) is an example of a convolution sum as defined in (2.28,2). With observed values x_t and w_t acting as substitutes for the random variables X_t and W_t in (2.27), the deterministic difference equation and the convolution sum in (6.118) are obtained.

$$x_t - ax_{t-1} = w_t \quad \text{and} \quad x_t = \sum_{u=0}^{\infty} a^u w_{t-u} \quad \text{for integer } t \tag{6.118}$$

$$y(t) = -a\dot{y}(t) + x(t) \quad \text{and} \quad y(t) = \int_{-\infty}^{t} f(t-u)x(u)\mathrm{d}u \quad \text{for real } t \tag{6.119}$$

$$\text{with} \quad f(t) \;=\; \begin{cases} 0 & \text{for } t < 0 \\ (1/a)\mathrm{e}^{-(1/a)t} & \text{for } t \geq 0 \end{cases}$$

The first order linear differential equation and its stationary solution resulting from Problem 2.11 are given in (6.119). The convolutions in (6.118) and (6.119) are transformations, as defined in (2.29).

Convolution theorem (6.48,6) stipulates that a convolution of two functions, e.g., $f(t)$ and $g(t)$, in the time domain is equivalent to a multiplication of their Fourier transforms, e.g., $F(s)$ and $G(s)$, in the frequency domain. Examining $F(s)$ and $G(s)$ in the frequency domain, it is often easier to understand the effects generated by convolving $f(t)$ and $g(t)$ in the time domain. In this section therefore, using definitions and propositions in (6.120) and (6.121), the moving average and the first order differences, as defined in (2.39), are analysed in the frequency domain.

6.8.1 Linear and Time-invariant (LTI) Systems

Systems as proposed in (6.118) and (6.119) are often described using the terms introduced in (6.120).

If the convolution integral (sum) as defined in (2.28)

$$\begin{aligned} y(t) &= \int_{-\infty}^{\infty} f(t-u)x(u)\mathrm{d}u = \big(f(t)\big) * \big(x(t)\big) \\ y_t \quad &= \sum_{-\infty}^{\infty} f(t-u)x(u) = (f_t) * (x_t) \end{aligned}$$

converges, then it is called a linear and time-invariant (LTI) system or an LTI filter. x is called input, y output, f impulse response function (sequence), and the Fourier transform of f is called transfer function or frequency response function. (6.120)

Impulse response function is a plausible term for f, as shown in the following rationale. Let $\delta(x)$, i.e., the delta or impulse function (6.60), be the input in the LTI filter (6.120). Applying (6.67) and (6.68), the output of the filter is $y(t) = \int_{-\infty}^{\infty} f(u)\delta(t-u)\mathrm{d}u = f(t)$, because $\delta(x)$ is the identity function pertaining to the convolution integral. Consequently, $f(t)$ is the response of the system to the impulse function.

The impulse response function determines the domain of the LTI filter, i.e., the set of all input functions or sequences, for which the convolution integral or sum (6.120) converges. A convergence depends on the impulse response function and on the input, e.g., a convolution sum exists when both, the impulse response sequence and input sequence, possess the properties required in (2.36).

The order of the impulse response function and of the input can be reversed in both convolution integral and sum because convolution is commutative as derived in (2.35) and Problem 2.4.

The convolution integral is shown to be LTI in (2.34), and plausibly, the convolution sum is also LTI. The LTI transformation (filter, operator) is defined in (2.29), (2.30), (2.31) and (2.32), examples being given in the remarks to (2.34).

An LTI filter can be described in both, the time and the frequency domain. Examples for the effects of an LTI filter observable in the time domain are the smoothing of a sequence using a moving average in Problem 2.15, and the removal of a trend in a time series by computing its first differences as demonstrated in Figs. 5.8 and 5.11. Both LTI filters, the moving average and the first order differences, are defined in (2.39). In the frequency domain, the effects of an LTI filter are described using definitions (6.121). There, the time t or u is assumed to be integer.

If t is integer, then the impulse response sequence and frequency response function in an LTI filter are a Fourier transform pair as defined in (6.49) and (6.50). This property is inherited by the input (x_t) and the output $y(s)$ of the filter, i.e., (x_t) is a sequence in $L^2(-\infty,\infty)$ having a Fourier transform $X(s) = \mathcal{F}_{-\mathrm{i}}(x_t) = \sum_{t=-\infty}^{\infty} x_t \mathrm{e}^{-\mathrm{i}2\pi st}$, such that $(x_t) = \mathcal{F}_{+\mathrm{i}}\big(F(s)\big) = \int_{-1/2}^{1/2} X(s)\mathrm{e}^{\mathrm{i}2\pi ts}\mathrm{d}s$. Consequently, (x_t) can be represented as a linear combination of trigonometric oscillations. Now assume that (x_t) contains the oscillation $\big(x_t^{(s_0)}\big) = X(s_0)\mathrm{e}^{\mathrm{i}2\pi s_0 t}$, $-1/2 \le s_0 \le 1/2$ a real constant and $t = \ldots, -1, 0, 1, \ldots$. What happens to $\big(x_t^{(s_0)}\big)$ in an LTI filter? An answer to this question can be obtained when the definitions and propositions in (6.121) are applied.

Let a real-valued sequence (x_t), $t = \ldots, -1, 0, 1, \ldots$, with Fourier transform $X(s) = \mathcal{F}_{-\mathrm{i}}(x_t)$, be the input in LTI filter (6.120) having impulse response sequence (f_t) and frequency response function $F(s) = \mathcal{F}_{-\mathrm{i}}(f_t)$, and let (x_t) and (f_t) have the properties required in (2.36,2). Then

1. $(y_t) = \sum_{-\infty}^{\infty} f_{t-u}x_u = (f_t) * (x_t) = (x_t) * f(_t)$ *for integer t, u*

   i.e., the output is in $L^2(-\infty,\infty)$, and its Fourier transform

2. $Y(s) = \sum_{-\infty}^{\infty} \big((f_t) * (x_t)\big)\mathrm{e}^{-\mathrm{i}2\pi st} = F(s)X(s)$ for $-\dfrac{1}{2} \le s \le \dfrac{1}{2}$ (6.121)

   is calculated using (6.48,6). Now define:

3.1 $|G(s)| = |F(s)| = \big(F(s)\overline{F(s)}\big)^{1/2}$ *is called gain function, and*

3.2 $P(s) = \big(-1/(2\pi)\big)\varphi(s)$ *is called phase shift function, with $|F(s)|$ being the absolute value and $\varphi(s)$ the phase of $F(s)$, as defined in (6.2,6). Using these definitions, the following product becomes:*

$$\begin{aligned} Y(s)\mathrm{e}^{\mathrm{i}2\pi st} &= \big(F(s)X(s)\big)\mathrm{e}^{\mathrm{i}2\pi st} = |F(s)|\big(X(s)\mathrm{e}^{\mathrm{i}2\pi st}\big)\mathrm{e}^{\mathrm{i}\varphi(s)} \\ &= |G(s)|\big(X(s)\mathrm{e}^{\mathrm{i}2\pi st+\varphi(s)}\big) \\ &= \big(|G(s)|X(s)\big)\mathrm{e}^{\mathrm{i}2\pi(s-P(s))t} \end{aligned}$$

From (6.121,2), it is concluded that $Y(s_0)\mathrm{e}^{\mathrm{i}2\pi s_0 t} = \big(F(s_0)X(s_0)\big)\mathrm{e}^{\mathrm{i}2\pi s_0 t}$, with $\big(x_t^{(s_0)}\big) = X(s_0)\mathrm{e}^{\mathrm{i}2\pi s_0 t}$ being a trigonometric oscillation with frequency s_0 in the input (x_t) of an LTI filter, as assumed above, and $F(s)$

its complex-valued frequency response function. More meaningful in applications, however, are the results obtained applying definitions (6.121,3): $\left(x_t^{(s_0)}\right)$ changes its absolute value and its phase when it becomes $\left(y_t^{(s_0)}\right) = \left(|G(s_0)|X(s_0)\right)\mathrm{e}^{\mathrm{i}2\pi(s_0-P(s_0))t}$ in the output. The sign of the gain function $G(s)$ in (6.121,3) is positive, and consequently $G(s)$ allows for computing an amplification ($G(s) > 1$) or damping ($G(s) < 1$) of a trigonometric oscillation in the input of an LTI filter. Plots of $G(s)$ and $P(s)$ are also called filter characteristics.

Using an LTI filter for example, the cosine oscillations $(x_t^{(s_i)})$, $i = 1, 2, 3$, on the left side in Fig. 6.29 are transformed into the oscillations $(y_t^{(s_i)})$ on the right side in this figure. These pairs of input and output sequences are used to demonstrate definitions (6.121).

Changes in the absolute value (6.2,4) of $\left(x_t^{(s_i)}\right)$ are obtained from the gain function of the LTI filter. For $G(s_i) > 1$, the absolute value of oscillation s_i is smaller in the input than in the output and s_i is amplified, as demonstrated in Fig. 6.29 plots (e) and (f). For $G(s_i) < 1$, the absolute value of oscillation s_i is larger in the input that in the output and s_i is damped, as demonstrated in Fig. 6.29, plots (a) and (b).

Changes in the phase (6.2,5) of $\left(x_t^{(s_i)}\right)$ are obtained in units of s from the phase shift function $P(s)$ of the LTI filter. In time units, changes in the phase amount to $P(s)/s = P(s)T$, on condition that $T = 1/s$, the period of oscillation $X(s)\mathrm{e}^{\mathrm{i}2\pi st}$ with frequency s in the input. Hence, a trigonometric oscillation is shifted in an LTI filter by a fraction $P(s)$ of its period. For example, the oscillation in Fig. 6.29 (a), having frequency $s_1 = 0.05$ and therefore period $T_1 = 20$, is shifted by $-5 < P(s_1)/T_1 < -4$, i.e., $15 < P(s_1)/T_1 < 16$ time units, as obtained from a comparison of plots (a) and (b) in Fig. 6.29.

Above, for the demonstrations in Fig. 6.29, a single trigonometric oscillation is used as input in an LTI filter. Usually, however, the input is a linear combination $(x_t) = \int_{-1/2}^{1/2} X(s)\mathrm{e}^{\mathrm{i}2\pi ts}\mathrm{d}s$ of trigonometric oscillations. In this case, the filter generates $(y_t) = \int_{-1/2}^{1/2} G(s)X(s)\mathrm{e}^{\mathrm{i}2\pi(s-P(s))t}\mathrm{d}s$ as output. In general, the output of an LTI filter is consequently a linear combination of the oscillations in the input that undergo amplifications or dampings in their absolute values as well as changes in their phases.

For example, the moving average and the first order difference (2.39) are LTI filters (6.121,1), with their impulse response sequence and frequency response function being a Fourier transform pair as defined in (6.49) and (6.50). In the time domain, the effects of moving averages are demonstrated in Figs. 2.9, 2.10 and 2.16, and the effects of first order differences in Figs. 2.16, 5.8 and 5.11. In the frequency domain, the effects of these LTI filters are analysed in Sect. 6.8.2.

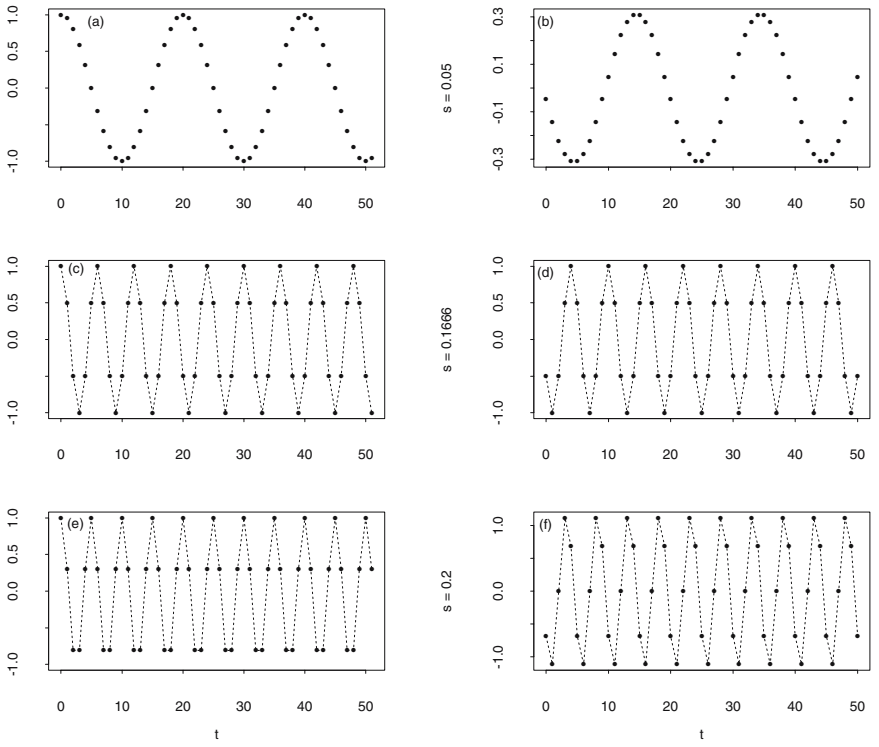


Fig. 6.29. Oscillations $x_t = \cos(2\pi st)$, $t = \ldots, -1, 0, 1, \ldots$, with $s = s_1 = 0.05$, $s = s_2 = 0.1666$ and $s = s_3 = 0.2$ (on the left) together with their first order differences $y_t = x_t - x_{t-1}$ (on the right). The oscillations are sequences and plotted with symbol •, the symbols are connected where necessary.

6.8.2 First Order Differences and Moving Average

Impulse response sequence $(f_u) = (f_0 = 1,\ f_1 = -1)$ and $f_u = 0$ for $u \neq 0, 1$, as in the first order differences (2.39,2), is plotted in Fig. 6.30 (a). The pertaining frequency response function

$$F(s) = \sum_{u=-\infty}^{\infty} f_u \mathrm{e}^{-\mathrm{i}2\pi su} \;=\; 1\mathrm{e}^{\mathrm{i}2\pi s(0)} - 1\mathrm{e}^{-\mathrm{i}2\pi s(+1)} \;=\; 1 - \mathrm{e}^{-\mathrm{i}2\pi s} \qquad (6.122)$$

is obtained as a Fourier transform of (f_t) as required in (6.49). In this case, (f_t) is mixed (6.32,3) and $F(s)$ is complex-valued, as concluded from (6.37). The cosine oscillations on the left side in Fig. 6.29, for example, are transformed by calculating their first order differences to obtain the trigonometric oscillations on the right side in Fig. 6.29.

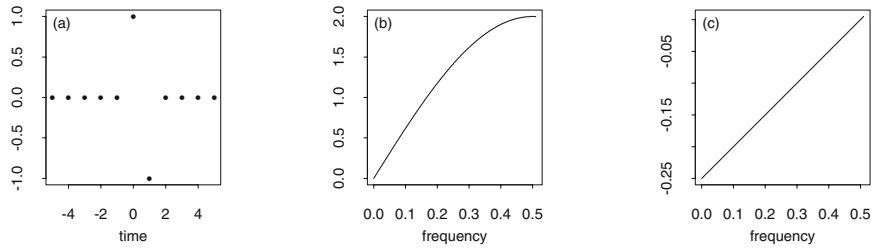


Fig. 6.30. First order differences (2.39,2) with frequency response function (6.122): impulse response sequence (a), gain (b) and phase shift (c) functions.

From (6.122), gain function $G(s) = 2\sin(\pi s)$ and phase shift function $P(s) = s/2 - 1/4$ of the first order differences are obtained in (6.123), when identities $\sin(x) = (1/2\mathrm{i})(\mathrm{e}^{\mathrm{i}x} - \mathrm{e}^{-\mathrm{i}x})$ and $\mathrm{i} = \mathrm{cis}(\pi/2) = \mathrm{e}^{\mathrm{i}2\pi(1/4)}$ and definitions (6.121,3) are applied.

$$\begin{aligned} F(s) &= \mathrm{e}^{-\mathrm{i}2\pi(s/2-s/2)} - \mathrm{e}^{-\mathrm{i}2\pi(s/2+s/2)} \qquad\qquad -1/2 \le s \le 1/2 \\ &= \mathrm{e}^{\mathrm{i}2\pi s/2}\mathrm{e}^{-\mathrm{i}2\pi s/2} - \mathrm{e}^{-\mathrm{i}2\pi s/2}\mathrm{e}^{-\mathrm{i}2\pi s/2} \\ &= \mathrm{e}^{-\mathrm{i}2\pi s/2}\Big(\mathrm{e}^{\mathrm{i}2\pi s/2} - \mathrm{e}^{-\mathrm{i}2\pi s/2}\Big) = \mathrm{e}^{-\mathrm{i}2\pi s/2}\Big(2\mathrm{i} \times \sin\big(2\pi(s/2)\big)\Big) \\ &= \mathrm{e}^{-\mathrm{i}2\pi s/2}\Big(\big(\mathrm{e}^{\mathrm{i}2\pi(1/4)}\big)2\sin\big(2\pi(s/2)\big)\Big) \\ &= 2\sin\big(2\pi(s/2)\big)\mathrm{e}^{-\mathrm{i}2\pi(s/2-1/4)} = G(s)\mathrm{e}^{-\mathrm{i}2\pi P(s)} \end{aligned} \tag{6.123}$$

$G(s)$ and $P(s)$ are plotted against $0 \le s \le 1/2$ in Fig. 6.30 (b) and (c).

Since $|G(s)| = |2\sin(\pi s)| = 1$ for $s = 1/6$, as well as $|G(s)| < 1$ for $|s| < 1/6$ and $|G(s)| > 1$ for $-1/2 \le s < -1/6$, and also $1/6 < s \le 1/2$, first order differences (i) damp, e.g., an oscillation with frequency $s_1 = 0.05$ as demonstrated in Fig. 6.29, plots (a) and (b), and (ii) amplify, e.g., an oscillation with frequency $s_3 = 0.2$ as demonstrated in Fig. 6.29, plots (e) and (f). Due to these properties, a time series being non-stationary in its first moment function often has stationary first order differences, as demonstrated in Figs. 5.8 and 5.11. Trends in the first moment function of a time series are very efficiently removed by calculating differences as in (2.39,2) or (2.40). This is the rationale for defining the ARIMA$[p, d, q]$ model in (5.55), with d the order of the differences.

First order differences, having a complex-valued frequency response function, also generate changes in the phases of the input oscillations. In Fig. 6.29 plots (a) and (b), for example, the cosine oscillation with frequency $s_1 = 1/20 = 0.05$ is shifted by $P(s = 0.05) = 1/40 - 1/4 = -9/40$ (with $P(s)$ as in (6.123)) of its period $T_1 = 20$, i.e., by -4.5 time units. This phase shift is seen in the remarks to Fig. 6.29, when plots (a) and (b) are compared.

The impulse response sequence $(f_u) = (1/3)\Pi_3^{(e)}(u)$ (6.52) in the moving average over three time points with identical weights, constructed as required

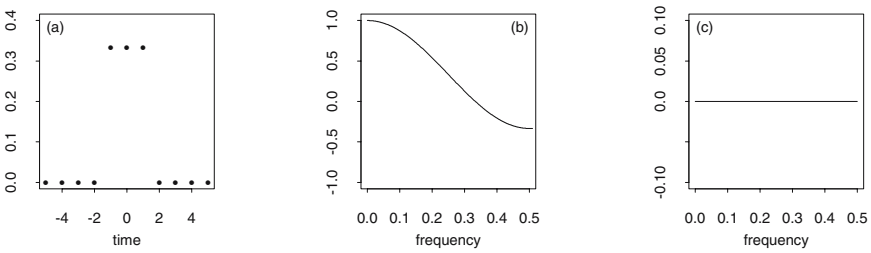


Fig. 6.31. Moving average over three time points as in (6.124): impulse response sequence (a), gain function (b) and phase shift function (c).

in (2.39,1), is plotted in Fig. 6.31 (a). Its Fourier transform

$$\begin{aligned} F(s) = \mathcal{F}_{-\mathrm{i}}(f_u) &= \sum_{u=-\infty}^{\infty} f_u \mathrm{e}^{-\mathrm{i}2\pi s u} \qquad -1/2 \le s \le 1/2 \\ &= (1/3)\left(\mathrm{e}^{-\mathrm{i}2\pi s(-1)} + \mathrm{e}^{-\mathrm{i}2\pi s(0)} + \mathrm{e}^{-\mathrm{i}2\pi s(+1)}\right) \\ &= (1/3)\left(1 + \mathrm{e}^{\mathrm{i}2\pi s} + \mathrm{e}^{-\mathrm{i}2\pi s}\right) = (1/3) + (2/3)\cos(2\pi s) \end{aligned} \tag{6.124}$$

is calculated as required in (6.49) to arrive at the frequency response function of the three-point moving average in (6.124). In (6.124), $F(s)$ is real-valued in accordance with (6.37), since (f_t) is even (6.32,1). $P(s)$ as defined in (6.121) becomes identically zero, as concluded from (6.2,5), on condition that $F(s)$ is real-valued. $F(s)$ and $P(s)$ are plotted in Fig. 6.31 (b) and (c) against $0 \le s \le 1/2$.

(6.121,3) implies that the three-point moving average, with frequency response function $F(s)$ as in (6.124) and plotted in Fig. 6.31 (b), does not change the absolute values of low-frequency trigonometric oscillations in its input, since $|G(s)| = |F(s)| \approx 1.0$ for $s \le 0.1$. Oscillations with frequencies $s \ge 0.1$, however, are markedly damped, and oscillations with frequency $s = 1/3$ are annihilated because $|G(s)| = |F(s)| = 0$ for $s = 1/3$. Oscillations with frequencies $s > 1/3$ in the input of the three-point moving average are damped and change their sign. These properties are demonstrated in Problem 6.31 where moving averages are applied to smooth trigonometric oscillations with a variety of frequencies.

Due to the properties of a moving average, a trend hidden in a time series by oscillations with high frequencies can become more clearly visible subsequent to smoothing. A smoothing favours guessing appropriate linear candidate models for the trend which, thereafter, are estimated applying theory and methods as introduced in Chap. 3.

In Fig. 3.3, for example, linear models are fitted to the yearly values in the Basel temperature series. In this case, a global model (as defined in (2.55,1), i.e., a model representative for the full time series) captures secular trends solely. Consequently, decadal fluctuations remain in the residuals of

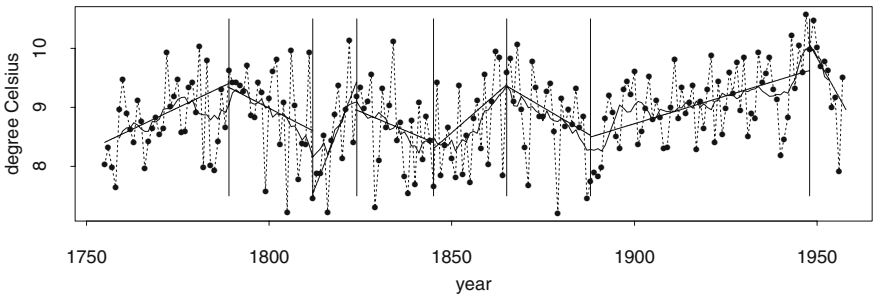


Fig. 6.32. Yearly values in the Basel temperature series (• with broken line) and nine-point moving average of this time series obtained by filtering with $(1/9)\Pi_9^{(e)}(t)$ (solid line), together with local linear models calculated from the monthly values in the time slices (end points with solid vertical lines) seen in the moving average.

this model, as is obvious when Figs. 5.7 and 5.8 are compared. Decadal fluctuations in a secular time series can be captured using local (2.55,2) linear models, as demonstrated in Fig. 6.32. In Fig. 6.32, local linear models are fitted to the Basel temperature series subsequent to filtering the yearly values in this series with $(1/9)\Pi_9^{(e)}(t)$. This moving average damps high-frequency oscillations, and in its output, therefore, time slices become visible allowing for local first order linear models to be fitted. These local models are then fitted to the monthly values in the Basel temperature series thereby increasing the number of values used for the estimation. In the end points of the intervals found by smoothing with $(1/9)\Pi_9^{(e)}(t)$, the local linear models jump: obviously due to being estimated from the monthly values in the preceding and subsequent intervals. Please estimate, in Problem 6.33, these local models such that they do not jump in the endpoints of the intervals.

The representation of the trend in the yearly values in the Basel temperature series as in Fig. 6.32 is not unique, because another filter, i.e., a 15-point moving average, generates another output and thus, other time slices.

Statistical inference of local properties in a time series is subject to restrictions, as demonstrated in Sect. 2.5.2. There it is shown that (i) the outcome of stastical tests is influenced by the selection of the time slices from whence the statistics are calculated and (ii) simulation experiments can be performed, under favourable circumstances, to obtain more realistic results not dependent on the time slices used. The conclusions obtained in Sect. 2.5.2 not only apply to estimates of local means but also to estimates of local linear models for the trend.

6.8.3 Low-pass, High-pass and Band-pass Filters

Both, the moving average and the first order differences, are defined in (2.39) by specifying their weights. These weights become, in (6.120) and (6.121), the impulse response sequence in an LTI filter, implying that an impulse response sequence and frequency response function constitute a Fourier transform pair as defined in (6.49) and (6.50).

Alternatively, such a system is defined by specifying its frequency response function in (6.125). The impulse response sequences (f_t) pertaining to the frequency response functions $F(s)$ as in (6.125) are band-limited; however, they are not time-limited, as is concluded from (6.109). Time-limited and band-limited sequences are defined in (6.108). Since impulse response sequences are band-limited, the filters in (6.125) cannot be implemented using time-limited sequences, therefore, they are said to be *ideal*.

Let (f_t) and $F(s)$ be a Fourier transform pair (impulse response sequence, frequency response function) as in (6.121). Then the LTI filter is called:

$$\begin{aligned}
&1.\ \textit{low-pass if } F(s) = \begin{cases} 1, & \textit{for } |s| \le s_0 \\ 0, & \textit{for } s_0 < |s| \le 0.5 \end{cases} \\
&2.\ \textit{high-pass if } F(s) = \begin{cases} 1, & \textit{for } s_0 \le |s| \le 0.5 \\ 0, & \textit{for } |s| < s_0 \end{cases} \\
&3.\ \textit{band-pass if } F(s) = \begin{cases} 1, & \textit{for } 0 < s_1 \le |s| \le s_2 < 0.5 \\ 0, & \textit{for } |s| < s_1 \textit{ and } s_2 < |s| \le 0.5 \end{cases}
\end{aligned} \tag{6.125}$$

Do approximate implementations of an ideal filter exist? The frequency response function $F(s)$, of e.g., the low-pass filter, is a rectangle function defined in $-1/2 \le s \le 1/2$ and periodic with period 1. For example, $G(s)/10$, $G(s)$ as in Fig. 6.23 (b), acts as frequency response function in a low-pass filter with $s_0 = 0.05$. The impulse response function (f_t) is the Fourier transform of $F(s)$ as defined by the integral in (6.50). Since $F(s)$ is even (6.32), (f_t) is calculated as a Fourier cosine transform (6.37).

$$\begin{aligned}
(f_t) &= \int_{-1/2}^{1/2} F(s)\mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}s = \int_{-s_0}^{s_0} \mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}s \qquad t = \ldots, -1, 0, 1, \ldots \\
&= \begin{cases} 1 \times 2s_0 & \text{for } t = 0 \\ 2\int_0^{s_0} \cos(2\pi st)\mathrm{d}s = 2\Big[\frac{1}{2\pi t}\sin(2\pi st)\Big]_{s=0}^{s=s_0} & \text{for } t \ne 0 \end{cases} \\
&= \begin{cases} 2s_0 & \text{for } t = 0 \\ \big(\sin(2\pi s_0 t)\big)/(\pi t) & \text{for } t \ne 0 \end{cases} \qquad t = \ldots, -1, 0, 1, \ldots
\end{aligned} \tag{6.126}$$

The impulse response function (f_t) of the low-pass filter thus obtained in (6.126) is not time-limited. However, it damps out with $1/t$ for $-\infty \leftarrow t$ and $t \to \infty$ and is thus in $L^2(-\infty, \infty)$, as concluded from (6.49) and (6.50).

Can the impulse response sequence (6.126) of the ideal low-pass filter be approximated using a time-limited sequence, i.e., a sequence being identically

zero outside a finite interval? An approximation is derived in (6.127) by assuming that $g_t^{\vdash t \dashv} \neq 0$ (with a few exceptions) for $-N \leq t \leq N$, and $g_t^{\vdash t \dashv} = 0$ for $|t| > N$, i.e., that $(g_t^{\vdash t \dashv})$ is a time-limited sequence in $L^2(-\infty, \infty)$. Using $(g_t^{\vdash t \dashv})$, a second sequence $h_{2N+1}(t)$ is constructed: $h_{2N+1}(t) = g_t^{\vdash t \dashv} - f_t$ for $-N \leq t \leq N$, and $h_{2N+1}(t) = f_t$ for $|t| > N$. Further assuming that the three pairs (f_t) and $F(s)$, $h_{2N+1}(t)$ and $H_{2N+1}(s)$, as well as $(g_t^{\vdash t \dashv})$ and $G^{\vdash t \dashv}(s)$, are Fourier transform pairs as required in (6.49) and (6.50), then $H_{2N+1}(s) = G^{\vdash t \dashv}(s) - F(s)$ is obtained, applying (6.48,5).

When approximating in the sense that the norm of differences becomes minimal as required in (6.5), Parseval's identity is applied. Parseval's identity is defined in (6.75) for Fourier transform pairs as in (6.29) and (6.30), and in the remarks to (6.75) for pairs as in (6.35). For pairs $(f_t) = \mathcal{F}_{-\mathrm{i}}\big(F(s)\big)$ as in (6.49) and (6.50), Parseval's identity becomes $\sum_{t=-\infty}^{\infty} f_t^2 = \int_{-1/2}^{1/2} |F(s)|^2 \mathrm{d}s$. When this version is applied, $\int_{-1/2}^{1/2} |H_{2N+1}(s)|^2 \mathrm{d}s = \sum_{t=-\infty}^{\infty} |h_{2N+1}(t)|^2 = \sum_{t=-N}^{N} |g_t^{\vdash t \dashv} - f_t|^2 + \sum_{t<-N} |f_t|^2 + \sum_{t>N} |f_t|^2$ is obtained. This result implies that the integral in (6.127) becomes minimal provided that $g_t^{\vdash t \dashv} = f_t$ for $|t| \leq N$.

Let $F(s)$ and (f_t) be, as required in (6.126), the frequency and the impulse response function of an ideal low-pass filter. Then, in the set of all sequences of length $2N+1$,

$$\text{the sequence } (f_t^{\vdash t \dashv}) = \begin{cases} (f_t) & \text{for } |t| \leq N, \\ 0 & \text{for } t < -N \text{ and } N < t \end{cases} \tag{6.127}$$

minimises

$$\int_{-1/2}^{1/2} \big|F^{\vdash t \dashv}(s) - F(s)\big|^2 \mathrm{d}s, \quad F^{\vdash t \dashv}(s) = \sum_{t=-N}^{N} f_t^{\vdash t \dashv} \mathrm{e}^{-\mathrm{i}2\pi s t}$$

(f_t), $|t| \leq N$, as in (6.127) is therefore the mean square approximation with length $2N+1$ of (f_t), $t = \ldots, -1, 0, 1, \ldots$, (f_t) and $F(s)$ being the impulse response sequence and frequency response function in a low-pass filter, as proposed in (6.126) and defined in (6.125). Mean square approximations of a variety of lengths of a low-pass filter are calculated in Problem 6.32.

A low-pass filter can alternatively be approximated using a time-limited sequence with the property that its Fourier transform comes as close as possible to zero for frequencies $|s| > s_0$, $0 < s_0 < 1/2$, as required in (6.125). This problem is akin to building an even data taper $h_N^{(e)}(t)$ as defined in (6.110) that minimises the leakage when a discrete Fourier transform is calculated, because (6.109) implies that there is no even data taper $h_N^{(e)}(t)$ with pertaining kernel $H_N^{(e)}(s) = 0$ for $|s| > s_0$, N a finite integer number, $0 < s_0 < 1/2$ and s_0 as small as possible.

This problem is solved by generating a band-limited sequence which is nearly time-limited: in the set of all sequences being limited to the band $-s_0 \leq s \leq s_0$, $s_0 < 1/2$ as required in (6.108,2), those with maximal

$\lambda = \sum_{t=-(N-1)/2}^{(N-1)/2} h_t^2 / \sum_{t=-\infty}^{\infty} h_t^2$ are sought for a given odd N, i.e., those sequences attaining their highest concentration in a time interval of length N. Band-limited sequences with maximal concentration are comprehensively dealt with in [125]. There, they are called *discrete prolate spheroidal sequences (DPSSs)*. DPSSs are used as data tapers in multi-taper spectral estimation [108]. Multi-taper spectral estimation is a relatively new procedure ([134] and [135]). It is, however, not a topic in this book. Multi-taper spectral estimation has some advantages (e.g., the one mentioned in the remarks to Fig. 10.14) when compared with the classic spectral estimation introduced in Chaps. 8, 9 and 10.

6.9 Supplements

Some properties of orthogonal trigonometric functions, of the discrete Fourier transform, of sums of trigonometric functions and of Fourier transform pairs are derived in Sects. 6.9.1, 6.9.2, 6.9.3 and 6.9.4. Thereafter, proofs of Bessel's inequality and Parseval's identity are given in Sect. 6.9.5. In Sects. 6.9.6 and 6.9.7, properties of the autocorrelation are derived and demonstrated. In Sect. 6.9.8, discrete Fourier transforms are calculated using R function `fft()`.

6.9.1 Orthogonal Trigonometric Functions

The properties of the trigonometric functions in the set (6.9) are summarised in (6.10), (6.11), (6.12) and (6.13).

The identities in (6.10) can be derived, for $k \neq 0$ and $k \neq N/2$, when the functions are written in exponential representation. Thereafter, the sums of the finite geometric series are obtained using $\sum_{k=0}^{N-1}(cw)^k = \big(1-(cw)^N\big)/\big(1-cw\big)$, w complex, c real. For example, the sum of the cosine functions is obtained

$$
\begin{aligned}
&\sum_{t=0}^{N-1} \cos\big(2\pi\frac{k}{N}t\big) = \frac{1}{2}\sum_{t=0}^{N-1}\Big(\mathrm{e}^{\mathrm{i}2\pi(k/N)t} + \mathrm{e}^{-\mathrm{i}2\pi(k/N)t}\Big) \qquad \begin{array}{l} k \text{ integer and} \\ k \neq 0, (N/2) \end{array} \\
&= \frac{1}{2}\sum_{t=0}^{N-1}\big(z^t + z^{-t}\big), \text{ with } z = \mathrm{e}^{\mathrm{i}2\pi(k/N)} \\
&= \frac{1}{2}\left(\frac{1-z^N}{1-z} + \frac{1-z^{-N}}{1-z^{-1}}\right) = \frac{1}{2}\left(\frac{1-\mathrm{e}^{\mathrm{i}2\pi k}}{1-z} + \frac{1-\mathrm{e}^{-\mathrm{i}2\pi k}}{1-z^{-1}}\right) \\
&= \frac{1}{2}\left(\frac{1-\big(\cos(2\pi k)+\mathrm{i}\times\sin(2\pi k)\big)}{1-z} + \frac{1-\big(\cos(2\pi k)-\mathrm{i}\times\sin(2\pi k)\big)}{1-z^{-1}}\right) \\
&= \frac{1}{2}\left(\frac{1-\big(1+\mathrm{i}0\big)}{1-z} + \frac{1-\big(1-\mathrm{i}0\big)}{1-z^{-1}}\right) \\
&= \frac{1}{2}\left(\frac{0}{1-z} + \frac{0}{1-z^{-1}}\right) = 0
\end{aligned}
\tag{6.128}
$$

on condition that $k \in (\ldots, -1, 0, 1, \ldots)$, $k \neq 0$ and $k \neq (N/2)$.

For $k = 0$ or $k = N/2$, the sums of the cosine functions are straightforwardly obtained in (6.129)

$$\sum_{t=0}^{N-1} \cos\big(2\pi \frac{k}{N} t\big) = \begin{cases} 0 & \text{for k = N/2} \\ N & \text{for k = 0} \end{cases} \tag{6.129}$$

because $k = N/2$ implies an even N, and in $\sum_{t=0}^{N-1} \cos(\pi t)$ therefore, pairs of subsequent terms add to zero. Herewith, the sum of the cosines in (6.10) has been derived. With a similar derivation, the sum of the sines is obtained.

The orthogonality relations in (6.12) are derived by applying the trigonometric identities and then calculating the sums of the exponential representations as demonstrated in (6.128). For $t = 0, 1, ..., N-1$ and $k, l = 0, 1, ..., m$, with m being the largest integer number such that $m \leq N/2$,

$$\begin{aligned} S_{k,l} &= \sum_{t=0}^{N-1} \cos\big(2\pi \frac{k}{N} t\big) \cos\big(2\pi \frac{l}{N} t\big) \\ &= \frac{1}{2} \sum_{t=0}^{N-1} \left(\cos\Big(\frac{2\pi t}{N}(k+l)\Big) + \cos\Big(\frac{2\pi t}{N}(k-l)\Big) \right) \\ &= \begin{cases} (1/2)\big(N+N\big) = N & \text{for } k = l = 0,\ N/2 \\ (1/2) \sum_{t=0}^{N-1} \cos\Big(((2\pi t)/N) 2k\Big) + N/2 = N/2 & \text{for } k = l \\ 0 & \text{for } k \neq l \end{cases} \\ &= \begin{cases} N & \text{for } k = l = 0,\ N/2 \\ N/2 & \text{for } k = l \\ 0 & \text{for } k \neq l \end{cases} \end{aligned} \tag{6.130}$$

is obtained because $\sum_{t=0}^{N-1} \cos\big(2\pi(k/N)t\big) = 0$ for $k \neq 0, N/2$, as shown above in (6.128). The orthogonality relations as in (6.11) and (6.13) are arrived at with similar derivations.

The infinite number of trigonometric functions in the set (6.14) have properties as summarised in (6.15), (6.16), (6.17), (6.18) and (6.19). The integrals (6.15) and (6.16) are the counterparts to the sums in (6.10) in the case of the number of functions being finite. These integrals are arrived at as shown in (6.131)

$$\begin{aligned} \int_{-T/2}^{T/2} \cos\big(2\pi \frac{k}{T} t\big) \mathrm{d}t &= \frac{T}{2\pi k} \sin\big(2\pi \frac{k}{T} t\big) \Big|_{-T/2}^{T/2} \\ &= \frac{T}{2\pi k} \sin(\pi k) - \frac{T}{2\pi k} \sin(-\pi k) = 0 \end{aligned} \tag{6.131}$$

for the cosine function and positive k, and, thereafter, lend themselves to prove the orthogononality relations as in (6.17), (6.18) and (6.19), subsequent to applying the trigononometric identities as demonstrated above in the derivation of (6.130).

6.9.2 Discrete Fourier Transform

(6.22,3) is derived by substituting F_k in definition (6.22,2) (or in (6.132)) with F_k as defined in (6.22,1) (or in (6.133)). Both Fourier transforms are periodic with period N, as concluded from the remarks to (6.22), and therefore applying (6.24), the sums are calculated for one period.

$$(f_u) = \sum_{k=0}^{N-1} F_k \mathrm{e}^{\mathrm{i}2\pi(u/N)k} \qquad \text{for } u = \ldots, -1, 0, 1, \ldots \tag{6.132}$$

$$(F_k) = \frac{1}{N} \sum_{t=0}^{N-1} f_t \mathrm{e}^{-\mathrm{i}2\pi(k/N)t} \qquad \text{for } k = \ldots, -1, 0, 1, \ldots \tag{6.133}$$

$$f_u = \sum_{k=0}^{N-1} \left(\frac{1}{N} \sum_{t=0}^{N-1} f_t \mathrm{e}^{-\mathrm{i}2\pi(k/N)t} \right) \mathrm{e}^{\mathrm{i}2\pi(u/N)k} \qquad \begin{array}{l}\text{for } k, u, t = \\ 0, \ldots, N-1\end{array} \tag{6.134}$$

$$= \frac{1}{N} \sum_{t=0}^{N-1} f_t \sum_{k=0}^{N-1} \mathrm{e}^{\mathrm{i}2\pi(k/N)(u-t)} \tag{6.135}$$

The sum in (6.135) results from applying (6.10) as follows

$$\begin{aligned} \sum_{k=0}^{N-1} \mathrm{e}^{\mathrm{i}2\pi(k/N)(u-t)} &= \sum_{k=0}^{N-1} \cos(2\pi \frac{u-t}{N} k) + \mathrm{i} \sum_{k=0}^{N-1} \sin(2\pi \frac{u-t}{N} k) \\ &= \begin{cases} 0 & \text{for } u - t \neq 0 \\ N & \text{for } u = t \end{cases} \end{aligned}$$

and, using this result,

$$\begin{aligned} f_u &= \frac{1}{N} \sum_{t=0}^{N-1} f_t \left\{ \begin{array}{ll} 0 & \text{for } u - t \neq 0 \\ N & \text{for } u = t \end{array} \right\} \qquad \text{with } u, t = 0, \ldots, N-1 \\ &= f_t \qquad \text{for } u = t \end{aligned} \tag{6.136}$$

is obtained. Consequently, (f_t) and (F_k) in (6.22,1,2) (or alternatively in (6.22,4)) are discrete Fourier transform pairs as required in (6.22,3).

6.9.3 Sums of Trigonometric Functions

Sums of trigonometric functions with frequencies $s \neq \ldots, -1, 0, 1, \ldots$ are arrived at in (6.137) and (6.138). The sums in (6.10) hold for the trigonometric functions as defined in (6.9) with frequencies $s_k = k/N$, $t, k = \ldots, -1, 0, 1, \ldots$, $0 \leq t \leq N-1$ and $0 \leq k \leq (N/2)$. Can sums of trigonometric functions be obtained when their frequencies are not restricted to $s_k = k/N$? For example, the sum of the cosine functions

$$\sum_{t=0}^{N-1} \cos(2\pi s t) = 1 + \cos(2\pi s) + \cos(4\pi s) + \ldots + \cos(2\pi(N-1)s)$$

is multiplied on both sides with $2\sin(2\pi(1/2)s)$ and, thereafter, applying identity $2\sin x\cos y = \sin(x-y)+\sin(x+y) = \sin(x+y)-\sin(y-x)$

$$\begin{aligned}
2\sin(2\pi\tfrac{1}{2}s)\sum_{t=0}^{N-1}\cos(2\pi st) &= 2\sin(2\pi\tfrac{1}{2}s)+\\
&\sin(2\pi(1+\tfrac{1}{2})s)-\sin(2\pi(1-\tfrac{1}{2})s)+\\
&\sin(2\pi(2+\tfrac{1}{2})s)-\sin(2\pi(2-\tfrac{1}{2})s)+\\
&\vdots\\
&\sin(2\pi(N-2+\tfrac{1}{2})s)-\sin(2\pi(N-2-\tfrac{1}{2})s)+\\
&\sin(2\pi(N-1+\tfrac{1}{2})s)-\sin(2\pi(N-1-\tfrac{1}{2})s)\\
&= \sin\big(2\pi(N-1+\tfrac{1}{2})s\big)+\\
&\qquad 2\sin(2\pi\tfrac{1}{2}s)-\sin(2\pi(1-\tfrac{1}{2})s)\\
&= \sin\big(2\pi(N-\tfrac{1}{2})s\big)+\sin\big(2\pi\tfrac{1}{2}s\big)
\end{aligned}$$

is obtained. The sum of the sine functions is arrived at by applying the identity $2\sin x\sin y = \cos(x-y)-\cos(x+y) = 2\sin y\sin x$.

Again applying the trigonometric identities, the sums obtained are written as products to obtain the results as in (6.137) and (6.138).

$$\begin{aligned}
\sum_{t=0}^{N-1}\cos(2\pi st) &= \frac{\sin\big(2\pi(N-\frac{1}{2})s\big)+\sin\big(2\pi\frac{1}{2}s\big)}{2\sin\big(2\pi\frac{1}{2}s\big)} \quad s\neq\ldots,-1,0,1,\ldots\\
&= \frac{\sin\big(2\pi\frac{N}{2}s\big)\cos\big(2\pi\frac{N-1}{2}s\big)}{\sin\big(2\pi\frac{1}{2}s\big)} \qquad (6.137)\\
\sum_{t=0}^{N-1}\sin(2\pi st) &= \frac{\cos\big(2\pi\frac{1}{2}s\big)-\cos\big(2\pi(N-\frac{1}{2})s\big)}{2\sin\big(2\pi\frac{1}{2}s\big)} \quad s\neq\ldots,-1,0,1,\ldots\\
&= \frac{\sin\big(2\pi\frac{N}{2}s\big)\sin\big(2\pi\frac{N-1}{2}s\big)}{\sin\big(2\pi\frac{1}{2}s\big)} \qquad (6.138)
\end{aligned}$$

Substituting s in (6.137) and (6.138) with k/N, the sums in (6.10) and (6.11) are arrived at on condition that $k\neq 0, N/2$. For even N, $\sin\big(2\pi(N/2)(k/N)\big) = \sin(\pi k) = 0$ with $k = 1,2,\ldots,m-1$. For odd N, $\sin\big(2\pi(N/2)(k/N)\big) = \sin(\pi k) = 0$ with $k = 1,2,\ldots,m$, m as in the remarks to (6.9) and (6.13).

6.9.4 Properties of Fourier Transform Pairs

Properties as proposed in (6.48) that apply to Fourier transform pairs as defined in (6.35) are derived in (6.139), (6.140), (6.141), (6.142) and (6.143). However, these properties are shared with Fourier transform pairs for cases

(6.20,1,2,4). For example, convolution theorem (6.48,7) is derived in (6.145) for Fourier transform pairs as in (6.49) and (6.50).

(6.48,1) directly follows from (6.35), and (6.48,2) directly from (6.37), because $\cos(2\pi st)$ is even whereas $\sin(2\pi st)$ is odd.

(6.48,3) is shown, for $a < 0$, by substituting $u = -at$. This substitution implies $t = -u/a$, $\mathrm{d}u/\mathrm{d}t = -a$, $\mathrm{d}t = (1/-a)\mathrm{d}u$ and integrating from ∞ to $-\infty$.

$$\begin{aligned}\mathcal{F}_{-\mathrm{i}}\big(f(-at)\big) &= \int_{-\infty}^{\infty} f(-at)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t = \int_{\infty}^{-\infty} f(u)\mathrm{e}^{-\mathrm{i}2\pi s\frac{-u}{a}}\frac{1}{-a}\,\mathrm{d}u \\ &= \frac{-1}{-a}\int_{-\infty}^{\infty} f(u)\mathrm{e}^{-\mathrm{i}2\pi\frac{s}{-a}u}\mathrm{d}u = \frac{1}{|-a|}\int_{-\infty}^{\infty} f(u)\mathrm{e}^{-\mathrm{i}2\pi\frac{s}{-a}u}\mathrm{d}u \\ &= \frac{1}{|-a|}F\Big(\frac{s}{-a}\Big) \end{aligned} \tag{6.139}$$

For $a > 0$, $\mathcal{F}_{-\mathrm{i}}\big(f(at)\big) = (1/a)F(s/a)$ is obtained. Thus, $\mathcal{F}_{-\mathrm{i}}\big(f(at)\big) = (1/|a|)F(s/a)$, for $a \neq 0$. When both sides are multiplied with $|a|^{1/2}$, $\mathcal{F}_{-\mathrm{i}}\big(|a|^{1/2}f(at)\big) = (1/|a|^{1/2})F(s/a)$ is arrived at, being a more symmetric form of (6.48,3).

(6.48,4) follows by substituting $u = t-b$, $\mathrm{d}u/\mathrm{d}t = 1$, $\mathrm{d}u = \mathrm{d}t$, $\mathrm{d}(t-b) = \mathrm{d}t$.

$$\begin{aligned}\mathcal{F}_{-\mathrm{i}}\big(f(t-b)\big) &= \int_{-\infty}^{\infty} f(t-b)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\ &= \int_{-\infty}^{\infty} f(t-b)\mathrm{e}^{-\mathrm{i}2\pi s(t-b+b)}\mathrm{d}(t-b) \\ &= \int_{-\infty}^{\infty} f(t-b)\mathrm{e}^{-\mathrm{i}2\pi s(t-b)}\mathrm{e}^{-\mathrm{i}2\pi bs}\mathrm{d}(t-b) \\ &= F(s)\mathrm{e}^{-\mathrm{i}2\pi bs} \end{aligned} \tag{6.140}$$

(6.48,5) holds because integration is linear:

$$\begin{aligned}H(s) &= \int_{-\infty}^{\infty} h(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t = \int_{-\infty}^{\infty}\big(af(t)+bg(t)\big)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\ &= \int_{-\infty}^{\infty} af(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t + \int_{-\infty}^{\infty} bg(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\ &= aF(s) + bG(s) \end{aligned} \tag{6.141}$$

(6.48,6) holds because the functions in a Fourier transform pair as defined in (6.35) are in $L^2(-\infty,\infty)$. Applying definition (2.28,1), and because $\int_{-\infty}^{\infty} f(t-u)\mathrm{e}^{\mathrm{i}2\pi s(t-u)}\mathrm{d}t = \int_{-\infty}^{\infty} f(r)\mathrm{e}^{\mathrm{i}2\pi sr}\mathrm{d}r$ holds for all u, convolution theorem (6.48,6) is obtained.

$$H(s) = \int_{-\infty}^{\infty} h(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t = \int_{-\infty}^{\infty}\Big(\int_{-\infty}^{\infty} g(u)f(t-u)du\Big)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)f(t-u)\mathrm{e}^{-\mathrm{i}2\pi s(u+t-u)}\mathrm{d}u\ \mathrm{d}t \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(u)\mathrm{e}^{-\mathrm{i}2\pi su}f(t-u)\mathrm{e}^{-\mathrm{i}2\pi s(t-u)}\mathrm{d}u\ \mathrm{d}t \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(t-u)\mathrm{e}^{-\mathrm{i}2\pi s(t-u)}\mathrm{d}t\right)g(u)\mathrm{e}^{-\mathrm{i}2\pi su}\mathrm{d}u \\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t\right)g(u)\mathrm{e}^{-\mathrm{i}2\pi su}\mathrm{d}u \\
&= \int_{-\infty}^{\infty} F(s)g(u)\mathrm{e}^{-\mathrm{i}2\pi su}\mathrm{d}u = F(s)G(s) \qquad (6.142)
\end{aligned}
$$

(6.48,7) also holds because the functions are in $L^2(-\infty,\infty)$. Substituting $f(t)$ with the +i-transform of $F(s)$, this convolution theorem is obtained as follows:

$$
\begin{aligned}
H(s) &= \int_{-\infty}^{\infty} h(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t = \int_{-\infty}^{\infty} f(t)g(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\
&= \int_{t=-\infty}^{\infty}\left(\int_{-\infty}^{\infty} F(r)\mathrm{e}^{+\mathrm{i}2\pi rt}\mathrm{d}r\right)g(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\
&= \int_{-\infty}^{\infty} F(r)\left(\int_{-\infty}^{\infty} g(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{e}^{\mathrm{i}2\pi rt}\mathrm{d}t\right)\mathrm{d}r \\
&= \int_{-\infty}^{\infty} F(r)\left(\int_{-\infty}^{\infty} g(t)\mathrm{e}^{-\mathrm{i}2\pi(s-r)t}\mathrm{d}t\right)\mathrm{d}r \\
&= \int_{-\infty}^{\infty} F(r)G(s-r)\mathrm{d}r = F(s) * G(s) \qquad (6.143)
\end{aligned}
$$

Convolution theorem (6.48,7) is obtained in (6.145) for Fourier transform pairs as defined in (6.49) and (6.50), i.e., for case (6.20,4). Assume that (f_t), (g_t) and (h_t), $t = \ldots,-1,0,1\ldots$, are sequences in $L^2(-\infty,\infty)$, and that $F(s)$, $G(s)$ and $H(s)$, s real, are functions in $L^2(-1/2,1/2)$, and periodic with period 1. Further assume that (f_t) and $F(s)$, (g_t) and $G(s)$ as well as (h_t) and $H(s)$ are Fourier transform pairs as defined in (6.49) and (6.50). Then, in contrast to derivation (6.143) for case (6.20,3), start in (6.144) with the convolution in the frequency domain.

In (6.144), convolution is defined differently than in (2.28,1) where it is defined for the case of non-periodic functions. Since $F(s)$, $G(s)$ and $H(s)$ are functions defined in $-1/2 \le s \le 1/2$ and periodic with period 1, the limits in the convolution integral become $-1/2$ and $1/2$, i.e., one integrates over one complete period.

$$
H(s) = F(s) * G(s) = \int_{-1/2}^{1/2} F(s-r)G(r)\mathrm{d}r \qquad (6.144)
$$

$$
\begin{aligned}
&= \int_{-1/2}^{1/2} F(s-r) \left(\sum_{t=-\infty}^{\infty} g(t) \mathrm{e}^{-\mathrm{i}2\pi rt} \right) \mathrm{d}r \\
&= \sum_{t=-\infty}^{\infty} \left(\int_{-1/2}^{1/2} F(s-r) \mathrm{e}^{\mathrm{i}2\pi(s-r)t} \mathrm{d}r \right) g(t) \mathrm{e}^{-\mathrm{i}2\pi st} \\
&= \sum_{t=-\infty}^{\infty} \left(\int_{-1/2}^{1/2} F(r) \mathrm{e}^{\mathrm{i}2\pi rt} \mathrm{d}r \right) g(t) \mathrm{e}^{-\mathrm{i}2\pi st} \\
&= \sum_{t=-\infty}^{\infty} (f_t)(g_t) \mathrm{e}^{-\mathrm{i}2\pi st} \qquad (6.145)
\end{aligned}
$$

In the second line, $G(s)$ is written as the Fourier transform of (g_t). Since $F(s)$ is in $L^2(-1/2, 1/2)$, integration and summation are interchanged and the third line results. The fourth line is obtained since $F(s)$, being periodic with period 1, can be integrated in an arbitrary interval of length 1, as concluded from the remarks to (6.24). Consequently, $(h_t) = (g_t)(f_t)$ because $H(s) = \mathcal{F}_{-\mathrm{i}}(h_t)$ as assumed above.

6.9.5 Bessel's Inequality and Parseval's Identity

In (6.74) and in (6.36) it is assumed that $f(t)$ is in $L^2(-T/2, T/2)$, implying that its norm $\|f(t)\|$, and therefore $(2/T) \int_{-T/2}^{T/2} \big(f(t)\big)^2 \mathrm{d}t$, exist. Since a square is positive,

$$
\begin{aligned}
0 &\le \int_{-T/2}^{T/2} \big(f(t) - f_M(t)\big)^2 \mathrm{d}t \le \int_{-T/2}^{T/2} \big(f(t)\big)^2 - 2f(t)f_M(t) + \big(f_M(t)\big)^2 \mathrm{d}t \\
&\le \int_{-T/2}^{T/2} \big(f(t)\big)^2 \mathrm{d}t - 2 \int_{-T/2}^{T/2} \big(f(t) f_M(t)\big) \mathrm{d}t + \int_{-T/2}^{T/2} \big(f_M(t)\big)^2 \mathrm{d}t \\
&\le S_1 - S_2 + S_3 \qquad (6.146)
\end{aligned}
$$

is obtained. The second term is integrated using the Fourier coefficients A_k and B_k and the Fourier partial sum $f_M(t)$ as in (6.28,2,3).

$$
\begin{aligned}
S_2 &= 2 \int_{-T/2}^{T/2} f(t) \left(\frac{A_0}{2} + \sum_{k=1}^{M} \big(A_k \cos(2\pi s_k t) + B_k \sin(2\pi s_k t) \big) \right) \mathrm{d}t \\
&= 2 \frac{A_0}{2} \int_{-T/2}^{T/2} f(t) \mathrm{d}t + \\
&\qquad 2 \sum_{k=1}^{M} \left(A_k \int_{-T/2}^{T/2} f(t) \cos(2\pi s_k t) \mathrm{d}t + B_k \int_{-T/2}^{T/2} f(t) \sin(2\pi s_k t) \mathrm{d}t \right) \\
&= 2 \frac{A_0}{2} \frac{T A_0}{2} + 2 \sum_{k=1}^{M} \left(\frac{T}{2} A_k^2 + \frac{T}{2} B_k^2 \right) = 2 \frac{T}{4} A_0^2 + 2 \frac{T}{2} \sum_{k=1}^{M} \big(A_k^2 + B_k^2 \big) \qquad (6.147)
\end{aligned}
$$

In the third term, the mixed products become identically zero since the trigonometric functions (6.14) are orthogonal:

$$\begin{aligned} S_3 &= \int_{-T/2}^{T/2} \left(\frac{A_0}{2} + \sum_{k=1}^{M} \big(A_k \cos(2\pi s_k t) + B_k \sin(2\pi s_k t) \big) \right)^2 \mathrm{d}t \\ &= \frac{T}{4} A_0^2 + \sum_{k=1}^{M} \left(A_k^2 \int_{-T/2}^{T/2} \cos(2\pi s_k t) \cos(2\pi s_k t) \mathrm{d}t + \right. \\ &\qquad \left. B_k^2 \int_{-T/2}^{T/2} \sin(2\pi s_k t) \sin(2\pi s_k t) \mathrm{d}t \right) \\ &= \frac{T}{4} A_0^2 + \sum_{k=1}^{M} \left(A_k^2 \frac{T}{2} + B_k^2 \frac{T}{2} \right) = \frac{T}{4} A_0^2 + \frac{T}{2} \sum_{k=1}^{M} \big(A_k^2 + B_k^2 \big) \,. \end{aligned} \tag{6.148}$$

Substituting, (6.74,2) is obtained in (6.149).

$$\begin{aligned} 0 &\le \int_{-T/2}^{T/2} \big(f(t) \big)^2 \mathrm{d}t - \frac{T}{4} A_0^2 - \frac{T}{2} \sum_{k=1}^{M} \big(A_k^2 + B_k^2 \big) \\ \frac{A_0^2}{2} + \sum_{k=1}^{M} \big(A_k^2 + B_k^2 \big) &\le \frac{2}{T} \int_{-T/2}^{T/2} \big(f(t) \big)^2 \mathrm{d}t \end{aligned} \tag{6.149}$$

In (6.149), the sequence of the partial sums in the series of the squared Fourier coefficients A_k and B_k does not decrease for $M = 1, 2, ...$ and is bounded since the integral exists, as assumed above. As a consequence, (6.74,3) holds.

6.9.6 Properties of the Autocorrelation

The $\star$-transformation as defined in (6.102) has properties similar to those of the convolution (i.e., the $*$-transformation) which is defined in (2.28).

Similar to the derivation (6.145), it is assumed that (f_t) and $F(s)$, (g_t) and $G(s)$ as well as (h_t) and $H(s)$ are Fourier transform pairs as defined in (6.49) and (6.50). In the first line of the following derivation, (h_t) is assumed to result from (f_t) and (g_t) by applying the $\star$-operator as defined in (6.102). The sixth line is arrived at because $\sum_{t=-\infty}^{\infty} f(t+u) \mathrm{e}^{-\mathrm{i}2\pi s(t+u)} = \sum_{v=-\infty}^{\infty} f(v) \mathrm{e}^{-\mathrm{i}2\pi s v}$ for every u, $t, u, v = \ldots, -1, 0, 1, \ldots$.

$$\begin{aligned} H(s) &= \sum_{t=-\infty}^{\infty} h_t \mathrm{e}^{-\mathrm{i}2\pi s t} \quad \text{with } (h_t) = (f_t) \star (g_t) \\ &= \sum_{t=-\infty}^{\infty} \left(\sum_{u=-\infty}^{\infty} g_u f_{t+u} \right) \mathrm{e}^{-\mathrm{i}2\pi s t} \\ &= \sum_{t=-\infty}^{\infty} \sum_{u=-\infty}^{\infty} g_u f_{t+u} \mathrm{e}^{-\mathrm{i}2\pi s (u+t-u)} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{t=-\infty}^{\infty} \sum_{u=-\infty}^{\infty} g_u \mathrm{e}^{+\mathrm{i}2\pi su} f_{t+u} \mathrm{e}^{-\mathrm{i}2\pi s(t+u)} \\
&= \sum_{u=-\infty}^{\infty} \left(\sum_{t=-\infty}^{\infty} f_{t+u} \mathrm{e}^{-\mathrm{i}2\pi s(t+u)} \right) g_u \mathrm{e}^{+\mathrm{i}2\pi su} \\
&= \sum_{u=-\infty}^{\infty} \left(\sum_{t=-\infty}^{\infty} f_t \mathrm{e}^{-\mathrm{i}2\pi st} \right) g_u \mathrm{e}^{+\mathrm{i}2\pi su} \\
&= \sum_{u=-\infty}^{\infty} F(s) g_u \mathrm{e}^{+\mathrm{i}2\pi su} = F(s)\overline{G(s)} \qquad (6.150)
\end{aligned}
$$

In (6.150), $F(s)\overline{G(s)}$ in the frequency domain corresponds to $(f_t) \star (g_t)$ in the time domain, whereas, in (6.145), $F(s)G(s)$ corresponds to $(f_t) * (g_t)$ in the time domain. Please note that $\overline{G(s)}$ is the Fourier +i-transform of (g_t) provided that $G(s)$ is the Fourier −i-transform of (g_t).

In the time domain, $*$- and $\star$-transformations can be distinguished clearly when they are performed using the paper strip device as demonstrated in (2.38). When calculating a convolution sum as required in (2.38), (f_t) is written on a sheet of paper and a copy (g_{t-u}) with the values in (g_t) in reversed order is written on a paper strip which is then displaced relatively to the sheet: the result is obtained by summing the products of pairs of values (f_t, g_{t-u}) appearing on the same line. An autocorrelation sum is calculated using the same procedure, but without reversing the order of the values in (g_t).

Using (6.150), the *autocorrelation theorem* is easily obtained. Let (f_t) be a real-valued sequence in $L^2(-\infty, \infty)$ with Fourier transform $F(s)$ as in (6.49) and (6.50). Then definition (6.103) implies $f \star f(t) = f \star f(-t)$, comparable to (2.4,3) implying (2.7,1) and (2.9,4). Thus being even as defined in (6.32), the autocorrelation has a real-valued Fourier transform $\mathcal{F}_{-\mathrm{i}}\big(f \star f(t)\big) = F(s)\overline{F(s)} = |F(s)|^2$, as obtained when, in (6.150), (g_t) is substituted with (f_t).

Further, $f \star f(t)$ is maximal in $t = 0$. This property is, in the case of time-limited sequences, easily derived, since the Cauchy-Schwarz inequality as defined in (6.4) also applies to sums. Let (f_t) be a time-limited sequence as in (6.108,1) with $f_t = 0$ for $t < t_2$ and $f_t = 0$ for $t > t_3$, $t_2 < t_3$. The autocorrelation of (f_t) becomes identically zero, i.e., $f \star f(t) = \sum_{t=-\infty}^{\infty} f_t f_{t+u} = 0$, on condition that all products in the autocorrelation sum become identically zero, i.e., for displacements $|u| > t_3 - t_2$. This becomes obvious when $f \star f(t)$ is calculated using the paperstrip device (2.38), however modified as required above.

With $t_1 < t_2 - (t_3 - t_2)$ and $t_4 > t_3 + (t_3 - t_2)$, $\sum_{t=t_1}^{t_4}(f_t) = \sum_{t=t_2}^{t_3}(f_t) = \sum_{t=t_1}^{t_4}(f_{t+u})$ and $\sum_{t=t_1}^{t_4}(f_t)^2 = \sum_{t=t_2}^{t_3}(f_t)^2 = \sum_{t=t_1}^{t_4}(f_{t+u})^2$ are obtained for displacements $|u| > t_3 - t_2$, since both sums contain identical terms. Consequently, both sequences, (f_t) and its displaced duplicate (f_{t+u}), $|u| >$

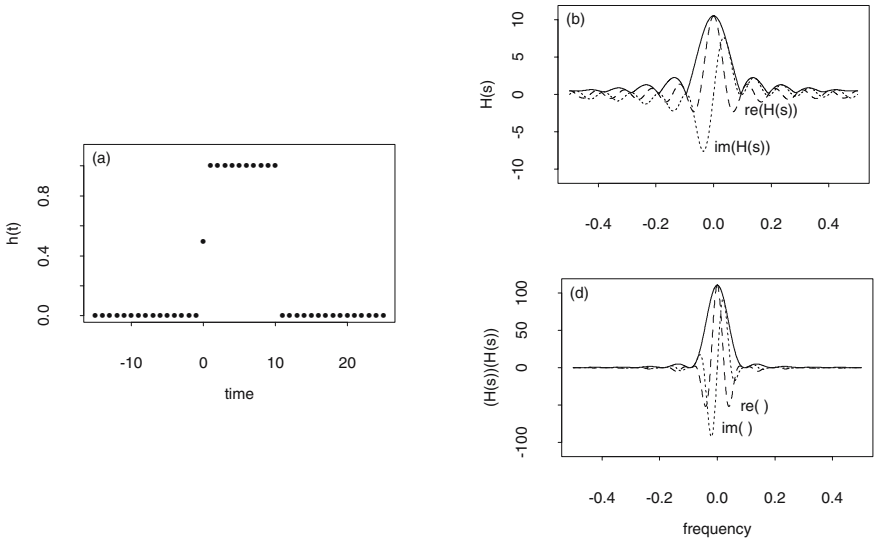


Fig. 6.33. Sequence (h_t) (a) together with its Fourier transform $H(s)$ as defined in (6.49) and (6.50) (b). The real and imaginary parts $\mathrm{re}(H(s))$ and $\mathrm{im}(H(s))$ are plotted with broken lines, $|H(s)|$ with a solid line. Plot (d) contains $H(s)H(s)$ with spectrum $|H(s)|^2$ (solid line).

$t_3 - t_2$, are in $L^2(t_1, t_4)$. Substituting, in a version of the Cauchy-Schwarz inequality for sequences, $\phi(u)$ with (f_t) and $\psi(u)$ with (f_{t+u}):

$$\left(\sum_{t=t_1}^{t_4} f_t f_{t+u}\right)^2 \leq \sum_{t=t_1}^{t_4} (f_t)^2 \sum_{t=t_1}^{t_4} (f_{t+u})^2 \leq \left(\sum_{t=t_1}^{t_4} (f_t)^2\right)^2 \tag{6.151}$$

$\sum_{t=t_1}^{t_4} f_t f_{t+u} \leq \sum_{t=t_1}^{t_4} (f_t)^2$ is obtained.

For example, sequence (h_t) as plotted in Fig. 6.33 (a) is time-limited with $t_2 = 0$ and $t_3 = 10$. Its autocorrelation $h \star h(t)$ becomes identically zero for displacements $|u| > 10 - 0 = 10$ as shown in Fig. 6.34 (c). Consequently, $t_1 < -10$ and $t_4 > 20$.

(6.151) also applies to a time series (x_t), $t = 1, 2, \ldots, N$, since its empirical covariance function $\hat{c}_X(\tau)$ is its autocorrelation normalised with $(1/N)$, as concluded from a comparison of definitions (2.1) and (6.103). Consequently, $\hat{c}_X(\tau)$ is maximal for lag $\tau = 0$. $\hat{c}_X(\tau)$ shares this property with the covariance function $c_X(\tau)$ of a stationary stochastic process (X_t), shown to be maximal for $\tau = 0$ in the remarks to (2.9).

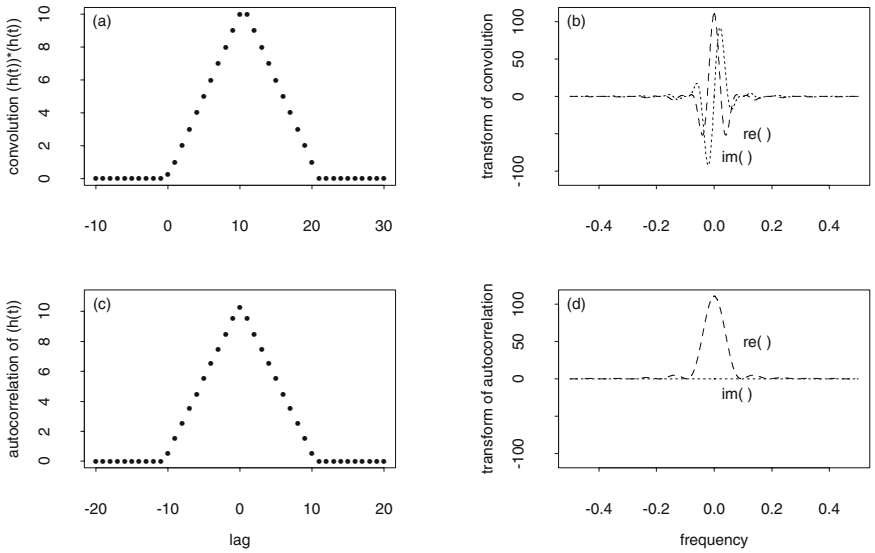


Fig. 6.34. Convolution of $h * h(t)$ (a) and autocorrelation $h \star h(t)$ (c), (h_t) as in Fig. 6.33 (a), together with their Fourier transforms as defined in (6.49) and (6.50), (b) and (d).

6.9.7 Demonstration of the Convolution Theorem

Convolution theorem (6.48,6) for Fourier transform pairs as defined in (6.49) and (6.50) is demonstrated in Figs. 6.33 and 6.34. In these figures, the sequences in the time domain are plotted on the left side (at the exterior of the intervals shown, the sequences are identically zero), and the functions in the frequency domain are plotted on the right side for interval $-1/2 \le s \le 1/2$, being periodic with period 1. Sequence (h_t) in Fig. 6.33 (a) is mixed, as required in (6.32), and therefore has, as concluded from (6.37), a complex-valued Fourier transform $H(s)$ as plotted in Fig. 6.33 (b). Using complex multiplication, $H(s)H(s)$ as plotted in Fig. 6.33 (d) is arrived at, with $|H(s)|^2$ being the spectrum of (h_t) as defined in (6.76).

$h*h(t)$, the self-convolution of (h_t), is mixed and thus has a complex-valued Fourier transform $\mathcal{F}_{-\mathrm{i}}\big(h*h(t)\big)$. $h*h(t)$ and $\mathcal{F}_{-\mathrm{i}}\big(h*h(t)\big)$ can be seen above in Fig. 6.34, plots (a) and (b). Compare Fig. 6.34 (b) with Fig. 6.33 (d) to obtain a demonstration of $\mathcal{F}_{-\mathrm{i}}\big(h*h(t)\big) = H(s)H(s)$ as required in (6.48,6).

$h\star h(t)$, the autocorrelation of (h_t) as defined in (6.103), has the Fourier transform $\mathcal{F}_{-\mathrm{i}}\big(h\star h(t)\big) = H(s)\overline{H(s)} = |H(s)|^2$, as derived in (6.150). $h\star h(t)$ and $|H(s)|^2$ are shown below in Fig. 6.34, plots (c) and (d). $h\star h(t)$ is even, and thus $|H(s)|^2$ is real-valued. Since (6.150) implies $\mathcal{F}_{-\mathrm{i}}\big(f\star f(t)\big) = |F(s)|^2$,

and since the spectrum of a function is the squared absolute value of its Fourier transform as defined in (6.76), it is obvious that the spectrum can be obtained by calculating the Fourier transform of the autocorrelation of a function. Usually, however, a spectrum is calculated as the square of the absolute value of its Fourier transform.

6.9.8 Autocorrelation and Discrete Fourier Transform in R

$h*h(t)$ and $h\star h(t)$ as in Fig. 6.34 (a) and (c) are easily calculated from $(h_t) = (1/2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1)$ for $t = 0, 1, 2, \ldots, 10$ and $h_t = 0$, for t at the exterior of this interval, as in Fig. 6.33 (a). (h_t) is convolved with itself using the paper strip device (2.38) to obtain $h*h(t)$, and, when the values in (h_t) are written on the paper strip without reversing their order, the autocorrelation $h\star h(t)$ is obtained. In R, a convolution is calculated with `filter()`, introduced in the remarks to Fig. 2.16, whereas an autocorrelation as defined in (6.103) can be calculated using `acf()` on condition that the result is re-scaled to account for factor $(1/N)$ in (2.1,3) which is lacking in (6.103), as is demonstrated in Problem 9.5. Without using these R functions, $h*h(t)$ and $h\star h(t)$ are obtained with the following R expressions:

```
N0 <- 5              #11 values not being identically zero
N1 <- 2*N0+1         #as in Fig. 6.33 (a)
h1 <- (rep(1,N1))    #in R vector h1, first non-zero value
h1[1] <- 1/2         #at first place
#convolve h1 with itself
h3 <- c(rep(0,N1),h1)   #h3 for convolution (h1)*(h1)
h3 <- c(h3,rep(0,N1))
h3 <- as.vector(filter(h3,h1,sides=1))
#first non-zero value at first place
h3 <- h3[(N1+1):(3*N1-1)]
#autocorrelation of h1
N2 <- 4*N0+1             #number of non-zero elements
h2 <- rep(0,N2)          #in the autocorrelation of h1
h2[N1] <- sum(h1*h1)     #displacement 0
for(i in 1:(N1-1) ) {    #---> to slow for large N <---
  h2[N1+i] <- sum( (h1[1:(N1-i)])*(h1[(1+i):N1]) )
  h2[N1-i] <- h2[N1+i]  #h2 is even
}
```

Fourier transforms $H(s)$, $\mathcal{F}_{-\mathrm{i}}\big(h*h(t)\big)$ and $\mathcal{F}_{-\mathrm{i}}\big(h\star h(t)\big)$, as plotted on the right in Figs. 6.33 and 6.34, are calculated applying (6.49) for real $-1/2 \le s \le 1/2$ from sequences (h_t), $h*h(t)$ and $h\star h(t)$ as plotted on the left in Figs. 6.33 and 6.34, the sequences being in $L^2(-\infty, \infty)$, $t = \ldots, -1, 0, 1, \ldots$. These sums are not as easily obtained as in (6.122) and (6.124), because these sequences have more than only two or three non-zero values. A well-known example for evaluating a Fourier sum as required in (6.49) is the derivation of

the Dirichlet kernel $D_N(s)$ in (6.55) from (6.54). (6.54), in turn, is calculated using (6.137) and (6.138). Can a Fourier sum as required in (6.49) be replaced, for practical purposes such as plotting, etc., by a discrete Fourier transform using (6.22) or (6.26)?

Discrete Fourier transforms, when computed using (6.26) from fairly long time slices including the parts with the non-zero values in (h_t), $h*h(t)$ and $h\star h(t)$ as plotted on the left in Figs. 6.33 and 6.34, are available at discrete frequencies s'_k. These frequencies and their intermediate distance $\Delta s'_k$ are defined in the remarks to (6.27). In a plot, these discrete Fourier transforms cannot be distinguished from $H(s)$, $\mathcal{F}_{-\mathrm{i}}\big(h*h(t)\big)$ or $\mathcal{F}_{-\mathrm{i}}\big(h\star h(t)\big)$ provided that $\Delta s'_k$ is sufficiently small.

$\Delta s'_k = 1/N' = 1/(2^p)$ becomes sufficiently small on condition that R vectors `h1, h2` and `h3`, with the non-zero values of these sequences as obtained above, are padded with 0 until a finite number of values $N' = 2^p$ is arrived at, $p = 1, 2, 3, \ldots$. The zero-padded sequences are then discretely Fourier transformed, with the results plotted on the right in Figs. 6.33 and 6.34.

In R, a discrete Fourier transform is computed using expression `Fvec <- fft(fvec,inverse=...)`. `fft(...,inverse=T)` calculates a $-$i-transform and `fft(...,inverse=F)` a $+i$-transform, as required in (6.22,4). `fvec`, the first argument in function `fft()`, is a real- or complex-valued vector with the sequence (f_t) to be transformed. `fft()` returns the discrete Fourier transform (F_k) of (f_t) in vector `Fvec`. `Fvec` is complex-valued. (F_k) is periodic with period N as shown in the remarks to (6.22), and $(F_{k'})$ is periodic with period N', when (f_t) is zero-padded as in (6.26). Consequently, $F(s_k)$ and $F(s'_k)$ are periodic with period 1 and are generally plotted in $-1/2 \le s \le 1/2$, as argued in the remarks to (6.24). As a further consequence, $\mathcal{F}_{+\mathrm{i}}(F_k)$ or $\mathcal{F}_{+\mathrm{i}}(F_{k'})$ are periodic with period N or N'. This symmetry is exploited by `fft()`, assuming that:

1. pairs $\big((f_t),(F_k)\big)$ or, subsequent to zero-padding (f_t), $\big((f_{t'}),(F_{k'})\big)$ are periodic with periods N or N', N or N' the number of values in argument vector `fvec`
2. `fvec[1]` $= f_0$ and `Fvec[1]` $= F_0$, i.e., the first value in both, argument and result vector, is the value of the sequences for $t = 0$ and $k = 0$, and
3. `fvec[1]` or `Fvec[1]` are followed by the values for $t > 0$ or $k > 0$, and thereafter by the values for $t < 0$ or $k < 0$, i.e., both half-periods of the sequences.

In this order, a discrete Fourier transform is obtained in `Fvec`, as demonstrated in the remarks to (6.26) and (6.27), provided that this order is adhered to in argument vector `fvec`, as demonstrated in the following examples.

As a first example, the discrete Fourier transform of (h_t), as in Fig. 6.33 (a) and as obtained above in R vector `h1`, is calculated by zero-padding `h1` on its right side; thus, `fft()` associates all non-zero values with non-negative $t = 0, 1, 2, \ldots$, since only zero-values are in the second half-period subsequent to zero-padding and therefore are associated with $t = \ldots, -2, -1$.

```
p <- 11                #h1, N1, N0 as above
N1V <- 2**p            #Delta sk becomes 1/(2**p)
H1V <- N1V - N1                    #number of zeros for padding
h1v <- c(h1,rep(0,H1V))            #argument vector for fft()
t1v <- 0:(N1V - 1)                 #for plotting h1v
H <- fft(h1v, inverse=T)           #-i-transformation
```

The zeros in the second half of `h1v` are then handled by `fft()` as if being the second-half period of the sequence to be transformed, i.e., pertaining to time points $t < 0$ as required in the definition of (h_t).

As a second example, vector `h3` containing $h * h(t)$ is also zero-padded on its right side prior to being transformed. One-sided padding is possible because the first non-zero value in vector `h3`, as constructed using the above R expressions from `h1`, is associated with time $t = 0$, as plotted in Fig. 6.34 (a). The values in the second half of `h3` after zero-padding, being associated with time points $t < 0$ by `fft()`, are then all identically zero, as required in the definition of $h*h(t)$ for $t < 0$.

In the third example, autocorrelations $h \star h(t)$ of (h_t) are not identically zero for negative lags. For this reason, vector `h3` as obtained above and plotted in Fig. 6.34 (c), is zero-padded on both sides to obtain $N' = 2^p$ in a first step. The zero-padded vector contains in its first half the $2^{(p-1)} - 1$ autocorrelations associated with time points $t < 0$, thereafter, at place $2^{(p-1)}$, the autocorrelation for $t = 0$, followed by the $2^{(p-1)}$ autocorrelations associated with time points $t > 0$. In the second step, the order is changed as required by `fft()`:

```
N2V <- 2**p                      #h2, N1, N2, p as above
H2Vright <- 2**(p-1) - (N1-1)  #one zero more on the right side
H2Vleft <- 2**(p-1) - N1       #than on the left one, N2 being odd
h2v <- c(rep(0,H2Vleft),h2)    #(as the sum for
h2v <- c(h2v,rep(0,H2Vright))  #even N in (6.24))
t2v <- -(2**(p-1)-1):(2**(p-1))      #for plotting
#h2v has now length N2V = 2**p, re-order for fft()
h2v <- c(h2v[(2**(p-1)):N2V], h2v[1:(2**(p-1)-1)])
HAH <- fft(h2v, inverse=T)           #-i-transformation
```

The discrete Fourier transform of $h \star h(t)$ is obtained in vector `HAH` with values ordered as in argument vector `h2v`: `HAH[1]` is the result for $s'_k = 0$, followed by the values for the positive and, thereafter, the negative frequencies. `HAH` is plotted by means of generating a vector with discrete frequencies s'_k obeying the order in `HAH`, or by rearranging `HAH` such that it can be plotted against the s'_k in normal order. The first possibility is chosen in the examples following (6.26) and (6.27).

More Fourier transform pairs as defined in (6.49) and (6.50) are plotted as examples in Fig. 6.35. Fig. 6.35 (a) shows sequence $(g_t) = (h_t)/\big(\sum_{t=-\infty}^{\infty} h_t\big)$, (h_t) as in Fig. 6.33 (a). Dividing by $\sum_{t=-\infty}^{\infty} h_t$ implies $\sum_{t=-\infty}^{\infty} g_t = 1$ and, as a further consequence, $G(s) = 1$ for $s = 0$, $G(s) = \mathcal{F}_{-\mathrm{i}}\big((g_t)\big)$, is obtained

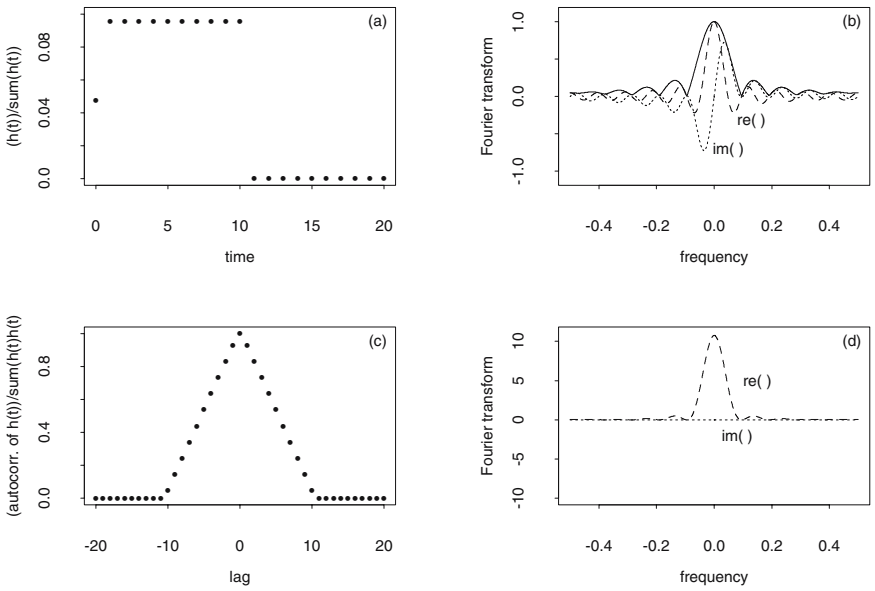


Fig. 6.35. Sequences $(h_t)/\big(\sum_{t=0}^{10} h_t\big)$ (a) and $h \star h(t)/\big(\sum_{t=0}^{10} h_t^2\big)$ (c), (h_t) as in Fig. 6.33 (a), together with their Fourier transforms as defined in (6.49) and (6.50), (b) and (d).

applying (6.48,1). A normalisation with the sum of the sequence is usually performed to obtain the impulse response function of a moving average as defined in (2.39) and discussed in (6.124).

Fig. 6.35 (c) shows $(f_t) = \big(h \star h(t)\big)/\Sigma(h^2)$, $\Sigma(h^2) = \sum_{t=-\infty}^{\infty} h_t^2$, (h_t) as plotted in Fig. 6.33 (a). Its Fourier transform $\big(1/\Sigma(h^2)\big)\mathcal{F}_{-\mathrm{i}}\big(h \star h(t)\big) = \big(1/\Sigma(h^2)\big)|H(s)|^2$ is obtained with (6.48,5) and the remarks to (6.48). Applying Parseval's identity, in its version as introduced in the remarks to (6.127) for Fourier transform pairs (6.49) and (6.50), $\Sigma(h^2) = \sum_{t=-\infty}^{\infty} h_t^2 = \int_{-1/2}^{1/2} |H(s)|^2 \mathrm{d}s$ is arrived at, implying $\big(1/\Sigma(h^2)\big) \int_{-1/2}^{1/2} |H(s)|^2 \mathrm{d}s = 1$. This result is also arrived at applying (6.48,1), because definition (6.103) implies that $h \star h(t)/\Sigma(h^2) = 1$ for $t = 0$.

6.10 Problems

6.1. Using R expressions as in the remarks to Fig. 6.1, the sawtooth-shaped sequence above in Fig. 6.1 is generated and the pertaining Fourier coefficients are calculated. Trigonometric oscillations with these Fourier coefficients are then added to the reconstruction of the sequence as plotted below in Fig. 6.1. Please reconstruct the sawtooth-shaped sequence using

1. the oscillations marked ∘ in Table 6.1 solely
2. the oscillations marked ∗ in Table 6.1 solely
3. all harmonics pertaining to the fundamental oscillation with frequency 5/140
4. all harmonics pertaining to the fundamental oscillation with frequency 28/140,

by replacing the R expressions for adding the oscillations with

```
#fundamental oscillation and harmonics for frequency 5/140
skk <- c(1*5+1,2*5+1,3*5+1,4*5+1,5*5+1,8*5+1)
#fundamental oscillation and harmonics for frequency 5/140 28/140
#skk <- c(1*28+1,2*28+1)
modsqsum <- 0.0        #sum of squared moduli of coefficients
for(i in 1:length(skk)) {
modsqsum <- modsqsum + (mod[skk[i]])*(mod[skk[i]])
}
xhracc <- rep(A[1], hrlen)  #accumulator
for(i in 1:length(skk)) {  #reconstruction
   xhr <- (rep(A[skk[i]],hrlen))*cos(2*pi*sk[skk[i]]*timehr) +
      (rep(B[skk[i]],hrlen))*sin(2*pi*sk[skk[i]]*timehr)
   xhracc <- xhracc + xhr
}
```

Compare the sums of the squared moduli of the coefficients pertaining to the oscillations used for the above reconstructions with the sum of the squared moduli of the coefficients pertaining to all oscillations used for the reconstruction below in Fig. 6.1.

6.2. Use (6.4) to show that the triangle inequality $\|\phi(t) + \psi(t)\| \leq \|\phi(t)\| + \|\psi(t)\|$ holds in $L^2(a, b)$.

6.3. R expressions in file `/path/fig66.R` generate functions in the sequences $y_k(x) = \mathrm{e}^{-(1/10)^k x^2}$, $k = 0, 1, 2, \ldots$ and $z_k(x) = \mathrm{e}^{-10^k x^2}$, $k = 1, 2, 3 \ldots$, some of which are plotted in Fig. 6.6. Please calculate $z_k^{(2)}(x) = \mathrm{e}^{-2^k x^2}$ for $x = -1.00, -9.99, \ldots, -0.01, 0.00, 0.01, \ldots, 9.99, 1.00$ and sum $\left(z_{k+1}^{(2)}(x) - z_k^{(2)}(x)\right)^2$ over all x for $k = 1, 2, \ldots, 20$. The results suggest that $z_k^{(2)}(x)$ is a Cauchy sequence in $L^2(-1.0, 1.0)$. For which k do the sums become less than $\epsilon = 10^{-12}$?

Is $y_k(x)$ a Cauchy sequence in $L^2(-\infty,\infty)$? Integrate the first line in the following derivation using the result of Problem 1.12

$$\begin{aligned}\int &= \int_{-\infty}^{\infty}\left(\mathrm{e}^{-(1/10^n)x^2}-\mathrm{e}^{-(1/10^m)x^2}\right)^2\mathrm{d}x\\ &= \int_{-\infty}^{\infty}\mathrm{e}^{-(2/10^n)x^2}\mathrm{d}x-2\int_{-\infty}^{\infty}\mathrm{e}^{-(1/10^n+1/10^m)x^2}\mathrm{d}x+\int_{-\infty}^{\infty}\mathrm{e}^{-(2/10^m)x^2}\mathrm{d}x\\ &= \sqrt{\pi}\left(\sqrt{10^n}/\sqrt{2}-2\sqrt{10^n\times 10^m}/\sqrt{10^n+10^m}+\sqrt{10^m}/\sqrt{2}\right)\end{aligned}$$

for $n=6$ and $m=4$, $n=8$ and $m=6$, $n=10$ and $m=8$, ... and show that this integral does not decrease with increasing n and m. Thus $y_k(x)$ is not a Cauchy sequence in $L^2(-\infty,\infty)$.

Then integrate $\int_{-\infty}^{\infty}\big(z_n(x)-z_m(x)\big)^2\mathrm{d}x$ for $m=1,\ldots,15$ and $n=m+2$, and show that this integral does not increase with increasing n and m. Thus $z_k(x)$ is a Cauchy sequence in $L^2(-\infty,\infty)$.

6.4. Please show that the sequence (F_k) (6.22) is periodic with period N: $(F_k)=(F_{k+pN})$, for integer k,p,N. Hint: F_k is the linear combination of f_t with the orthogonal trigonometric functions (6.9). Trigonometric functions are periodic with the periods in the remarks to (6.1).

6.5. Please zero-pad the sequences plotted above in Figs. 6.1 and 6.4 using the following R expressions, and then calculate the Fourier transforms.

```
#x is the sequence above in Fig. 6.1 being constructed using
#the R expressions in the remarks to this figure
N <- length(x)       #N' = 2N, N' and N as in the remarks to (6.27)
x0 <- c(x,rep(0,N))  #zero padded, x0 now has length 2N
x0t <- fft(x0, inverse=T)
s0k <- c( (0:N)/(2*N), -((N-1):1)/(2*N) )    #frequencies sk'
#now the sequence above in Fig. 6.4
x <- scan("/path/fig64.dat")
N <- length(x)
x0 <- c(x,rep(0,N))    #zero padded
x0t <- fft(x0, inverse=T)
s0k <- c( (0:N)/(2*N), -(N:1)/(2*N) )
```

Compare plots of the transforms and spectra with the plots in Figs. 6.2 and 6.4.

6.6. Generate sequences of lengths $l_1=2^{15}=32768$, $l_2=32769$, $l_3=32001$ and calculate Fourier transforms using `fft()`. Compare the durations of the calculations.

6.7. $f(x)=|x|$, $-2\le x\le 2$, is an even (6.32) function in $-2\le x\le 2$. A periodic version of $f(x)$ has period 4. Calculate the Fourier coefficients pertaining to $f(x)$ and approximate using (6.28,3) with $M=3$. Plot the

function together with its approximation by inserting the Fourier terms in the following `for()`-loop:

```
i <- 1
t1 <- -200:200
t2 <- t1/100
f1 <- <- abs(t2)
f3 <- -200:200
plot(t2,f1,type="n",xlab=" ",ylab=" ")
lines(t2[1:401],f1[1:401])
for(i in 1:401) {f3[i] <- ... + ... + ... }   #Fourier series
lines(t2,f3)
```

As an alternative, a Fourier series can be calculated using appropriate R vectors as arguments in the cosine functions, as demonstrated by the R expressions for the calculation of the Fourier representation of the saw-tooth shaped sequence in Fig. 6.1 (c).

6.8. Please find the defining equations of the periodic functions plotted in Fig. 6.7. Hint: the function in plot (c) is quadratic. Calculate the pertaining Fourier coefficients, and thereafter approximate the functions using (6.28,3) with $M = 3$.

6.9. Derive (6.37,4) for the case (6.20,3), i.e., for $f(t)$ and $F(s)$ being a Fourier transform pair (6.35). $f(t) = \text{even}(f(t)) + \text{odd}(f(t))$ is obtained using (6.32,6) and $\mathrm{e}^{-\mathrm{i}2\pi st} = \cos(2\pi st) - \mathrm{i} \times \sin(2\pi st)$ using (6.2,6). Substituting in (6.35)

$$\begin{aligned}
\overline{F(s)} &= \int_{t=-\infty}^{\infty} \Big(\text{even}\big(f(t)\big) + \text{odd}\big(f(t)\big)\Big)\Big(\cos(2\pi st) + \mathrm{i} \times \sin(2\pi st)\Big)\mathrm{d}t \\
&\dots \\
&= 2\int_{t=0}^{\infty} \text{even}\big(f(t)\big)\cos(2\pi st)\mathrm{d}t + \mathrm{i} \times 2\int_{t=0}^{\infty} \text{odd}\big(f(t)\big)\sin(2\pi st)\mathrm{d}t \\
&= \text{re}\big(F(s)\big) + \mathrm{i} \times \text{im}\big(F(s)\big) \qquad \text{where } F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)
\end{aligned}$$

is obtained, the last line using (6.2,1,2).

6.10. (6.40,2) holds since $\sin(x) < x < \tan(x)$, x being the arc in the unit circle. (6.40,4) is obtained by showing that $\int_{-\infty}^{\infty}\big(1/(bs)\big)(\sin(bs)\mathrm{d}s = \pi/|2b|$, since a product of two odd functions is even (6.32,4): with $b = 2\pi c$, the integral evaluates to $2\pi/|2 \times 2\pi c| = 1/|2c|$.

Hint: Show that $g'(u) = \mathrm{e}^{pu}\sin(qu)$, on condition that $g(u) = \big(1/(p^2 + q^2)\big)\mathrm{e}^{pu}\big(p\sin(qu) - q\cos(qu)\big)$, $g'(u)$ being the derivative of $g(u)$. Using this result, the two-fold integral $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{e}^{-xby}\sin(by)\mathrm{d}x\mathrm{d}y$ can be evaluated in two ways. Thereafter, $\int_{-\infty}^{\infty}\big(1/(bs)\big)(\sin(bs)\mathrm{d}s$ is easily calculated.

6.11. Calculate the Fourier transform of the shifted rectangle function $f(t) = a$, for $c - b \le t \le c + b$, $f(t) = 0$, for t at the exterior of this interval.

6.12. Calculate the Fourier transform $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ of the triangle function $f(t)$,

$$f(t) = \begin{cases} a - (1/b)|t| & \text{for } -b \le t \le b \\ 0 & \text{for } |t| > b \end{cases}$$

$$F(s) = \begin{cases} 2\frac{\sin(2\pi bs)}{2\pi s}(a-1) + \frac{1}{b}\Big(\frac{\sin(\pi bs)}{\pi s}\Big)^2 & \text{for } s \neq 0 \\ ab & \text{for } s = 0 \end{cases}$$

and plot this Fourier transform pair for $a = 1$ and $b = 1$. Compare the Fourier transform of the triangle function with the Fourier transform of the rectangle function (6.38) using (6.48,6). Hint:

$$F(s) = \int_{-b}^{b} \big(a - (1/b)|t|\big)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t = 2\int_0^b \big(a - (1/b)t\big)\cos(2\pi st)\mathrm{d}t$$
$$= 2a\int_0^b \cos(2\pi st)\mathrm{d}t - 2\frac{1}{b}\int_0^b t\cos(2\pi st)\mathrm{d}t = G(s) - H(s)$$

$G(s)$ is the Fourier transform of the even rectangle function $g(t)$ with height a and width $2b$, and $H(s)$ is obtained when integrating by parts. For $a = 1$ and $b > 0$, the Fourier transform pair

$$f(t) = \begin{cases} 1 - (1/b)|t| & \text{for } -b \le t \le b \\ 0 & \text{for } |t| > b \end{cases} \qquad F(s) = \begin{cases} \frac{1}{b}\Big(\frac{\sin(\pi bs)}{\pi s}\Big)^2 & \text{for } s \neq 0 \\ b & \text{for } s = 0 \end{cases}$$

is obtained.

6.13. Show that

$$\mathcal{F}_{+\mathrm{i}}\big(G(s) = \mathrm{e}^{-|s|}\big) = \frac{2}{1 + (2\pi t)^2} = g(t)$$

$$\mathcal{F}_{+\mathrm{i}}\left(F(s) = \begin{cases} \mathrm{e}^{-s} & \text{for } s \ge 0 \\ 0 & \text{for } s < 0 \end{cases}\right) = \frac{1 - \mathrm{i}2\pi t}{1 + (2\pi t)^2} = f(t)$$

and compare with the Fourier transform pairs (6.44) and (6.46).

6.14. Is the following proposition correct? The Fourier transform pair as defined in (6.49) and (6.50) is a special case of the Fourier transform pair as defined in (6.29) and (6.30) on condition that $T = 1$ and the roles of time and frequency are reversed.

6.15. Evaluate the first sum in (6.53)

$$\mathcal{F}_{-\mathrm{i}}\big(\Pi_N(t)\big) = \sum_{t=0}^{N-1} 1 \times \mathrm{e}^{-\mathrm{i}2\pi st} = \sum_{t=0}^{N-1} \cos(2\pi st) - \mathrm{i}\sum_{t=0}^{N-1} \sin(2\pi st)$$

using (6.137) and (6.138).

6.16. $F_N(s) = \mathcal{F}_{-\mathrm{i}}\big(\Lambda^{(e)}_{2N+1}(t)\big) = \mathcal{F}_{+\mathrm{i}}\big(\Lambda^{(e)}_{2N+1}(t)\big)$, since the Fourier transform of the even triangle sequence (6.56) is the Fejer kernel (6.57). The Fourier transform of the even triangle sequence can also be calculated using direct summation, however, these calculations are more intricate than the arguments in the lines preceding (6.57):

$$\begin{aligned} F_N(s) &= \sum_{t=-N}^{N} \left(1 - \frac{|t|}{N}\right) \mathrm{e}^{\mathrm{i}2\pi st} \\ &= \frac{1}{N} \sum_{t=-N}^{N} (N - |t|) z^t = \frac{1}{m} \sum_{t=-(m-1)}^{m-1} (m - |t|) z^t \end{aligned}$$

with $m = N$, since $m - |t| = 0$ for $|t| = m$. This sum is obtained when the sums over the diagonals of the following matrix are calculated: the main diagonal sums to mz^0, the first subdiagonal above to $(m-1)z^{-1}$, below to $(m-1)z^1$, etc.

$$\begin{array}{c} \\ 0 \\ 1 \\ \dots \\ u \\ \dots \\ m-1 \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & -1 & \dots & -v & \dots & -(m-1) \end{array} \\ \begin{pmatrix} z^0 z^0 & z^0 z^{-1} & & z^0 z^{-v} & & z^0 z^{-(m-1)} \\ z^1 z^0 & z^1 z^{-1} & & z^1 z^{-v} & & z^1 z^{-(m-1)} \\ & & & & & \\ z^u z^0 & z^u z^{-1} & & z^u z^{-v} & & z^u z^{-(m-1)} \\ & & & & & \\ z^{m-1} z^0 & z^{m-1} z^{-1} & & z^{m-1} z^{-v} & & z^{m-1} z^{-(m-1)} \end{pmatrix} \end{array}$$

Sum this matrix over the rows and the columns

$$\begin{aligned} F_N(s) &= \frac{1}{m} \sum_{u=0}^{m-1} \sum_{v=0}^{m-1} z^u z^{-v} = \frac{1}{m} \sum_{u=0}^{m-1} z^u \sum_{v=0}^{m-1} z^{-v} = \frac{1}{m} \sum_{u=0}^{m-1} z^u \frac{1 - z^{-m}}{1 - z^{-1}} \\ &\dots \\ &= \frac{1}{N} \left(\frac{\sin(\pi N s)}{\sin(\pi s)} \right)^2 \qquad s \neq \dots, -1, 0, 1, \dots \end{aligned}$$

to obtain the Fejer kernel.

6.17. For $s = \dots, -1, 0, 1, \dots$ the sum

$$\begin{aligned} F_N(s) &= \frac{1}{m} \sum_{t=-(m-1)}^{m-1} (m - |t|) \big(\cos(2\pi st) + \mathrm{i} \times \sin(2\pi st) \big) \qquad s = \dots, -1, 0, 1, \dots \\ &\dots \\ &= 2m - 1 - (m-1) = m \end{aligned}$$

is arrived at.

6.18. Show that $f_q(t) = q/\big(\pi(t^2 + q^2)\big)$ is a defining sequence for $\delta(t)$ provided that $q = 1, 1/2, 1/3, \dots$. Plot the functions for some q.

6.19. Given are the sine and cosine functions

$$f(t) = a\cos(2\pi s_1 t) \quad \text{and} \quad g(t) = b\sin(2\pi s_2 t)$$

a, b, s_1, s_2 real constants. Please calculate their Fourier transforms,

$$F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big) = (a/2)\delta(s - s_1) + (a/2)\delta(s + s_1)$$
$$G(s) = \mathcal{F}_{-\mathrm{i}}\big(g(t)\big) = (b/2)\delta(s - s_2) - (b/2)\delta(s + s_2)$$

$\delta(s)$ being the delta function as defined in (6.60). Hint: Derive the Fourier transforms as the limiting case of two $(\sin x)/x$ functions as defined in (6.39). Then use (6.40,4) to obtain $\int_{-\infty}^{\infty} F_b(s)\mathrm{d}s = 1$ on condition that $F_b(s) = (2b\sin(2\pi bs))/(2\pi bs)$ for $s \neq 0$ and $F_b(s) = 2b$ for $s = 0$. If b increases, then (i) $F_b(s)$ increases in the absolute values of the minima and maxima and (ii) increases its frequency. In the limit for $b \to \infty$, $F_b(s) \to \infty$ for $s = 0$ is obtained, however, $F_b(s)$ is not identically zero for $s \neq 0$, and, consequently, $F_b(s)$ is not reconcilable with (6.60,1). Ignoring this deficiency, assume that $F_b(s)$ is a defining sequence for $\delta(s)$ [32].

6.20. Define $\delta(s)$ using sequence $f_q(t)$, $q = 1, 2, \ldots$,

$$f_q(t) = \frac{q}{\pi(t^2 + q^2)} \qquad \frac{\mathrm{d}f_q(t)}{\mathrm{d}t} = \frac{-2tq}{\pi(t^2 + q^2)^2}$$

with first derivatives $\mathrm{d}f_q(t)/\mathrm{d}t$. Plot $f_q(t)$ and $\mathrm{d}f_q(t)/\mathrm{d}t$ for some q. Show that $\int_{-\infty}^{\infty} t(\mathrm{d}f_q(t)/\mathrm{d}t)\mathrm{d}t = -1$.

6.21. The first order linear differential equation $x(t) + ax'(t) = w(t)$ is solved in Problem 2.11. The stationary solution $x(t) = \int_{-\infty}^{t} g(t - u)w(u)\mathrm{d}u$, with $g(t) = 0$, for $t < 0$, and $g(t) = (1/a)\ \mathrm{e}^{-(1/a)t}$, for $t \geq 0$, is obtained. Evaluate the convolution integral for some $w(t)$: rectangle and triangle functions, sine and cosine functions, the unit step function, i.e., $w(t) = 0$ for $t < 0$ and $w(t) = 1$ for $t \geq 0$, and the delta function, using integration and (6.48,6).

6.22. Use the Fourier series obtained in Problem 6.7 together with Parseval's identity (6.74) to sum $1/(1^4) + 1/(2^4) + \ldots + 1/(n^4) + \ldots$.

6.23. Show by direct integration that

$$\mathcal{F}_{-\mathrm{i}}\left(2\mathrm{e}^{-|at|}\cos(2\pi rt)\right) = \frac{2a}{a^2 + \big(2\pi(s - r)\big)} + \frac{2a}{a^2 + \big(2\pi(s + r)\big)}$$

6.24. In Figs. 6.17, 6.19, 6.20 and 6.21, $G(s'_k)$ is calculated for frequencies with spaces $\Delta s'_k = 1/N'$. $\Delta s'_k \approx (1/2)\Delta s_k = 1/N$ because $N = 201, 301, 2401$ and $N' = 401, 601, 2801$. Draw $\Delta s_k = 1/301$ and $\Delta s'_k = 1/601$ in the plot in Fig. 6.21.

6.25. File `/path/problem625.R` contains R expressions which are used to calculate $f(t)$ and $F(s)$ as well as (g_t) and $G(s_k)$ as required in (6.77), (6.78), (6.79) and (6.80).

a) Fig. 6.15 contains plots of $f(t)$, $F(s)$, (g_t) and $G(s_k)$ for $s_1 = 0.1$, $s_2 = 0.7$, $a_1 = 0.5$, $a_2 = 0.15$ as well as $N = 41$. Use other values for s_1 and s_2. For which s_1 and s_2 does aliasing occur?

b) Fig. 6.17 contains plots of $f(t)$, $F(s)$, (g_t) and $G(s'_k)$ for $s_1 = 0.01$, $s_2 = 0.02$, $a_1 = 0.002$, $a_2 = 0.0001$ as well as $N = 201$ and $N' = 401$; Figs. 6.19 and 6.20 plots $F(s)$ and $G(s'_k)$ for $N = 301$ and $N' = 601$, for $N = 2401$ and $N' = 4801$ as well as for $N = 65535$ and $N' = 65536$. Reconstruct these plots using larger and smaller a_1 and a_2.

6.26. Sequence $g_N(t) = 1/(2N)$ for $-N < t < N$, $g_N(t) = 1/(4N)$ for $t = N$ or $t = -N$, and $g_N(t) = 0$ for $|t| > N$ is plotted for $N = 5$ in Fig. 6.36 (e). Compute the equivalent, σ-, and autocorrelation widths pertaining to this sequence for $N = 5$. Thereafter compute these widths for the even rectangle and triangle sequences in Fig. 6.36, plots (c) and (a).

6.27. Draw the widths at half height (6.97) in the plots (b), (d) and (f) in Fig. 6.36. Which function has the smallest width at half height? Which the largest? Which of the sequences $\Pi_N^{(e)}(t)$ and $2N\big(g_N(t)\big)$, $g_N(t)$ in Problem 6.26, do you prefer as an even data taper (6.110)?

6.28. Use the R expressions in file `/path/problem628.R` to generate sequence (g_t) of the damped trigonometric oscillations plotted in Fig. 6.27 with a broken line. Then compute even cosine tapers $\widehat{h_N^{(e)}}(t)$ (6.111,2) for $N = 65535$ and $p_1 = 0.001$, $p_2 = 0.01$, $p_3 = 0.05$, $p_4 = 0.10$, $p_5 = 0.15$, etc. Thereafter, multiply (g_t) with $\widehat{h_N^{(e)}}(t)$ and add zeros to obtain zero-padded sequences with length $N' = 65536$ prior to arriving at $G^{(p_i)}(s'_k)$, $i = 1, 2, \ldots$ by applying a discrete Fourier transform as defined in (6.26) and (6.27).

6.29. Show that a band-limited sequence (6.109) is not time-limited.

6.30. Generate sequences of trigonometric oscillations with a variety of frequencies other than those of the sequences plotted in Fig. 6.29. Thereafter calculate the differences of the sequences generated and plot input and output of the difference filter. Hint: the sequences plotted in Fig. 6.29 can be generated using R expressions as in `/path/diff.R`. In these R expressions, you can substitute `filter()` for `diff()`, if you prefer `filter()`.

6.31. Apply R expressions as in `/path/diff.R` to generate discrete-time trigonometric oscillations having a variety of frequencies. Thereafter smooth the oscillations thus obtained with $(1/3)\Pi_3^{(e)}(u)$, i.e., the impulse response function in (6.124). Repeat the smoothing operation using $(1/5)\Pi_5^{(e)}(u)$.

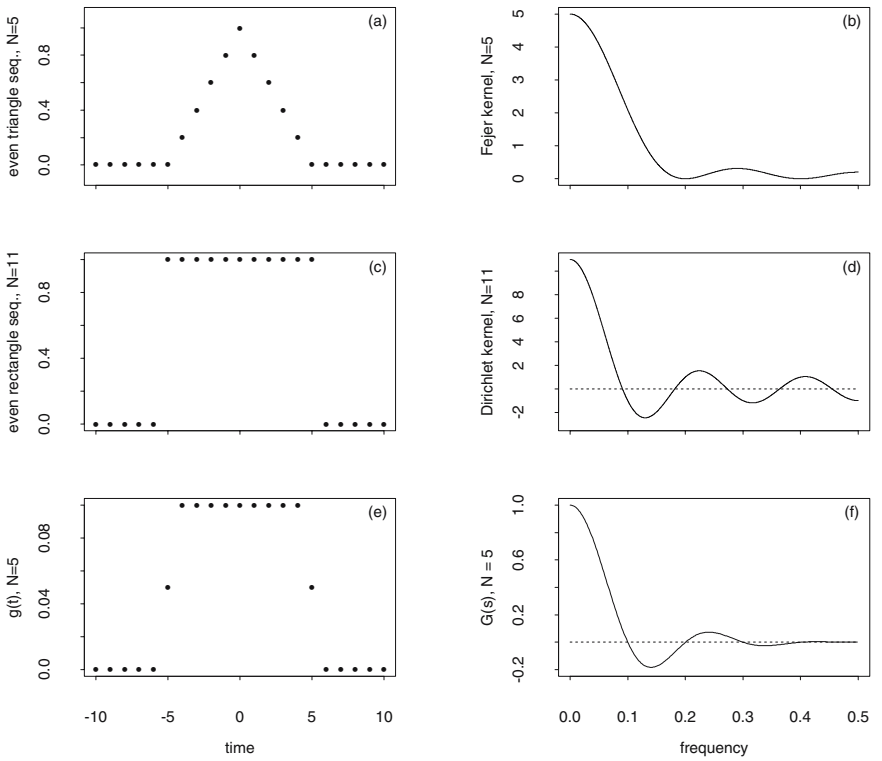


Fig. 6.36. Fourier transform pairs as defined in (6.49) and (6.50). Even triangle sequence $\Lambda^{(e)}_{2N+1}(t)$ (6.56) for $N = 5$ (a), even rectangle sequence $\Pi^{(e)}_N(t)$ (6.52) for $N = 11$ (b), $g_N(t)$ (Problem 6.26) for $N = 5$ (c).

6.32. Approximate a low pass filter as defined in (6.125) with $s_0 = 0.05$ and $s_0 = 0.2$ using time-limited impulse response functions as obtained in (6.127). A solution is easily obtained by slightly modifying the R expressions used for the generation of the plots in Fig. 6.23. These R expressions are available in `/path/fig623.R`.

6.33. The local linear models for the yearly values in the Basel temperature series as plotted in Fig. 6.32 jump at the limits of the time slices. Smaller jumps are obtained by displacing the limits slightly. Can local models with no jumps be estimated? Hint: Linear models as formulated in R are introduced in Sect. 3.2.1 and the constant term in a linear model can be excluded (i.e., the model is forced through origin) with `-1`.

7 Fourier Representation of a Stationary Stochastic Process

In Chap. 6, a deterministic function $f(t)$ with integer or real argument t is shown to have a Fourier representation $F(s_k)$ or $F(s)$ on condition that $f(t)$ is in $L^2(0, N-1)$ or $L^2(\infty, \infty)$ (for integer t) or in $L^2(-T/2,\, T/2)$ or $L^2(-\infty, \infty)$ (for real t). If $f(t)$ has an $F(s_k)$ or $F(s)$ it is then possible to distribute its power on the frequencies of the trigonometric oscillations in $F(s_k)$ or $F(s)$ by calculating the spectrum of $f(t)$ as defined in (6.76).

Now let $x(t)$, $-\infty < t < \infty$, be a realisation of a discrete- or continuous-time stationary random function $X(t)$. Can then $x(t)$ be represented as a Fourier series or Fourier integral? Obviously, this is not possible because, in general, $x(t)$ is neither periodic (therefore, Fourier representations cannot be obtained for intervals $T/2 \le x \le T/2$ (or $0 \le x \le N-1$ for integer t)) nor in $L^2(-\infty, \infty)$. $x(t)$ is not in $L^2(-\infty, \infty)$ since its norm is not finite in $-\infty < t < \infty$, on condition that $X(t)$ is a stationary and ergodic random function: $\int_{-T/2}^{T/2} \big(x(t)\big)^2 \mathrm{d}t \overset{T\to\infty}{\longrightarrow} +\infty$. Consequently, Fourier representations of realisations $x(t)$ do not exist and thus, it is not possible to obtain a Fourier representation of $X(t)$ by way of calculating the mean of the Fourier representations of $x(t)$. Despite this rationale, a stationary stochastic process does have a Fourier representation, as summarized in the Fourier representation theorem (7.51) and comprehensively dealt with in [113].

The Fourier representation of a stationary stochastic process is, in principle, a Fourier-Stieltjes integral of a complex-valued random function with real parameter. For this reason, an introduction to stochastic calculus and some examples of stochastic processes having a real parameter are given in Sects. 7.2 and 7.1. These include the Poisson process which is often used to model exponential inter-arrival (waiting) times and also the orthogonal increment process whose stationary derivative is used in a stochastic model for Brownian motion.

In Sect. 7.3.2, the claim that a stationary random function has a Fourier representation is formulated as an existence theorem, a proof, however, is given only for a special case, namely the harmonic process, in Sect. 7.3.1. In Sect. 7.4, spectra of linear processes are obtained from the coefficients in AR[p]-, MA[q]- or ARMA[p, q] models.

Sect. 7.5 contains the supplements and Sect. 7.6 the problems.

7.1 Independent and Poisson-distributed Events

A Poisson process, as introduced in Sect. 7.1.1, is a suitable model for the number of (i) alpha particles emitted by a radio-active substance, (ii) cars arriving at a toll station, (iii) incoming calls at a telephone call centre, (iv) spots on a sheet of paper produced by a paper mill, (v) crystal impurities in a semiconductor, (vi) failures of a data aquisition system, etc. A deeper insight into these processes is given in Sect. 7.1.2, and a further example in Sect. 7.1.3.

7.1.1 Poisson Process

Independent random events occurring with probability λ per time unit generate k points in the half-open (half-closed) interval $0 < u \leq t$, t, u real. For this interval, $(0, t]$ is written in this chapter. What then is the probability of obtaining k points in $(0, t]$, $k = 1, 2, 3, \ldots$? The answer follows.

Preparing for a heuristic derivation of the above probability, $(0, t]$ is divided into n subintervals of length $\Delta t = t/n$. Since every subinterval has the same probability of receiving one point, the probability of an event occurring in a subinterval is $\lambda \Delta t$. The probability that two events will occur within the same subinterval is $(\lambda \Delta t)^2$, since the events are independent, and the probability of three events occurring is $(\lambda \Delta t)^3$. If the product $\lambda \Delta t$ is small, then the probabilities of two or more events occurring are small enough to be neglected. Thus, the experiment results in k points in n subintervals with every subinterval having the same probability to hold a point. This corresponds to a sequence of n independent Bernoulli trials, and the probability to find k points in n subintervals is given by the binomial distribution (7.1).

$$\Pr(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}, \quad p = \lambda \Delta t, \quad n \Delta t = t \tag{7.1}$$

$$\Pr\big(N(t) = k\big) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^k}{k!}, \quad \begin{array}{l} \text{where } N(t) \text{ is the number} \\ \text{of events in interval } (0, t] \end{array} \tag{7.2}$$

Letting $\Delta t \to 0$, the discretisation inherent in this experiment is undone: $p \to 0$ such that $p/\Delta t = \lambda$ remains constant and $n \to \infty$ such that $n \Delta t = t$ remains constant as well. The result is the Poisson approximation (7.2) of the binomial distribution (7.1), as shown in an introduction to Statistics.

The above experiment generates a pattern of points in $(0, t]$ that could be described by the random function $Y(u) = 1$ for $0 < u \leq t$ with a point, and $Y(u) = 0$ for those $0 < u \leq t$ without a point. $Y(u)$ would be equal to 1 for a few time points in $(0, t]$ and equal to 0 for the majority of time points in $(0, t]$, in which case the essential features of the experiment, i.e., that the points are generated by random and independent events, would not be captured by $Y(u)$. Hence, the parameter of the random function is chosen to be the interval $(0, t]$ and the Poisson process defined in (7.3,2,3) becomes a special case of the count process defined in (7.3,1).

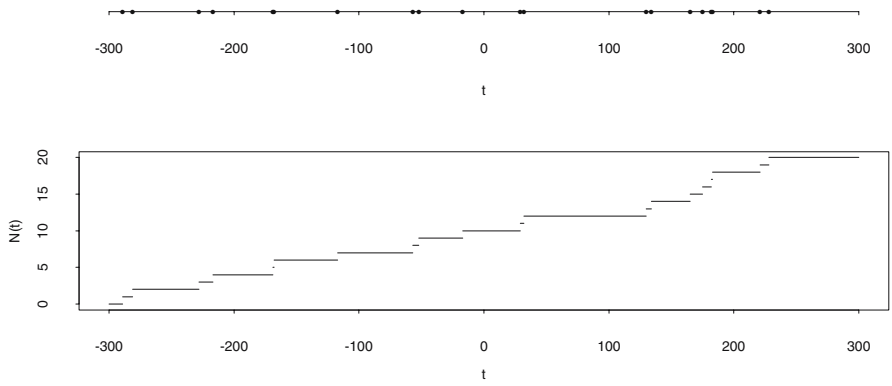


Fig. 7.1. A realisation of a general count process $N(t)$ defined in (7.3,1). Above, the events are plotted as points on the time axis; below, the step function $N(t)$.

$N(t)$ *is a (stochastic) count process on condition that*
1. $N(0) = 0$ *and* $N(t)$ *is the number of events (points) in* $(0, t]$, $(0, t]$ *being an interval closed on its right side.*
   $N(t)$ *is a Poisson process provided that:* (7.3)
2. $\Pr\big(N(t) = k\big) = \big(\mathrm{e}^{-\lambda t}(\lambda t)^k\big)/(k!)$ *and*
3. *increments* $N(t_1) - N(0) = N(t_1)$, $N(t_2) - N(t_1) = N(t_2)$, ... , $N(t_n) - N(t_{n-1}) = N(t_n)$ *are independent from each other.*

Increment $N(t_n) - N(t_{n-1})$ of a count process is the number of events (points) within interval $(t_n, t_{n-1}]$, $0 < t_n < t_{n-1} < t$. Every realisation of a count process is a step function on the real numbers with jumps of height 1, an example being given in Fig. 7.1. A count process similar to the one in (7.3,1) can be defined for a two-dimensional parameter: $N(x, y)$ is the number of events in $(0, 0, x, y]$, $N(0, 0) = 0$, and $N(x_2, y_2) - N(x_1, y_2) - N(x_2, y_1) + N(x_1, y_1)$ is the number of events in $(x_1, y_1, x_2, y_2]$. A Poisson process $N(t)$ is a count process with increments $\Delta N(t) = N(t + \Delta t) - N(t)$. Δt can then be chosen such that $\Delta N(t) = 0$ or $\Delta N(t) = 1$, provided that the $\Delta N(t)$ are independent from each other as required in (7.3,3).

If, in an application, it is supposed that the observed events are governed by a Poisson process, then an estimated Poisson distribution is compared with the empirical distribution of the events. Let $k_1, k_2, \ldots k_m$ be observed events in m intervals of duration T (when the parameter is one-dimensional) or in m plots with area F (when the parameter is two-dimensional). Then $\sum_{i=1}^{m} k_i = K$ is the total number of events and mT the total duration of the experiment or mF the total area under observation. $\hat{\lambda} = K/(mT)$ or $\hat{\lambda} = K/(mF)$ is an estimator for λ since, in (7.2), λ is the number of events per time unit or area unit. A Poisson distribution with $\hat{\lambda}$ is then compared with the empirical distribution using a probability plot, i.e., the quantiles of

the observed events are plotted versus the quantiles of a Poisson distribution with $\hat{\lambda}$. In Problem 7.3, for example, the locations with observations of the tree-line in the Valais alps, as in Fig. 3.10, are shown to be a realisation of a Poisson process.

7.1.2 Exponential Waiting Times

The number of alpha-particles, cars, telephone calls, etc. given as examples in the introduction to Sect. 7.1 do, obviously, not result from a sequence of independent Bernoulli trials, as has been assumed above in the heuristic derivation of (7.2). Rather, it is more realistic to assume that phenomena in nature and activities in society do exist which generate Poisson distributed events. A deeper insight into these phenomena and activities is gained from a derivation of the Poisson process based on the assumption that an event does not continue to have an effect on the next event.

This property is illustrated using a well-known Poisson process: the age, i.e., the lifetime X. Then $(X > u)$ is the event that a person lives longer than u from the time of his or her birth, and $\big(X > (t+u)\big)$ is the event that the remaining lifetime is larger than u on condition that a person is of age t. The conditional probability of living longer than $t+u$ when age t has been reached, is $\Pr\big(X > (t+u)\big)/\Pr(X > t)$. This is the probability for the remaining lifetime. It is equal to the probability for the total lifetime, $\Pr\big(X > (t+u)\big)/\Pr(X > t) = \Pr(X > u)$, on condition that there are no aftermaths: the past has no influence on the future, i.e., no conclusions regarding the future can be drawn from experiences made in the past. It is said that lifetime is memoryless.

The above result regarding the lifetime holds, in general, for memoryless random variables X which are usually inter-arrival times of random events. If the above probabilities are written using distribution functions then $F(t+u) = F(t)F(u)$ is obtained, an equation which is satisfied by exponential functions $\mathrm{e}^{-\lambda t}$. Consequently, a random variable is *memoryless* on condition that its distribution is exponential, as defined in (7.4).

$$F(t) = \Pr(X \le t) = \begin{cases} 1 - \mathrm{e}^{-\lambda t} & \text{for } t \ge 0 \\ 0 & \text{for } t < 0 \end{cases} \tag{7.4}$$

$$f(t) = \begin{cases} \lambda \mathrm{e}^{-\lambda t} & \text{for } t \ge 0 \\ 0 & \text{for } t < 0 \end{cases} \tag{7.5}$$

$$\mathrm{E}X = \int_0^\infty t\lambda \mathrm{e}^{-\lambda t}\mathrm{d}t \;=\; \frac{1}{\lambda} \tag{7.6}$$

$$\mathrm{Var}X = \int_0^\infty t^2\lambda \mathrm{e}^{-\lambda t}\mathrm{d}t - (\mathrm{E}X)^2 \;=\; \frac{1}{\lambda^2} \tag{7.7}$$

For example, 240 independent events are simulated in $(0, 300]$ such that their inter-arrival times are identically and exponentially distributed, as defined in (7.4), with $\lambda = 0.8$. These simulated events are indicated with

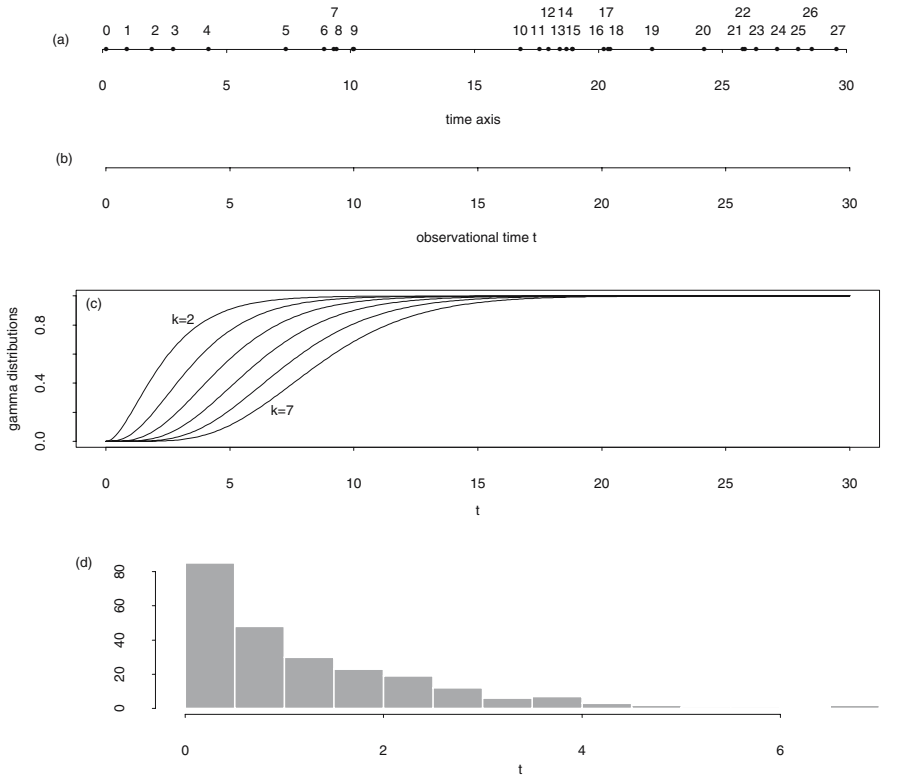


Fig. 7.2. Derivation of the Poisson distribution without discretising the time. Random points with exponential (as defined in (7.4), $\lambda = 0.8$) waiting times in $(0, 30]$, (symbol $\bullet$ in plot (a) together with their indices), observational time begins at point with index $k = 0$ (in plot (b)), gamma distributions $F_k(t)$ as in (7.8) and (7.9) (in plot (c)), and histogram of the waiting times X_k for 240 points in interval $(0, 300]$ (in plot (d)).

$0, 1, \ldots, k-1, k, k+1 \ldots, 239$, and plotted for subinterval $(0, 30]$ in Fig. 7.2 (a) with $\bullet$. The axis, below in Fig. 7.2 (b), constructed such that its origin coincides with event 0, is used to measure the arrival times, e.g., the event with index 9 arrives at $t \approx 10$. The inter-arrival times X_k are the times elapsed between the events as measured on this axis. Consequently, $X_0 = 0$ is the waiting time for event 0, and X_k, i.e., the time elapsed between the events with indices $k-1$ and k, is the inter-arrival time for event k. Under the above assumptions, the X_k are (exponentially) iid. The histogram of the realised inter-arrival times is plotted in Fig. 7.2 (d).

Are events Poisson distributed on condition that their inter-arrival times are exponentially distributed as defined in (7.4)? Let $Y_1, Y_2, Y_3, \ldots, Y_k$ be

the times elapsed (beginning with the occurrence of event 0) until events $0, 1, \ldots, k$ occur. Then $Y_0 = X_0 = 0$, $Y_1 = X_1$, $Y_2 = X_1 + X_2$, $Y_3 = X_1 + X_2 + X_3, \ldots, Y_k = \sum_{i=1}^{k} X_i$ are obtained, where X_i is the inter-arrival time of event i. Consequently, one waits no longer than t until event k occurs, $(Y_k \leq t)$, on condition that $\big((X_1 + \ldots X_k) \leq t\big)$ or $\big((X_1 + \ldots + X_k + X_{k+1}) \leq t\big)$ or $\big((X_1 + \ldots + X_k + X_{k+1} + \ldots) \leq t\big)$ occur. In a first step, probability $F_k = \Pr(Y_k \leq t)$ is determined. This probability is given by the gamma distribution in (7.8). The gamma distributions (7.8) and (7.9) are derived from the properties of the X_k in the remarks to (7.98) and (7.99).

$$\Pr(Y_k \leq t) = F_k(t) = 1 - \mathrm{e}^{-\lambda t}\left(1 + \frac{\lambda t}{1!} + \ldots + \frac{(\lambda t)^{k-1}}{(k-1)!}\right) \tag{7.8}$$

$$\Pr(Y_{k+1} \leq t) = F_{k+1}(t) = 1 - \mathrm{e}^{-\lambda t}\left(1 + \frac{\lambda t}{1!} + \ldots + \frac{(\lambda t)^{k-1}}{(k-1)!} + \frac{(\lambda t)^{k}}{k!}\right) \tag{7.9}$$

As an example, gamma distributions $F_k(t)$ as in (7.8) and (7.9), for $\lambda = 0.8$ and $k = 2, 3, 4, 5, 6, 7$, are plotted in Fig. 7.2 (c). From plots (a) and (b) in this figure it is obvious that events k with inter-arrival times X_k, $k = 1, 2, 3, 4$ are realised in $0 < t \leq 5$. The probability of this event is approximately 0.55, as shown in Fig. 7.2 (c).

In the second step, the Poisson distribution (as required in (7.3,2,3)) of the events is derived from the gamma distributions obtained in (7.8) and (7.9). (7.3,3) holds since the events are assumed to be independent, which implies that also their increments are independent on condition that Δt is chosen as required in the remarks to (7.3). It remains to show that $\Pr\big(N(t) = k\big) = \big(\mathrm{e}^{-\lambda t}(\lambda t)^k\big)/(k!)$.

A derivation using the following elementary results of probability theory is straightforward. Let A and B be events reconcilable with the probability axioms. Then $A \subset B$ means that every realisation of A induces a realisation of B and, consequently, $\Pr(A) \leq \Pr(B)$ holds for the probabilities of these events. Moreover, $A \cup B$ means that either A or B occur apart or, as the third possibility, A and B occur jointly, $A \cap B$ means that A and B occur jointly, and $B \setminus A$ means that B occurs excluding an occurrence of A. For arbitrary A and B, $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$ is obtained directly from the addition axiom, and further, $\Pr(B \setminus A) = \Pr(B) - \Pr(A)$ for $A \subset B$.

Using the above results, $\Pr\big(N(t) = k\big)$ can be derived from the gamma distributions of the random variables Y_k in (7.8) and (7.9) on condition that $N(t) = k$ in (7.3,2) can be formulated using Y_k. The definition of Y_k implies that $(Y_k \leq t)$ occurs on condition that $(Y_{k+1} \leq t)$, $k = 1, 2, \ldots$, occurs: $(Y_{k+1} \leq t) \subset (Y_k \leq t)$ and from Fig. 7.2 (c) it is obvious that $F_{k+1}(t) \leq F_k(t)$, $F_k(t)$ and $F_{k+1}(t)$ as required in (7.8) and (7.9). Consequently, $(Y_k \leq t) \setminus (Y_{k+1} \leq t)$ occurs on condition that k (but not a larger number of) events are in interval $(0, t]$, i.e., k (but not a larger number of) inter-arrival times are realised in $(0, t]$. This, thus, is the event $N(t) = k$ in (7.3,2), because t in (7.8) and (7.9) is measured starting with the occurrence of event 0, i.e.,

owing to the origin of the time axis being chosen such that it is reconcilable with t in (7.3), as demonstrated in Fig. 7.2.

$\Pr\big(N(t) = k\big)$ can be calculated from the probabilities of $(Y_k \leq t)$ and $(Y_{k+1} \leq t)$ in (7.8) and (7.9): subtracting $F_{k+1}(y)$ from $F_k(y)$, the Poisson density in (7.3,2) is arrived at. Thus, the Poisson process is derived in a continuous-time model from independent and exponentially distributed inter-arrival times of random events. In contrast, time is discretised in the heuristic derivation of Poisson approximation of the binomial distribution in (7.2).

Exponential inter-arrival times are often assumed in *queuing theory* (e.g., [15] or [4]). Are exponential inter-arrival times a suitable model for the intervals between the peaks of aerosol concentrations measured on-road with instruments mounted on a mobile measurement platform?

7.1.3 On-road Measurements of Aerosols

Road traffic pollutes the air by, amongst other things, emitting aerosol particles into the ambient air and lower troposphere. Aerosols are solid or liquid particles with aerodynamic diameter (a unifying variable for the size of particles having different shapes and densities) in the range from 1 νm (1 $\nu\text{m} = 10^{-9}$ m, one nanometer) to 100 μm (1 $\mu\text{m} = 10^{-6}$ m, one micrometer or micron), which are suspended in a gas (here the air). Aerosols in the ambient air are inhaled by humans and animals and are likely to cause adverse health effects, especially those ultrafine particles (with aerodynamic diameter less than 50 νm) are suspected of having a much higher impact (relative to their mass) than larger particles. Hence, in order to describe the quality of the air with respect to adverse health effects, the surface of the aerosols in a unit volume of air should be measured together with their mass concentration. The "active surface" of the aerosols is measured with a diffusion charging (DC) sensor in the unit $(\mu\text{m})^2/\text{cm}^3$.

For example, air quality variables were measured on November 21, 2001, starting at 11:29:01, with a mobile laboratory [25]. The DC measurements obtained on that day are plotted in Fig. 7.3.

The DC time series shows peaks of large values superimposed on the background load of aerosols. The background load is low when the mean distance of the DC sensor from the sources of pollutants is large, for instance when the mobile laboratory is on a car park which is not much frequented (in the time slice from 4300 through to 5300 seconds (s) from the start of the measurements) or when it is on roads with little traffic (from 11000 through to 13000 s). When the background load is low, a peak is recorded once the mobile laboratory encounters a source of pollutants (usually also mobile). A high background load is measured when the mean distance of the DC sensor from the sources of pollutants is small, for instance when the mobile laboratory drives (or crawls along, or stands still) in heavy traffic. When the background load is high, a peak is measured when the plume of pollutants emitted by a single powerful source (or the plumes emitted by two or more

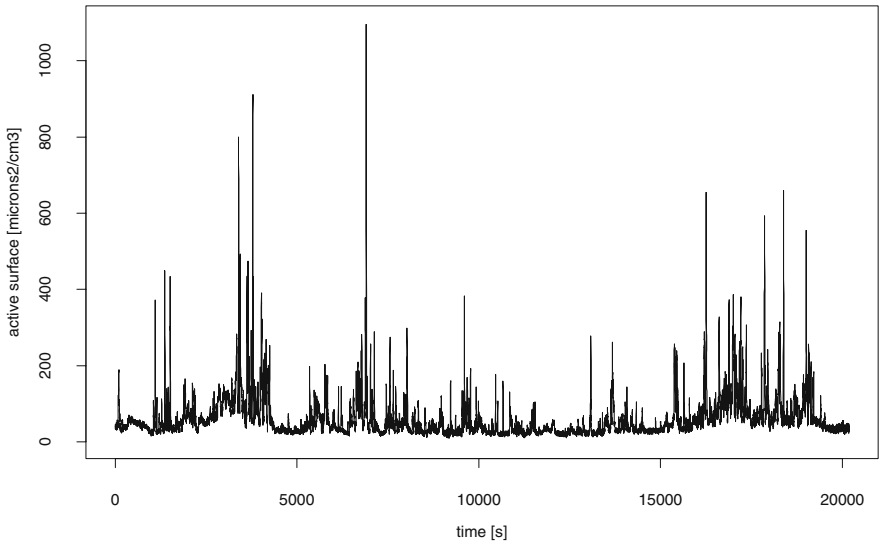


Fig. 7.3. "Active surface" of aerosols (in $(\mu m)^2/cm^3$) measured with a diffusion charging sensor (DC values) mounted on a mobile laboratory on November 21, 2001, starting at 11:29:01. Time in seconds (s)

sources simultaneously) reach(es) the mobile laboratory. This is the most straightforward model which "explains" the peaks in the DC measurements.

The background load is obtained by computing the 0.05 quantiles of the DC measurements with a moving interval of 61 s duration. These quantiles are plotted in Fig. 7.3 for two time slices of 250 s duration each. A too large result (and thus less plausible) is obtained by computing moving averages.

The times when peaks occur are obtained in two steps. In the first step, .90 quantiles of the DC measurements are computed in a moving interval of 31 s duration. These quantiles are plotted in Fig. 7.4 for two time slices of 250 s duration each. A peak occurs whenever the DC measurements exceed their moving .90 quantiles. The duration of the moving interval and the percentage of the quantiles were arrived at subsequent to a few experiments. When using a moving interval of a longer duration, the .90 quantiles become too smooth and many peaks clearly discernible in the DC measurements are not captured, whereas, when using an interval of shorter duration than 31 s, the .90 quantiles are not smooth enough and many small fluctuations are wrongly identified as peaks. Using a moving interval of 31 s duration and .85 quantiles, the peaks are not confined sharply enough, whereas some peaks are lost when .95 quantiles are calculated in this moving interval.

In the second step, the times when the peaks occur are calculated by comparing, for one time point after the other, the DC measurements with

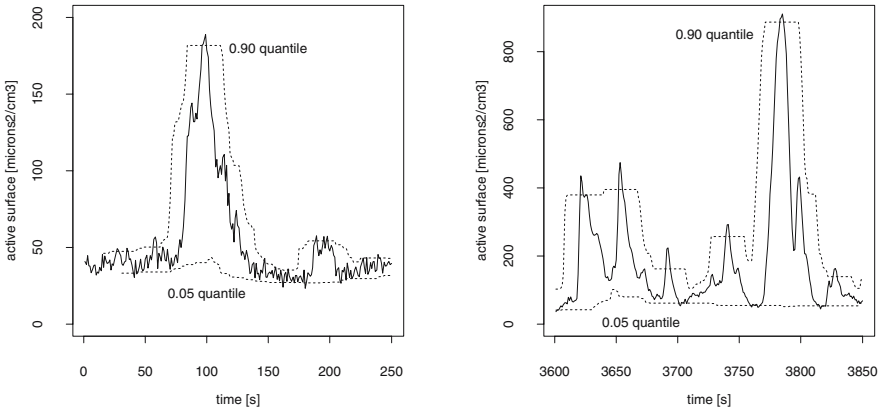


Fig. 7.4. DC values as in Fig. 7.3 (solid line) together with their 0.05 and 0.90 quantiles calculated in moving intervals of 61 and 31 s duration (broken lines) in higher resolution for two time slices.

their moving .90 quantiles. The first comparison is performed for a time point t with the DC-measurement below the moving .90 quantile. The comparisons are repeated for $t+1, t+2, t+\ldots$, until a DC measurement above the quantiles is found for time t_a. Thereafter, the comparisons are repeated until a DC measurement below the quantiles is found for t_b and the initial condition (requiring that the DC-measurement be below the moving .90 quantile) is restored. t_a and t_b thus found confine a peak. This procedure is repeated until all peaks are found. From a pair (t_a, t_b), the time $t_p = t_a + (t_b - 1 - t_a)/2$ when the peak occurs is obtained, assuming that a peak is symmetric in interval $t_a \leq t \leq t_b$ to a vertical line in t_p. For example, the peak occurring at approximately 100 s in Fig. 7.4 (a) is confined by $t_a = 97$ and $t_b = 100$, and thus $t_p = 98$ s. Using this procedure, peaks are obtained at integer seconds and, in-between, at half seconds; with the minimal inter-arrival time being 2 s, as shown in Problem 7.2.

Calculate the .90 quantiles of the DC measurements and thereafter apply the above procedure to obtain the time points when peaks occur, using R expressions in `/path/aerosol.R`: 1037 peaks are found with the times of their occurrence in R vector `peak_times`. These peaks are due to the events described previously, in the remarks to Fig. 7.3, and are assumed to be governed by a Poisson process. As shown in Sect. 7.1.2, events stem from a Poisson process on condition that their inter-arrival times are independent and exponentially distributed. Under this assumption, 1036 inter-arrival times of the peaks (inter-peak times) are obtained with

```
inter_peak_times <- diff(peak_times)    #as in Problem 7.1
```

which thereafter are assumed to be a realisation of a stationary stochastic process. From their empirical correlation function, as plotted in Fig. 7.5 (c),

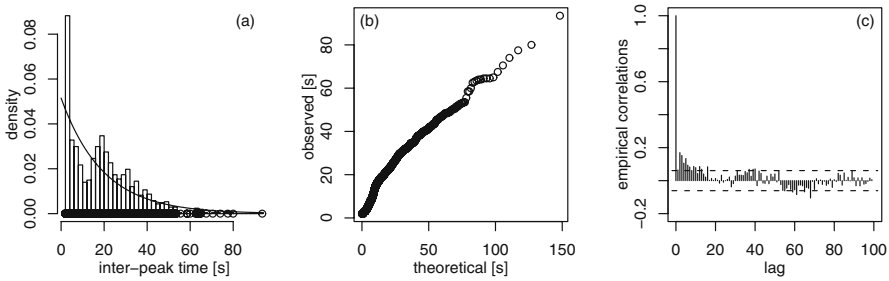


Fig. 7.5. Inter-arrival times of the peaks (inter-peak times) in the DC values as plotted in Fig. 7.3. (a): normalised histogram with fitted exponential density ($\hat{\lambda} = 0.0515\ s^{-1}$, the inverse of the mean inter-peak time), (b): probability plot, and (c): empirical correlation function.

it is concluded that the inter-peak times are weakly correlated for small lags, and therefore are not independent. Thus, the inter-peak times are only approximately reconcilable with (7.3,3).

Do they stem from an exponential distribution? The mean of the inter-peak times is 19.41 s. Hence, $\hat{\lambda} = 1/19.41 = 0.0515\ s^{-1}$ is an estimator for λ in (7.4), on condition that the parameter of the distribution is estimated by the inverse of the mean inter-arrival time ($1/\lambda$ in (7.6) is the expectation of an exponentially distributed random variable). The normalised histogram and the probability plot of the inter-peak times are also shown in Fig. 7.5, plots (a) and (b). The exponentially distributed inter-arrival times X_k used in the derivation of the Poisson distribution in the continuous-time model shown in Fig. 7.2 are real and positive. The sampling interval for the measurement of the DC values as in Fig. 7.3, however, is 1 s and the time points when the peaks occur are calculated such that all inter-peak times are either greater than or equal to 2 s. Consequently, it is assumed that those inter-peak times lasting less than 2 s appear, in Fig. 7.5 (a), as if they had a 2 s duration and thus inflate the first class of the histogram. The histogram and the probability plot in Fig. 7.5, used as diagnostics, indicate that the inter-peak times stem from two exponential distributions and therefore suggest that the mean inter-peak time is larger when the background load is low rather than when it is high (Problem 7.2).

Since the inter-peak times in the DC-measurements are only approximately reconcilable with (7.3,2,3), it is assumed that an event which produces a peak in the DC measurements has only a small influence on the event producing the next peak or any of the peaks occurring further on: traffic jams and other situations which generate large aerosol loads, as described in the remarks to Fig. 7.3, are only approximately memoryless (examples for memoryless random variables are given in the remarks to (7.4)).

7.2 Continuous-time Stochastic Processes

The Poisson process introduced in Sect. 7.1 is a continuous-time stochastic process, i.e., it has a real parameter. Its definition is relatively straightforward since neither derivatives nor integrals of a stochastic process, i.e., no random variables in the limit, are involved. There are, however, continuous-time stochastic processes that are defined using random variables obtained as limiting cases of appropriate defining sequences.

For example, Langevin's equation $m(\mathrm{d}\boldsymbol{v}(t)/\mathrm{d}t)+\beta\boldsymbol{v}(t) = \boldsymbol{F}(t)$ (introduced in the remarks concluding Sect. 7.5.2) is a continuous-time AR[1] process (7.104) describing the random motion of a small particle (with diameter $\approx$ 1 μm) in a fluid known as Brownian motion. The velocity of the particle $\boldsymbol{v}(t)$, its derivative and a randomly fluctuating force $\boldsymbol{F}(t)$ are three-dimensional random functions with real parameter t. Once this differential equation is solved, integrals of random functions are obtained. An integral of a random function is also part of the Fourier representation of a stationary stochastic process.

Derivatives and integrals of random functions are defined as limiting cases of sequences of random variables. Hence, the convergence of a sequence of random variables in vector space $L^2(\Omega, F)$ is introduced in Sect. 7.2.1. Using this convergence, the derivative and the integral of a continuous-time stochastic process are defined in Sect. 7.2.2. Thereafter, these definitions are illustrated with two examples in Sect. 7.2.3: a continuous-time stochastic process is constructed using the discrete-time white noise process as a template, and the orthogonal increment process is introduced. Sect. 7.2.3 also contains the definition of the complex-valued random function.

7.2.1 Random Variables in $L^2(\Omega, F)$

A sequence of random variables is shown to converge in mean square to a limiting random variable by constructing an appropriate Cauchy sequence, on condition that the set of all random variables (together with a scalar product and further restrictions) is a complete linear vector space. Examples of complete vector spaces and Cauchy sequences are given in the remarks to definitions (6.6).

The set of random variables X (or Y) with scalar product $\mathrm{E}(XY)$ *and finite norm* $\mathrm{E}(X^2) < \infty$ *is called* $L^2(\Omega, F)$. *$L^2(\Omega, F)$ is a linear vector space and $L^2(\Omega, F)$ is complete.* (7.10)

Ω in $L^2(\Omega, F)$ is the symbol for the set of elementary events from which events are constructed such that their probabilities are in agreement with the axioms, and F is the symbol for the distribution functions pertaining to random variables used for describing events together with their probabilities. It is easy to prove that $L^2(\Omega, F)$ is a linear vector space because linear combinations and expectations of random variables are often used. In contrast,

it is difficult to derive that $L^2(\Omega, F)$ is complete: every Cauchy sequence (defined in (6.6,1) for functions in linear vector space $L^2(-\infty, \infty)$) that can be constructed from random variables being in $L^2(\Omega, F)$ has to be shown to converge to a limiting random variable also being in $L^2(\Omega, F)$ [21].

Exploiting the favourable properties of $L^2(\Omega, F)$, defined in (7.10), the mean square convergence (i.e., a convergence in the norm as introduced in (6.5) for functions in linear vector space $L^2(-\infty, \infty)$) of a sequence of random variables is defined in (7.11):

$$\begin{aligned} &\textit{A sequence of random variables } X_n \in L^2(\Omega, F) \textit{ converges in} \\ &\textit{mean square on condition that there is a random variable} \\ &X \in L^2(\Omega, F) \textit{ such that } \lim_{n\to\infty} \mathrm{E}\big((X_n - X)^2\big) = 0. \end{aligned} \tag{7.11}$$

For example, mean square convergence is required in (2.19), and therefore (2.19) is proved, on condition that a Cauchy sequence can be constructed from the linear combination of the white noise variables in (2.18) and (2.19), since $L^2(\Omega, F)$ is complete.

Let sequence $Z_n(t)$ be the n-th partial sum in the infinite linear combination (2.19): $Z_n(t) = \sum_{u=0}^{n} a_u W_{t-u}$ with $a_u = a^u$ and $|a| < 1$. Is $Z_n(t)$ a Cauchy sequence? An ϵ arbitrarily small is assumed and N such that $\sigma_W^2 \sum_{u=N}^{\infty} a_u^2 = \sigma_W^2 \sum_{u=N}^{\infty} a^{2u} = \sigma_W^2\big((a^{2N})/(1-a^2)\big) < \epsilon$ is obtained, with σ_W^2 the variance of the W_t. Thereafter, for $q > p > N$, the expectation of the squared difference of $Z_p(t)$ and $Z_q(t)$ in (7.12) is calculated.

$$\mathrm{E}\big(|Z_p(t) - Z_q(t)|^2\big) = \mathrm{E}\left(\left|\sum_{u=0}^{q} a_u W_{t-u} - \sum_{u=0}^{p} a_u W_{t-u}\right|^2\right) \tag{7.12}$$

$$= \mathrm{E}\left(\left|\sum_{u=p+1}^{q} a_u W_{t-u}\right|^2\right) = \mathrm{Var}\left(\left|\sum_{u=p+1}^{q} a_u W_{t-u}\right|\right) \tag{7.13}$$

$$= \sum_{u=p+1}^{q}\sum_{v=p+1}^{q} a_u a_v \mathrm{Cov}\big(W_{t-u}, W_{t-v}\big) = \sigma_W^2\left(\sum_{u=p+1}^{q} a_u^2\right) \tag{7.14}$$

$$= \sigma_W^2 \sum_{u=p+1}^{q} a^{2u} \le \sigma_W^2 \sum_{u=N}^{\infty} a^{2u} = \sigma_W^2 \frac{a^{2N}}{1-a^2} < \epsilon \tag{7.15}$$

The second = in (7.13) is obtained applying (1.15,4), since $\mathrm{E}W_t = 0$ (as required in (2.13)) implies $\mathrm{E}|\sum_{u=p+1}^{q} a_u W_{t-u}| = 0$. (7.14) is arrived at using (1.15,10) since the random variables in a white noise process as defined in (2.10) or (2.11) are not correlated. Consequently, $Z_n(t)$ is a Cauchy sequence and (2.19) holds.

7.2.2 Derivatives and Integrals of Random Functions

The derivative of a continuous-time stochastic process, i.e., a random function with real parameter as introduced in the remarks to definition (2.2), is a

random variable. Hence, it is not yet defined, since derivatives of deterministic functions (solely) are defined in an introduction to Calculus. Integrals of continuous-time random functions are not yet defined for the same reason. Such an integral, for example, is obtained when $x(t)$ and $y(t)$ in LTI filter (6.120), both being deterministic functions, are substituted with stochastic processes having a real parameter.

In (7.16) and (7.17), continuous random functions $X(t)$ with real parameter t and their derivatives are defined using the mean square convergence (7.11), i.e., it is assumed that $X(t)$ and possible limiting random variables are in $L^2(\Omega, F)$, as defined in (7.10).

In (7.16,1), a random function (X_t) with real parameter t has a jump discontinuity in t_0, which is approached from both its right and left sides. If the expectations in the limit as defined in (7.16,1) exist then (i) $X(t)$ is said to be mean-square continuous from the right provided that t is on the right side of t_0, and (ii) mean-square continuous from the left provided that t is on the left side of t_0.

Let $X(t)$ be a stochastic process with real parameter $t \in T$ and t_0 be an arbitrary time point in T. Then:

1. *$X(t)$ is mean-square continuous from the right (from the left) in t_0 where it has a jump discontinuity on condition that* $\lim_{t\to t_0^+}\ (\lim_{t\to t_0^-})\ \mathrm{E}\Big(\big(X(t) - X(t_0)\big)^2\Big) = 0.$ (7.16)
2. *A random function $X(t)$ is mean-square continuous in t_0, on condition that it is mean-square continuous both from the right and from the left:* $\lim_{t\to t_0} \mathrm{E}\Big(\big(X(t) - X(t_0)\big)^2\Big) = 0.$

If a random function $X(t)$ is mean-square continuous as required in (7.16,2) then, in the mean of all its realisations $x(t)$, the square of the differences $\big(x(t)-x(t_0)\big)^2$ is identically zero for $t \to t_0$; however, realisations (x_t) having a jump discontinuity in t_0 are not excluded. Most continuous-time random functions are mean-square continuous. A counter-example is easily constructed: random function $V(t)$ with second moment as in (7.22) is not mean-square continuous.

Assuming, as in (7.16), that t_0 is an arbitrary time point in the parameter domain T, and also taking advantage of the favourable properties of $L^2(\Omega, F)$, mean-square differentiable random functions $X(t)$ are defined in (7.17) by means of a Cauchy sequence.

$X(t)$ is mean-square differentiable in t_0 on condition that $\dfrac{X(t_0+\tau_n) - X(t_0)}{\tau_n}$, *with $\tau_n \to 0$ for $n = 1, 2, \ldots$, is a Cauchy sequence in $L^2(\Omega, F)$. $\dot{X}_{t_0}$ is the random variable in the limit:* $\lim\limits_{\tau\to 0} \dfrac{X(t_0+\tau) - X(t_0)}{\tau} \overset{\text{ms}}{\to} \dot{X}_{t_0}.$ (7.17)

$\dot{X}(t)$ with $\dot{X}(t=t_0)=\dot{X}_{t_0}$ is a random function, i.e., the derivative of $X(t)$, since t_0 is in the parameter domain, i.e., $t_0 \in T$; and, because $\dot{X}_{t_0}$ is in $L^2(\Omega, F)$, (7.18) is arrived at.

$$\lim_{\tau\to 0} \mathrm{E}\left(\frac{\big(X(t_0+\tau)-X(t_0)\big)^2}{\tau^2}\right) = \mathrm{E}\Big(\big(\dot{X}_{t_0}\big)^2\Big) \tag{7.18}$$

Proposition (7.18) and definition (7.16) contain expectations of quadratic functions of the continuous-time random function $X(t)$. Expectations of a quadratic function of $X(t)$ are also used to define, in (2.4), the second moment functions of an $X(t)$ assumed to be stationary. Are the second moment functions of a stationary random function related with its mean-square continuity and differentiability?

If a random function $X(t)$ is stationary and mean-square continuous in t_0, then its correlation function $r_X(\tau)$ is continuous (in the sense of the usual definition) for all τ. This result is arrived at in two steps. In a first step, substituting $\tau = t - t_0$, $\lim_{\tau\to 0} r_X(\tau) = 1$ is obtained, because $0 = \mathrm{E}\big((X(t) - X(t_0))^2\big) = \sigma_X^2(t) + \sigma_X^2(t_0) - 2c_X(t, t_0) = 2\sigma_X^2(1 - r_X(t_0+\tau, t_0)$. Consequently, $r_X(\tau)$ is continuous (in the usual sense) in $\tau_0 = 0$. This result is used to derive, in a second step, that $r_X(\tau)$ is continuous for all τ [113].

If a random function $X(t)$ is stationary and mean-square differentiable in t_0, then (i) the derivative of its correlation function is identically zero for lag $\tau = 0$, $\dot{r}_X(0) = 0$, and (ii) its second derivative $\ddot{r}_X(0)$ exists for $\tau = 0$. If $X(t)$ is differentiable in t_0, then $\lim_{\tau\to 0}(1/\tau^2)\mathrm{E}(X(t_0+\tau) - X(t_0))^2 = \lim_{\tau\to 0}(1/\tau^2)\big(\sigma_X^2(t_0+\tau) + \sigma_X^2(t_0) - 2c_X(\tau)\big) = \lim_{\tau\to 0}(2\sigma_X^2/\tau^2)(1 - r_X(\tau)) = 2\sigma_X^2 \lim_{\tau\to 0}(1/\tau^2)(1 - r_X(\tau))$ converges. Since $\sigma_X^2 < \infty$, also $\lim_{\tau\to 0}(1/\tau^2)(1 - r_X(\tau))$ converges. $\lim_{\tau\to 0}(1/\tau^2)(1 - r_X(\tau))$ diverges, except for $\dot{r}_X(0) = 0$, because $\dot{r}_X(0) = \lim_{\tau\to 0}(1/\tau)(1 - r_X(\tau))$, and the first result is arrived at. From this result, the second one is derived using a Taylor series $r_X(\tau) = r_X(0) + \tau\dot{r}_X(0) + (\tau^2/2)\ddot{r}_X(0) + \ldots$ for $r_X(\tau)$: the first term becomes $r_X(0) = 1$, the second one $\dot{r}_X(0) = 0$, the third one contains the second derivative $(\tau^2/2)\ddot{r}_X(0)$ of the correlation function and, neglecting the further terms, $r_X(\tau) \to 1 + (\tau^2/2)\ddot{r}_X(0)$ and $(r_X(\tau) - 1)/\tau^2 \to \ddot{r}_X(0)/2$ are arrived at. These results and the convergence of $\lim_{\tau\to 0}(1/\tau^2)(1 - r_X(\tau))$ obtained above imply the convergence of $\ddot{r}_X(0)$, i.e., that the second derivative of the correlation function exists for lag $\tau = 0$. Further, if $\ddot{r}_X(0)$ exists, then $\ddot{r}_X(\tau)$ exists for all τ [113].

For example, the correlation function of the stationary AR[1] process with real parameter as obtained in (7.114) is continuous (in the usual sense) in $\tau = 0$; for this lag, however, its derivative from the left is not equal to the one from the right. Consequently, the AR[1] process with real parameter as defined in (7.104) is mean-square continuous but not mean-square differentiable.

Using mean square convergence, as above in the definitions of continuous and differentiable random functions, the stochastic Riemann and Stieltjes

integrals are defined in (7.20) and (7.21) subsequent to a refresher course on the Riemann and Stieltjes integrals for deterministic functions.

The Riemann integral $\int_a^b f(x)\mathrm{d}x$, $f(x)$ being a real-valued deterministic function with real argument x, is the integral usually encountered in Physics and Engineering and is defined, in an introduction to Calculus, by taking the limit of a Riemann sum. A Riemann sum of a function $f(x)$ in interval $a \leq x \leq b$ is a sum of the areas of rectangles of widths $x_i - x_{i-1}$ with $x_{i-1} \leq x_i$ and heights $f(x_i^*)$ with $x_i \leq x_i^* \leq x_{i-1}$, where the x_i are arbitrary points partitioning an interval $a \leq x \leq b$ into subintervals. If the Riemann sum has a finite limit for arbitrarily fine partitions of $a \leq x \leq b$, i.e., for the maximal $x_i - x_{i-1}$ becoming arbitrarily small, then $f(x)$ is said to be integrable in $a \leq x \leq b$ and the limit is called the Riemann integral of $f(x)$ in $a \leq x \leq b$.

The Riemann-Stieltjes integral is defined in (7.19) and its properties are given in, e.g., [126]. Let $g(t)$ and $F(t)$ be real-valued deterministic functions defined in interval $a \leq t \leq b$, and let $g(t)$ be continuous in $a \leq t \leq b$ whereas $F(t)$ may have jump discontinuities in $a \leq t \leq b$. The end points of the interval become $t_0 = a$ and $t_n = b$, and points $t_1, \ldots, t_{n-1}$ with $t_{i-1} \leq t_i$ are assumed to be in $a \leq t \leq b$ such that the interval is partitioned into n subintervals. Thereafter, the Riemann-Stieltjes sum in (7.19) is built. If the Riemann-Stieltjes sum converges on condition that n increases such that maximal width of the subinterval decreases, then the limit is called *Riemann-Stieltjes integral* or *Stieltjes integral* of $g(t)$ with respect to $F(t)$. $g(t)$ is called the integrand and $F(t)$ the integrator.

$$\lim_{\substack{\max(t_i - t_{i-1}) \to 0 \\ n \to \infty}} \sum_{i=1}^{n} g(t_i)\big(F(t_i) - F(t_{i-1})\big) \;=\; \int_a^b g(t)\mathrm{d}F(t) \tag{7.19}$$

A Stieltjes integral is often used to define the expectation $\mathrm{E}X$ of a random variable X having probability distribution $F(x)$. $\mathrm{E}X$ is defined by substituting, in (7.19), x for t, $F(x)$ for the integrator $F(t)$ and x for $g(t)$ and thereafter integrating in interval $-\infty \leq x \leq \infty$. Thus, the Stieltjes integral $\mathrm{E}X = \int_{-\infty}^{\infty} x\mathrm{d}F(x)$ converges. A proof exploits the properties of $F(x)$ (bounded, monotonically non-decreasing, from the right continuous in possible jump discontinuities) and $g(x) = x$ (continuous).

If a random variable Y has a continuous probability distribution $F(y)$, then $\mathrm{E}Y = \int_{-\infty}^{\infty} y\mathrm{d}F(y) = \int_{-\infty}^{\infty} y(\mathrm{d}F(y)/\mathrm{d}y)\mathrm{d}y = \int_{-\infty}^{\infty} yf(y)\mathrm{d}y$ and the Stieltjes integral becomes the usual Riemann integral. For example, the expectation of a normally-distributed Y, $\mathrm{E}Y = \int_{-\infty}^{\infty} y(2\pi)^{-(1/2)}\mathrm{e}^{-(y^2/2)}\mathrm{d}y = 0$, becomes identically zero since the normal density is an even function as defined in (6.32,1).

If a random variable X has a discrete distribution $F(x) = \sum_{x_i \leq x} f(x_i)$, then $\mathrm{E}X = \int_{-\infty}^{\infty} x\mathrm{d}F(x) = \sum_i xf(x_i)$ and the Stieltjes integral becomes a sum. For example, a Bernoulli distributed random variable X has distribution $\Pr(X \leq x) = F(x)$, where $F(x) = 0$ for $x < 0$, $F(x) = q$ for $0 \leq x < 1$ and $F(x) = q + p = 1$ for $1 \leq x$, i.e., $F(s)$ has jump discontinuities in $x = 0$

and in $x = 1$ with heights q and p. $F(x)$ is therefore a non-decreasing step function (a step function $h(x)$ with real argument x is a linear combination $h(x) = \sum_{i=1}^{n} c_i h_i(x)$ with $h_i(x) = 1$ for $a_i \leq x < b_i$ and $h_i(x) = 0$ at the exterior of this interval, where a_i, b_i and c_i are real constants for $i = 1, 2, \ldots, n$). The density $f(x)$ is drawn as spikes of heights q and p in the jump discontinuities. The Stieltjes integral for the expectation of X therefore becomes $\mathrm{E}X = \int_{-\infty}^{\infty} x \mathrm{d}F(x) = 0 \times q + 1 \times p = p$.

The Stieltjes integral allows for defining the expectation of random variables with continuous, discrete or mixed (a mixed probability distribution is not constant in the intervals confined by its jump discontinuities, i.e., it is not a step function) probability distributions without discriminating between the cases.

Riemann and Stieltjes integrals for deterministic functions are defined above; for random functions, these integrals are defined in (7.20) and (7.21). Let, within interval $a \leq t \leq b$, a deterministic function $g(t)$ be continuous and a random function $X(t)$ be mean-square continuous (7.16). Partition $a \leq t \leq b$ into n subintervals by points $t_1, \ldots, t_{n-1}$, $t_{i-1} \leq t_i$, with $t_0 = a$ and $t_n = b$, thereafter build the sum in (7.20) and let n increase such that the maximal width of the subintervals decreases. If, under these assumptions, a random variable $G \in L^2(\Omega, F)$ exists in the limit such that (7.20) holds, then G is called the *Riemann integral* of random function $X(t)$. For this convergence, usually $\int_a^b g(t)X(t)\mathrm{d}t \overset{\text{ms}}{\to} G$ or $G = \int_a^b g(t)X(t)\mathrm{d}t$ is written.

$$\lim_{\substack{\max(t_i - t_{i-1}) \to 0 \\ n \to \infty}} \mathrm{E}\left(\sum_{i=1}^{n} g(t_i)X(t_i)(t_i - t_{i-1}) - G \right)^2 = 0 \tag{7.20}$$

If a random function $X(t)$ is mean-square continuous from the right in jump discontinuities found in interval $a \leq t \leq b$ then the question is whether a random variable $G \in L^2(\Omega, F)$ exists such that (7.21) holds. If this random variable exists in the limiting case then it is called *Stieltjes integral* of $g(t)$ with respect to $X(t)$: $\int_a^b g(t)\mathrm{d}X(t) \overset{\text{ms}}{\to} G$ or $G = \int_a^b g(t)\mathrm{d}X(t)$.

$$\lim_{\substack{\max(t_i - t_{i-1}) \to 0 \\ n \to \infty}} \mathrm{E}\left(\sum_{i=1}^{n} g(t_i)\big(X(t_i) - X(t_{i-1})\big) - G \right)^2 = 0 \tag{7.21}$$

Using a Stieltjes integral (7.21), the orthogonal increment process is defined in (7.102) and the Fourier representation of a stationary stochastic process is defined in (7.51).

7.2.3 Examples

Does a white noise process with real parameter exist? The discrete-time white noise process (W_t) with integer parameter t is defined in (2.10) or (2.11). The properties of (W_t) are transferred, in the following experiment, to a

continuous-time stochastic process $V(t)$ with real parameter t, provided that its first moment function is required to be constant, and its second moment function is required to be identically zero for non-zero lags: $\mathrm{E}V(t) = \mu_V$ and $\mathrm{Cov}\big(V(t), V(u)\big) = \sigma_V^2$ for $t = u$ and $\mathrm{Cov}\big(V(t), V(u)\big) = 0$, for $t \neq u$, t and u being time points in the parameter domain of $V(t)$. If this condition is written applying the Kronecker δ, as defined in the remarks to (6.7), together with the constant variance function $\sigma_V^2(t) = \sigma_V^2$ of $V(t)$, then the covariance function in (7.22) is obtained:

$$\mathrm{Cov}\big(V(t), V(u)\big) = \sigma_V^2 \delta_{tu} \qquad \text{or} \qquad c_V(\tau) = \begin{cases} \sigma_V^2 & \text{for } \tau = 0 \\ 0 & \text{for } \tau \neq 0 \end{cases} \tag{7.22}$$

$V(t)$ is required above to be stationary: its expectation function $\mathrm{E}V(t) = \mu_V$ is constant and its covariance function $c_V(\tau)$ as defined in (7.22) depends only on the lag. The jump discontinuity in $c_V(\tau)$ implies that $V(t)$ is not mean-square continuous (7.16) since a mean-square continuous stochastic process has a covariance function which is continuous for $\tau = 0$, as derived in the remarks to (7.16) and (7.18). A realisation $v(t)$ of $V(t)$ fluctuates with extreme amplitudes because neighbour variables pertaining to $V(t)$, i.e., random variables associated with real time points being close neighbours on the time axis, are required to be non-correlated in (7.22).

In contrast to $V(t)$ with covariance function as in (7.22), the orthogonal increment process is defined in (7.23): now, the increments of the stochastic process are required to be non-correlated.

$A(t)$ is called a random function with orthogonal increments (orthogonal increment process) on condition that the first and second moments of differences of $A(t)$ have the properties:

1. $\mathrm{E}\big(A(t) - A(s)\big) = \mu_{\mathrm{d}A}(t - s)$, *with* $s < t$,
2. $\mathrm{Cov}\big(A(t) - A(s), A(v) - A(u)\big) = \sigma_{\mathrm{d}A}^2 \big|(s,t] \cap (u,v]\big|$, *with* $s < t$, $u < v$ *and* $\big|(s,t] \cap (u,v]\big|$ *the distance between the endpoints of the intersection of the half-open intervals* $(s,t]$ *and* $(u,v]$.

(7.23)

Examples of intervals $(s,t]$ and $(u,v]$, for which the differences of $A(t)$ are built in (7.23), are shown in Fig. 7.6: the first interval $(s,t]$ is plotted on the horizontal axis and the second $(u,v]$ on the vertical one. In Fig. 7.6 (b), the intersection of both intervals is empty and, therefore, the distance between the endpoints of the intersection becomes identically zero, whereas, in Fig. 7.6 (c), this distance is $t-u$. In Fig. 7.6 (a), $s = 0$ and $u = 0$ implying that $\min(t, v)$ is the distance between the endpoints of the intersection

For example, the Poisson process (7.3) is a monotonically non-decreasing orthogonal increment process as defined in (7.23) because its increments are independent and positive. In general, however, an orthogonal increment process does not monotonically increase or decrease. An orthogonal increment process $A(t)$ with $A(t)$ being normally distributed for all t is called a *Wiener process.*

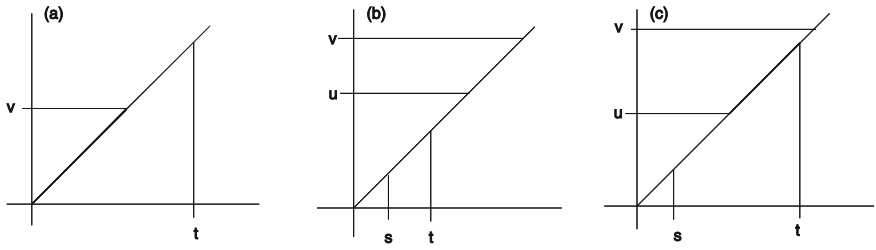


Fig. 7.6. Orthogonal increment process. Three cases of semi-open intervals $(s,t]$ and $(u,v]$ in definition (7.23). The projection of the emphasized distance on one of the axes is the intersection of the intervals. In plot (a), $s=0$ and $u=0$.

The moment functions of the orthogonal increment process are easily obtained with $s=0$ and $v=0$ in definition (7.23):

$$\begin{aligned} \mathrm{E}\big(A(t)\big) &= \mu_{\mathrm{d}A} \times t, \\ \mathrm{Var}\big(A(t)\big) &= \sigma^2_{\mathrm{d}A} \times t \quad \text{and} \\ \mathrm{Cov}\big(A(t), A(v)\big) &= \sigma^2_{\mathrm{d}A} \times \big(\min(t,v)\big). \end{aligned} \tag{7.24}$$

The moment functions in (7.24) are compared with those in (2.67) and (2.69). From this comparison it becomes obvious that the moment functions of the continuous-time orthogonal increment process $A(t)$ are identical with those of the discrete-time random walk process (X_t): $\mu_{\mathrm{d}A}$ and $\sigma^2_{\mathrm{d}A}$ in (7.24) play the roles of μ_W and σ^2_W in (2.67) and (2.69). As a first consequence, both expectation functions $\mathrm{E}A(t) = \mu_{\mathrm{d}A} \times t$ and $\mathrm{E}X_t = \mu_W \times t$ increase or decrease linearly with time t: both processes are not stationary in the first moment function, except for $\mu_{\mathrm{d}A} = 0$ and $\mu_W = 0$. As a second consequence, both variance functions $\mathrm{Var}A(t) = \sigma^2_{\mathrm{d}A} \times t$ and $\mathrm{Var}X_t = \sigma^2_W \times t$ increase linearly with increasing t, and both covariance functions are proportional, with factors $\sigma^2_{\mathrm{d}A}$ and σ^2_W, to the distance of the intersection of $(0,t]$ and $(0,v]$ in Fig. 7.6 (a). This behaviour of the second moment functions implies that both processes are not stationary.

Difference quotients $Y(t)$ of the orthogonal increment process are defined in (7.25), and their covariance function $c_Y(\tau) = \mathrm{Cov}(Y(t), Y(t+\tau))$ is obtained in (7.26) directly from definition (7.23).

$$Y(t) = (1/\Delta t)\big(A(t+\Delta t) - A(t)\big) \tag{7.25}$$

$$\mathrm{Cov}\big(Y(t), Y(t+\tau)\big) = \begin{cases} (\sigma_{\mathrm{d}A}/\Delta t)^2(\Delta t - \tau) & \text{for } |\tau| \le \Delta t \\ 0 & \text{for } |\tau| > \Delta t \end{cases} \tag{7.26}$$

$\mathrm{Cov}\big(Y(t), Y(t+\tau)\big)$ in (7.26) becomes a sequence of triangle functions, as in (6.63), with decreasing widths on condition that Δt decreases stepwise. Since a sequence of triangle functions is a defining sequence for the delta function

(6.60), $\mathrm{Cov}\big(Y(t), Y(t+\tau)\big) = \sigma_{\mathrm{d}A}^2 \delta(\tau)$ is arrived at in the limiting case for $\Delta t \to 0$ and, consequently, $\mathrm{Var}\big(Y(t)\big) \to \infty$ for $\Delta t \to 0$. In the limiting case for $\Delta t \to 0$ therefore, the difference quotients $Y(t)$ as defined in (7.25) have an infinite variance and thus are not in $L^2(\Omega, F)$. In contrast, the covariance $\mathrm{Cov}\big(V(t), V(u)\big) = \sigma_V^2 \delta_{tu}$ in (7.22), with δ_{tu} being the Kronecker delta (to be clearly distinguished from the Dirac delta in (6.60)), becomes the finite variance σ_V^2 of the process $V(t)$ for $t = u$.

As a third example, the complex-valued random function is defined in (7.27).

$$
\begin{aligned}
&Z(t) = U(t) + \mathrm{i}V(t) \textit{ is called a complex-valued random function.}\\
&\mathrm{E}Z(t) = \mathrm{E}U(t) + \mathrm{i} \times \mathrm{E}V(t) = \mu_Z(t) \textit{ is its expectation function,}\\
&\mathrm{Var}Z(t) = (|Z(t) - \mathrm{E}\mu_Z|^2) = \sigma_Z^2(t) > 0 \textit{ its variance function, and}\\
&\mathrm{Cov}\big(Z(t), Z(u)\big) = \mathrm{E}\big((\overline{Z(t) - \mu_Z(t)})(Z(u) - \mu_Z(u))\big) = c_Z(t,u)\\
&\textit{its covariance function.}
\end{aligned}
\tag{7.27}
$$

In this chapter, $Z(t)$ denotes, as is usual in time series literature, a complex-valued random function. In Chap. 4 and in definition (4.6), however, $Z(x)$ denotes a spatial random function, as is usual in geostatistical literature. As a consequence, the meaning of $Z(t)$ or $Z(x)$ depends on the context.

A real-valued random function is stationary on condition that it has the properties required in (2.8), whereas a complex-valued one is stationary on condition that it satisfies (7.28).

$$
\begin{aligned}
&\textit{Let } Z(t) = U(t) + \mathrm{i}V(t) \textit{ be a complex-valued random function.}\\
&\textit{Then } Z(t) \textit{ is stationary on condition that:}\\
&1.\ U(t) \textit{ and } V(t) \textit{ are stationary, (consequently } \mathrm{E}Z(t) = \mu_Z,\\
&\quad \textit{a complex constant), and}\\
&2.\ \mathrm{Cov}\big(Z(t), Z(t+\tau)\big) = c_Z(\tau) \textit{ and } c_Z(-\tau) = \overline{c_Z(\tau)}.
\end{aligned}
\tag{7.28}
$$

Consequently, the covariance matrix $\boldsymbol{C}_Z$ of a complex-valued random function is a complex-valued Toeplitz matrix (in the remarks to (5.15) it is mentioned that the Levinson-Durbin recursion (5.14) is a procedure to solve an equation system with a Toeplitz matrix); $\boldsymbol{C}_Z$ is, however, not symmetric since $c_Z(-\tau) = \overline{c_Z(\tau)}$.

The covariance matrix of a real-valued stochastic process $X(t)$ is positive semidefinite (2.7). This property (i) implies that the variance of linear predictions as defined in Chaps. 4 and 5 is non-negative and (ii) is also a precondition for the spectral representation of a stationary stochastic process introduced in Sect. 7.3. As in the real-valued case, the covariance matrix $\boldsymbol{C}_Z$ of a complex-valued stochastic process $Z(t)$ is also positive semidefinite because, in (7.29), linear combinations are not negative for arbitrary complex coefficients d_j:

$$
\mathrm{Var}\left(\sum_{j=1}^{n} d_j Z(t_j)\right) = \sum_{j=1}^{n}\sum_{k=1}^{n} d_j \overline{d_k} \boldsymbol{C}_Z(t_j, t_k) \geq 0. \tag{7.29}
$$

Applying the definitions introduced in Sects. 7.2.1 and 7.2.2, you are now in a position to (i) solve Langevin's equation $m(\mathrm{d}\boldsymbol{v}(t)/\mathrm{d}t) + \beta\boldsymbol{v}(t) = \boldsymbol{F}(t)$ introduced in the remarks concluding Sect. 7.5.2 and (ii) formulate theorem (7.51) which asserts that almost all stationary stochastic processes have a Fourier representation.

7.3 Fourier Representation of a Stationary Stochastic Process

Let $x(t)$, $-\infty < t < \infty$, be a realisation of a discrete- or continuous-time stationary random function $X(t)$. Can then $x(t)$ be represented as a Fourier series or Fourier integral? Obviously, this is not possible for the reason that, in general, $x(t)$ is neither periodic (therefore, Fourier representations cannot be obtained in intervals $L^2(-T/2, T/2)$ (or in $L^2(0, N-1)$ for integer t)), nor in $L^2(-\infty, \infty)$. $x(t)$ is not in $L^2(-\infty, \infty)$ because $\int_{-T/2}^{T/2}\big(x(t)\big)^2\mathrm{d}t \overset{T\to\infty}{\longrightarrow} +\infty$, i.e., because its norm is not finite in $-\infty < t < \infty$, since $X(t)$ is a stationary and ergodic random function. Consequently, Fourier representations of realisations $x(t)$ do not exist and thus it is not possible to obtain a Fourier representation of $X(t)$ by way of calculating the mean of the Fourier representations of $x(t)$.

Above, $\int_{-T/2}^{T/2}\big(x(t)\big)^2\mathrm{d}t \overset{T\to\infty}{\longrightarrow} +\infty$ is implied by $(1/T)\int_{-T/2}^{T/2}\big(X(t)\big)^2\mathrm{d}t \overset{T\to\infty}{\longrightarrow} \mathrm{E}\big((X(t))^2\big)$ with probability one, on condition that $X(t)$ is stationary and ergodic. A sequence of random variables X_n converges with probability one to a random variable G, on condition that $\Pr(X_n \text{ converges to G}) = 1$. The assertion that an event has probability one is the strongest statement that can be made in probability theory and, therefore, convergence with probability one implies mean-square convergence.

Despite the above rationale, a stationary stochastic process does have a Fourier representation. The Fourier representation of a stationary stochastic process is comprehensively (including a variety of proofs) dealt with in [113]. In this section, a Fourier representation is shown to exist for the special case of the harmonic process, and, thereafter, it is formulated in (7.51) as an existence theorem in the general case.

7.3.1 The Harmonic Process and its Fourier Representation

The finite Fourier series $f_m(t)$ as defined in (6.21,3) converges in $L^2(0, N-1)$ to the deterministic function $f(t)$, i.e., $f_m(t) = (f_t) = \sum_{k=0}^{m}\big(A_k\cos(2\pi s_k t) + B_k\sin(2\pi s_k t)\big)$. As a consequence, (f_t) can be represented as a linear combination of the trigonometric functions (6.9) with Fourier coefficients A_k and B_k (6.21,2). These trigonometric functions are orthogonal in $L^2(0, N-1)$ as proposed in (6.11), (6.12) and (6.13). This linear combination is written using

complex notation in (6.22): $(f_t) = \sum_{k=0}^{N-1} F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k}$ and, since it is periodic, $\sum_{k=-(N-1)/2}^{(N-1)/2} F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k}$ for odd N or $\sum_{k=-(N/2-1)}^{N/2} F_k \mathrm{e}^{\mathrm{i}2\pi(t/N)k}$ for even N in (6.24), with complex-valued F_k, is obtained from the real-valued A_k and B_k by applying (6.25). Such a linear combination is also easily defined for real t, where $f(t) = \sum_{k=0}^{m}\big(A_k \cos(2\pi s_k t) + B_k \sin(2\pi s_k t)\big)$ is obtained. For example, the Fourier representation in Fig. 6.1 converges to the saw-tooth shaped sequence for $t = 0, 1, \ldots, 140$, whereas, for real $0 \le t \le 139$, it interpolates the sequence.

Do linear combinations of trigonometric functions with random coefficients exist? If so, then what properties do they have?

Let $0 < s_j < 1/2$ be ordered ($s_j < s_{j+1}$) frequencies of trigonometric oscillations with absolute values R_j and phases P_j. Further, let R_j be real constants and P_j independent random variables having a uniform (or rectangular) distribution in $[-\pi, \pi]$, and n be a finite integer number. Then

$$X_t = \sum_{j=1}^{n}\Big(R_j \cos(2\pi s_j t + P_j)\Big) = \sum_{j=-n}^{n} D_j \mathrm{e}^{\mathrm{i}2\pi s_j t}, \tag{7.30}$$

with $t = \ldots, -1, 0, 1, \ldots$, is called a harmonic process. Assuming $s_0 = 0$, $s_{-j} = -s_j$, $R_0 = 0$, and $R_{-j} = R_j$ and using the identities in the remarks to (6.29),

$$D_j = \begin{cases} (R_j/2)\mathrm{e}^{\mathrm{i}P_j} & \text{for } j = 1, \ldots, n, \\ 0 & \text{for } j = 0, \\ (R_j/2)\mathrm{e}^{-\mathrm{i}P_j} & \text{for } j = -n, \ldots, -1 \end{cases}$$

is obtained.

A random variable has a continuous uniform (or rectangular) distribution in $a \le y \le b$ on condition that its density $f(y) = 1/(b-a)$ for $a \le y \le b$ and $f(y) = 0$ at the exterior of this interval. Since $s_j < 1/2$ is required and $\Delta t = 1$, no aliasing occurs, as is concluded from (6.92). A realisation of a harmonic process is plotted, for example, in Fig. 7.7.

In its complex version, the harmonic process (7.30) is similar to the discrete Fourier transform (6.22); a closer look, however, allows for drawing the following distinctions:

1. The trigonometric functions in (7.30) are not orthogonal and thus do not have the properties proposed in (6.11), (6.12) and (6.13), because the frequencies s_j in (7.30) are not the Fourier frequencies s_k as defined in (6.21,1). However, the s_j are ordered.
2. The coefficients in (7.30), i.e., $D_j = (R_j/2)\mathrm{e}^{\mathrm{i}P_j}$ and $D_{-j} = (R_j/2)\mathrm{e}^{-\mathrm{i}P_j} = \overline{D_j}$ for $j = 1, \ldots, n$, as well as $D_0 = 0$, are a complex-valued random function as defined in (7.27) having parameter $j = -n, \ldots, -1, 0, 1, \ldots, n$.

The moment functions of the harmonic process as proposed in (7.31) are derived in (7.116) and (7.119).

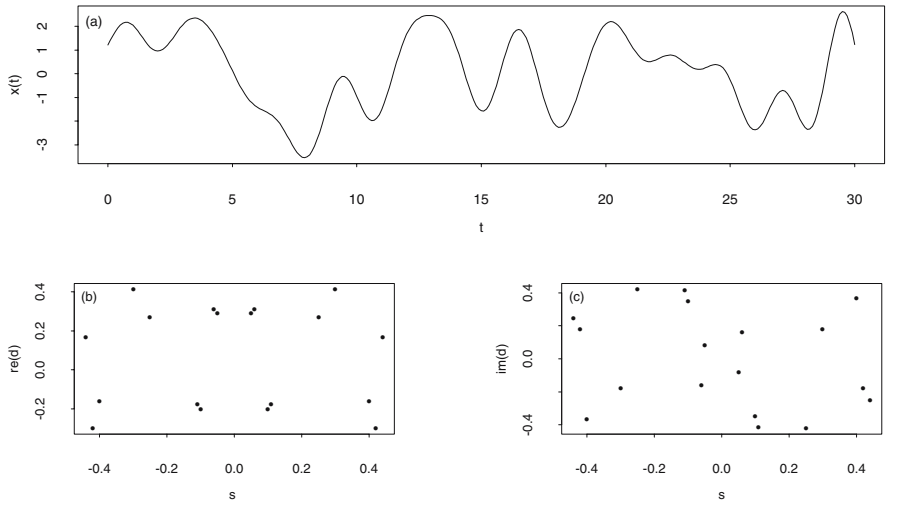


Fig. 7.7. Above (a), a realisation of a harmonic process (7.30) with $n = 9$, below, (b) and (c), the real $\mathrm{re}(d_j)$ and imaginary parts $\mathrm{im}(d_j)$ of the realised d_j.

$$\mathrm{E}X_t = 0 \qquad c_X(t,u) = c_X(\tau) = \sum_{j=1}^{n} \left(R_j^2/2\right) \cos(2\pi s_j \tau) \tag{7.31}$$

The harmonic process is stationary because its expectation function is constant and its covariance function depends solely on the lag $\tau = t - u$. Understandably, this result is obtained because each of the oscillations in (7.30) is stationary.

The complex-valued D_j in (7.30) are random variables for $j \neq 0$. Their expectations and variances are derived in (7.120) and (7.121) with the results in (7.32). Furthermore, pairs D_j and D_k, $j,k = -n,\ldots,-1,1,\ldots,n$ and $j \neq |k|$ are independent, as concluded directly from definition (7.30), implying that the covariances in (7.33) are identically zero.

$$\mathrm{E}D_j = 0 \text{ and } \mathrm{Var}D_j = \frac{R_j^2}{4}, \qquad j = -n,\ldots,-1,1,\ldots,n \tag{7.32}$$

$$\mathrm{Cov}\left(D_j, D_k\right) = 0, \qquad j,k = -n,\ldots,-1,1,\ldots,n,\ j \neq |k| \tag{7.33}$$

$$\mathrm{Cov}\left(D_j, D_{-j}\right) = 0, \qquad j = -n,\ldots,-1,1,\ldots,n \tag{7.34}$$

Although D_j and D_{-j} are not independent, their covariance is identically zero as derived in (7.125) and summarized in (7.34).

Using the expectations of D_j being identically zero and the covariances of pairs (D_j, D_k) being identically zero as well, (1.15,10) is applied to obtain (easily) the variance of the harmonic process, because only the terms for $j = k$ remain in (7.35), all other terms becoming identically zero.

$$\begin{aligned}\mathrm{Var}X_t &= \mathrm{Var}\left(\sum_{j=-n}^{n} D_j \mathrm{e}^{\mathrm{i}2\pi s_j t}\right)\\ &= \sum_{j=-n}^{n}\sum_{k=-n}^{n} \mathrm{e}^{\mathrm{i}2\pi s_j t}\mathrm{e}^{-\mathrm{i}2\pi s_k t}\mathrm{Cov}\big(D_j, D_k\big) \qquad (7.35)\\ &= \sum_{j=-n}^{n} \mathrm{E}\big(|D_j|^2\big) = \sum_{j=-n}^{n} \frac{R_j^2}{4}, \quad R_{-j} = R_j \text{ and } R_0 = 0 \qquad (7.36)\end{aligned}$$

In (7.36), the variance of the harmonic process is the sum of the variance of the D_j since $\mathrm{Var}D_j = \mathrm{E}(|D_j|^2) = R_j^2/4$. The $R_{-n}^2/4$, …, $R_{-1}^2/4$, $R_1^2/4$, …, $R_n^2/4$, $R_{-j} = R_j$, are associated with frequencies $-s_n$, …, $-s_1$, s_1, …, s_n in definition (7.30). Consequently, definition (7.37) distributes the variance of the harmonic processes over the frequencies $-1/2 \le s \le 1/2$, and the spectrum $C_X(s)$ of the harmonic process $X(t)$ is arrived at.

Let (X_t) be a harmonic process as defined in (7.30) and $-1/2 \le s \le 1/2$ be real frequencies. Then:

1. $C_X(s) = \sum_{j=-n}^{n}(R_j^2/4)\delta(s - s_j)$ *is called the spectrum of* (X_t), *and*
2. $C_X^{(I)}(s) = \sum_{j,s_j \le s}(R_j^2/4)$ *is called the integrated spectrum of* (X_t).

(7.37)

Since $C_X(s)$ in (7.37) is even (6.32,1), (6.37) and (6.66) are applied to obtain its Fourier cosine transform: $\mathcal{F}_{+\mathrm{i}}\big(C_X(s)\big) = 2\int_0^{1/2}\sum_{j=1}^{n}(R_j^2/4)\delta(s-s_j)\times \cos(2\pi st)\mathrm{d}s = 2\sum_{j=1}^{n}(R_j^2/4)\cos(2\pi s_j t)$. Substituting t with τ, $\mathcal{F}_{+\mathrm{i}}\big(C_X(s)\big) = \sum_{j=1}^{n}(R_j^2/2)\cos(2\pi s_j\tau) = c_X(\tau)$, i.e., the covariance function (7.31) of the harmonic process is arrived at. $c_X(\tau)$ is, however, not in $L^2(-\infty,\infty)$, and therefore, both functions $c_X(\tau)$, τ integer, as well as $C_X(s)$, s real, are not a Fourier transform pair as proposed in (6.49) and (6.50). Consequently, $c_X(\tau)$ and $C_X(s)$ are neither reconcilable with (6.20,4) (although t is discrete and s continuous) nor with any of the other cases in (6.20). Some properties of this Fourier transform pair will be given in the remarks to (7.70).

Prior to continuing the analysis of the harmonic process, an alternative form is proposed in (7.38), which is a stochastic version of the finite Fourier series (6.21,3). A harmonic process (X_t) is a linear combination of the trigonometric functions in (7.38)

$$X_t = \sum_{j=1}^{n}\big(A_j\cos(2\pi s_j t) + B_j\sin(2\pi s_j t)\big),\ t = \ldots,-1,0,1,\ldots \qquad (7.38)$$

$$\mathrm{Var}A_j = \mathrm{Var}B_j = R_j^2/2, \quad j = 1,\ldots,n \qquad (7.39)$$

$$0 = \mathrm{E}A_j = \mathrm{E}B_j = \mathrm{Cov}(A_j, B_j), \qquad j = 1,\ldots,n \qquad (7.40)$$

$$0 = \mathrm{Cov}(A_j, B_k) = \mathrm{Cov}(B_j, B_k) = \mathrm{Cov}(A_j, A_k), \quad \begin{matrix} j,k = 1,\ldots,n \\ j \ne k \end{matrix} \qquad (7.41)$$

where the coefficients A_j and B_j are random variables. On the one hand, definition (7.30) implies (7.38) since $A_j = R_j \cos(P_j)$ and $B_j = -R_j \sin(P_j)$ are obtained, R_j and D_j as in (7.30), by applying identity $\cos(x + y) = \cos(x)\cos(y) - sin(x)\sin(y)$. If the harmonic process is defined as in (7.30) then, because the D_j are independent, the moments of the A_j and B_j as proposed in (7.39), (7.40) and (7.41) are derived in Problem 7.5. Using these results, the moment functions of (X_t) in (7.38) are arrived at in Problem 7.6, where (X_t) is shown to be stationary. On the other hand, additional assumptions are necessary to derive (7.30) from (7.38).

If definitions (6.76) and (7.37) are compared then it becomes obvious that the power of a deterministic function is distributed over the frequencies in (6.76), whereas, in (7.37), the variance of the harmonic process is distributed over the frequencies. In (6.76), the deterministic function is required to have a Fourier representation, i.e., a Fourier transform as defined in (6.22,1,4), (6.30), (6.35) or (6.49) with the spectrum being the squared absolute value of the Fourier transform. Does the harmonic process also have a Fourier representation? Is its spectrum the squared absolute value of its Fourier representation?

These questions are answered by constructing a complex-valued random function $Z(s)$ from the complex-valued (D_j), $j = -n, \ldots, -1, 0, 1, \ldots, n$ in (7.30). Their properties in (7.32), (7.33) and (7.34) imply that (i) pairs (D_j, D_k) are not correlated and (ii) sequence (D_j) is stationary in the first moment function and non-stationary in the second. Moreover, the complex-valued random coefficients (D_j) are associated with frequencies (s_j), being in interval $0 < s_j < 1/2$ and ordered, because $s_j < s_{j+1}$ is required in (7.30). Exploiting these properties, a complex-valued stochastic process $Z(s)$ with real parameter s is built from the (D_j) in (7.42) and (7.43) such that its increments $\mathrm{d}Z(s)$ are orthogonal and stationary in their first moment function:

$$Z(s) = \begin{cases} 0 & \text{for } 0 \le s < s_1 \\ D_1 & \text{for } s_1 \le s < s_2 \\ \sum_{j=1}^{k} D_j & \text{for } s_k \le s < s_{k+1} \\ \sum_{j=1}^{n} D_j & \text{for } s_n \le s < 1/2 \end{cases} \tag{7.42}$$

$$\mathrm{d}Z(s) = \begin{cases} Z(s) - Z(s - \mathrm{d}s) & \text{for } 0 \le s < 1/2 \\ 0 & \text{for } s = 1/2 \\ \overline{\mathrm{d}Z(-s)} & \text{for } -1/2 \le s \le 0 \end{cases} \tag{7.43}$$

$Z(s)$ is complex-valued with complex-valued increments $\mathrm{d}Z(s)$, with $-\mathrm{d}s$ being a decrement and $\mathrm{d}s$ an increment in frequency s and $|\mathrm{d}s|$ being small. In frequencies $s = s_j$, $j = -n, \ldots, -1, 1, \ldots, n$, as defined in (7.30), $Z(s)$ has jump discontinuities $\mathrm{d}Z(s_j) = D_j$ such that $Z(s)$ is mean-square continuous from the right as defined in (7.16,1). Furthermore, $\mathrm{d}Z(s) = 0$ for $s = 1/2$ is required in (7.43), asserting that $Z(s)$ is continuous in $s = 1/2$. The construction of $Z(s)$ is commented on in the remarks to (7.70) and a realisation is plotted in Fig. 7.8.

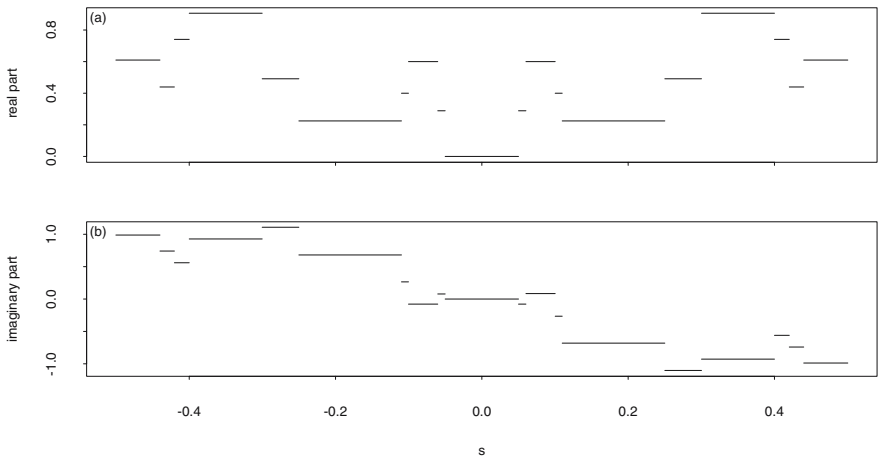


Fig. 7.8. A realisation of $Z(s)$ as defined in (7.42) constructed using the realisaton of the D_j in (7.7): in frequencies s_j, $Z(s)$ has jump discontinuities with steps $\mathrm{re}(d_j)$ and $\mathrm{im}(d_j)$. The real and imaginary parts of $Z(s)$ are shown in plots (a) and (b).

If $|\mathrm{d}s|$ is chosen small enough then an interval $[s, s+\mathrm{d}s]$ or $[s-\mathrm{d}s, s]$ contains at most one frequency s_j; however, most intervals are empty, i.e., they do not contain an s_j. Such an interval is associated with increment $\mathrm{d}Z(s)$ and, therefore, $\mathrm{d}Z(s) = D_j$ for non-empty and $\mathrm{d}Z(s) = 0$ for empty intervals. Using this result and (7.32), the expectation function of the increments $\mathrm{d}Z(s)$ for $-1/2 \le s \le 1/2$ is obtained in (7.44).

$$\mathrm{E}\big(\mathrm{d}Z(s)\big) = 0 \tag{7.44}$$

$$\begin{aligned}\mathrm{Cov}\big(\mathrm{d}Z(s), \mathrm{d}Z(r)\big) &= \mathrm{E}\big(\overline{\mathrm{d}Z(s)}, \mathrm{d}Z(r)\big) \\ &= \sum_{j=1}^{n} \mathrm{E}\big(|D_j|^2\big)\delta(s-s_j)\delta(s-r)\mathrm{d}s\mathrm{d}r \\ &= \sum_{j=1}^{n} (R_j^2/4)\delta(s-s_j)\delta(s-r)\mathrm{d}s\mathrm{d}r \end{aligned} \tag{7.45}$$

If the intersection of two intervals $[s, s+\mathrm{d}s]$ and $[r, r+\mathrm{d}r]$, having increments $\mathrm{d}Z(s)$ and $\mathrm{d}Z(r)$, does not contain an s_j, i.e., if it is empty, the covariance of the increments becomes identically zero. If, in contrast, the intersection is not empty, it contains an s_j, and the covariance function in (7.45) is obtained. When the result in (7.45) is compared with covariance function $\mathrm{Cov}\big(\mathrm{d}A(t), \mathrm{d}A(u)\big) = \sigma^2_{\mathrm{d}A}\delta(t-u)\mathrm{d}t\mathrm{d}u$ used in (7.102) to define the orthogonal increment process, it becomes obvious that $\mathrm{E}\big(|D_j|^2\big)\delta(s-s_j) = (R_j^2/4)\delta(s-s_j)$ plays the role of $\sigma^2_{\mathrm{d}A}$. Consequently, increments $\mathrm{d}Z(s)$ in (7.43) are not stationary in the second moment function whereas increments

$\mathrm{d}A(t)$ in (7.102) are stationary since their covariance function depends solely on the displacement $t - u$.

Not all properties of the orthogonal increment process $A(t)$ as defined in (7.102) or (7.23) are inherent in $Z(s)$ as defined in (7.42); however, $Z(s)$ is (like $A(t)$) continuous from the right side in its jump discontinuities as required in (7.16). For this reason and also for the reason that $Z(s)$ is constant between its jump discontinuities, the sum in (7.30) can be written using the stochastic Stieltjes integral as defined in (7.21): the integrand $\mathrm{e}^{\mathrm{i}2\pi st}$ is integrated with respect to the complex-valued integrator $Z(s)$. Since $0 < s_j < 1/2$ is required in (7.30), all jump discontinuities of $Z(s)$ are in $-1/2 < s < 1/2$ and the Stieltjes integral is evaluated in $-1/2 \le s \le 1/2$ to obtain the Fourier representation of the harmonic process:

$$X_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st} \mathrm{d}Z(s) \qquad t = \ldots, -1, 0, 1, \ldots \tag{7.46}$$

The Fourier representation in (7.46) is a Fourier-Stieltjes integral with respect to a complex-valued random function with orthogonal increments and an expectation function being identically zero, properties enabling a (at least formally) straightforward calculation of the variance of the harmonic process using its Fourier representation.

Using (7.46), the variance in (7.47) is obtained by applying definitions (7.27) and proposition (1.15,10), since $\mathrm{E}\big(dZ(s)\big) = 0$. Substituting the covariance function obtained in (7.45) for $\mathrm{E}\big(\overline{\mathrm{d}Z(r)}, \mathrm{d}Z(s)\big)$, the integral with the delta functions as in (7.49) is arrived at. Due to $\delta(s - r)$, the integrand becomes identically zero for $s \neq r$ and thus the integral is evaluated for the case $s = r$, and using (6.67), the integral in (7.50) is arrived at.

$$\mathrm{Var}X_t = \mathrm{E}\big(\overline{X_t} X_t\big) = \mathrm{E}\left(\int_{-1/2}^{1/2} \mathrm{e}^{-\mathrm{i}2\pi rt} \overline{\mathrm{d}Z(r)} \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st} \mathrm{d}Z(s) \right) \tag{7.47}$$

$$= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi(s-r)t} \mathrm{E}\Big(\overline{\mathrm{d}Z(r)}, \mathrm{d}Z(s) \Big) \tag{7.48}$$

$$= \sum_{j=1}^{n} \left(\int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi(s-r)t} \frac{R_j^2}{4} \delta(s - s_j) \delta(s - r) \mathrm{d}s \mathrm{d}r \right) \tag{7.49}$$

$$= \sum_{j=1}^{n} \left(\int_{-1/2}^{1/2} \frac{R_j^2}{4} \delta(s - s_j) \mathrm{d}s \right) = \sum_{j=-n}^{n} \frac{R_j^2}{4} \tag{7.50}$$

Evaluating the integral in (7.50), a variance is obtained which is identical to the one in (7.35) and also to the one obtained when integrating the spectrum $C_X(s)$ in (7.37).

7.3.2 Fourier Representation of a Stationary Stochastic Process

Every stationary random function has a Fourier representation. Besides being stationary, a random function with real parameter is required to be mean-square continuous as defined in (7.16): this excludes cases like $V(t)$ with covariance function $c_V(\tau)$ as proposed in (7.22). Existence theorem (7.51) can be found, together with a variety of proofs, in [113].

Let $X(t)$ or (X_t) be a stationary stochastic process with real (in this case, $X(t)$ is required to be mean-square continuous (7.16)) or integer parameter t, and with an expectation function being identically zero. Then a complex-valued random function $Z_X(s)$ exists such that the following Fourier-Stieltjes integrals converge in $L^2(\Omega, F)$ (defined in (7.10)):

$$X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}Z_X(s) \qquad t \text{ real}$$

$$(X_t) = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}Z_X(s) \qquad t \text{ integer}$$

The increments $\mathrm{d}Z_X(s)$ of $Z_X(s)$ have the following properties:

1. $\mathrm{E}\big(\mathrm{d}Z_X(s)\big) = 0$ *for all s s, i.e., their mean is identically zero,*
2. $\mathrm{Cov}\big(\mathrm{d}Z_X(s), \mathrm{d}Z_X(r)\big) = 0$ *for $s \neq r$, i.e., they are orthogonal, and*
3. $\mathrm{E}\big(|\mathrm{d}Z_X(s)|^2\big) \stackrel{\mathrm{def}}{=} \mathrm{d}C_X^{(I)}(s)$ *for all s. $C_X^{(I)}(s)$ is bounded, non-decreasing and continuous from the right in possible jump discontinuities. $C_X^{(I)}(s)$ is called the integrated spectrum of $X(t)$ or (X_t).*

(7.51)

Priestley in [113] also makes the Fourier representation theorem plausible in the case of a continuous-time stationary process $X(t)$. Borrowing from [113], it becomes plausible that the power (in the physical sense as in the remarks to (6.76)) of the process pertaining to frequencies between s and $s + \mathrm{d}s$ is the expectation of the squared increments $\mathrm{E}\big(|\mathrm{d}Z_X(s)|^2\big)$ in (7.51,3). The rationale starts with a realisation of $X(t)$ in (7.52) and arrives at the expectation in (7.60) from which the result then follows in the limiting case.

Let $x(t)$ be a realisation of the stationary random function $X(t)$ in interval $-T/2 \leq t \leq T/2$, and let $x_T(t)$ be a periodic function with period T such that $x_T(t) = x(t)$ in $-T/2 \leq t \leq T/2$ and $x_T(t+nT) = x_T(t)$ at the exterior of this interval, $n = \ldots, -1, 0, 1, \ldots$. The Fourier series of $x_T(t)$ is written complex $\sum_{k=-\infty}^{\infty} F_k \mathrm{e}^{\mathrm{i}2\pi s_k t}$ as in (6.29) with $F_k = (1/T)\int_{-T/2}^{T/2} x_T(t)\mathrm{e}^{-\mathrm{i}2\pi s_k t}\mathrm{d}t$, $k = \ldots, -1, 0, 1, \ldots$, and (7.52) is obtained since $\Delta s_k = 1/T$, as argued in the remarks to (6.35).

$$x_T(t) = \sum_{k=-\infty}^{\infty} \left(\int_{-T/2}^{T/2} x_T(t)\mathrm{e}^{-\mathrm{i}2\pi s_k t}\mathrm{d}t \right) \mathrm{e}^{\mathrm{i}2\pi s_k t}\Delta s_k \tag{7.52}$$

$$g_T(s_k) = \int_{-T/2}^{T/2} x_T(t)\mathrm{e}^{-\mathrm{i}2\pi s_k t}\mathrm{d}t \tag{7.53}$$

$$x_T(t) = \sum_{k=-\infty}^{\infty} g_T(s_k)\mathrm{e}^{\mathrm{i}2\pi s_k t}\Delta s_k = \sum_{k=-\infty}^{\infty} \frac{g_T(s_k)}{T}\mathrm{e}^{\mathrm{i}2\pi s_k t} \tag{7.54}$$

For the integral in (7.52), $g_T(s_k)$ as defined in (7.53) is written in (7.54) where $1/T$ is substituted with $\Delta s_k = s_{k+1} - s_k$ to arrive at a Fourier representation of $x_T(t)$, i.e., a realisaton (x_t) confined in the interval $-T/2 \le t \le T/2$. Does a Fourier representation of (x_t) exist, $t = \ldots, -1, 0, 1, \ldots$?

If, as in the remarks to (6.35), the interval $-T/2 \le t \le T/2$ becomes wider with increasing T, then $x(t)$ is required to be in $L^2(-\infty, \infty)$ in the limiting case for $T \to \infty$ thus asserting the existence of the integral in (7.53). $x(t)$ however, being a realisation of a stationary process $X(t)$, is not in $L^2(-\infty, \infty)$ because it does not decay and thus does not become identically zero for $-\infty \leftarrow t$ and $t \to \infty$. In a physical rationale similar to those in the remarks to (6.76), the energy pertaining to trigonometric oscillations with frequencies in an interval being Δs_k wide, i.e., $|g_T(s_k)|^2 \Delta s_k$ as obtained by substituting $\Delta s_k = 1/T$ in (7.54), increases with increasing T, as is concluded from Parseval's identity in (6.74,2). In the limiting case for $T \to \infty$, the dissipated energy becomes infinite and Δs_k becomes arbitrarily small. Consequently, infinite energy is dissipated to generate non-damped sinusoidal oscillations with frequencies between s and $s + \mathrm{d}s$, i.e., in an arbitrarily small interval, this conclusion being reconcilable with the Fourier transform pairs obtained in Problem 6.19. Consequently, such oscillations do not exist.

Do oscillations with frequencies between s and $s + \mathrm{d}s$ and having a finite power $\big(|g_T(s)|^2/T\big)\mathrm{d}s$ exist in the limiting case for $T \to \infty$? If such oscillations exist then a power spectrum, i.e., the power of the realisation as a function of the frequencies, is obtained in the limiting case. For all realisations $x(t)$ of the random function $X(t)$, the mean of $\big(|g_T(s)|^2/T\big)$ is calculated and the limit in (7.55) is arrived at:

$$\lim_{T\to\infty} \mathrm{E}\left(\frac{|G_T(s)|^2}{T}\right). \tag{7.55}$$

This limit is, on condition that it exists, the mean power of $X(t)$ pertaining to trigonometric oscillations with frequencies between s and $s + \mathrm{d}s$.

Now, $\big(|G_T(s)|^2\big)/T$ is redefined by substituting, in (7.52) and (7.53), realisation $x_T(t)$ with random function $X_T(t)$ and by assuming that the Fourier-Stieltjes-Integral $G_T(s_k) = \int_{-T/2}^{T/2} \mathrm{e}^{-\mathrm{i}2\pi s_k t}\mathrm{d}X_T(t)$ thus obtained does exist. Hence, (7.56) is arrived at, being a stochastic version of (7.54):

$$X_T(t) = \sum_{k=-\infty}^{\infty} G_T(s_k)\mathrm{e}^{\mathrm{i}2\pi s_k t}\Delta s_k = \sum_{k=-\infty}^{\infty} \frac{G_T(s_k)}{T}\mathrm{e}^{\mathrm{i}2\pi s_k t}. \tag{7.56}$$

If $\big(G_T(s_k)\big)/T = G_T(s_k)\Delta s_k$ is such that $G_T(s)$ is integrable in the limiting case for $T \to \infty$ and $\Delta s_k \to 0$, then

$$\Delta Z_T(s_k) = \int_{-\infty}^{s_k} G_T(s)\mathrm{d}s - \int_{-\infty}^{s_{k-1}} G_T(s)\mathrm{d}s \tag{7.57}$$

$$\approx G_T(s_k)\Delta s_k = G_T(s_k)/T \tag{7.58}$$

$$X_T(t) = \sum_{k=-\infty}^{\infty} \Delta Z_T(s_k)\mathrm{e}^{\mathrm{i}2\pi s_k t}\Delta s_k \tag{7.59}$$

is obtained and, taking the expectations,

$$\mathrm{E}\big(|\Delta Z_T(s_k)|^2\big) \approx \mathrm{E}\big(|G_T(s_k)|^2(\Delta s_k \times \Delta s_k)\big) = \mathrm{E}\left(\frac{|G_T(s_k)|^2}{T}\right)\Delta s_k \tag{7.60}$$

is finally arrived at. Assuming that these expectations exist in the limiting case for $T \to \infty$ and $\Delta s_k \to 0$, (7.60) is compared with (7.51): it now becomes plausible that, in the limiting case, the expectation of the squared increments $\mathrm{E}(|\mathrm{d}Z_X(s)|^2)$ in (7.51,3) is the power of the stationary process $X(t)$ pertaining to those trigonometric oscillations in its Fourier representation which have frequencies between s and $s + \mathrm{d}s$.

7.3.3 Covariance Function and Spectrum

If a continuous-time stochastic process $X(t)$ is stationary, mean-square continuous from the right in possible jump discontinuities and has an identically zero expectation function, as required in (7.51), then its covariance function $c_X(\tau)$ can be obtained as a Fourier-Stieltjes integral from its integrated spectrum $C_X^{(I)}(s)$:

$$c_X(\tau) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi\tau s}\mathrm{d}C_X^{(I)}(s), \qquad \tau \text{ real.} \tag{7.61}$$

The Fourier-Stieltjes integral (7.61) is defined for all integrated spectra having properties (7.51,3), i.e., also for integrated spectra with jump discontinuities. If, in contrast, an integrated spectrum is differentiable for all s, $C_X(s) = \mathrm{d}\big(C_X^{(I)}(s)\big)/\mathrm{d}s$, then the usual Fourier integral (7.62)

$$c_X(\tau) = \int_{-\infty}^{\infty} C_X(s)\mathrm{e}^{\mathrm{i}2\pi\tau s}\mathrm{d}s, \qquad \tau \text{ real} \tag{7.62}$$

$$C_X(s) = \int_{-\infty}^{\infty} c_X(\tau)\mathrm{e}^{-\mathrm{i}2\pi s\tau}\mathrm{d}\tau = 2\int_0^{\infty} c_X(\tau)\cos(2\pi s\tau)\mathrm{d}\tau \tag{7.63}$$

is obtained. To this Fourier transform pertains the inverse transform (7.63). Hence, $c_X(\tau)$ and $C_X(s)$ are a Fourier transform pair as defined in (6.35). The second = in (7.63) applies on condition that $X(t)$ is a real-valued random function with an even covariance function since (2.7,1) implies (6.32). If $X(t)$ is complex-valued then its complex-valued covariance function is not even, as concluded from the remarks to (7.28).

A discrete-time stochastic process (X_t) with integer parameter t as required in (7.51) has a covariance function $c_X(\tau)$ defined for integer lags τ and an integrated spectrum $C_X^{(I)}(s)$ defined in interval $-1/2 \le s \le 1/2$ (and being periodic with period 1). Since $C_X^{(I)}(s)$ is confined to $-1/2 \le s \le 1/2$, i.e., the period pertaining to the sampling interval $\Delta t = 1$, there is no aliasing, as argued in the remarks to (7.30). Hence, the Fourier-Stieltjes integral in (7.64) and, provided that $C_X^{(I)}(s)$ is differentiable, the usual Fourier integral in (7.65), are evaluated from $-1/2$ until $1/2$.

$$c_X(\tau) = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi\tau s} \mathrm{d}C_X^{(I)}(s) \qquad \tau = \ldots, -1, 0, 1, \ldots \tag{7.64}$$

$$c_X(\tau) = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi\tau s} C_X(s) \mathrm{d}s \qquad \tau = \ldots, -1, 0, 1, \ldots \tag{7.65}$$

$$C_X(s) = \sum_{\tau=-\infty}^{\infty} c_X(\tau) \mathrm{e}^{-\mathrm{i}2\pi s\tau} = 2 \sum_{\tau=0}^{\infty} c_X(\tau) \cos(2\pi s\tau) \tag{7.66}$$

$c_X(\tau)$ and $C_X(s)$ in (7.65) and (7.66) are a Fourier transform pair as defined in (6.49) and (6.50). This Fourier transform pair is discussed in Sect. 7.5.4.

(7.61) through to (7.66) are derived from the spectral representation theorem (7.51). For example, the Fourier-Stieltjes integral (7.61) is obtained as follows:

$$\begin{aligned} c_X(\tau) &= \mathrm{Cov}\big(X(t), X(u)\big) = \mathrm{E}\big(\overline{X(t)} X(u)\big) \qquad \tau = t - u \qquad (7.67) \\ &= \mathrm{E}\left(\int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}2\pi st} \overline{\mathrm{d}Z_X(s)} \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi ru} \mathrm{d}Z_X(r) \right) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}2\pi st} \mathrm{e}^{\mathrm{i}2\pi ru} \mathrm{E}\Big(\overline{\mathrm{d}Z_X(s)} \mathrm{d}Z_X(r) \Big) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}2\pi st} \mathrm{e}^{\mathrm{i}2\pi ru} \Big(\mathrm{d}C_X^{(I)}(s) \delta(s-r) \Big) \mathrm{d}r \qquad (7.68) \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}2\pi(t-u)s} \Big(\mathrm{d}C_X^{(I)}(s) \Big) \\ &= \int_{-\infty}^{\infty} \mathrm{e}^{-\mathrm{i}2\pi\tau s} \mathrm{d}C_X^{(I)}(s) \qquad (7.69) \end{aligned}$$

(7.68) directly follows from (7.51) since the variance and covariance functions of increments $\mathrm{d}Z_X(s)$ as stipulated in (7.51,2) and (7.51,3) can be summarised writing $\mathrm{E}\big(\overline{\mathrm{d}Z_X(s)} \mathrm{d}Z_X(r)\big) = \delta(s-r) \mathrm{d}C_X^{(I)}(s) \mathrm{d}r$ due to the following reasons. Firstly, for $r \neq s$, $\mathrm{Cov}\big(\mathrm{d}Z_X(r), \mathrm{d}Z_X(s)\big) = \mathrm{E}\big(\overline{\mathrm{d}Z_X(s)} \mathrm{d}Z_X(r)\big) = 0$, since the increments are orthogonal, and secondly, for $r = s$, i.e., for lags becoming identically zero, $\mathrm{Cov}\big(\mathrm{d}Z_X(r), \mathrm{d}Z_X(s)\big) = \mathrm{d}C_X^{(I)}(s) \delta(s-r) \mathrm{d}r$, as concluded from the remarks to (7.26) and (7.102).

The above derivation of (7.69), despite being concise and smart, does only to some extent illustrate the properties of both the covariance function and the integrated spectrum. Since these properties are in close relationship to each other, they are enumerated in (7.70). Functions $\varphi(t)$ and $F(s)$ as in (7.70,2) are a Fourier-Stieltjes transform pair. $\varphi(t)$ in (7.70,3) shares its properties with the covariance function $c_X(\tau)$ of a stationary random function $X(t)$, as is concluded from (2.7) and (2.8), and $F(s)$ in (7.70,1) shares its properties with the integrated spectrum $C_X^{(I)}(s)$ of $X(t)$ as required in (7.51,3). The properties of $\varphi(t)$ imply those of $F(s)$, and vice-versa, as proposed in (7.70,4) and (7.70,2).

1. *Let $F(s)$ be a real-valued and non-decreasing function, being continuous from the right in possible jump discontinuities and having a lower bound $F(-\infty) = \lim_{-\infty \leftarrow s} F(s)$ and an upper bound $F(\infty) = \lim_{s \to \infty} F(s)$. Then the*
2. *Fourier-Stieltjes integral $\varphi(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi ts} \mathrm{d}F(s)$ exists.*
3. *Let $\varphi(t)$ be bounded($|\varphi(t)| \leq \int_{-\infty}^{\infty} dF(s) = F(\infty) - F(-\infty)$), positive semi-definite($\sum_{j=1}^{n} \sum_{k=1}^{n} d_j \overline{d_k} \varphi(t_j - t_k) \geq 0$), and let $\varphi(-t) = \overline{\varphi(t)}$ (or $\varphi(-t) = \varphi(t)$, for real-valued $\varphi(t)$).*
4. *For each $\varphi(t)$ with properties (7.70,3) there exists an $F(s)$ with properties (7.70,1) such that (7.70,2) is satisfied.*

(7.70)

The properties of $F(s)$ as required in (7.70,1) assert the existence of the Fourier-Stieltjes integral $\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi ts} \mathrm{d}F(s)$. Properties (7.70,3) of $\varphi(t)$ are a consequence of integral (7.70,2) and easily obtained. In contrast, it is difficult to prove (7.70,4), i.e., that, given a function $\varphi(t)$ having the properties (7.70,3), there is a function $F(s)$ such that $\varphi(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi ts} \mathrm{d}F(s)$ [126].

If $t = \ldots, -1, 0, 1, \ldots$ in (7.70), then $F(s)$ has properties (7.70,1) duly modified to accommodate the integer argument of $\varphi(t)$: the Fourier-Stieltjes integral is evaluated in the interval $-1/2 \leq s \leq 1/2$ and it is required that $F(s)$ be bounded in $-1/2 \leq s \leq 1/2$ and continuous from the left in $1/2$ with limit $F(1/2) = \lim_{s \to 1/2} F(s)$. The properties of $\varphi(t)$ in (7.70,3) do not undergo changes in this case. For example, bearing in mind that $\mathrm{E}\big(|\mathrm{d}Z_X(s)|^2\big) = \mathrm{d}C_X^{(I)}(s)$ as defined in (7.51), a complex-valued random function $Z(s)$ is constructed in (7.42) using the harmonic process (X_t) in (7.30) such that its integrated spectrum $C_X^{(I)}(s)$ possesses the properties required above: $\mathrm{d}Z(1/2) = 0$ is required in (7.42) in order to obtain $\mathrm{E}\big(|\mathrm{d}Z(1/2)|^2\big) = 0 = \mathrm{d}C_X^{(I)}(1/2)$ and therefore the desired integrated spectrum which is continuous in $1/2$.

Properties (7.70) of pairs of Fourier-Stieltjes transforms are known as *Herglotz's theorem* (for integer t) or *Bochner's theorem* (for real t): Given a positive semi-definite function $\varphi(t)$, a bounded and non-decreasing function $F(s)$ exists such that $\varphi(t)$ is the Fourier-Stieltjes integral of $F(s)$.

7.3.4 Types of Spectra

An integrated spectrum is a real-valued function that is bounded, does not decrease, and is continuous from the right in possible jump discontinuities. Functions having these properties can be classified as in (7.71).

Let $C_X^{(I)}(s)$ be the integrated spectrum (7.51,3) pertaining to a stationary stochastic process $X(t)$. $C_X^{(I)}(s)$ is called:

1. *differentiable on condition that $C_X(s) = \mathrm{d}\big(C_X^{(I)}(s)\big)/\mathrm{d}s$ for all s (in this case $C_X(s)$ is called a continuous spectrum or a spectral density)*
2. *a discrete spectrum or a line spectrum on condition that $C_X^{(I)}(s)$ has jump discontinuities*
3. *a mixed spectrum on condition that $C_X^{(I)}(s)$ is differentiable except for frequencies s_i, $i = 1, \ldots, n$, n a small number of jump discontinuities.*

(7.71)

In contrast to definitions (7.71), only two types of spectra either discrete or mixed, or continuous are discerned in (7.61) and (7.62) as well as in (7.64) and (7.65).

To give a first example, the line spectrum of the harmonic process $X(t)$ (7.30) is derived from the covariance function $c_X(\tau)$ (7.31) pertaining to $X(t)$. $c_X(\tau)$ is not in $L^2(-\infty, \infty)$ because $|c_X(\tau)|$ does not damp out for $-\infty \leftarrow \tau$ and $\tau \to \infty$. Hence, the Fourier transform $C_X(s) = \mathcal{F}_{-i}\big(c_X(\tau)\big)$ in (7.72)

$$C_X(s) = \sum_{\tau=-\infty}^{\infty} \left(2 \sum_{j=1}^{n} \Big(\big(R_j^2/4\big) \Big) \cos(2\pi s_j \tau) \right) \cos(2\pi s \tau) \tag{7.72}$$

$$= 2 \sum_{j=1}^{n} \Big(\big(R_j^2/4\big) \big(\delta(s - s_j) + \delta(s + s_j)\big) \Big) \tag{7.73}$$

$$= \begin{cases} R_j^2/4 & \text{for } s = s_{-n}, \ldots, -1, 1, \ldots, s_n \\ 0 & \text{for all other } s \end{cases} \qquad -1/2 \leq s \leq 1/2 \tag{7.74}$$

is calculated using the result of Problem 6.19 to obtain (7.73). Thereafter, by applying the rules pertaining to the evaluation of integrals with delta functions as derived from (6.64), the spectrum of the harmonic process is arrived at in (7.74). Please compare this result with the definition given in (7.37). Also using the rules derived from (6.64), the covariance function of the harmonic process is obtained from its spectrum in the remarks to (7.37).

The second and third example stochastic processes are stationary with integer parameter $t = \ldots, -1, 0, 1, \ldots$, and both processes have covariance functions that converge absolutely, as defined in (2.24), for increasing lags and thus are in $L^2(-\infty, \infty)$. From the covariance functions of these processes, their continuous spectra are obtained by calculating the Fourier transforms as required in (7.63) or (7.66).

Example process no. 2 is the white noise process (W_t) as defined in (2.10) or (2.11) having expectation function $\mu_W(t) = \mu_W$, variance function $\sigma_W^2(t) = \sigma_W^2$ and covariance function $c_W(\tau) = \sigma_W^2$ for $\tau = 0$, and $c_W(\tau) = 0$ for $\tau \neq 0$. $\tau = \ldots, -1, 0, 1, \ldots$ implies that $c_W(\tau)$ and $C_W(s)$ are a Fourier transform pair as defined in (6.49).

$$C_W(s) = \sum_{\tau=-\infty}^{+\infty} c_W(\tau)\mathrm{e}^{-\mathrm{i}2\pi s\tau} \qquad \tau = \ldots, -1, 0, 1, \ldots \tag{7.75}$$

$$= \sum_{\tau=-\infty}^{-1} 0 \times \mathrm{e}^{-\mathrm{i}2\pi s\tau} + \sigma_W^2 \times \mathrm{e}^{-\mathrm{i}2\pi s0} + \sum_{\tau=1}^{+\infty} 0 \times \mathrm{e}^{-\mathrm{i}2\pi s\tau}$$

$$= \sigma_W^2 \qquad -1/2 \leq s \leq 1/2 \tag{7.76}$$

The spectrum of the white noise process is constant for $-1/2 \leq s \leq 1/2$. Consequently, all frequencies contribute identically to the variance of the process, or, using the terminology introduced in Sect. 6.4, the total energy of the process is uniformly distributed over all frequencies (the continuous uniform probability distribution is defined in the remarks to (7.30)).

The spectrum of the electromagnetic radiation emitted by the sun is approximately constant in the visible range (wave-lengths between approximately 400 and 700 νm). In the visible range therefore, the sun radiates with approximately identical intensity for each colour band (e.g., yellow light has wave-lengths between 575 and 585 νm) and thus sunlight appears to be white to the human eye. Consequently, the stochastic process (W_t) defined in (2.10) or (2.11) is called a white noise process, since its spectrum $C_W(s) = \sigma_W^2$ is constant as shown in (7.76).

Example process no. 3 is the AR[1] process (X_t), $t = \ldots, -1, 0, 1, \ldots$, as constructed in (2.13) from a white noise process (W_t) and thus having covariance function $c_X(\tau)$ in (2.16) on condition that $\mu_W = 0$. If $|a| < 1$ and $N \to \infty$ then (X_t) becomes stationary with $c_X(\tau) = \sigma_W^2\big(1/(1-a^2)\big)a^{|\tau|}$. By Fourier transforming $c_X(\tau)$

$$C_X(s) = \sigma_W^2 \frac{1}{1-a^2}\left(\sum_{\tau=-\infty}^{+\infty} a^{|\tau|}\mathrm{e}^{-\mathrm{i}2\pi s\tau}\right) \tag{7.77}$$

$$= \sigma_W^2 \frac{1}{1-a^2}\left(\sum_{\tau=-\infty}^{-1} a^{|\tau|}\mathrm{e}^{-\mathrm{i}2\pi s\tau} + 1 + \sum_{\tau=1}^{\infty} a^{|\tau|}\mathrm{e}^{-\mathrm{i}2\pi s\tau}\right)$$

$$= \sigma_W^2 \frac{1}{1-a^2}\left(1 + \sum_{\tau=1}^{\infty} a^{\tau}\mathrm{e}^{\mathrm{i}2\pi s\tau} + \sum_{\tau=1}^{\infty} a^{\tau}\mathrm{e}^{-\mathrm{i}2\pi s\tau}\right)$$

$$= \sigma_W^2 \frac{1}{1-a^2}\left(1 + \sum_{\tau=1}^{\infty}\left(a\mathrm{e}^{\mathrm{i}2\pi s}\right)^{\tau} + \sum_{\tau=1}^{\infty}\left(a\mathrm{e}^{-\mathrm{i}2\pi s}\right)^{\tau}\right)$$

$$= \sigma_W^2 \frac{1}{1-a^2}\left(1 + a\mathrm{e}^{\mathrm{i}2\pi s}\sum_{\tau=0}^{\infty}\left(a\mathrm{e}^{\mathrm{i}2\pi s}\right)^{\tau} + a\mathrm{e}^{-\mathrm{i}2\pi s}\sum_{\tau=0}^{\infty}\left(a\mathrm{e}^{-\mathrm{i}2\pi s}\right)^{\tau}\right) \tag{7.78}$$

$$= \sigma_W^2 \frac{1}{1-a^2}\left(1 + \frac{a\mathrm{e}^{\mathrm{i}2\pi s}}{1-a\mathrm{e}^{\mathrm{i}2\pi s}} + \frac{a\mathrm{e}^{-\mathrm{i}2\pi s}}{1-a\mathrm{e}^{-\mathrm{i}2\pi s}}\right)$$

$$= \sigma_W^2 \frac{1}{1+a^2-2a\cos(2\pi s)} \qquad -1/2 \le s \le 1/2 \tag{7.79}$$

the sums in (7.78) are arrived at which thereafter can be easily evaluated since a geometric series $1+z+z^2+z^3+\ldots = 1/(1-z)$ converges for $|z| < 1$. The final result in (7.79) follows because $1-a\mathrm{e}^{-\mathrm{i}\theta}-a\mathrm{e}^{\mathrm{i}\theta}+a\mathrm{e}^{-\mathrm{i}\theta}a\mathrm{e}^{\mathrm{i}\theta} = 1-2a\cos\theta+a^2$.

The spectrum of the AR[1] model as arrived at in (7.79) is not constant even though this model is constructed from a white noise process having a constant spectrum as derived in (7.76). Spectra of linear processes (and therefore also the spectrum of the AR[1] model) are dealt with to a certain extent in Sect. 7.4.

Example process no. 4, $X(t) = U(t) + V(t)$, has a mixed spectrum as defined in (7.71,3) on condition that $U(t)$ and $V(t)$ are stochastic processes with the following properties: (i) the expectation function of both processes is identically zero, i.e., $\mathrm{E}U(t) = \mathrm{E}V(t) = 0$, (ii) they are not correlated, i.e., $\mathrm{Cov}\big(U(t), V(u)\big) = 0$ for $t \neq u$ and for $t = u$, and (iii), they are stationary and thus have a spectrum, $U(t)$ a continuous one and $V(t)$ a discrete one. Taking advantage of these properties, the covariance function $c_X(\tau)$ of $X(t)$ is obtained in (7.80)

$$\begin{aligned} c_X(\tau) &= \mathrm{Cov}\big(U(t)+V(t), U(t+\tau)+V(t+\tau)\big) \\ &= \mathrm{E}\Big(\big(U(t)+V(t)\big)\big(U(t+\tau)+V(t+\tau)\big)\Big) \\ &= \mathrm{E}\big(U(t)U(t+\tau) + U(t)V(t+\tau) \\ &\qquad + V(t)(U(t+\tau) + V(t)V(t+\tau)\big) \\ &= \mathrm{Cov}\big(U(t), U(t+\tau)\big) + 0 + 0 + \mathrm{Cov}\big(V(t), V(t+\tau)\big) \\ &= c_U(\tau) + c_V(\tau) \end{aligned} \tag{7.80}$$

and thereafter the spectrum of $X(t)$ in (7.81), by applying (6.48,5).

$$\begin{aligned} C_X(s) &= \int_{-\infty}^{\infty} \big(c_U(\tau) + c_V(\tau)\big)\mathrm{e}^{-\mathrm{i}2\pi s\tau}\mathrm{d}\tau \\ &= \int_{-\infty}^{\infty} c_U(\tau)\mathrm{e}^{-\mathrm{i}2\pi s\tau}\mathrm{d}\tau + \int_{-\infty}^{\infty} c_V(\tau)\mathrm{e}^{-\mathrm{i}2\pi s\tau}\mathrm{d}\tau \\ &= C_U(s) + C_V(s) \end{aligned} \tag{7.81}$$

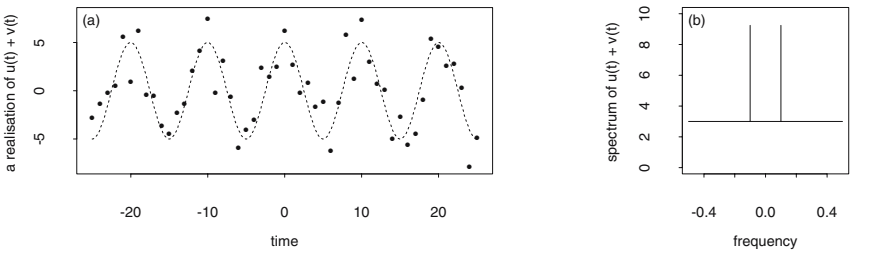


Fig. 7.9. On the left, a realisation $(x_t) = (u_t) + (v_t)$ (•) with $v_t = 5\cos(0.2\pi t)$ (broken line) in the interval $-25, \ldots, -1, 0, 1, \ldots, 25$, i.e., a realisation of the process having a mixed spectrum as in (7.82) with $\sigma_U^2 = 3$, $R_l = 5$, $s_l = 0.1$, (plotted on the right).

Since $V(t)$ has a discrete spectrum $C_V(s)$ and thus $c_V(\tau)$ is not in $L^2(-\infty, \infty)$, the Fourier transform $\int_{-\infty}^{\infty} c_V(\tau)\mathrm{e}^{-\mathrm{i}2\pi s\tau}\mathrm{d}\tau$ is calculated using the delta function, as demonstrated above in (7.72) in the case of the harmonic process.

For example, let (U_t) be a white noise process as defined in (2.10) or (2.11) having a spectrum $C_U(s) = \sigma_U^2$ (7.76), and let (V_t) be a harmonic process as defined in (7.30) having only one trigonometric oscillation, i.e., $V_t = R_l\cos(2\pi s_l t + P_l)$, $0 < s_l < 1/2$, with spectrum $C_V(s) = R_l^2/4$, for $s = -s_l, s_l$ and $C_V(s) = 0$, for $s \neq -s_l$ or $s \neq s_l$, as obtained in (7.74). Then, using (7.81) derived above, the spectrum of $(X_t) = (U_t) + (V_t)$

$$C_X(s) = \begin{cases} \sigma_U^2 + R_l^2/4 & \text{for } s = -s_l, s_l \\ \sigma_U^2 & \text{for } s \neq -s_l \text{ or } s \neq s_l \end{cases} \qquad -1/2 \leq s \leq 1/2 \tag{7.82}$$

is arrived at in (7.82). $C_X(s)$ is a constant σ_U^2 with added spikes of height $R_l^2/4$ at frequencies $s = -s_l$ and $s = s_l$, as plotted in Fig. 7.9 (b). A realisation of $(X_t) = (U_t) + (V_t)$ is plotted in Fig. 7.9 (a).

7.3.5 Bandwidth of a Spectrum

The bandwidth of a spectrum as defined in (7.83) is the width of its narrowest peak or valley or the smallest distance between the spectral lines.

1. *The bandwidth* $\mathcal{B}_{\triangle}\big(C_X(s)\big)$ *of a continuous spectrum* $C_X(s)$ *(a spectral density as defined in (7.71,1)) is the width at half height (6.97) of* $C_X(s)$.
2. *The bandwidth* $\mathcal{B}_{\triangle}\big(C_X^{(I)}(s)\big)$ *of a discrete (or line) spectrum is the smallest distance between the jump discontinuities of the pertaining integrated spectrum* $C_X^{(I)}(s)$.

(7.83)

Consequently, the bandwidth of a continuous spectrum is determined by means of a non-logarithmic plot, as demonstrated in the remarks to (6.97).

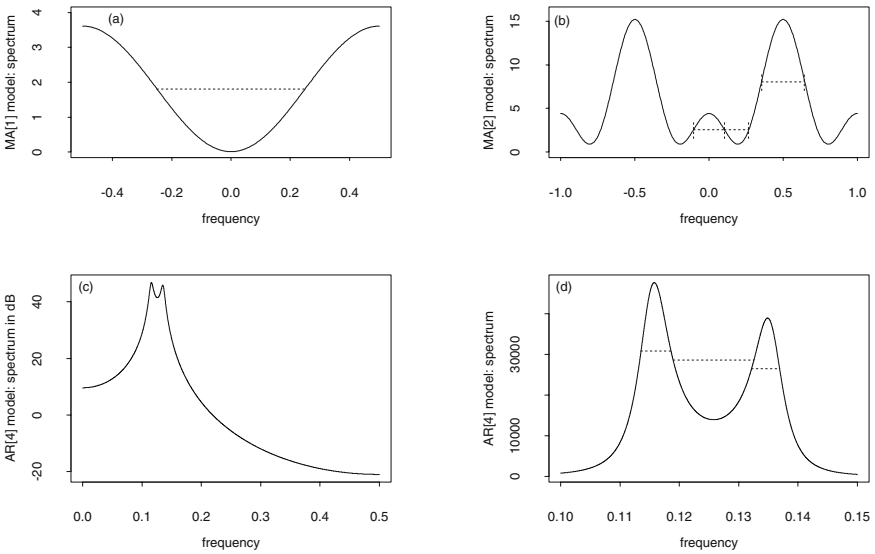


Fig. 7.10. Examples of continuous spectra with their bandwidths as defined in (7.83).

Plots (a) and (b) above in Fig. 7.10, for example, show two continuous spectra of linear processes in non-logarithmic scale: the one of an MA[1] model with $b = 0.9$ and $\sigma_W^2 = 1$ and the one of an MA[2] model with $b_1 = 0.9$, $b_2 = -2.0$ and $\sigma_W^2 = 1$. Below in Fig. 7.10, the spectrum of an AR[4] model as defined in Fig. 7.14 is plotted in logarithmic (c) and non-logarithmic scale (d). These spectra are derived in Sect. 7.4.

In the non-logarithmic spectral densities, the widths at half height of the peaks and valleys are drawn. Often, the width at half height can only be determined in adequate plots, as demonstrated by the following examples: (i) in Fig. 7.10 (a) and (b), the spectra are plotted for intervals $-1/2 \leq t \leq 1/2$ and $-1 \leq t \leq 1$ in order to capture their forms, seeing that the spectra of discrete-time stochastic processes are even functions and periodic with period 1, and (ii) in Fig. 7.10 (d), the spectrum under analysis is plotted for the frequency band containing its peaks. The widths at half height of the peaks in Fig. 7.10 (d) are obtained using the valley between the peaks as base height: 0.00506 is the width at half height of the left peak and 0.00474 the one of the right peak, whereas the width at half height of the valley is 0.01364. Consequently, the bandwidth of the spectrum of the AR[4] process in Fig. 7.10 is 0.00474, applying the remark to (6.97).

7.4 Spectra of Linear Processes

A linear process (2.25) is the output of a stochastic version of the LTI filter (6.120) on condition that a white noise process is the input. Consequently, the linear process inherits the continuous spectrum of the white noise process, as shown in Sect. 7.4.1. In Sect. 7.4.2, examples of spectra of linear processes are given.

7.4.1 Stochastic Filters

A stochastic version of the linear and time-invariant (LTI) filter, introduced in Sect. 6.8.1, is defined in (7.84). If the input in the stochastic LTI filter as defined in (7.84) is a discrete-time stochastic process then the parameter of both, input and output, are assumed to be integer, and the convolution integral is substituted with a convolution sum. Both a convolution sum or a convolution integral are, as proposed in the remarks closing Sect. 2.4.1, linear, time-invariant and commutative, albeit on condition that input and output are stochastic processes, since (2.28), (2.29), (2.33) and (2.35) are not restricted to deterministic functions. If the expectation function of the stationary input is not identically zero then the de-meaned process $X(t)-\mu_X$ is convolved. Under the conditions required in (7.84) the convolution integral or sum converge in mean square, as defined in (7.11).

$$
\begin{aligned}
&Y(t) = \int_{-\infty}^{\infty} f(t-u)X(u)\mathrm{d}u = \int_{-\infty}^{\infty} f(u)X(t-u)\mathrm{d}u \quad t, u \text{ real} \\
&\text{is called a stochastic LTI filter on condition that:} \\
&1.\ \text{the real-valued input } X(t) \text{ is a stationary random function} \\
&\quad \text{with moment functions } \mu_X(t) = \mu_X = 0 \text{ and } c_X(\tau) \\
&2.\ \text{the impulse response function } f(t) \text{ is in } L^2(-\infty,\infty), \text{ and} \\
&\quad \text{consequently, the frequency response or transfer function} \\
&\quad F(s) = \textstyle\int_{-\infty}^{\infty} f(t)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \text{ is obtained using (6.35), and} \\
&3.\ \textstyle\int_{-\infty}^{\infty} |F(s)|^2 \mathrm{d}C_X^{(I)}(s) < \infty.
\end{aligned}
\tag{7.84}
$$

If, for example, an ARMA$[p,q]$ model (5.36) is constructed using (5.38) as the output of a stochastic filter, then the convolution sum converges since (i) the input is a white noise process as defined in (2.11) with moment functions as required in (7.84,1), (ii) the impulse response function (5.38) is in $L^2(-\infty,\infty)$ as required in (7.84,2), and (iii) the Stieltjes integral in (7.84,3) (integrated in the interval $-1/2 \le s \le 1/2$, because $t = \ldots,-1,0,1,\ldots$) converges as is concluded from (6.49) and (7.76).

The input (7.84,1) in a stochastic LTI filter has a Fourier representation $X(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}Z_X(s)$ as required in (7.51), and also the output has a Fourier representation $Y(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}Z_Y(s)$ on condition that $Y(t)$ is stationary. If it is further assumed that convolution theorem (6.48,6) also

applies to a filter with random functions as input and output then $\mathrm{d}Z_Y(s) = F(s)\mathrm{d}Z_X(s)$ is obtained, with $F(s) = \mathcal{F}_{-i}\big(f(t)\big)$. This result is the stochastic version of (6.121,2) and is arrived at in (7.87).

A derivation of (7.87) begins by substituting the Fourier representation of $X(t)$ as input in the stochastic filter (7.84) to obtain the convolution integral (7.85). For all t, in (7.85), the Fourier-Stieltjes integral is multiplied with $f(t-u)$ and the convolution integral is evaluated. Since a Riemann or a Stieltjes integral as required in (7.20) and (7.21) is linear, the derivation is continued by using (6.142) as a template and (7.86) is arrived at, provided that the integrals exist.

$$
\begin{aligned}
Y(t) &= \int_{-\infty}^{\infty} f(t-u)\left(\int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi su}\mathrm{d}Z_X(s)\right)\mathrm{d}u && (7.85)\\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t-u)\mathrm{e}^{\mathrm{i}2\pi s(u-t+t)}\mathrm{d}Z_X(s)\mathrm{d}u\\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(t-u)\mathrm{e}^{-\mathrm{i}2\pi s(t-u)}\mathrm{d}u\right)\mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}Z_X(s)\\
&= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} f(v)\mathrm{e}^{-\mathrm{i}2\pi sv}\mathrm{d}v\right)\mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}Z_X(s) \qquad v = t-u\\
&= \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi st}F(s)\mathrm{d}Z_X(s) && (7.86)\\
\mathrm{d}Z_Y(s) &= F(s)\mathrm{d}Z_X(s) && (7.87)
\end{aligned}
$$

The Fourier representation (7.86) of the output process (on condition that it converges as is shown below) implies (7.87), i.e., the orthogonal increments $\mathrm{d}Z_Y(s)$ pertaining to the output process are proportional to those pertaining to the input process. Hence, the first moment function of $\mathrm{d}Z_Y(s)$ obtained in (7.88) is identically zero.

$$
\mathrm{E}\big(\mathrm{d}Z_Y(s)\big) = \mathrm{E}\big(F(s)\mathrm{d}Z_X(s)\big) = F(s)\mathrm{E}\big(\mathrm{d}Z_X(s)\big) = 0 \qquad (7.88)
$$

From definitions (7.27) and (7.51,3), the second moment function of $\mathrm{d}Z_Y(s)$ is arrived at in (7.89),

$$
\begin{aligned}
&\mathrm{Cov}\big(\mathrm{d}Z_Y(r), \mathrm{d}Z_Y(s)\big) = \mathrm{Cov}\big(F(r)\mathrm{d}Z_X(r), F(s)\mathrm{d}Z_X(s)\big)\\
&= \overline{F(r)}F(s)\mathrm{E}\big(\overline{\mathrm{d}Z_X(r)}, \mathrm{d}Z_X(s)\big)\\
&= \begin{cases} \big(|F(s)|^2\big)\mathrm{E}\big(|\mathrm{d}Z_X|^2\big) = \big(|F(s)|^2\big)\mathrm{d}C_X^{(I)}(s) = \mathrm{d}C_Y^{(I)}(s) & \text{for } r = s\\ \overline{F(r)}F(s) \times 0 = 0 & \text{for } r \neq s. \end{cases} \qquad (7.89)
\end{aligned}
$$

a result which implies that (i) $Z_Y(s)$ has orthogonal increments $\mathrm{d}Z_Y(s)$, and (ii) $C_Y^{(I)}(s)$ is non-decreasing and mean-square continuous from the right because $\mathrm{d}C_Y^{(I)}(s) = |F(s)|^2\mathrm{d}C_X^{(I)}(s)$ and $|F(s)|^2 \geq 0$ for all s.

Hence, $Z_Y(s)$ inherits the properties required in (7.51) from $Z_X(s)$, the orthogonal increment process pertaining to the input $X(t)$ in the stochastic

filter, on condition that the Fourier-Stieltjes integral (7.86) exists for all s. This integral is shown to converge in the following paragraph.

From the property of the stochastic filter required in (7.84,3), it is concluded by applying (7.70,2) that the following integrals

$$c_Y(0) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi 0s} \mathrm{d}C_Y^{(I)}(s) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi 0s} |F(s)|^2 \mathrm{d}C_X^{(I)}(s) < \infty$$
$$c_Y(\tau) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi\tau s} \mathrm{d}C_Y^{(I)}(s) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi\tau s} |F(s)|^2 \mathrm{d}C_X^{(I)}(s) \leq c_Y(0)$$

exist, and consequently, $c_Y(\tau)$ and $C_Y^{(I)}(s)$ are a Fourier-Stieltjes transform pair as required in (7.61). To this pair belongs, as is deduced from (7.51), a stationary stochastic process $Y(t)$ such that $c_Y(\tau)$ is the covariance function of $Y(t)$ and $C_Y^{(I)}(s)$ is the spectrum of $Y(t)$. Thus, the Fourier-Stieltjes integral (7.86) converges for all s.

The results derived above are summarized in (7.90).

Let random functions $X(t)$ and $Y(t)$ be the input and output of a stochastic filter as defined in (7.84) having frequency response function $f(t)$. Then:

1. $\mathrm{d}Z_Y(s) = F(s)\mathrm{d}Z_X(s)$
2. $Y(t) = \int_{-\infty}^{\infty} \mathrm{e}^{\mathrm{i}2\pi st} F(s)\mathrm{d}Z_X(s)$
3. $\mathrm{d}C_Y^{(I)}(s) = |F(s)|^2 \mathrm{d}C_X^{(I)}(s)$ *and*
4. $C_Y(s) = |F(s)|^2 C_X(s)$, *on condition that $C_X(s)$ is continuous.*

(7.90)

If the input process $X(t)$ has an integer parameter t (the example in the remarks to (7.84) is a process with integer parameter) then the argument of the impulse response function is integer and the integral in (7.90,2) becomes $\int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st} F(s)\mathrm{d}Z_X(s)$.

$\mathrm{d}C_Y^{(I)}(s)(s) = |F(s)|^2 \mathrm{d}C_X^{(I)}(s)$ in (7.90,3) implies that the spectrum of the output process for frequency s depends only on the spectrum of the input process and the squared absolute value of the transfer function for this frequency (and not on any other frequencies). Thus, the effects of the filter on an oscillation of a given frequency s depend on s solely and, therefore, the frequency domain is often preferred when a stochastic LTI filter is described.

This is also true in the case of a deterministic LTI filter as defined in (6.120) and (6.121,1), because (6.121,2) is the deterministic version of proposition (7.90,1), i.e., that a convolution of two functions in the time domain corresponds to a multiplication of their Fourier transform in the frequency domain. Hence, the results obtained in Sect. 6.8.2, where the moving average and differences are analysed, also hold in the case of stochastic processes with integer parameter.

7.4.2 Examples

The stationary and invertible ARMA$[p, q]$ model (5.36) is, due to its MA$[\infty]$ representation (5.38), a weighted sum of a white noise process with the sequence of weights being in $L^2(-\infty, \infty)$. Hence, it is a linear process as defined in (2.25). From a comparison of (2.25) and (7.84) it is concluded that a linear process is a discrete-time stochastic filter: input and output processes have parameter $t = \ldots, -1, 0, 1, \ldots$ which implies that the definitions and propositions undergo minor re-adjustments, as mentioned in the remarks to (7.84) and (7.90).

The input in a linear process (shown above to be a stochastic filter) is a white noise process (W_t) in (2.10) or (2.11) with covariance function $c_W(\tau) = \sigma_W^2$ for $\tau = 0$, and $c_W(\tau) = 0$ for $\tau \neq 0$, implying spectrum $C_W(s) = \sigma_W^2$, $-1/2 \leq s \leq 1/2$, as obtained in (7.76). Hence, (W_t) has the properties required in (7.84,1). The weights of the white noise variables are required to be absolutely convergent in (2.25) and (5.38) and, for this reason, the impulse response sequence (f_t) of the linear process is in $L^2(-\infty, \infty)$, as specified in (7.84,2). Consequently, its frequency response function $F(s)$ as required in (6.49) is in $L^2(-1/2, 1/2)$ and is periodic with period 1 at the exterior of $-1/2 \leq s \leq 1/2$. Moreover, the Stieltjes integral in (7.84,3) becomes $\int_{-1/2}^{1/2} |F(s)|^2 \sigma_W^2 \mathrm{d}s < \infty$ and thus, a linear process has the desirable property (7.84,3).

Since a linear process is shown above to be a stochastic filter having all properties required in (7.84), its spectrum is obtained, using (7.90,4), from its weights and the variance of the input white noise process. This method is usually more straightforward than calculating the spectrum as the Fourier transform of the covariance function applying (7.66), as demonstrated by the following examples.

An MA[1] process (X_t), with $X_t = W_t - bW_{t-1}$ as defined in (2.21), is the convolution of impulse response sequence $(b_t) = (1, -b)$ for $t = 0, 1$ and $b_t = 0$ for $t = \ldots, -2, -1$ as well as $t = 2, 3, \ldots$ with a white noise process (W_t) as defined in (2.10) or (2.11). (X_t) has covariance function $c_X(-1) = -b\sigma_W^2$, $c_X(0) = (1 + b^2)\sigma_W^2$, $c_X(1) = -b\sigma_W^2$ and $c_X(\tau) = 0$, for $|\tau| > 1$, from which its spectrum is obtained in (7.91)

$$\begin{aligned} C_X(s) &= \sigma_W^2(1 + b^2) + \sigma_W^2(-2b\cos(2\pi s) \\ &= \sigma_W^2\big(1 + b^2 - 2b\cos(2\pi s)\big) \end{aligned} \tag{7.91}$$

using (7.66). For example, a time slice from a realisation of the MA[1] process (X_t) with $b = 0.9$ and $\sigma_W^2 = 1$ is plotted in Fig. 7.11 (a) together with its spectrum $C_X(s)$ (plots (b) and (c)). This spectrum has bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big) = 0.5$, calculated as required in (7.83). Its minimum is approximately 0.01, its maximum approximately 3.6, and therefore its dynamic range (6.96) is approximately 25 dB. Or, easier, its dynamic range is read from the plot (c) in Fig. 7.11.

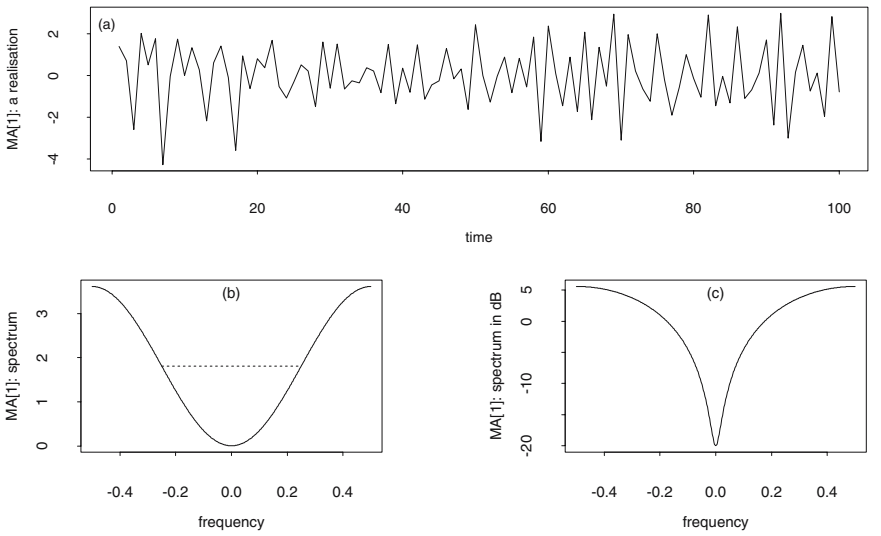


Fig. 7.11. MA[1] process with $b = 0.9$ and $\sigma_W^2 = 1$. Time slice of a realisation (above, (a)) and spectrum as obtained in (7.91) or (7.93) together with its bandwidth (below, (b)), and in logarithmic scale (below, (c)).

Alternatively, the spectrum of the above example MA[1] process is calculated using (7.90,4): the frequency response function is the Fourier $-i$-transform of the impulse response function, $\mathcal{F}_{-\mathrm{i}}(b_t) = 1\mathrm{e}^{-\mathrm{i}2\pi s0} + (-b)\mathrm{e}^{-\mathrm{i}2\pi s1} = 1 - b\mathrm{e}^{-\mathrm{i}2\pi s} = 1 - b\big(\cos(2\pi s) - b\mathrm{i}\sin(2\pi s)\big)$, and thereafter, its squared absolute value $|\mathcal{F}_{-\mathrm{i}}(b_t)|^2 = \mathcal{F}_{+\mathrm{i}}(b_t)\mathcal{F}_{-\mathrm{i}}(b_t)$ is multiplied with the spectrum of the white noise process to obtain, in (7.93),

$$
\begin{aligned}
C_X(s) &= \mathcal{F}_{+\mathrm{i}}(b_t)\mathcal{F}_{-\mathrm{i}}(b_t)C_W(s) && (7.92)\\
&= \big(1-b\cos(2\pi s)-b\mathrm{i}\sin(2\pi s)\big)\big(1-b\cos(2\pi s)+b\mathrm{i}\sin(2\pi s)\big)\sigma_W^2\\
&= \Big(1 - 2b\cos(2\pi s) + b^2\Big(\big(\cos(2\pi s)\big)^2 + \big(\sin(2\pi s)\big)^2\Big)\Big)\sigma_W^2\\
&= \big(1 + b^2 - 2b\cos(2\pi s)\big)\sigma_W^2 \quad -1/2 \le s \le 1/2 && (7.93)
\end{aligned}
$$

a result being identical to the one achieved in (7.91).

An MA[q] process (X_t) is obtained in (5.28) by convolving a white noise process with frequency response function $(b_t) = (1, -b_1, \ldots, -b_q)$ for $t = 0, 1, 2, \ldots, q$, and $b_t = 0$ for $t = \ldots, -2, -1$ as well as $t = q+1, q+2, \ldots$: $X_t = \sum_{u=0}^{q} b_u W_{t-u} = (b_t) * (W_t)$. Applying (7.90,4) or else (6.48,6), $C_X(s)$ is then obtained as the product of $C_W(s) = \sigma_W^2$, $-1/2 \le s \le 1/2$, and the squared absolute value of the transfer function:

$$\mathcal{F}_{-\mathrm{i}}(b_t) = 1\mathrm{e}^{-\mathrm{i}2\pi s0} + (-b_1)\mathrm{e}^{-\mathrm{i}2\pi s1} + (-b_2)\mathrm{e}^{-\mathrm{i}2\pi s2} +$$

$$\ldots + (-b_q)\mathrm{e}^{-\mathrm{i}2\pi sq} \tag{7.94}$$

$$C_X(s) = \mathcal{F}_{+\mathrm{i}}(b_t)\mathcal{F}_{-\mathrm{i}}(b_t)C_W(s) \quad -1/2 \le s \le 1/2 \tag{7.95}$$

For example, the spectrum of the MA[2] process with $b_1 = 0.9$ and $b_2 = -2.0$ is plotted in Fig. 7.10 (a).

An AR[p] process $X_t = a_1X_{t-1} + \ldots + a_pX_{t-p} + W_t$, $t = \ldots, -1, 0, 1, \ldots$ and (W_t) a white noise process, is stationary as required in (5.2) on condition that its coefficients $(a_t) = (1, -a_1, \ldots, -a_p)$ for $t = 0, 1, \ldots, p$ and $a_t = 0$ for $t = \ldots, -2, -1$ as well as $t = p+1, p+2, \ldots$, are invertible. Using convolution sums, a stationary AR[p] process becomes $(W_t) = (a_t) * (X_t)$ or, in its MA[∞] representation, $(X_t) = (a_t)^{-1} * (W_t)$. Since $(a_t)^{-1} * (a_t) = (e_t)$, the identity sequence as defined in (2.41) and applying (6.48,6), $\mathcal{F}_{-\mathrm{i}}\big((a_t)^{-1}\big)\mathcal{F}_{+\mathrm{i}}(a_t) = 1 = \mathcal{F}_{-\mathrm{i}}(e_t)$ is obtained. Hence, $\mathcal{F}_{-\mathrm{i}}\big((a_t)^{-1}\big) = 1/\mathcal{F}_{+\mathrm{i}}(a_t)$ is the frequency response function pertaining to the impulse response sequence $(a_t)^{-1}$.

An AR[1] process has coefficients $(a_t) = (1, -a)$ for $t = 0, 1$ and $(a_t) = 0$ for $t = \ldots, -2, -1$ as well as $t = 2, 3, \ldots$, with Fourier transform $\mathcal{F}_{-\mathrm{i}}(a_t) = 1 - a\mathrm{e}^{-\mathrm{i}2\pi s}$ and, therefore, the Fourier transform of the inverse is $\mathcal{F}_{-\mathrm{i}}\big((a_t)^{-1}\big) = 1/(1 - a\mathrm{e}^{\mathrm{i}2\pi s})$. Applying (7.90,4), the spectrum of the AR[1] process is arrived at in (7.96).

$$\begin{aligned} C_X(s) &= \frac{1}{(1 - a\mathrm{e}^{\mathrm{i}2\pi s})(1 - a\mathrm{e}^{-\mathrm{i}2\pi s})}\, C_W(s) \qquad -1/2 \le s \le 1/2 \\ &= \frac{1}{1 + a^2 - 2a\cos(2\pi s)}\, \sigma_W^2 \end{aligned} \tag{7.96}$$

This spectrum in (7.96) is identical to the one obtained in (7.79) as Fourier transform of the covariance function pertaining to the AR[1] model. From (7.96) it is obvious that the power of the process is associated predominantly with low frequency oscillations for $0 < a < 1$, whereas, for $-1 < a < 0$, the higher frequencies dominate in the spectrum. Thus the spectrum of an AR[1] process is very easily obtained applying (7.90,4). In the same straightforward way, the spectrum of an AR[2] process is obtained in (7.97), bearing in mind that $-1/2 \le s \le 1/2$:

$$\begin{aligned} C_X(s) &= \frac{1}{(1 - a_1\mathrm{e}^{\mathrm{i}2\pi s} - a_2\mathrm{e}^{\mathrm{i}4\pi s})(1 - a_1\mathrm{e}^{-\mathrm{i}2\pi s} - a_2\mathrm{e}^{-\mathrm{i}4\pi s})}\, C_W(s) \\ &= \frac{1}{1 + {a_1}^2 + {a_2}^2 + 2(a_1a_2 - a_1)\cos(2\pi s) - 2a_2\cos(4\pi s)}\, \sigma_W^2. \end{aligned} \tag{7.97}$$

For example, time slices from realisations of AR[2] processes $(Y_t^{(1)})$ and $(Y_t^{(2)})$ are plotted above in Figs. 7.12 and 7.13. In both example processes, the innovations are a white noise process (W_t) with $\mu_X = 0$ and $\sigma_W^2 = 1$. From the realisations, as plotted, it becomes obvious that $(Y_t^{(1)})$ with $a_1 = 0.3$ and $a_2 = -0.7$ has oscillations with lower frequencies and smaller amplitudes than process $(Y_t^{(2)})$ with $a_1 = -1.6$ and $a_2 = -0.9$. Quantitatively, this

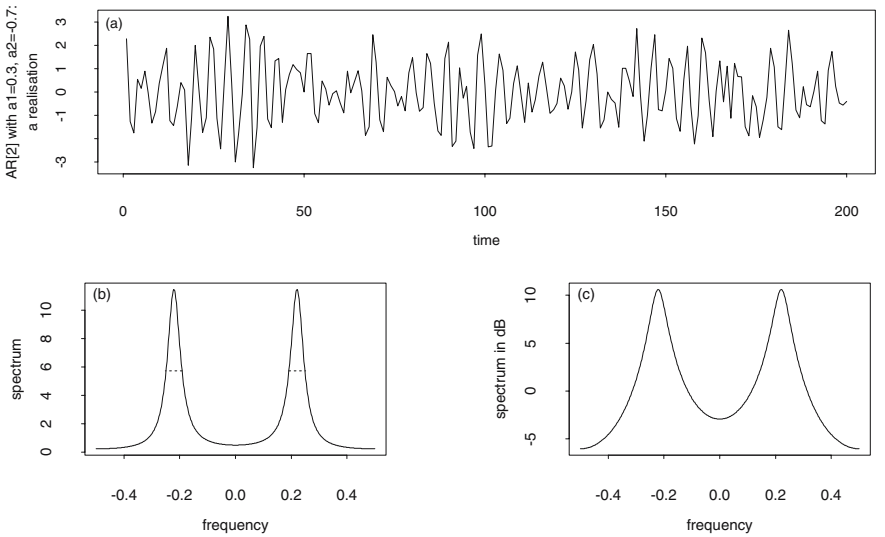


Fig. 7.12. AR[2] model $Y_t^{(1)} = 0.3Y_{t-1}^{(1)} - 0.7Y_{t-2}^{(1)} + W_t$. Time slice of a realisation (above) and spectrum as obtained in (7.97) together with its bandwidth (below, (b)) and in logarithmic scale (below, (c)).

discrepancy is seen in plots (b) and (c), in both Figs. 7.12 and 7.13: $C_{Y^{(1)}}(s)$ has a peak approximately 11.5 high located in frequency 0.2, whereas the peak in $C_{Y^{(2)}}(s)$ attains an approximative height 350 for frequency 0.4, the two frequencies also being approximations. Both spectra are also distinct in their dynamic ranges (6.96) being read directly in the logarithmic plots as well in their bandwidths (7.83) drawn in the non-logarithmic plots.

Example process no. 3 is the AR[4] model (X_t), with $X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9838X_{t-4} + W_t$ and (W_t) being a white noise process having expectation $\mu_W = 0$ and variance $\sigma_W^2 = 1$, as introduced in [108]. A realisation of this process is plotted above in Fig. 7.14, with its spectrum as plotted in Figs. 7.10 (c) and 7.14 (b), obtained by applying (7.90,4) in the same way as the spectra of the AR[1] and AR[2] models are obtained above in (7.96) and (7.97). This procedure is implemented in R with the following expressions:

```
a <- c(2.7607, -3.8200, 2.6535, -0.9238)   #coefficients
s1 <- (0:4000)/8000 #4001 frequencies between 0 and 0.5
# -1i*2*pi*s1 is a complex vector having the same length as s1
sw2 <- rep(1,4001)  #4001 variances of the white noise process
spec <- sw2/( (1 - a[1]*exp(-1i*2*pi*s1) - a[2]*exp(-1i*4*pi*s1)
       - a[3]*exp(-1i*6*pi*s1) - a[4]*exp(-1i*8*pi*s1) ) *
              (1 - a[1]*exp( 1i*2*pi*s1) - a[2]*exp( 1i*4*pi*s1)
       - a[3]*exp( 1i*6*pi*s1) - a[4]*exp( 1i*8*pi*s1)   )  )
```

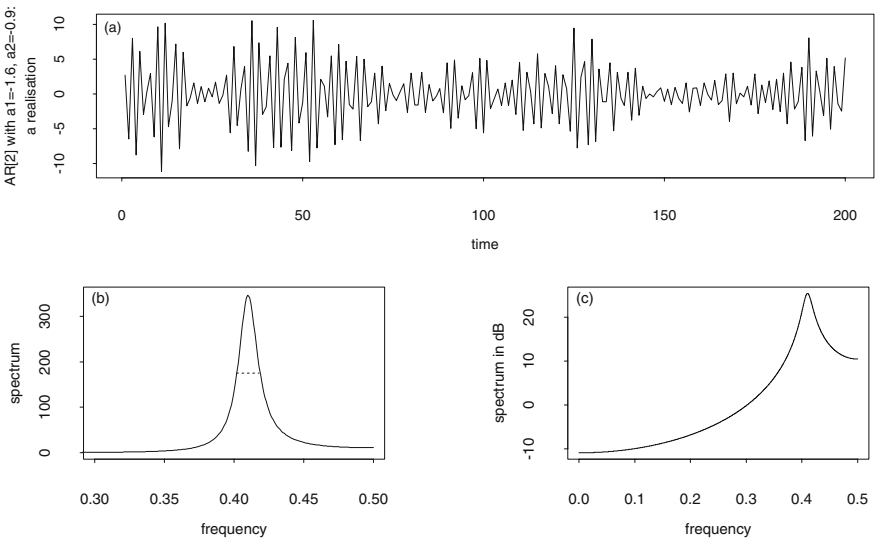


Fig. 7.13. AR[2] model $Y_t^{(2)} = -1.6Y_{t-1}^{(2)} - 0.9Y_{t-2}^{(2)} + W_t$. Time slice of a realisation (above) and spectrum as obtained in (7.97) together with its bandwidth (below, (b)) and in logarithmic scale (below, (c)).

```
#alternative: if the order of the process is larger and
#if the coefficients are in vector a, then:
#i <- 1
#q <- as.complex(rep(1+0i,4001))        #accumulator
#for(i in 1:length(a)) {                #adding
# q <- q -(a[i]*exp(-1i*(2*i)*pi*s1))
#}
#spec <- sw2/ ((Mod(q))*(Mod(q)))       #spectrum
#alternative end here
#in vector spec is now the spectrum non-logarithmic
specdb <- 10*log10(spec)      #in dB
```

With minor changes in R vectors `s1` and `sw2` this spectrum is obtained for 5001 frequencies in the interval $0.10 \le s \le 0.15$ in Figs. 7.10 (d) and 7.14 (c). These plots show that the spectrum is continuous, as required in (7.90,4). The widths at half height of its peaks and valleys are obtained in Fig. 7.10 (d) as required in (6.97); from these, its bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big) = 0.00474$ is obtained applying (7.83) and (6.97) as width at half height of its right peak. Its dynamic range is approximately 65 dB, as read in Fig. 7.14 (b).

Having a small bandwidth and a large dynamic range, this AR[4] model is able to generate instructive example realisations when the bias of the periodogram is discussed in Sect. 8.3 and also when estimators for a continuous spectrum are introduced in Chap. 9.

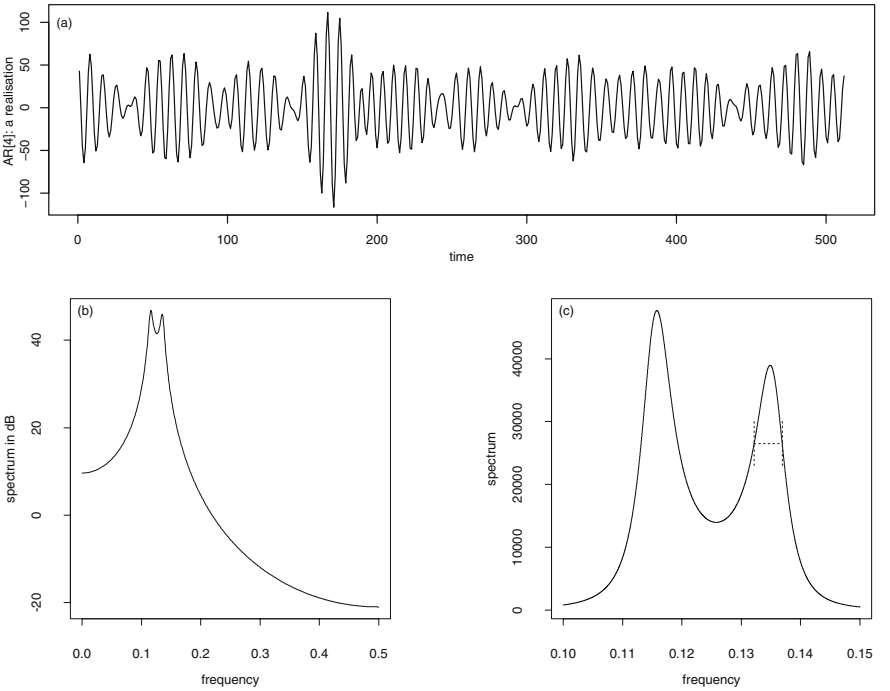


Fig. 7.14. AR[4] model $X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9838X_{t-4} + W_t$, with $\mu_W = 0$ and $\sigma_W^2 = 1$. A realisation in the interval $0 \leq 6 \leq 511$ (above) and spectrum of the process in logarithmic (b) and non-logarithmic scale together with its bandwidth (c).

7.5 Supplements

The gamma distribution is introduced by evaluating two-fold integrals of the exponential density in Sect. 7.5.1, and in Sect. 7.5.2, stochastic Stieltjes integrals are evaluated with respect to an orthogonal increment process. In Sect. 7.5.3, the harmonic process is shown to be stationary by deriving its moment functions. In Sect. 7.5.4, it is shown that the spectrum of a discrete-time stochastic process is the expectation in the limit for $N \to \infty$ of the empirical covariance function obtained from an observed realisation. Finally, in Sect. 7.5.5, an answer is given to the question: which of the two functions, the covariance function or the spectrum, is preferably used to describe the second moment function of a stochastic process?

7.5.1 Gamma Distribution

Y_k in $\Pr(Y_k \leq t)$ and $\Pr(Y_{k+1} \leq t)$ are obtained in (7.8) and (7.9) as the sum of mutually independent random variables $X_1, X_2, \ldots, X_i$ having an

identical exponential distribution (7.4). The probability distributions $F_2(y)$ pertaining to $Y_2 = X_1 + X_2$, $F_3(y)$ pertaining to $Y_3 = X_1 + X_2 + X_3$, ..., and $F_k(y)$ pertaining to $Y_k = X_1 + \ldots + X_k$, are arrived at, as required in (1.5), by evaluating a two-fold, three-fold, ..., k-fold integral of the exponential density. Such a two-fold integral, however only in the case of two independent and normally distributed random variables, is given in (1.18) with the region of integration to be sketched in Problem 1.8. This integral becomes in (1.20) a convolution integral as defined in (2.28,1). A two-fold integral similar to the one in (1.18) is also obtained in the first line of the following derivation (the densities are now exponential and x_1, x_2 are substituted with t, u).

$$
\begin{aligned}
F_2(y) &= \iint\limits_{t+u\le y} \lambda \mathrm{e}^{-\lambda t}\lambda \mathrm{e}^{-\lambda u}\mathrm{d}t\mathrm{d}u = \int_0^y \int_0^{y-t} \lambda \mathrm{e}^{-\lambda t}\lambda \mathrm{e}^{-\lambda u}\mathrm{d}t\mathrm{d}u \\
&= \int_0^y \lambda \mathrm{e}^{-\lambda t}\left(1 - \mathrm{e}^{-\lambda(y-t)}\right)\mathrm{d}t \;=\; \ldots \\
&= 1 - \mathrm{e}^{-\lambda y}(1+\lambda y) \text{ implying } f_2(y) = \lambda^2 \mathrm{e}^{-\lambda y} y \qquad (7.98) \\
F_3(y) &= \iint\limits_{t+u\le y} \lambda \mathrm{e}^{-\lambda t}\lambda^2 \lambda \mathrm{e}^{-\lambda u} u\mathrm{d}t\mathrm{d}u = \int_0^y \int_0^{y-t} \lambda \mathrm{e}^{-\lambda t}\lambda^2 \lambda \mathrm{e}^{-\lambda u} u\mathrm{d}t\mathrm{d}u \\
&= \int_0^y \lambda \mathrm{e}^{-\lambda t}\left(1 - \mathrm{e}^{-\lambda(y-t)}\big(1+\lambda(y-t)\big)\right)\mathrm{d}t \;=\; \ldots \\
&= 1 - \mathrm{e}^{-\lambda y}\left(1 + \lambda y + \frac{1}{2}(\lambda y)^2\right) \text{ implying } f_3(y) = \lambda^3 \mathrm{e}^{-\lambda y}\frac{1}{2}y^2 \qquad (7.99)
\end{aligned}
$$

Continuing the above derivations, the *gamma distributions* in (7.100) and (7.101) are arrived at, the name originating from the gamma function $\Gamma_n(x) = \int_0^\infty x^{n-1}\mathrm{e}^{-x}\mathrm{d}x$.

$$
F_k(y) = 1 - \mathrm{e}^{-\lambda y}\left(1 + \frac{\lambda y}{1!} + \ldots + \frac{(\lambda y)^{k-1}}{(k-1)!}\right) \qquad (7.100)
$$

$$
F_{k+1}(y) = 1 - \mathrm{e}^{-\lambda y}\left(1 + \frac{\lambda y}{1!} + \ldots + \frac{(\lambda y)^{k-1}}{(k-1)!} + \frac{(\lambda y)^k}{k!}\right) \qquad (7.101)
$$

7.5.2 Integrating with Respect to an Orthogonal Increment Process

As a supplement to definition (7.23) of the orthogonal increment process $A(t)$, a formalism is introduced which allows for obtaining the covariance function of $A(t)$ (which is, in (7.23), part of the definition) by formal integration.

In the first step, $A(t)$ is alternatively defined in (7.102) by means of a Stieltjes-integral where the infinitesimal increments $\mathrm{d}A(t)$ of the integrator $A(t)$ are required to be orthogonal (as are the differences in (7.23)) and, additionally, stationary ("homogeneous"). The covariance of the increments

$\mathrm{d}A(t)$ in (7.102), i.e., $\mathrm{Cov}\big(\mathrm{d}A(t), \mathrm{d}A(u)\big)$, is arrived at in the limiting case for $\Delta t \to 0$ in (7.26) on condition that the infinitesimal increments $\mathrm{d}A(t)$ and $\mathrm{d}A(u)$, multiplied with $\mathrm{d}t$ and $\mathrm{d}u$, are substituted for $Y(t)$ and $Y(u)$, i.e., the difference quotients of the orthogonal increment process $A(t)$ as defined in (7.23).

Let $A(t)$ be a right-continuous random function with increments $\mathrm{d}A(t) = A(t) - A(t - \mathrm{d}t)$ and $\mathrm{d}A(u) = A(u) - A(u - \mathrm{d}u)$ such that $\mathrm{Cov}\big(\mathrm{d}A(t), \mathrm{d}A(u)\big) = \sigma^2_{\mathrm{d}A}\delta(t-u)\mathrm{d}t\mathrm{d}u$, and $\mathrm{E}\big(\mathrm{d}A(t)\big) = \mu_{\mathrm{d}A}$.
Then $A(t) = \int_0^t \mathrm{d}A(u)$ or $A(t_2) - A(t_1) = \int_{t_1}^{t_2} \mathrm{d}A(u)$ is called a process with orthogonal and stationary increments. (7.102)

In the second step, applying (1.15), (1.16) and (6.67) as well as the region of integration on the left in Fig. 7.6,

$$\begin{aligned}
\mathrm{Cov}\big(A(t), A(u)\big) &= \mathrm{Cov}\left(\int_0^t \mathrm{d}A(x), \int_0^u \mathrm{d}A(y)\right) \\
&= \int_0^t \int_0^u \mathrm{Cov}\big(\mathrm{d}A(x), \mathrm{d}A(y)\big) \\
&= \sigma^2_{\mathrm{d}A} \int_0^t \int_0^u \delta(x-y)\mathrm{d}x\mathrm{d}y \\
&= \sigma^2_{\mathrm{d}A} \int_0^{\min(t,u)} \mathrm{d}x = \sigma^2_{\mathrm{d}A}\min(t,u)
\end{aligned} \tag{7.103}$$

the result in (7.103) (being identical to the covariance in (7.24)) is arrived at. This is the first example for the evaluation of a stochastic Stieltjes integral as defined in (7.21) with respect to an orthogonal increment process.

As a second example, the covariance function of the continuous-time AR[1] process is calculated from its definition in (7.104) which is obtained by substituting a stochastic differential equation for the difference equation in (2.13): $X(t)$ and $\dot{X}(t)$ for X_t and X_{t-1}, as well as $\mathrm{d}A(t)$ (as defined in (7.102)) for the innovations W_t.

Let $\mathrm{d}A(t)$ be orthogonal and stationary increments (7.102). Then $X(t)$, with $X(t)$ being the stationary solution of the differential equation $X(t) - a\dot{X}(t) = \mathrm{d}A(t)$, is called a continuous-time AR[1] process (or AR[1] process with real parameter). (7.104)

For the first order linear differential equation $x(t) + a\dot{x}(t) = w(t)$, with $x(t)$ and $w(t)$ being deterministic functions, the stationary solution in (6.119) is obtained in Problem 2.11: the convolution integral in (6.119) is a LTI filter as defined in (6.120). As in the deterministic case, the solution in (7.105) is the sum of (i) the general solution of $X(t) - a\dot{X}(t) = 0$ and (ii) a particular solution of $X(t) - a\dot{X}(t) = \mathrm{d}A(t)$, i.e., the stochastic differential equation defining the continuous-time AR[1] process.

$$X(t) = B\mathrm{e}^{-(1/a)t} + \int_{-\infty}^{t} g(t-u)\mathrm{d}A(u) \qquad \text{with} \tag{7.105}$$

$$g(t) = \begin{cases} 0 & \text{for } t < 0 \\ (1/a)\mathrm{e}^{-(1/a)t} & \text{for } t \geq 0 \end{cases}$$

$$X(t) = \int_{0}^{t} (1/a)\mathrm{e}^{-(1/a)(t-u)}\mathrm{d}A(u), \quad \text{for } t \to \infty \text{ and } a > 0 \tag{7.106}$$

For $a > 0$, the general solution becomes identically zero on condition that $t \to \infty$ and thus $X(t)$ becomes stationary, since in the Stieltjes (convolution) integral (7.106), the increments $(1/a)\mathrm{e}^{-(1/a)(t-u)}\mathrm{d}A(u)$ are weighted stationary $\mathrm{d}A(u)$ and the weight function decays exponentially. In practical applications therefore, the increments going far back can be neglected on condition that a long time has elapsed since the initialisation of the process. The very same rationale is used in the remarks to (2.16) and (2.17) to derive the moment functions of the discrete-time AR[1] model, and also in the remarks to (2.27).

The covariance function in (7.107) of the stationary continuous-time AR[1] process is evaluated as follows. (7.108) is obtained applying (1.15,10), (7.109) follows since $\mathrm{Cov}\big(\mathrm{d}A(x), \mathrm{d}A(y)\big) = \sigma^2_{\mathrm{d}A}\delta(x-y)\mathrm{d}x\mathrm{d}y$ by definition, and, integrating as above in the derivation of (7.103), (7.110) is arrived at provided that $\tau > 0$.

$$c_X(\tau) = \mathrm{Cov}\left(\int_0^t \frac{1}{a}\mathrm{e}^{-(1/a)(t-x)}\mathrm{d}A(x), \int_0^t \frac{1}{a}\mathrm{e}^{-(1/a)(t+\tau-y)}\mathrm{d}A(y)\right) \tag{7.107}$$

$$= \frac{1}{a^2}\int_0^t \int_0^{t+\tau} \mathrm{e}^{-(1/a)(t-x)}\mathrm{e}^{-(1/a)(t-y)}\mathrm{Cov}\big(\mathrm{d}A(x), \mathrm{d}A(y)\big) \tag{7.108}$$

$$= \sigma^2_{\mathrm{d}A}\frac{1}{a^2}\mathrm{e}^{-(1/a)\tau}\int_0^t \int_0^{t+\tau} \mathrm{e}^{-(1/a)(t-x)}\mathrm{e}^{-(1/a)(t-y)}\delta(x-y)\mathrm{d}x\mathrm{d}y \tag{7.109}$$

$$= \sigma^2_{\mathrm{d}A}\frac{1}{a^2}\mathrm{e}^{-(1/a)\tau}\int_0^t \mathrm{e}^{-(1/a)(t-x)}\mathrm{e}^{-(1/a)(t-x)}\mathrm{d}x \qquad \tau > 0 \tag{7.110}$$

$$= \sigma^2_{\mathrm{d}A}\frac{1}{a^2}\mathrm{e}^{-(1/a)\tau}\int_0^t \mathrm{e}^{-(2/a)(t-x)}\mathrm{d}x \tag{7.111}$$

$$= \sigma^2_{\mathrm{d}A}\frac{1}{a^2}\mathrm{e}^{-(1/a)\tau}\frac{a}{2}\Big(1 - \mathrm{e}^{-(2/a)t}\Big) \tag{7.112}$$

$$= \sigma^2_{\mathrm{d}A}\frac{1}{2a}\mathrm{e}^{-(1/a)|\tau|} = \mathrm{Cov}\big(X(t), X(t+\tau)\big) \qquad t \to \infty \tag{7.113}$$

$$\sigma^2_X = c_X(0) = \frac{1}{2a}\sigma^2_{\mathrm{d}A} \qquad\qquad r_X(\tau) = \mathrm{e}^{-(1/a)|\tau|} \tag{7.114}$$

(7.112) follows, using the substitution $t - x = u$, and the covariance function $c_X(\tau)$ in (7.113) is arrived at since (i) the continuous-time AR[1] process is stationary for $t \to \infty$, and (ii) the covariance function of a stationary process is an even function as defined in (6.32,1) due to its properties (2.7).

The covariance together with the correlation functions $c_X(\tau)$ and $r_X(\tau)$, τ real, of the continuous-time AR[1] $X(t)$ process are continuous in $\tau = 0$ and for all τ. This is in line with the remarks to (7.18) where it is required that the second moment functions of a mean-square continuous random function be continuous in the usual sense. In $\tau = 0$, however, the derivative from the left of the covariance function is not identical with the derivative from the right and, consequently, $\dot{c}_X(0)$ and $\dot{r}_X(0)$ do not exist. From this result and the remarks to (7.18) it is concluded that $X(t)$ is not mean-square differentiable.

For example, *Langevin's equation* $m(\mathrm{d}\boldsymbol{v}(t)/\mathrm{d}t) + \beta\boldsymbol{v}(t) = \boldsymbol{F}(t)$ is a continuous-time AR[1] process (7.104) describing the random motion of a small particle (with diameter $\approx 1\ \mu$m) in a fluid *(Brownian motion),* cf. [104] or [81]). In Langevin's equation, m is the mass of a particle, $\boldsymbol{v}(t)$ its velocity, i.e., the derivative $(\mathrm{d}\boldsymbol{x}/\mathrm{d}t)$ of its position $\boldsymbol{x}(t)$ in a rectangular coordinate system, and $\boldsymbol{F}(t)$ a randomly fluctuating collision force due to the bombardment of the particle by molecules. On the molecular scale, $\boldsymbol{F}(t)$ is a very complicated term, on the scale of the particle size however, $\boldsymbol{F}(t)$ is assumed to be the stationary derivative of an orthogonal increment process $A(t)$, i.e., the difference quotients of $A(t)$ as obtained in (7.25) in the limiting case for $\tau = 0$.

7.5.3 Moments of the Harmonic Process

The moment functions of the harmonic process (7.30) are obtained by direct integration since $\mathrm{E}V = \int_{-\infty}^{\infty} g(u)\mathrm{d}F(u)$ on condition that $V = g(U)$ and $\mathrm{E}U = \int_{-\infty}^{\infty} u\mathrm{d}F(u)$, with U and V being random variables and $F(u)$ the probability distribution of U. In (7.30), random variables P_j have a uniform distribution $f(p_j) = 1/(2\pi)$ for $-\pi \le p_j \le \pi$ and $f(p_j) = 0$ at the exterior of this interval. $f(p_j)$ is continuous and thus, by evaluating the Riemann integral in (7.115), the first moment of a random variable X_t in the harmonic process is obtained.

$$\mathrm{E}X_t = \sum_{j=1}^{n} R_j \int_{-\pi}^{\pi} \cos(2\pi s_j t + p_j)\frac{1}{2\pi}\mathrm{d}p_j \tag{7.115}$$

$$= \sum_{j=1}^{n} \frac{1}{2\pi} R_j \Big[\sin(2\pi s_j t + p_j)\Big]_{p_j=-\pi}^{p_j=\pi} = \sum_{j=1}^{n} \frac{1}{2\pi} R_j \times 0 = 0 \tag{7.116}$$

This result holds for all t, and thus it is concluded from (7.116) that the expectation function of the harmonic process as defined in (7.30) is identically zero.

Applying (1.15,10), the second moment function of the harmonic process in (7.117) becomes the two-fold sum in (7.118).

$$c_X(t,u) = \mathrm{Cov}\big(X_t, X_{t+\tau}\big) \tag{7.117}$$

$$= \sum_{j=1}^{n} \sum_{k=1}^{n} R_j R_k \mathrm{E}\Big(\cos(2\pi s_j t + P_j)\cos\big(2\pi s_k(t+\tau) + P_k\big)\Big) \qquad (7.118)$$

When the terms in this sum are arranged in a matrix, then, obviously, the expectations in the subdiagonals become identically zero because the random phases P_j in (7.30) are independent, i.e., for $j \neq k$, and

$$\begin{aligned} &\mathrm{E}\Big(\cos(2\pi s_j t + P_j)\cos\big(2\pi s_k(t+\tau) + P_k\big)\Big) \\ &= \mathrm{E}\big(\cos(2\pi s_j t + P_j)\big)\mathrm{E}\Big(\cos\big(2\pi s_k(t+\tau) + P_k\big)\Big) = 0 \end{aligned}$$

is obtained. The expectations in the main diagonal are calculated using a Riemann integral, i.e., for $j = k$, and

$$\begin{aligned} &\mathrm{E}\Big(\cos(2\pi s_j t + P_j)\cos\big(2\pi s_j(t+\tau) + P_j\big)\Big) \\ &= \mathrm{E}\left(\frac{1}{2}\Big(\cos\big(2\pi s_j t + P_j + 2\pi s_j(t+\tau) + P_j\big) + \cos(2\pi s_j \tau)\Big)\right) \\ &= \frac{1}{4\pi}\int_{-\pi}^{\pi}\Big(\cos(4\pi s_j t + 2\pi s_j \tau + 2p_j) + \cos(2\pi s_j \tau)\Big)\mathrm{d}p_j \\ &= 0 + \frac{1}{4\pi}\cos(2\pi s_j \tau)\int_{-\pi}^{\pi}\mathrm{d}p_j = (1/2)\cos(2\pi s_j \tau) \end{aligned}$$

is arrived at. From these results, the covariance function of the harmonic process follows in (7.119):

$$c_X(t,u) = c_X(\tau) = \sum_{j=1}^{n}\big(R_j^2/2\big)\cos(2\pi s_j \tau). \qquad (7.119)$$

Consequently, the harmonic process (7.30) is stationary, since its expectation function (7.116) is constant and its covariance function (7.119) depends solely on the lag.

The expectations of the complex-valued random variables D_j in (7.30) are calculated using Riemann integrals (as already demonstrated in the derivations of (7.116) and (7.119)). Since $\mathrm{e}^{\mathrm{i}p_j} = \cos(p_j) + \mathrm{i}\sin(p_j)$ as in (6.2,6),

$$\begin{aligned} \mathrm{E}D_j &= \mathrm{E}\big((R_j/2)\mathrm{e}^{\mathrm{i}P_j}\big), \qquad j = 1,\ldots,n \\ &= \frac{R_j}{2}\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}p_j}\frac{1}{2\pi}\mathrm{d}p_j = \frac{R_j}{4\mathrm{i}\pi}\Big[\mathrm{e}^{\mathrm{i}p_j}\Big]_{-\pi}^{\pi} = \frac{R_j}{4\mathrm{i}\pi}[-1-(-1)] = 0 \qquad (7.120) \\ \mathrm{E}D_{-j} &= 0, \qquad j = 1,\ldots,n \end{aligned}$$

the expectations of the D_j are identically zero.

The variances of the D_j are obtained using (1.15,1), (7.27) as well as $|\mathrm{e}^{\mathrm{i}P_j}|^2 = \big(\cos(P_j)\big)^2 + \big(\sin(P_j)\big)^2 = 1$:

$$\begin{aligned}
\mathrm{Var}D_j &= \mathrm{Var}\big((R_j/2)\mathrm{e}^{\mathrm{i}P_j}\big) \qquad j = 1, \ldots, n \\
&= \frac{R_j^2}{4}\mathrm{Var}\big(\mathrm{e}^{\mathrm{i}P_j}\big) = \frac{R_j^2}{4}\mathrm{E}\big(|\mathrm{e}^{\mathrm{i}P_j}|^2\big) = \frac{R_j^2}{4} \qquad (7.121) \\
\mathrm{Var}D_{-j} &= \mathrm{Var}\big((R_j/2)\mathrm{e}^{-\mathrm{i}P_j}\big) = \frac{R_j^2}{4}, \qquad j = 1, \ldots, n
\end{aligned}$$

The covariances $\mathrm{Cov}\big(D_j, D_k\big)$ of the D_j in (7.30) are also obtained by taking advantage of the favourable properties of the P_j. Since P_j and P_k are independent for $j \neq k$, (7.122) and (7.123) are obtained. Since, however, D_j and D_{-j} are not independent, $\mathrm{Cov}(D_j, D_{-j}) = 0$ in (7.124) is a result of the covariances as defined in (7.27), subsequent to substituting $\mathrm{E}D_j = \mathrm{E}D_{-j} = 0$.

$$\begin{aligned}
\mathrm{Cov}\big(D_j, D_k\big) &= 0, \qquad j, k = 1, \ldots, n,\ j \neq k, \qquad (7.122) \\
\mathrm{Cov}\big(D_j, D_k\big) &= 0, \qquad j, k = -n, \ldots, -1,\ j \neq k, \qquad (7.123) \\
\mathrm{Cov}\big(D_j, D_{-j}\big) &= \mathrm{E}\big(\overline{D_j}D_{-j}\big) = \mathrm{E}\big(\overline{(R_j/2)\mathrm{e}^{\mathrm{i}P_j}}(R_j/2)\mathrm{e}^{-\mathrm{i}P_j}\big) \qquad (7.124) \\
&= \frac{R_j^2}{4}\mathrm{E}\big(\mathrm{e}^{-\mathrm{i}P_j}\mathrm{e}^{-\mathrm{i}P_j}\big) = \frac{R_j^2}{4}\int_{-\pi}^{\pi}\mathrm{e}^{-\mathrm{i}2p_j}\frac{1}{2\pi}\mathrm{d}p_j \\
&= \frac{R_j^2}{4} \times \left(\frac{-1}{4\mathrm{i}\pi}\right)\Big[\mathrm{e}^{-2\mathrm{i}p_j}\Big]_{-\pi}^{\pi} \\
&= \frac{R_j^2}{4} \times \left(\frac{-1}{4\mathrm{i}\pi}\right)\Big(\mathrm{e}^{-2\mathrm{i}\pi} - \mathrm{e}^{+2\mathrm{i}\pi}\Big) \\
&= \frac{R_j^2}{4} \times \left(\frac{1}{2\pi}\right)\sin(2\pi) = 0 \qquad (7.125)
\end{aligned}$$

7.5.4 Spectrum of a Discrete-time Stochastic Process

If the covariance function $c_X(\tau)$ of a stationary stochastic process (X_t), with $t = \ldots, -1, 0, 1, \ldots$ and $\mu_X = 0$, converges absolutely as required in (2.24), then (i) the empirical covariance function $\hat{c}_X(\tau)$, computed applying (2.1) from a realisation $(x_t) = (x_0, x_1, \ldots, x_{N-1})$ of (X_t), has properties (2.58) and (ii) $c_X(\tau)$ is in $L^2(-\infty, \infty)$. Thus, $c_X(\tau)$ and $C_X(s)$ in (7.65) and (7.66) are a Fourier transform pair as defined in (6.49) and (6.50). In contrast to realisation $x_T(t)$ in (7.53), (i) the time is discrete, i.e., $t = \ldots, -1, 0, 1, \ldots$, and (ii) the observations are confined to the interval $0 \leq t \leq N-1$. Assuming that (x_t) becomes identically zero at the exterior of the observational interval, the Fourier transform $g_N(s)$ of (x_t) is obtained as in (7.126):

$$g_N(s) = \sum_{t=-\infty}^{\infty} x_t \mathrm{e}^{-\mathrm{i}2\pi st} = \sum_{t=0}^{N-1} x_t \mathrm{e}^{-\mathrm{i}2\pi st} \qquad (7.126)$$

(x_t) and $g_N(s)$ are a pair of Fourier transforms as defined in (6.49) and (6.50), implying that $g_N(s)$ is in $L^2(-1/2, 1/2)$ and periodic with period 1.

The spectrum of (x_t) is calculated as required in (6.76) and demonstrated in Sect. 6.1, i.e., $|g_N(s)|^2 = \overline{g_N(s)} g_N(s)$ in the case above, or alternatively, as mentioned in the remarks to (6.103) and demonstrated in Fig. 6.34, by Fourier transforming the autocorrelation of a time series as stipulated in (7.127).

$$|g_N(s)|^2 = \sum_{t=0}^{N-1} \big((x_t) \star (x_t)\big) \mathrm{e}^{-\mathrm{i}2\pi st} \tag{7.127}$$

$$= \sum_{\tau=-(N-1)}^{N-1} \left(\sum_{t=0}^{N-1-|\tau|} \big((x_t)(x_{t+|\tau|})\big) \right) \mathrm{e}^{-\mathrm{i}2\pi s\tau} \tag{7.128}$$

$$= N \sum_{\tau=-(N-1)}^{N-1} \hat{c}_X(\tau) \mathrm{e}^{-\mathrm{i}2\pi s\tau} \tag{7.129}$$

$$\frac{|g_N(s)|^2}{N} = \sum_{\tau=-(N-1)}^{N-1} \hat{c}_X(\tau) \mathrm{e}^{-\mathrm{i}2\pi s\tau} \tag{7.130}$$

(7.128) results by applying definition (6.103), and (7.129) is obtained by comparing definitions (6.103) and (2.1,3) under the assumption $\mu_X = 0$. In (7.129), the spectrum $|g_N(s)|^2$ of the time series (x_t) is N times the Fourier transform (6.49) of the empirical covariance function $\hat{c}_X(\tau)$ on condition that $\hat{c}_X(\tau) = 0$ for $|\tau| \geq N$.

In (7.130), under the above assumptions, $\hat{c}_X(\tau)$ estimates the covariance function $c_X(\tau)$ of the stationary discrete-time stochastic process (X_t) without bias, on condition that $N \to \infty$, and thereby the right side in (7.131) is obtained. On the left side in (7.131) is the expectation of the power of the time series in the limiting case for $N \to \infty$.

$$\lim_{N\to\infty} \mathrm{E}\left(\frac{|g_N(s)|^2}{N}\right) = \lim_{N\to\infty} \mathrm{E}\left(\sum_{\tau=-N}^{N} \hat{c}_X(\tau) \mathrm{e}^{-\mathrm{i}2\pi s\tau} \right) \tag{7.131}$$

$$= \lim_{N\to\infty} \left(\sum_{\tau=-(N-1}^{N-1} (1 - \tau/N) c_X(\tau) \mathrm{e}^{-\mathrm{i}2\pi s\tau} \right)$$

$$= \sum_{\tau=-\infty}^{\infty} c_X(\tau) \mathrm{e}^{-\mathrm{i}2\pi s\tau} \tag{7.132}$$

Applying (2.58), (7.132) is obtained, and, when this result is compared with (7.66), it is concluded that $\big(|g_N(s)|^2\big)/N = C_X(s)$ for $N \to \infty$.

(7.132) suggests that the spectrum of a stationary stochastic process, on condition that it has an integer parameter and a covariance function that converges absolutely, can be estimated by calculating the Fourier transform of its empirical covariance function which is computed as required in (2.1,3) from an observed time series. Such an estimate, however, inherits the

unfavourable properties of the empirical covariance function enumerated in (2.58). In Chaps. 8 and 9, therefore, alternative estimators for the spectrum are introduced.

7.5.5 Spectrum and/or Covariance Function?

Which of the two, the continuous spectrum $C_X(s)$ or the covariance function $c_X(\tau)$, is more adequate for describing the second moment function of a stationary stochastic process $X(t)$ having spectrum $C_X(s)$? It does not matter, because both functions, $c_X(\tau)$ and $C_X(s)$, are a Fourier transform pair as defined in (7.65) and (7.66), and thus, ultimately, contain the same information. This answer, however, applies only in cases where the process under analysis has a simple spectrum, for example in the case of a white noise process (W_t), $t = \ldots, -1, 0, 1, \ldots$, with covariance function $c_W(\tau) = \sigma_W^2$, $\tau = 0$, and $c_W(\tau) = 0$ for $\tau \neq 0$, and spectrum $C_W(s) = \sigma_W^2$ in (7.76), both being simple functions.

If the process $X(t)$ under analysis has a less simple spectrum $C_X(s)$, then a plot of $C_X(s)$ will give a more detailed picture of its second moment function than will a plot of $c_X(\tau)$ for the following reason. If $c_X(\tau)$ is known then it is difficult to obtain diagnostics for the shape of $C_X(s)$, i.e., the number of its peaks together with their frequencies, its dynamic range and its bandwidth, directly from $c_X(\tau)$, i.e., without Fourier transforming $c_X(\tau)$.

The number and the frequencies of possible peaks in $C_X(s)$ remain difficult to obtain from periodic or quasi-periodic fluctuations in $c_X(\tau)$. The difficult tracking of footprints generated by periodic fluctuations in observations, indeed, gives reason to calculate the Fourier transform of the observations and thereafter their spectrum, as is demonstrated as a first example using the sawtooth-shaped sequence in Fig. 6.1 and its spectrum in Fig. 6.2. Three more examples are given in Fig. 7.15. There, plots of the covariance functions and the spectra are shown which pertain to the AR[2] models $(Y_t^{(1)})$ and $(Y_t^{(2)})$ (with realisations plotted above in Figs. 7.12 and 7.13) and to the AR[4] model (X_t) (with a realisation plotted above in Fig. 7.14).

$(Y_t^{(1)})$ is dominated by oscillations with frequencies $s \approx 0.2$, and $(Y_t^{(2)})$ by oscillations with frequencies $s \approx 0.4$, as is obvious from the spectra of these processes as plotted in Fig. 7.15 (b) and (d). In the pertaining covariance functions shown in plots (a) and (b), it is, however, difficult to identify the the frequencies of the dominating oscillations.

A diagnostic for the bandwidth (7.83) of $C_X(s)$ can in principle be obtained from the width at half height (6.97) of $c_X(\tau)$, because $C_X(s)$ is narrow on condition that $c_X(\tau)$ is wide, as implied by the similarity theorem (6.98) or the fundamental uncertainty relationship (6.107). If, in addition, $c_X(\tau) > 0$, a diagnostic for the bandwidth of $C_X(s)$ can be obtained from the equivalent width (6.99) or σ-width (6.101) of $c_X(\tau)$, since $c_X(\tau) \leq c_X(0)$ in the remarks to (6.151). To obtain such a diagnostic, however, it is assumed that

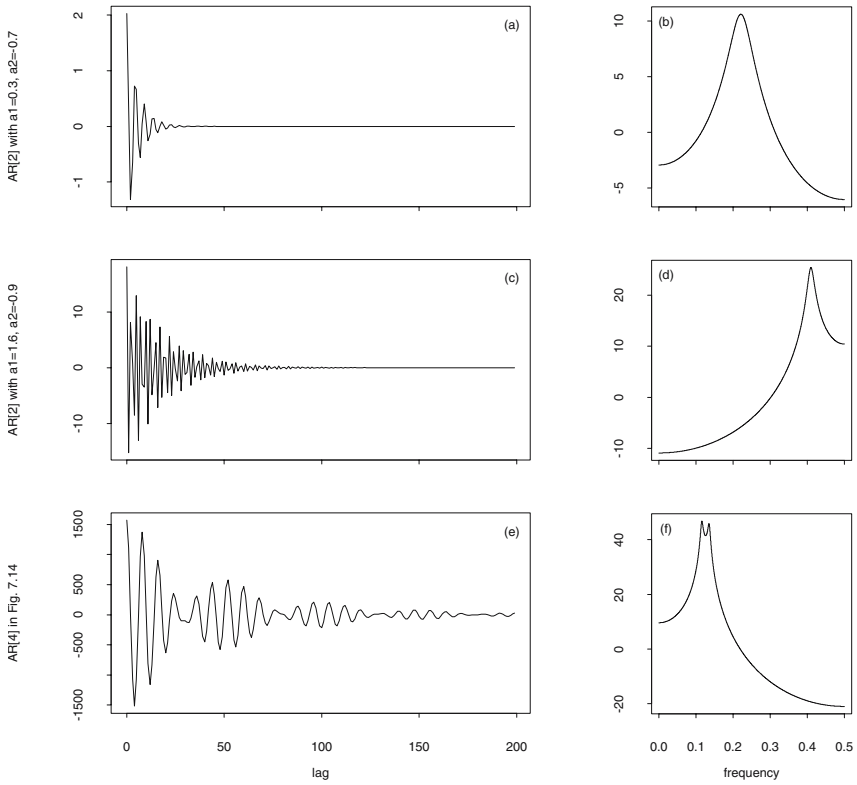


Fig. 7.15. Covariance functions (on the left) and spectra in dB (on the right) of the AR[2] models $(Y_t^{(1)})$ and $(Y_t^{(2)})$ in Figs. 7.12 and 7.13 and of the AR[4] model in Fig. 7.14, from top to down.

$C_X(s)$ has only one peak. Under the further assumption that a narrow function has a large dynamic range and that, in contrast, a wide function has a small dynamic range, the dynamic range of $C_X(s)$ can be approximated from the dynamic range of $c_X(\tau)$. In practical applications, nevertheless, it is difficult and often impossible to obtain diagnostics for the bandwidth and the dynamic range of $C_X(s)$ even when both bandwidth and dynamic range of $c_X(\tau)$ are known.

For example, from the variance $c_X(0)$ and the covariance function $c_X(\tau)$ of the AR[4] model (X_t) as plotted in Fig. 7.15 (e), it is not possible to foresee whether the spectrum $C_X(s)$ of this process will have the narrow and high twin peaks as shown in Fig. 7.15 (f). And it is also impossible to guess the bandwidth or the dynamic range of the spectra in Fig. 7.15 (b) and (d) from the covariance functions in Fig. 7.15 (a) and (c).

These examples demonstrate that it often pays when a known covariance function of a linear process is transformed using (7.66) in the continuous spectrum of the process: the remuneration is that a deeper insight into the process can be attained in the frequency domain. This Fourier transform is not a technical trick but makes good mathematical sense owing to the existence theorem (7.51). When calculating this Fourier transform, however, precautions have to be taken against a possible aliasing, and, if the covariance function is not known for $\tau = \ldots, -1, 0, 1, \ldots$, against a possible leakage, as demonstrated in Sects. 6.5 and 6.7.

Further pitfalls have to be circumvented when an empirical covariance function $\hat{c}_X(\tau)$, calculated from a realisation (x_t) observed in time slice $t = 0, 1, \ldots, N-1$ of a linear process (X_t), is Fourier transformed in order to arrive at an estimator for $C_X(s)$, i.e., the spectrum of (X_t). These difficulties are enumerated in the remarks to (7.132) and comprehensively dealt with in Chaps. 8 and 9.

7.6 Problems

7.1. Simulate N points having a uniform (rectangular) distribution (defined in the remarks to (7.30)) in interval $x_1 \leq x \leq x_2$. Sort the points, calculate the inter-point distances, and thereafter estimate their distribution.

```
x <- runif(240,0,300)   #240 points in interval [0,300]
plot(x,x*0)             #--> lambda = 0.8
difx <- diff(sort(x))   #inter-point distances
par(mfrow=c(1,3))
hist(difx,nclass=40,xlab=" ",probability=T)
lambda <- 1/mean(difx)
lines(sort(difx),dexp(sort(difx),rate=lambda))
#x <- (0:(max(difx)*10))/10  #for higher resolution plot,
#fx <- lambda*exp(-lambda*x) #use (7.5)
#lines(x,fx)                 #
plot(qexp(ppoints(difx),rate=lambda),  #probability plot
     sort(difx),xlab="theoretical",ylab="simulated")
difxts <- ts(difx,start=1,frequency=1,)
acf(difxts,lag.max=100,type="correlation",plot=T)
```

Are the inter-point distances independent? exponentially distributed?

7.2. The time points when peaks occur in the DC measurements plotted in Fig. 7.3, as well as their inter-peak times, are obtained by applying R expressions in `/path/aerosol.R`. The shortest inter-peak time is 2 s. Why?

For Splus-users only: obtain the time points when peaks occur using `peaks()`.

Identify periods with low and high background aerosol loads in the DC measurements. Thereafter calculate the inter-peak times for both types of periods. Are the inter-peak times exponentially distributed in both cases?

7.3. Tree-line observations in the Valais alps are plotted in Fig. 3.10. Are the locations of these observations random, i.e., do they have a uniform distribution? If so, the number of observations in non-overlapping subregions is Poisson distributed as required in (7.2) and (7.3,2,3). Partition the map in Fig. 3.10 into 32 subregions by drawing equi-spaced straight lines, 7 in vertical and 3 in horizontal direction. Then count the number of observations in those subregions that intersect the observational area as defined by the polygon. Assume, for simplicity, that the subregions have unit area. Hint: histograms and probability plots as shown, for example, in Fig. 7.5, are obtained in Problem 7.1.

7.4. Integrate $c_V(\tau) = \sigma_V^2$, for $\tau = 0$ and $c_V(\tau) = 0$, for $\tau \neq 0$, τ real, in the interval $-1 \leq \tau \leq 1$. Sketch the Riemann sum pertaining to this integral. Thereafter integrate $c_U(\tau) = \sigma_U^2 \delta(\tau)$ in the interval $-1 \leq \tau \leq 1$, $\delta(\tau)$ being the delta function as defined in (6.60).

7.5. Exploit the properties of the P_j in (7.30) to obtain the moments of the A_j and B_j in (7.39), (7.40) and (7.41). Calculate the expectations as demonstrated in (7.120).

7.6. Show that the random function defined in (7.38) is stationary. First, derive the moment functions of a random trigonometric oscillation, i.e., the expectation and the covariance functions of the stochastic process $X(t) = A\cos(2\pi st) + B\sin(2\pi st)$, s being constant and real, A and B being random variables having a normal distribution with expectation $\mu = 0$ and variance σ^2. Under these assumptions, (7.40) implies that A and B are independent. These calculations result in

$$\begin{aligned} \mathrm{E}X(t) &= 0 & c_X(t,u) &= \sigma^2 \cos\big(2\pi s(t-u)\big) \\ c_X(\tau) &= \sigma^2 \cos(2\pi s\tau) & c_X(0) &= \mathrm{Var}X(t) = \sigma^2 \end{aligned}$$

and, therefore, $X(t)$ is stationary. Further on, the random function in (7.38) is stationary because all its oscillations are stationary.

The $X(t)$, being a linear combination of the normally distributed random variables A and B, are normally distributed with $\mu_X = 0$ and $\sigma_X^2 = \sigma^2$. Calculate their density for time point t and, applying (1.12), the joint density for two time points t_1 and t_2 having lag h.

7.7. Apply R expressions as in `/path/harmproc.R` to generate simulations of harmonic processes.

7.8. The linear process (2.25) is a stochastic LTI filter as defined in (7.84) having a one-sided sequence of weights, i.e., an impulse response function that is not even. One-sided weight sequences are preferred when constructing linear processes. Plausible reasons are given in the remarks to (2.26) and, using one-sided sequences, the output process becomes causal as required

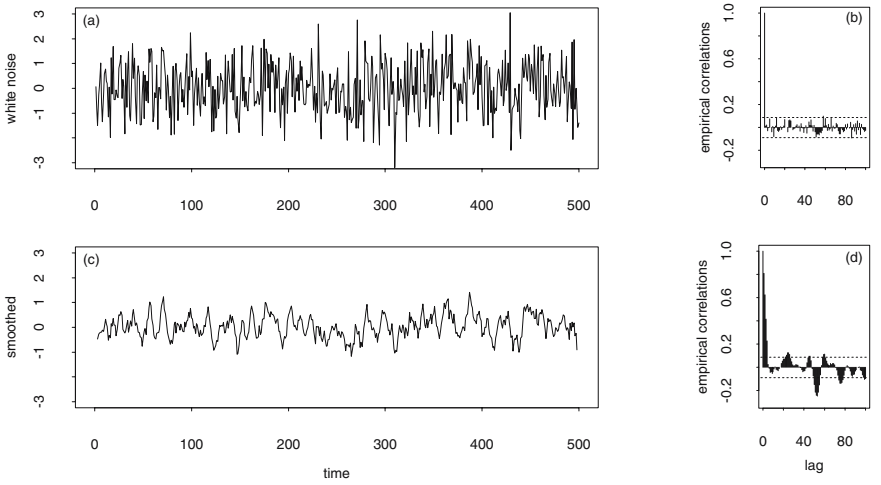


Fig. 7.16. An LTI filter as defined in (7.84) with a two-sided frequency response function. A realisation of a white noise process in the time slice $1 \leq t \leq 500$ as shown in plot (a) is smoothed with the frequency response function $(1/5)\Pi_5^{(e)}$ as defined in (6.52) to obtain an output as shown in plot (b). On the right side, in plots (b) and (d), the empirical covariance functions of the input and output process are shown.

in (5.3). Linear processes having a one-sided impulse response function are estimated in Chap. 5, and predictions are calculated from the estimates. Example spectra of such processes are given in Sect. 7.4.2.

In (7.84), however, the impulse response function is not restricted to being one-sided. For example, let `wts` be the time slice from the realisation of a white noise process as plotted in Fig. 7.16 (a). Then, using R expression

```
xts <- filter(wts,c(1/5,1/5,1/5,1/5,1/5),
                          method="convolution",sides=2)
```

the output in Fig. 7.16 (c) is obtained. Simulate more examples of LTI filters with two-sided impulse response functions having a realisation of a white noise process as input. Then calculate the empirical covariance functions of the outputs.

7.9. The continuous spectrum shown in Fig. 7.17 (b) is made available with

```
a717spec <- scan("/path/fig717.dat")
s1 <- -4096:4096
s1 <- s1/8192
N1 <- 4097
N2 <- 8193
plot(s1[N1:N2],10*log10(a717spec[N1:N2]),
        type="n",xlab="frequency",ylab="spectrum in dB")
```

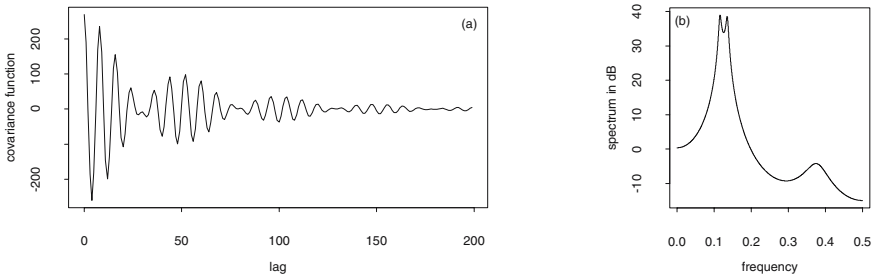


Fig. 7.17. Covariance function $c_X(\tau)$ and spectrum $C_X(s)$ as in (7.65) of a linear process (X_t) having a continuous spectrum with three peaks.

```
lines(s1[N1:N2],10*log10(a717spec[N1:N2]))
```

in R vector `a717spec`. Calculate the covariance function pertaining to this spectrum using a discrete Fourier transform (6.22,4). Hint: Apply the R expressions in the remarks to Fig. 6.35. Thereafter, plot the covariance function for lags larger than $\tau = 200$, and compare with the covariance function and the spectrum in Fig. 7.15 (e) and (f).

7.10. Let $X(t)$ be a real-valued and stationary random function $X(t)$ having covariance function $c_X(\tau)$ and spectrum $C_X(s)$ such that $c_X(\tau)$ and $C_X(s)$ are a Fourier transform pair as required in (7.63): $C_X(s) = \int_{-\infty}^{\infty} c_X(\tau) \mathrm{e}^{-\mathrm{i}2\pi s\tau} \mathrm{d}\tau$. Then construct $Z(t) = \mathrm{e}^{-\mathrm{i}2\pi s_0 t} X(t)$. $Z(t)$ is a complex-valued random function. Please obtain its covariance function $c_Z(\tau) = \mathrm{e}^{+\mathrm{i}2\pi s_0 \tau} c_X(\tau)$ as well as its spectrum $C_Z(s) = C_X(s + s_0)$.

7.11. Let the set of all real-valued random functions be a subset of the complex-valued random functions. Then the complex-valued process $Z(t) = \mathrm{e}^{-\mathrm{i}2\pi s_0 t} X(t) = \mathcal{L}\big(X(t)\big)$ constructed in Problem 7.10 is a linear transformation as defined in (2.30). Is $\mathcal{L}$ not only linear but also time-invariant?

7.12. Let $X(t)$ be a real-valued random function and $a \neq 0$ as well as $b \neq 0$ be real-valued constants. Then: is $\mathcal{L}\big(X(t)\big) = Y(t) = a + bX(t)$ linear and time-invariant?

8 Does a Periodogram Estimate a Spectrum?

The spectrum of a deterministic function $f(t)$ in $L^2(-\infty, \infty)$ is obtained by computing its Fourier transform $F(s)$ and thereafter $|F(s)|^2$, i.e., the squared absolute value of $F(s)$. $F(s)$ is arrived at using definition (6.35) and is in $L^2(-\infty, \infty)$, on condition that t is real. If $f(t)$ is observed at discrete time points in a finite interval then a sequence of observations is obtained which are thereafter Fourier transformed, etc., to arrive at the spectrum of $f(t)$. In this second case, the results are subject to aliasing and leakage, due to the discretisation and the finite observational period imposed. These pitfalls exposed in Sect. 6.5 can be circumvented using the diagnostics introduced in Sect. 6.7.2. Aliasing and leakage are also encountered when the spectrum of a stationary discrete-time stochastic process is estimated using the estimators introduced in this chapter (the periodogram), in Chap. 9 (estimators for a continuous spectrum) and in Chap. 10 (estimators for a discrete spectrum).

In Sect. 8.1, the periodogram is defined, an example is given, and it becomes obvious that the empirical spectra calculated in Sect. 6.1.3 from four example time series are identical with the periodograms of these time series. In Sect. 8.2, the properties of a periodogram calculated from observations of a white noise process are derived and tests for white noise are constructed. The statistics derived allow for testing the null hypothesis that the observations stem either from a white noise process (the null hypothesis) or from two alternatives: (i) a process with a non-flat continuous spectrum or (ii) a process having a mixed spectrum similar to the one in (7.82).

In Sect. 8.3, the expectation function of the periodogram is obtained under the assumption that it is calculated from observations of a discrete-time stationary stochastic process. The derived expectation function implies that, in general, the periodogram does not estimate the spectrum without bias. In Sect. 8.4, the second moment functions and the probability distribution of the periodogram are calculated assuming that the observations stem from a process having a continuous spectrum [21]. The properties derived for this case are, however, shared by periodograms calculated from observations stemming from most stationary and ergodic stochastic processes.

The supplements and the problems follow in Sects. 8.5 and 8.6.

8.1 The Periodogram

Borrowing from the definitions of the Fourier coefficients in (6.21), estimators $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ are defined in (8.1,1). Using these estimators, the periodogram $I_X(s_k)$ is defined for the Fourier frequencies in (8.1,2,3).

Let (x_t), $t = 0, \ldots, N-1$, be a time-slice in a realisation of a stationary discrete-time stochastic process (X_t) and s_k, k and m as defined in (6.21). Then, using the estimators

1. *in case of N being even:*

$$\widehat{A}_X(s_k) = \begin{cases} (2/N)\sum_{t=0}^{N-1} X_t \cos(2\pi s_k t) & k=1,\ldots,m-1 \\ (1/N)\sum_{t=0}^{N-1} X_t \cos(2\pi s_k t) & k=0,m \end{cases}$$

$$\widehat{B}_X(s_k) = \begin{cases} (2/N)\sum_{t=0}^{N-1} X_t \sin(2\pi s_k t) & k=1,\ldots,m-1 \\ 0 & k=0,m \end{cases},$$

in case of N being odd:

$$\widehat{A}_X(s_k) = \begin{cases} (2/N)\sum_{t=0}^{N-1} X_t \cos(2\pi s_k t) & k=1,\ldots,m \\ (1/N)\sum_{t=0}^{N-1} X_t \cos(2\pi s_k t) & k=0 \end{cases}$$

$$\widehat{B}_X(s_k) = \begin{cases} (2/N)\sum_{t=0}^{N-1} X_t \sin(2\pi s_k t) & k=1,\ldots,m \\ 0 & k=0 \end{cases},$$

the periodogram $I_X(s_k)$ of (x_t) is defined for the s_k, i.e., the Fourier frequencies: (8.1)

2. $$I_X(s_k) = \begin{cases} (N/4)\big((\widehat{A}_X(s_k))^2 + (\widehat{B}_X(s_k))^2\big) & s_k \neq -1/2, 0, 1/2 \\ N\big((\widehat{A}_X(s_k))^2\big) & s_k = -1/2, 0, 1/2 \end{cases}.$$

Substituting the above expressions for $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$, the distinction of cases is no longer required:

3. $$I_X(s_k) = \frac{1}{N}\left(\sum_{t=0}^{N-1} X_t \cos(2\pi s_k t)\right)^2 + \frac{1}{N}\left(\sum_{t=0}^{N-1} X_t \sin(2\pi s_k t)\right)^2$$

Finally, the periodogram is defined for $-1/2 \leq s \leq 1/2$:

4. $$I_X(s) = \frac{1}{N}\left(\sum_{t=0}^{N-1} X_t \cos(2\pi s t)\right)^2 + \frac{1}{N}\left(\sum_{t=0}^{N-1} X_t \sin(2\pi s t)\right)^2$$

In (8.1,4), the periodogram $I_X(s)$ is defined for all real frequencies $-1/2 \leq s \leq 1/2$ and is assumed to be periodic with period 1. These are the properties of spectrum $C_X(s)$ in (7.65) and also of the integrated spectrum $C_X^{(I)}(s)$ in (7.64), both pertaining to a stationary stochastic process (X_t) with integer parameter t. Consequently, $I_X(s)$ is a candidate estimator for the spectrum (and, subsequent to being integrated, for the integrated spectrum) of a discrete-time stochastic process having the properties required in (7.51).

$I_X(s_k)$ and $I_X(s)$ estimate a function, and not a parameter. This property is shared, for example, by the empirical covariance function $\hat{c}_X(\tau)$ which

estimates the covariance function $c_X(\tau)$, $\tau = \ldots, -1, 0, 1, \ldots$, of a stationary stochastic process (X_t) with $\mu_X = 0$ and absolutely convergent $c_X(\tau)$ as summarised in (2.58). The unfavourable properties of $\hat{c}_X(\tau)$ enumerated in (2.58) are inherited by its Fourier transform $\mathcal{F}_{-\mathrm{i}}\big(\hat{c}_X(\tau)\big)$ which renders it inappropriate for estimating the spectrum of (X_t), as concluded from the results obtained in Sect. 7.5.4. Is the periodogram a more adequate estimator for the spectrum? An answer is given in Sects. 8.3 and 8.4, subsequent to calculating an example periodogram in the following paragraphs and to deriving the properties of a periodogram calculated from observations stemming from a white noise process in Sect. 8.2.

In R, a periodogram as defined in (8.1,3) is calculated using `spec.pgram()`, `spec.univariate()` or `fft()`. For example, assuming that the NAO index (Jones), in the period from 1824 through to 2000 as plotted in Fig. 2.10, has been made available as R time series `naojonests` in Problem 2.2, the following R expressions

```
x <- as.vector(naojonests) - mean(as.vector(naojonests))
N <- length(x)
m <- floor(N/2)   #m <- (N/2) for even N, m <- (N-1)/2 for odd N
sk <- (0:m)/N           #Fourier frequencies
xt <- fft(x,inverse=T)  #-i-Transformation
spec <- (1/N)*( (Re(xt))*(Re(xt)) + (Im(xt))*(Im(xt)) )
plot(c(0,0.5), c(-30.0,10.0), type="n",xlab="frequency",ylab="dB")
for(k in 1:m) {
  lines( c(sk[k+1],sk[k+1]), c(-30.0,10*log10((spec)[k+1])) ) }
```

(i) calculate the periodogram of the NAO index (Jones) applying definitions (8.1,3) and (6.22) and (ii) plot the values thus obtained using the vertical solid lines in Fig. 8.1. Thereafter, using

```
spec.nao <- spec.pgram(naojonests,pad=0,taper=0,
                       fast=F,demean=T,detrend=F,plot=F)
points(spec.nao$freq,10*log10(spec.nao$spec))
```

the periodogram values plotted with ∘ in Fig. 8.1 are obtained. Unsurprisingly, both results are identical. Details on R functions `spec.pgram()` and `spec.univariate()` are given in Sect. 9.5.6. The above example demonstrates that a periodogram is easily calculated for the Fourier frequencies using `fft()` and definition (8.1), given an observed time series. Calculations using `fft()` are demonstrated in Sects. 6.1.1, 6.3.1 and 6.9.8.

Usually, a time series is de-meaned prior to calculating the periodogram, as demonstrated above. As a consequence, $I_X(0) \approx 0$ (substituting $s_k = 0$ in (8.1,1,2), $I_X(0) = N\hat{\mu}_X^2$ is arrived at, with $\hat{\mu}_X$ as defined in (2.1,1)), and, taking the logarithm, a small negative number is obtained, since $-\infty \leftarrow \log_{10}\big(I_X(0)\big)$ for $I_X(0) \to 0$. Consequently, $I_X(s_k = 0)$ is usually omitted when the periodogram is plotted.

The periodogram of the NAO index (Jones) in Fig. 8.1 is identical with the empirical spectrum calculated from this time series as plotted in Fig. 6.5

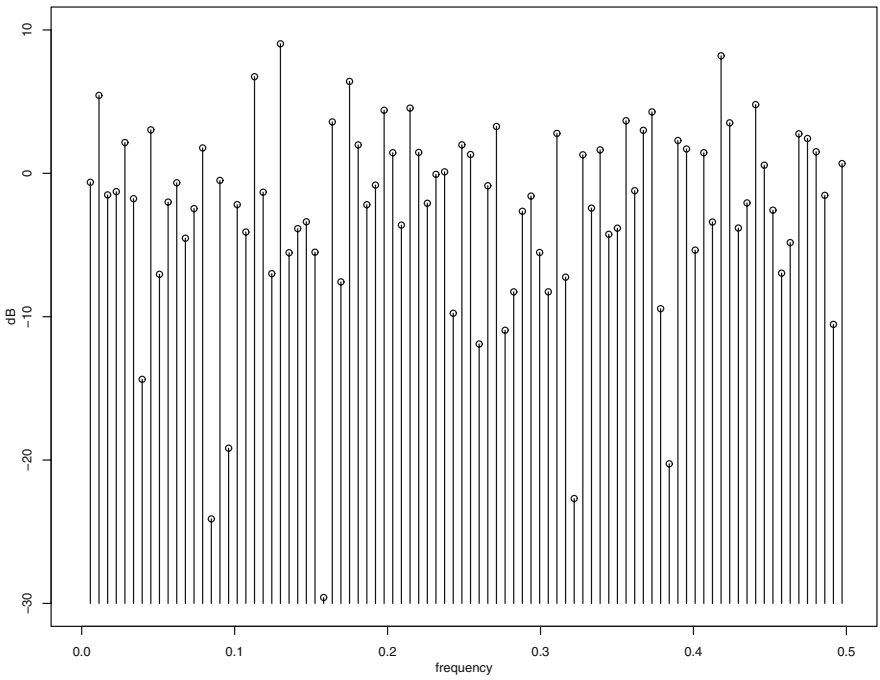


Fig. 8.1. Periodogram of the NAO index (Jones) calculated from the time series plotted in Fig. 2.10 using `fft()` (vertical solid lines) and `spec.pgram()` (symbol ∘).

(a), both showing large fluctuations even in values being close neighbours, a property shared by all empirical spectra calculated in Sect. 6.1.3 and plotted in Fig. 6.5. In Sect. 6.1.3, it is assumed that the wild fluctuations in the example empirical spectra are due to the observations stemming from stochastic processes (and not from deterministic variables) and therefore that the probabilistic properties of the variables observed have to be accounted for when the spectrum of the process is estimated from a realisation.

More specifically, since the observations of the NAO index (Jones) are shown to be a realisation of a normal white noise process in Sects. 2.3.1 and 2.5.1, the problem arises as to whether and how an estimator for the continuous and flat spectrum of the NAO index (Jones) can be constructed from the wildly fluctuating periodogram shown in Fig. 8.1. This problem can be solved using the answers to the following questions: Can the moment functions of $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ be computed exploiting the properties of the orthogonal trigonometric functions in (6.9), on condition that (X_t) is assumed to be a white noise process as defined in (2.10) and (2.11)? Using these results, can the moment functions $\mathrm{E}I_X(s_k)$, $\mathrm{Var}I_X(s_k)$ and $\mathrm{Cov}\big(I_X(s_k), I_X(s_l)\big)$, $s_k \neq s_l$, now be obtained?

8.2 Calculated from a Realisation of a White Noise Process

It is assumed that the process under analysis is (W_t), i.e., a white noise process as defined in (2.10) and (2.11) having expectation function $\mathrm{E}W_t = 0$ and covariance function $\mathrm{Cov}(W_t, W_u) = \sigma_W^2$ for $t = u$ and $\mathrm{Cov}(W_t, W_u) = 0$ for $t \neq u$. From this covariance function, the constant spectrum $C_W(s) = \sigma_W^2$ of the white noise process is derived in (7.76).

8.2.1 Its Properties

From a time slice in a realisation of this model, i.e., (w_t), $t = 0, \ldots, N-1$, estimators $\widehat{A}_W(s_k)$ and $\widehat{B}_W(s_k)$, s_k being the Fourier frequencies as defined in (6.21,1), are calculated as required in (8.1,1). Assuming $s_k \neq 1/2, 0, 1/2$, the expectations of these estimators are obtained in (8.2) and (8.3)

$$\mathrm{E}\widehat{A}_W(s_k) = \mathrm{E}\left(\frac{2}{N}\sum_{t=0}^{N-1} W_t \cos(2\pi s_k t)\right) = \frac{2}{N}\sum_{t=0}^{N-1} \cos(2\pi s_k t)\mathrm{E}W_t = 0 \tag{8.2}$$

$$\mathrm{E}\widehat{B}_W(s_k) = 0 \tag{8.3}$$

and their variances in (8.4),

$$\begin{aligned}\mathrm{Var}\widehat{A}_W(s_k) &= \mathrm{Var}\left(\frac{2}{N}\sum_{t=0}^{N-1} W_t \cos(2\pi s_k t)\right)\\ &= \frac{4}{N^2}\sum_{t=0}^{N-1}\sum_{t=0}^{N-1} \cos(2\pi s_k t)\cos(2\pi s_k u)\mathrm{Cov}\big(W_t, W_u\big)\\ &= \frac{4}{N^2}\sigma_W^2 \sum_{t=0}^{N-1}\big(\cos(2\pi s_k t)\big)^2 = \frac{2}{N}\sigma_W^2 = \mathrm{Var}\widehat{B}_W(s_k)\end{aligned} \tag{8.4}$$

using the rules in (1.15,10), the properties of the trigonometric functions in (6.9), the covariance function of the white noise process, and the elementary trigonometric identity $\cos(x)\cos(y) = (1/2)\big(\cos(x+y) + \cos(x-y)\big)$.

A similar derivation results in the covariances of these estimators for identical and distinct Fourier frequencies in (8.5) and (8.6). Consequently, (i) estimators $\widehat{A}_W(s_k)$ and $\widehat{B}_W(s_k)$ are non-correlated for all s_k and (ii) sequences $\widehat{A}_W(s_k)$ and $\widehat{B}_W(s_k)$ are, individually, white noise processes as defined in (2.10).

$$\begin{aligned}&\mathrm{Cov}\big(\widehat{A}_W(s_k), \widehat{B}_W(s_k)\big)\\ &\quad = \mathrm{Cov}\left(\frac{2}{N}\sum_{t=0}^{N-1}\cos(2\pi s_k t)W_t,\ \frac{2}{N}\sum_{t=0}^{N-1}\sin(2\pi s_k t)W_t\right)\end{aligned}$$

$$= \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} \frac{2}{N} \cos(2\pi s_k t) \frac{2}{N} \sin(2\pi s_k u) \mathrm{Cov}\big(W_t, W_u\big)$$

$$= \frac{4}{N^2} \sum_{t=0}^{N-1} \cos(2\pi s_k t) \sin(2\pi s_k t) \mathrm{Cov}\big(W_t, W_t\big)$$

$$= \frac{4}{N^2}\, 0\, \sigma_W^2 = 0 \tag{8.5}$$

$$\mathrm{Cov}\big(\widehat{A}_W(s_k), \widehat{A}_W(s_l)\big) = \mathrm{Cov}\big(\widehat{B}_W(s_k), \widehat{B}_W(s_l)\big) = 0 \text{ for } k \neq l \tag{8.6}$$

Using the moments of $\widehat{A}_W(s_k)$ and $\widehat{B}_W(s_k)$ obtained above, further properties of $I_W(s_k)$, i.e., the periodogram calculated from a realisation of a white noise process (W_t), are derived under the additional assumption that (W_t) is a normal process as defined in (2.3). Under this additional assumption, $\widehat{A}_W(s_k)$ and $\widehat{B}_W(s_k)$ are linear combinations of normally distributed random variables and thus are also normally distributed, as concluded from the remark to (1.34). Consequently, pairs $\big(\widehat{A}_W(s_k), \widehat{B}_W(s_l)\big)$, $\big(\widehat{A}_W(s_k), \widehat{A}_W(s_l)\big)$, $\big(\widehat{B}_W(s_k), \widehat{B}_W(s_l)\big)$, $k \neq l$, as well as $\big(\widehat{A}_W(s_k), \widehat{B}_W(s_k)\big)$, are pairs of stochastically independent random variables. This result implies that a periodogram, being constructed from the above estimators, is independent for distinct Fourier frequencies, as summarised in (8.7,1,2).

Let $I_W(s_k)$ be the periodogram calculated using (8.1,1,2,3) from (w_t), $t = 0, \ldots, N-1$, with (w_t) being a realisation of a $(0, \sigma_W^2)$ normally distributed white noise process (W_t) and s_k the Fourier frequencies in (6.21,1). Then:

1. *$I_W(s_k)$ and $I_W(s_l)$ are independent for $k \neq l$*
2. *$\widehat{A}_W(s_k)$ is independent from $\widehat{B}_W(s_k)$*
3. *the normalised periodogram*
   $(2/\sigma_W^2) I_W(s_k)$ is $\chi^2_{(2)}$ distributed for $s_k \neq 1/2, 0, 1/2$
   $(1/\sigma_W^2) I_W(s_k)$ is $\chi^2_{(1)}$ distributed for $s_k = -1/2, 0, 1/2$
4. *$\mathrm{E} I_W(s_k) = \sigma_W^2$ for all s_k*
5. *$\mathrm{Var} I_W(s_k) = \begin{cases} \sigma_W^4 & \text{for } s_k \neq 1/2, 0, 1/2 \\ 2\sigma_W^4 & \text{for } s_k = -1/2, 0, 1/2 \end{cases}$.*

(8.7)

Using the moments of $\widehat{A}_W(s_k)$ and $\widehat{B}_W(s_k)$ obtained above in (8.2), (8.3), (8.4), (8.5) and (8.6), the distribution of $I_X(s_k)$, s_k being the Fourier frequencies, is derived as follows. In the periodogram written as in (8.8),

$$I_W(s_k) = (N/4)\Big(\big(\widehat{A}_W(s_k)\big)^2 + \big(\widehat{B}_W(s_k)\big)^2\Big) \qquad s_k \neq -1/2, 0, 1/2$$

$$= \big((N/4)^{1/2} \widehat{A}_W(s_k)\big)^2 + \big((N/4)^{1/2} \widehat{B}_W(s_k)\big)^2 \tag{8.8}$$

random variables $(N/4)^{1/2} \widehat{A}_W(s_k)$ and $(N/4)^{1/2} \widehat{B}_W(s_k)$ are obtained by multiplying $\widehat{A}_W(s_k)$ and $\widehat{B}_W(s_k)$ with $(N/4)^{1/2}$. They thus have variances

$(N/4)^{2/2}(2/N)\sigma_W^2 = (1/2)\sigma_W^2$, as obtained using (1.15,6) for $s_k \neq 1/2, 0, 1/2$. Hence, subsequent to normalising both estimators $\widehat{A}_W(s_k)$ and $\widehat{B}_W(s_k)$ with $(2/\sigma_W^2)^{1/2}$, $(0,1)$-normally distributed random variables are obtained which are independent. Consequently, $(2/\sigma_W^2)I_W(s_k)$ is $\chi^2_{(2)}$ distributed, i.e., the normalised periodogram of a normal white noise process has a chi-square distribution with 2 degrees of freedom, in agreement with the first stipulation in (8.7,3). The expectation of a random variable having a $\chi^2_{(n)}$ distribution is n, i.e., the number of degrees of freedom, its variance twice this number. Consequently, $\mathrm{E}\big((2/\sigma_W^2)I_W(s_k)\big) = 2$ and $\mathrm{Var}\big((2/\sigma_W^2)I_W(s_k)\big) = 4$ are arrived at, as required in (8.7,4) and by the first stipulation in (8.7,5).

For frequencies $s_k = -1/2, 0, 1/2$, however, $I_W(s_k) = \big(N^{1/2}\widehat{A}_W(s_k)\big)^2$ with variance $N(1/N)\sigma_W^2 = \sigma_W^2$; and, consequently, $(1/\sigma_W^2)^{1/2}\widehat{A}_W(s_k)$ is normally distributed with zero expectation and unit variance. This implies that $(1/\sigma_W^2)I_W(s_k)$ has a $\chi^2_{(1)}$ distribution and thus has expectation 1 and variance 2, reconcilable with the second stipulations in (8.7,3,5).

(8.7,3) also applies to a periodogram calculated from a non-normal white noise process on condition that $N \to \infty$ [21].

The $\chi^2_{(n)}$ distribution is a gamma distribution having order $n/2$ and parameter $\lambda = 1/2$. Gamma distributions are generated from exponential distributions in (7.98), (7.99), (7.100) and (7.101). For example, $f_{k\lambda}(x) = 0$ for $x < 0$ and $f_{k\lambda}(x) = \big(\lambda^k/\Gamma(k)\big)x^{k-1}\mathrm{e}^{-\lambda x}$ for $x \geq 0$ is the gamma density having order k and parameter λ. Using this result, it is shown in Problem 8.1 that a $\chi^2_{(n)}$ distribution is an exponential distribution as defined in (7.4). It is right-skewed and therefore a few extremely large values are likely to occur in a periodogram. In a logarithmic plot of a periodogram, however, some extremely small values are usually seen, because a random variable $J = 10\log_{10}(I)$ has a left-skewed distribution on condition that I has a right-skewed distribution (Problem 8.2).

8.2.2 Tests

The null hypothesis of the tests introduced in this section is that the observations stem from a white noise process as defined in (2.10) and (2.11). The alternative is that the observations do not stem from a white noise process. Tests for two alternatives are derived: (i) where the observations stem from a process with a non-constant continuous spectrum (examples in Sect. 7.4.2), or (ii) where the observations stem from a process having a mixed spectrum being constant with superimposed spikes (an example is given in (7.82)).

In the first case, it is assumed that the observations stem from a stationary process (X_t) having a continuous spectrum $C_X(s)$ with a small dynamic range. $C_X(s)$ is either constant (the null hypothesis) or non-constant (the alternative). Such a test can be obtained from the cumulated periodogram $I_X^{(\Sigma)}(s_k)$ subsequent to a normalisation: in (8.9), $I_X^{(\Sigma)}(s_k)$ is the sum of $I_X(s_k)$ as defined in (8.1,3) for Fourier frequencies being less than or equal to s_k, and

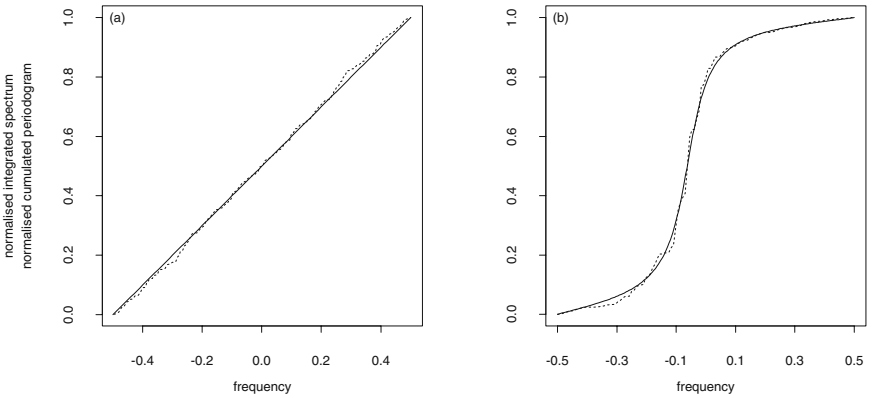


Fig. 8.2. Normalised cumulated periodograms (broken lines) and normalised integrated spectra (solid lines) as defined in (8.9) and (8.10) for a (0,1)-normally distributed white noise process (on the left) and for $Y_t^{(1)} = 0.3Y_{t-1}^{(1)} - 0.7Y_{t-2}^{(1)} + W_t$ with $\mu_W = 0$ and $\sigma_W^2 = 1$, i.e., the AR[2] model in Fig. 7.12.

$I_X^{(\Sigma/\sigma)}(s_k)$ is obtained as quotient of $I_X^{(\Sigma)}(s_k)$ and $N\hat{c}_X(0)$ (the empirical variance $\hat{c}_X(0) = \hat{\sigma}_X^2$ of the observations as defined in (2.1,2) multiplied with N, i.e., the number of observations). The equation in Problem 8.3 implies that $I_X^{(\Sigma/\sigma)}(s_k)$ is the quotient of (i) the sum of $I_X(s_k)$ for Fourier frequencies being less than or equal to s_k and (ii) the sum of all $I_X(s_k)$.

$$I_X^{(\Sigma/\sigma)}(s_k) = \frac{I_X^{(\Sigma)}(s_k)}{N\hat{\sigma}_X^2} = \frac{\sum_{-1/2<r_l\le s_k} I_X(r_l)}{\sum_{j=-(m-1)}^{m} I_X(s_j)} \tag{8.9}$$

$$\frac{C_X^{(I)}(s)}{\sigma_X^2} = \frac{\int_{-1/2}^{s} C_X(r)\mathrm{d}r}{\int_{-1/2}^{1/2} C_X(s)\mathrm{d}s} \tag{8.10}$$

In (8.10), the integrated spectrum $C_X^{(I)}(s)$ of (X_t) is normalised with the variance $c_X(0) = \sigma_X^2$ of (X_t), and the normalised integrated spectrum $C_X^{(I)}(s)/\sigma_X^2$ is obtained using (7.64) and (7.65) as well as (6.48,1). $C_X^{(I)}(s)/\sigma_X^2$ has the properties of a probability distribution seeing that $C_X(s)/\sigma_X^2$ has the properties of a probability density.

Two examples of normalised integrated spectra are plotted with solid lines in Fig. 8.2, i.e., those of a white noise process with $\mu_W = 0$ and $\sigma_W^2 = 1$ (on the left) and of $(Y_t^{(1)})$, the AR[2] model in Fig. 7.12 (on the right). A time slice in a realisation of $(Y_t^{(1)})$ and its spectrum $C_{Y^{(1)}}(s)$ as obtained in (7.97) are plotted in Fig. 7.12. The integrated spectrum of a white noise process (W_t), $C_W^{(I)}(s) = \int \sigma_W^2 \mathrm{d}s = \sigma_W^2 s + a$, $-1/2 \le s \le 1/2$ and a the constant of integration, is obtained by integrating the constant spectrum $C_X(s) = \sigma_W^2$ of (W_t) obtained in (7.76). $C_W^{(I)}(s)$ is normalised with the variance σ_W^2 as

required in (8.10) and, assuming $a = 1/2$, $(1/\sigma_W^2)C_W^{(I)}(s) = s + 1/2$, $-1/2 \leq s \leq 1/2$, is obtained, as plotted in Fig. 8.2 (a). $(1/\sigma_W^2)C_W^{(I)}(s)$ resembles a rectangular (in the remarks to (7.30)) probability distribution. The integrated spectrum of $(Y_t^{(1)})$ in Fig. 8.2 (b) results from a less easy integration.

From time slices of length $N = 256$ in realisations of both models, normalised cumulated periodograms are calculated as required in (8.9) and are plotted with broken lines in Fig. 8.2. Both step functions thus obtained are close to the integrated spectra. When the normalised cumulated periodograms are calculated from more than $N = 256$ simulated values in Problem 8.4, the absolute values of the maximal differences (normalised cumulated periodograms − normalised integrated spectra) decrease with increasing N. This result is to be expected, since the bias of a periodogram $I_X(s_k)$ reduces with increasing N as is concluded from the remarks to (8.21): the cumulated periodogram is an estimator for the integrated spectrum having a decreasing bias on condition that (i) N increases and (ii) the dynamic range of the spectrum is small $\left(\mathrm{dyn}\big(C_{Y^{(1)}}(s)\big) \approx 15 \text{ dB}\right.$ in Fig. 7.12), as is concluded from the remarks to Figs. 8.4, 8.5, 8.6 and 8.7.

The maximal absolute value z of differences $I_X^{(\Sigma/\sigma)}(s_k) - C_X^{(I)}(s_k)/\sigma_X^2$, as defined in (8.11), therefore lends itself as a candidate for a test statistic. In (8.11), (i) $I_X^{(\Sigma/\sigma)}(s_k)$ is the normalised cumulated periodogram calculated from a time slice in a realisation (x_t) stemming from a stationary stochastic process (X_t) with $\mu_X = 0$ and σ_X^2 and having a continuous spectrum $C_X(s)$ with a small dynamic range, and (ii) $C_X^{(I)}(s_k)/\sigma_X^2$ is the normalised integrated spectrum of (X_t). Since $I_X(0) \approx 0.0$ and $I_X(s_k)$ is even (6.32,1), the test statistic z is that of the m differences for $0 < s_k \leq 1/2$ which has the largest absolute value, with m and s_k as defined in (6.21).

$$z = \max_{0<k\leq m} \left| I_X^{(\Sigma/\sigma)}(s_k) - C_X^{(I)}(s_k)/\sigma_X^2 \right| \tag{8.11}$$

In definition (8.11), $C_X^{(I)}(s)/\sigma_X^2$ has the properties of a continuous distribution function $F(y)$ of random variable Y, and $I_X^{(\Sigma/\sigma)}(s_k)$ has the properties of a cumulated histogram, i.e., an empirical distribution $F_n(y)$ obtained from n observations of Y under the iid. assumptions (1.2), as is concluded from (8.7) on condition that (X_t) is a normal white noise process. Under this assumption (the null hypothesis) therefore, in the usual $z = \sup|F_n(y) - F(y)|$ (the Kolmogorov-Smirnov statistic as found in an introductory textbook), $F_n(y)$ can be substituted with $I_X^{(\Sigma/\sigma)}(s_k)$, $F(y)$ with $C_X^{(I)}(s_k)/\sigma_X^2$, and n with m. z in $z = \sup|F_n(y) - F(y)|$ and also in (8.11) is a realisation of random variable Z, which has a distribution $K(z)$ that does not depend on $F(y)$, as was shown by Kolmogorov and Smirnov, provided that $F(y)$ is a continuous probability distribution. The supremum sup is used when it is not clear whether a maximum exists. In the Kolmogorov-Smirnov statistic, sup is the upper bound of the set $|F_n(y) - F(y)|$.

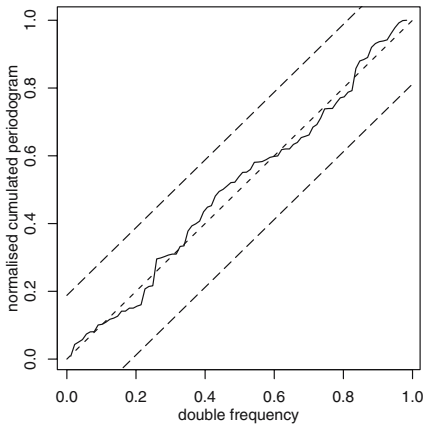


Fig. 8.3. The normalised cumulated periodogram of the NAO index is calculated from the periodogram in Fig. 8.1 (without the periodogram values for $s_k = 0$ and $s_k = 1/2$) and thereafter plotted against $2 \times s_k$, $0 < s_k \leq 1/2$ (solid line). The normalised integrated spectrum under the null hypothesis (the NAO index stems from a white noise process) is plotted as diagonal, together with the subdiagonals defining a .95 confidence interval of z (the Kolmogorov-Smirnov statistic) in (8.11).

Hence, assuming that the process under analysis (X_t) has a continuous spectrum $C_X(s)$, a Kolmogorov-Smirnov test for white noise can be performed when a time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of (X_t) is available. The null hypothesis is that $C_X(s)$ is constant, the alternative that it is not constant. This test is performed by calculating the periodogram $I_X(s_k)$, and then plotting the normalised cumulated periodogram $I_X^{(\Sigma/\sigma)}(s_k)$ as defined in (8.9) against $2 \times s_k$, i.e., twice the Fourier frequencies. Thereafter, the normalised integrated spectrum under the null hypothesis is drawn as the diagonal, as demonstrated in Fig. 8.2 (a), and, in vertical distance $|z|$ from the integrated spectrum, two parallel lines are drawn. z is obtained from the Kolmogorov-Smirnov statistic $K(z)$, given an appropriate significance level.

For example, does the NAO index (Jones) plotted in Fig. 2.10 stem from a white noise process? Using its periodogram obtained in Fig. 8.1, a test as described above is performed in Fig. 8.3. The normalised cumulated periodogram of the NAO index (Jones) fluctuates about the diagonal (the normalised integrated spectrum of a white noise process) safely within the .95 confidence interval obtained from $K(z = 0.188) = .95$, the distribution of the test statistic. Consequently, the null hypothesis is assumed, i.e., that the observations of the NAO index (Jones) are a time slice from a realisation of a white noise process. The outcome of this test is in line with the properties of the NAO index (Jones) time series obtained in Sects. 2.3.1, 2.5.1, 2.5.2, and Problem 2.33.

When (8.11) is applied to identify a white noise process the alternative is that the observations stem from a process with a non-constant continuous spectrum. There are, however, processes with a non-constant spectrum that is mixed. If the spectrum of the process under analysis is either constant (the null hypothesis) or constant with superimposed spikes as demonstrated in (7.82) (the alternative), then a test can be performed using the statistics in (8.13) and (8.15).

In this second case, it is assumed that the observations (x_t), $t = 0, \ldots, N-1$, stem from a stationary process (X_t) having a mixed spectrum: $C_X(s) = C_U(s) + C_V(s)$, with $C_U(s)$ being the flat spectrum of a white noise process (U_t) and $C_V(s)$ the spectrum (7.74) of a harmonic process (V_t), as defined in (7.30). An example of such a spectrum is given in (7.82) and Fig. 7.9. It is further assumed that (i) both (U_t) and (V_t) have the properties required for the derivation of (7.80) and (7.81), and (ii) $s_j = s_k$ for frequencies s_j, $j = 1, \ldots, n$, in (7.30) and Fourier frequencies s_k in (6.21,1). If N increases then Δs_k decreases, and thus assumption (ii) becomes more likely.

Under the null hypothesis, this model degenerates to a white noise process $(X_t) = (U_t) = (W_t)$ having a constant spectrum and, consequently, the null hypothesis is rejected on condition that the periodogram contains a value that is significantly larger than the average value. A test performing this comparison can be constructed using the properties of the normalised periodogram calculated from a realisation of a white noise process, as obtained in (8.7). From $\chi^2_{(n)}$, $n = 1, 2$, in (8.7,3), the probability distribution in (8.12) is arrived at on condition that the observations (x_t), $t = 0, \ldots, N-1$, stem from a white noise process.

$$\Pr\left(\frac{2I_W(s_k)}{\sigma_W^2} \le y\right) = \int_0^y f_{\chi^2_{(2)}}(z)\mathrm{d}z = \int_0^y \frac{1}{2}\mathrm{e}^{-z/2}\mathrm{d}z = 1 - \mathrm{e}^{-y/2} \qquad (8.12)$$

$$I_W(s_k), \sigma_W^2, s_k, m \quad \text{as in (8.7)}, \; s_k \neq 0, 1/2$$

Using (8.12), the probability of $I_W^{(max)}$ being larger than y, $\Pr(I_W^{(max)} > y)$ is obtained in (8.14), $I_W^{(max)}$ being the maximal value in the periodogram as defined in (8.13). This is the probability that all periodogram values are less than or equal to y which is obtained by calculating the probability of the complementary event, using (8.7,1).

$$I_W^{(max)} = \Big(\max_{1 \le k \le m} 2I_W(s_k)\Big) \Big/ \sigma_W^2 \qquad (8.13)$$

$$\begin{aligned} \Pr(I_W^{(max)} > y) &= 1 - \Pr\Big(\big(I_W(s_1)/\sigma_W^2 \le y\big) \cap \big(I_W(s_2)/\sigma_W^2 \le y\big) \cap \ldots \\ &= \qquad \ldots \cap \big(I_W(s_m)/\sigma_W^2 \le y\big)\Big) = 1 - \big(1 - \mathrm{e}^{-y/2}\big)^m \end{aligned} \qquad (8.14)$$

The statistic $I_W^{(max)}$ in (8.13) was introduced in 1898 by Schuster, the above introduction is borrowed from [113].

$I_W^{(max)}$ can be calculated on (the additional) condition that the variance σ_W^2 of the white noise process is known. Often, however, σ_W^2 is not known and has to be estimated by calculating the empirical variance using either the result obtained in Problem 8.3 or directly (2.1,2). Consequently, Schuster's statistic $I_W^{(max)}$ in (8.13) can be applied provided that N is large and it is accounted for that the empirical variance is not a robust estimate.

The estimation of σ_W^2 can be circumvented when the g statistic in (8.15) is calculated from the periodogram. This statistic is the quotient of the maximal periodogram value and the sum of all periodogram values.

$$g = \frac{\Big(\max\limits_{1\le k\le m} 2I_W(s_k)\Big)\Big/\sigma_W^2}{\sum\limits_{k=1}^{m}\big(2I_W(s_k)/\sigma_W^2\big)} = \frac{\max\limits_{1\le k\le m} I_W(s_k)}{\sum\limits_{k=1}^{m} I_W(s_k)} \tag{8.15}$$

$$\Pr(g > z) = \sum_{l=1}^{q}(-1)^{l-1}\binom{m}{l}(1-lz)^{m-1} \tag{8.16}$$

$$\approx 1-(\alpha/m)^{(1/(m-1))} \tag{8.17}$$

The probability of the g statistic as given in (8.16) was derived in 1929 by Fisher, the approximation in (8.17) is used in [101], and the above introduction is borrowed from [113] (as those of Schuster's test). Above in (8.15), (8.16) and (8.17), $I_W(s_k)$, s_k, m, N and k are as required in (8.7), and q is the largest integer number being less than $1/z$.

Fisher's g statistic in (8.15) is applied as follows. A significance level α is chosen such that $\Pr(g > z) = \alpha$ under the null hypothesis. The null hypothesis is discarded (i.e., the observations are assumed to stem from a process (X_t) having a mixed spectrum $C_X(s) = C_U(s) + C_V(s)$ as introduced above) on condition that g is larger than the critical value z.

For example, the NAO index (Jones) time series as plotted in Fig. 2.10 is found to stem from a white noise process in Fig. 8.3, where it is tested against the alternative that it stems from a process having a continuous and non-constant spectrum. Is this result confirmed when the alternative is that this time series stems from a process with a mixed spectrum? Fisher's statistic $g = 0.079$ for the NAO index (Jones) is obtained using (8.15) as the quotient of (i) the maximum 8.00 (attained at frequency $s_k = 0.1299 = 23/177$) of the periodogram in Fig. 8.1 and (ii) the sum 100.95 of all periodogram values. For $m = 87$ and significance levels 0.05 and 0.01, critical values 0.083 and 0.100 are obtained and therefore the alternative is discarded. This result confirms that the NAO index (Jones) stems from a white noise process, in agreement with the results obtained in Sects. 2.3.1, 2.5.1 and 2.5.2, Problem 2.33 and Fig. 8.3.

Fisher's g statistic provides for the most powerful test against the alternative that the observations stem from model (X_t) having a mixed spectrum $C_X(s) = C_U(s) + C_V(s)$ as defined above, where $C_V(s)$ is the spectrum (7.74)

of a harmonic process (V_t) as defined in (7.30) with $n = 1$, i.e., having only one oscillation with frequency $s_l = s_k$, $k = 1, \ldots, m$, coincident with one of the Fourier frequencies s_k.

If $n > 1$ oscillations with unknown frequencies are assumed to be hidden in the observed time series, then a multiple test procedure can be constructed from Fisher's statistic. For example, the g statistic and the corresponding p value can be calculated for each of the n' large values found in the periodogram. Thereafter, the p values are ordered to obtain a sequence $\big(p^{(1)}, p^{(2)}, \ldots, p^{(n')}\big)$. The null hypothesis is rejected for all oscillations with statistics $\big(g^{(1)}, g^{(2)}, \ldots, g^{(i_q)}\big)$ and p values $\big(p^{(1)}, p^{(2)}, \ldots, p^{(i_q)}\big)$ where i_q is the largest i for which $p^{(i)} \leq (i/n')q$. This procedure controls the false discovery rate at level q [12]. It is implemented in GeneTS [147], an R package available from the Comprehensive R Archive Network [114].

Alternatives to this multiple test procedure can be found in [108]. For example, Siegel proposed in [124] a correction of Fisher's statistic to account for multiple oscillations.

Above in this section, two periodogram tests for white noise are introduced: (i) the test against the alternative that the spectrum of the process under analysis is continuous and non-constant using statistic (8.11), and (ii) the test against the alternative that the spectrum of the process is constant with superimposed spikes generated by the oscillations in a harmonic process using statistic (8.15). In both cases, the probability distributions of the test statistics are derived from properties (8.7) of the periodogram under the null hypothesis that the observed time series stems from a normal white noise process. If, however, the null hypothesis is discarded, then the underlying process is assumed not to be a white noise process, i.e., a stochastic process having either a non-constant continuous (examples are given in Sect. 7.4.2) or a mixed spectrum (as shown for the simplest case with only one oscillation in Fig. 7.9). Assume that a periodogram is calculated from an observed time series stemming from such a process. Does this periodogram then have all properties (8.7)? Or only certain ones? Answers to these questions are given in Sects. 8.3 and 8.4.

8.3 Its Expectation Function

From (8.29) it is concluded that a periodogram as defined in (8.1,4) is the Fourier transform of the empirical covariance function as defined in (2.1,3), each calculated from a time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of a discrete-time stationary stochastic process (X_t) with $\mu_X = 0$, on condition that $\hat{c}_X(\tau) = 0$ for $|\tau| \geq N$ is assumed. Taking the expectations, (8.18) readily follows, and using $\mathrm{E}\big(\hat{c}_X(\tau)\big)$, i.e., the expectation function of the empirical covariance function as obtained in (2.57), the expectation function of the periodogram is arrived at in (8.19). This result holds for all discrete-time

stationary stochastic processes (X_t) seeing that only stationarity as defined in (2.8) is required to obtain $\mathrm{E}\big(\hat{c}_X(\tau)\big)$ in (2.57).

$$\mathrm{E}I_X(s) = \mathrm{E}\left(\sum_{\tau=-(N-1)}^{N-1} \hat{c}_X(\tau)\cos(2\pi s\tau)\right) \tag{8.18}$$

$$= \sum_{\tau=-(N-1)}^{N-1} \big(\mathrm{E}\hat{c}_X(\tau)\big)\cos(2\pi s\tau)$$

$$= \sum_{\tau=-(N-1)}^{N-1} \big((1-|\tau|/N)\big)\big(c_X(\tau)\big)\cos(2\pi s\tau) \tag{8.19}$$

$\mathrm{E}I_X(s)$, $I_X(s)$ being the periodogram as defined in (8.1,4), is the Fourier transform of the covariance function $c_X(\tau)$, which is weighted using the even triangle sequence $\Lambda^{(e)}_{2N+1}(t)$ as defined in (6.56). In (8.19), $(1-|\tau|/N)c_X(\tau)$ converges to $c_X(\tau)$ for $N \to \infty$, and thus $\hat{c}_X(\tau)$ is an asymptotically unbiased estimator of $c_X(\tau)$.

If (X_t) has a continuous spectrum $C_X(s)$, then $C_X(s)$ and $c_X(\tau)$ are a Fourier transform pair as defined in (7.65) and (7.66) as well as in (6.49) and (6.50). This Fourier transform pair is discussed in Sect. 7.5.4. In this case therefore, $\mathrm{E}\big(I_X(s)\big)$ converges to $C_X(s)$ for $N \to \infty$. Consequently, the periodogram is an unbiased estimator for a continuous spectrum $C_X(s)$ on condition that $N \to \infty$.

If (X_t) has a non-continuous spectrum, i.e., a line or a mixed spectrum as defined in (7.71,2,3), then $c_X(\tau)$ is calculated in (7.61) as a Fourier-Stieltjes integral from the integrated spectrum $C^{(I)}_X(s)$ of (X_t) and, vice-versa, $C_X(s)$ can be obtained from a Fourier transform of $c_X(\tau)$, as demonstrated for the case of the harmonic process in (7.72). Thereafter, $C_X(s)$ can be integrated to arrive at $C^{(I)}_X(s)$. Since $\hat{c}_X(\tau)$ is an asymptotically unbiased estimator of $c_X(\tau)$, even though (X_t) has a non-continuous spectrum, the cumulated periodogram is an unbiased estimator for $C^{(I)}_X(s)$ on condition that $N \to \infty$.

The results inferred above can also be derived from (8.35) (repeated in (8.20)) using definition (7.51,3) to arrive at (8.21), since $F_N(s)$ is a defining sequence for $\delta(s+n)$, $n = \ldots, -1, 0, 1, \ldots$, i.e., a delta function being periodic with period 1 as introduced in the remarks to (6.63), and $\delta(s)$ is the identity element under convolution (as argued in the remarks to (6.68)). Obviously, (8.19) and (8.21) are equivalent, as implied by convolution theorems (6.48,6,7), because $(1-|\tau|/N) = \Lambda^{(e)}_{2N+1}$ and $F_N(s)$ are a Fourier transform pair as derived in (6.57), and $c_X(\tau)$ and $\mathrm{d}C^{(I)}_X(s)$ are also a Fourier transform pair as concluded from the cases given in Sect. 7.3.3.

$$\mathrm{E}I_X(s) = \int_{-1/2}^{1/2} F_N(s-s')\mathrm{E}\big(|\mathrm{d}Z_X(s')|^2\big) \quad \begin{matrix}\text{convolution}\\ \text{as in (6.144)}\end{matrix} \tag{8.20}$$

$$= \int_{-1/2}^{1/2} F_N(s-s')\mathrm{d}C_X^{(I)}(s) = F_N(s) * \mathrm{d}C_X^{(I)}(s) \tag{8.21}$$

$$\begin{aligned}\mathrm{bias}\big(I_X(s)\big) &= -\mathrm{E}I_X(s) + \mathrm{d}C_X^{(I)}(s) \\ &= \mathrm{d}C_X^{(I)}(s) - F_N(s) * \mathrm{d}C_X^{(I)}(s)\end{aligned} \tag{8.22}$$

For finite N, however, convolution integral (8.21) implies $\mathrm{E}\big(I_X(s)\big) \neq \mathrm{d}C_X^{(I)}(s)$. Consequently, a bias has to be taken into account when the spectrum of a stationary stochastic process is estimated using the periodogram: the spectrum is over- or underestimated in the mean when many estimations are performed. The bias of the periodogram as obtained in (8.22) clearly depends on $\mathrm{d}C_X^{(I)}(s)$ and $F_N(s)$, the kernel used to convolve $\mathrm{d}C_X^{(I)}(s)$ in (8.21). The Fejer kernel $F_N(s)$ is defined in (6.57) with its properties enumerated in (6.58) and (6.59). Since $F_N(s)$ depends on N, the length of the observed time series, the bias also depends on N. Subsequent to taking the logarithms,

$$\begin{aligned}\log\big(\mathrm{d}C_X^{(I)}(s)\big) - \log\big(\mathrm{E}I_X(s)\big) &= \log\left(\frac{\mathrm{d}C_X^{(I)}(s)}{\mathrm{E}I_X(s)}\right) = -\log\left(\frac{\mathrm{E}I_X(s)}{\mathrm{d}C_X^{(I)}(s)}\right) \\ &= -\log\left(1 + \frac{\mathrm{E}I_X(s) - \mathrm{d}C_X^{(I)}(s)}{\mathrm{d}C_X^{(I)}(s)}\right) \\ &\approx -\frac{\mathrm{E}I_X(s) - \mathrm{d}C_X^{(I)}(s)}{\mathrm{d}C_X^{(I)}(s)}\end{aligned} \tag{8.23}$$

(8.23) becomes large on condition that $\mathrm{d}C_X^{(I)}(s)$ is small. This result is obvious because $\log\big(\mathrm{E}I_X(s)\big) \neq \mathrm{E}\big(\log I_X(s)\big)$ as is concluded from Problem 8.2.

For example, a realisation of $\big(Y_t^{(2)}\big)$, the AR[2] model introduced in Fig. 7.13, is plotted in Fig. 8.4 (a). Time slices from this simulation are used to calculate periodograms from values realised in the intervals $t = 101, \ldots, 164$ and $t = 101, \ldots, 356$ as required in (8.1,3). These periodograms are shown below in Fig. 8.4, plots (b) and (c), together with the model spectrum having bandwidth 0.02 and dynamic range 35 dB, both approximated from Fig. 7.13 (b) and (c) by applying definitions (7.83) and (6.96). This periodogram demonstrates (despite its large fluctuations, a property shared with the periodogram of the NAO index (Jones) estimated in Sects. 8.1 and 6.1.3) that (i) both periodograms are larger than the spectrum at low frequencies, and (ii) the periodogram for $N = 256$ is closer to the peak of the spectrum than the one for $N = 64$.

The simulation experiment in Fig. 8.4 and the similar ones performed in Problem 8.6 demonstrate that the bias of the periodogram (i) depends on N, i.e., the length of the simulated time series, (ii) becomes negligible with increasing N at the peak of the spectrum, and (iii) persists at frequencies a larger distance away from a peak where the spectrum is small. These are properties similar to those of the distortions due to leakage demonstrated in

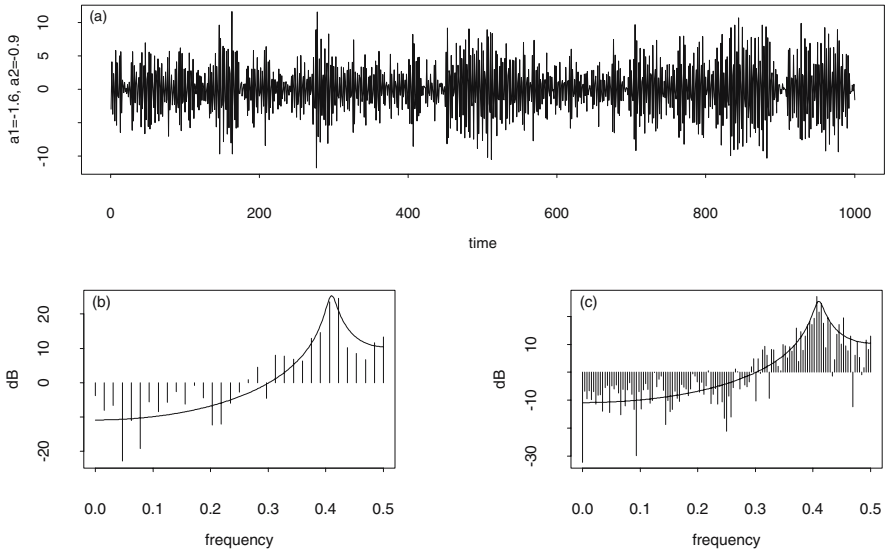


Fig. 8.4. Above, a realisation $(y_t^{(2)})$ of AR[2] model $Y_t^{(2)} = -1.6Y_{t-1}^{(2)} - 0.9Y_{t-2}^{(2)} + W_t$ with $\mu_W = 0$ and $\sigma_W^2 = 1$, as introduced in Fig. 7.13; below periodograms calculated from values realised in the intervals $t = 101, \ldots, 164$ (on the left) and $t = 101, \ldots, 356$ (on the right), together with the model spectrum $C_{Y^{(2)}}(s)$.

Sect. 6.5.3. The similarity becomes obvious when convolutions $D_N(s) * G(s)$ in (6.95) and $F_N(s) * \mathrm{d}C_X^{(I)}(s)$ in (8.21) are compared: the differences in the distortions are due to the different kernels. Thus, the bias in the periodogram is said to be due to leakage. When $F_N(s)$ and $D_N(s)$ are compared, the following differences are found:

1. $F_N(s) \geq 0$ for all s, and therefore $F_N(s) * \mathrm{d}C_X^{(I)}(s) \geq 0$ (since $\mathrm{d}C_X^{(I)}(s) \geq 0$), whereas $D_N(s)$ has negative side lobes.
2. $F_N(s)$ has a slightly wider main lobe than $D_N(s)$: $\mathcal{B}_{\triangle}^{\cap}\big(D_N(s)\big) \approx 6/(5N)$ and $\mathcal{B}_{\triangle}^{\cap}\big(F_N(s)\big) \approx 7/(8N)$, as obtained in the remarks 6.97. Above in Fig. 8.5, for example, $\mathcal{B}_{\triangle}^{\cap}\big(F_{64}(s)\big) \approx 0.015$ and $\mathcal{B}_{\triangle}^{\cap}\big(F_{256}(s)\big) \approx 0.0034$, whereas $\mathcal{B}_{\triangle}^{\cap}\big(D_{301}(s)\big) \approx 0.004$ in Fig. 6.24. Since the difference in the widths of $D_N(s)$ and $F_N(s)$ is small, diagnostics (6.115) and (6.116) to detect the leakage due to the convolution with the Dirichlet kernel in (6.95) also allow for approximating the length of the observational period such that the resolution in the frequency domain remains high enough when the spectrum is convolved with the Fejer kernel in (8.21). This approximation is however restricted to spectra with known bandwidths (7.83). If the bandwidth of the spectrum to be estimated is not known, then the procedures introduced in Sect. 9.3.5 can be applied (despite the estimator being the periodogram, and not a direct spectral estimator).

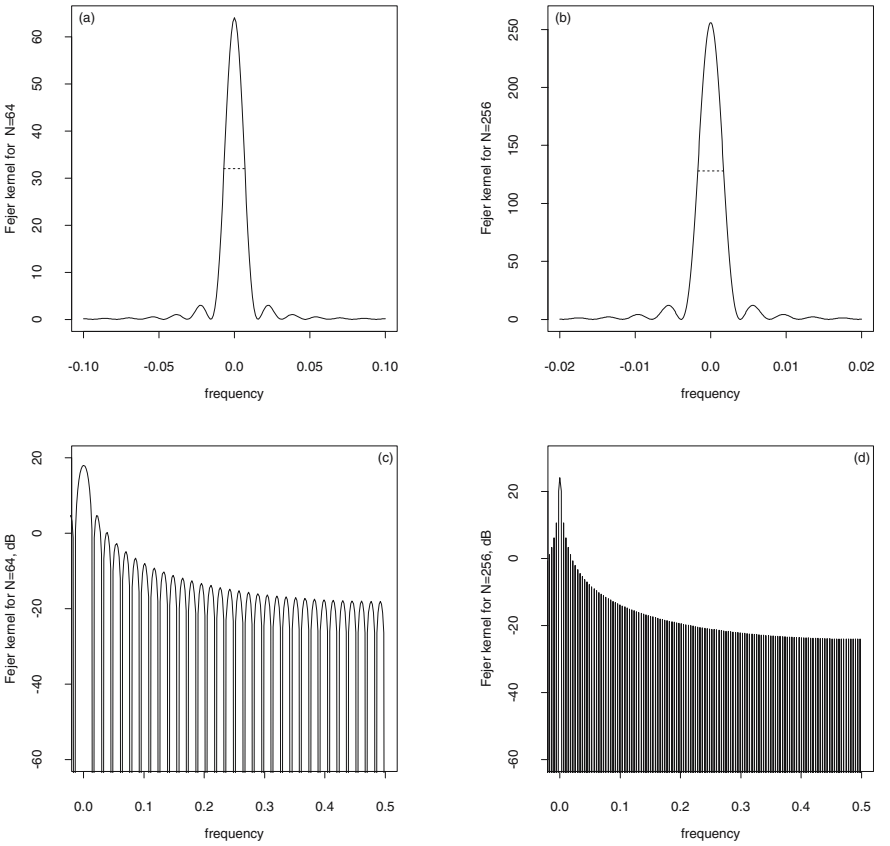


Fig. 8.5. Fejer kernels $F_N(s)$ for $N = 64$ (on the left) and $N = 256$ (on the right): plots showing main lobes and widths at half height (above) and plots in dB (below).

3. $F_N(s)$ has smaller side lobes than $D_N(s)$. For example, Fejer kernels for $N = 64$ and $N = 246$ as plotted below in Fig. 8.5 have side lobes smaller than -15 dB for large frequencies, whereas the side lobes of Dirichlet kernels are approximately 0 dB even for large N as can be seen in Fig. 6.22. Consequently, (i) the leakage due to the side lobes of the kernel is smaller in (8.21) than in (6.95) and (ii) when applying diagnostic (6.117) to convolutions (8.21) too pessimistic conclusions are drawn.
4. $F_N(s)$, $N = 1, 2, \ldots$, is a defining sequence for $\delta(s+n)$, $n = \ldots, -1, 0, 1, \ldots$, i.e., a delta function being periodic with period 1 as introduced in the remarks to (6.63), having the implications in the remarks to (8.21).

When performing simulation experiments such as those in Fig. 8.4, the bias of the periodogram due to leakage generated by convolution (8.21) becomes apparent when the known spectrum of the process is plotted together

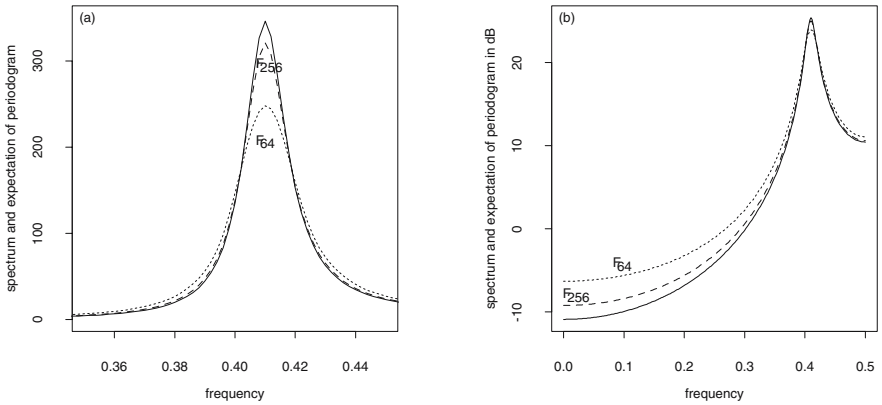


Fig. 8.6. Spectrum $C_{Y^{(2)}}(s)$ of AR[2] model in Fig. 8.4 (solid line) together with $\mathrm{E}I_{Y^{(2)}}(s)$ resulting from the convolution of $C_{Y^{(2)}}(s)$ with Fejer kernels $F_{64}(s)$ and $F_{256}(s)$ as required in (8.21) (broken lines).

with periodograms calculated from time slices of simulated realisations, as demonstrated in Fig. 8.4. In addition to simulation experiments, (8.21) lends itself to demonstrating the bias due to leakage given $\mathrm{d}C_X^{(I)}(s)$ and N, i.e., the spectrum of the process and the number of simulated values used for the calculation of the periodogram: because both functions on the left side in (8.21) are known, $\mathrm{E}\big(I_X(s)\big)$ is easily obtained, and, when plotted together with $\mathrm{d}C_X^{(I)}(s)$, a possible bias will become visible, as demonstrated in Figs. 8.6 and 8.7 for the case of continuous spectra.

In Fig. 8.6 for example, $F_{64}(s) * C_{Y^{(2)}}(s)$ and $F_{254}(s) * C_{Y^{(2)}}(s)$ as required in (8.21), $C_{Y^{(2)}}(s)$ being the spectrum of the AR[2] model introduced in Fig. 7.13, are plotted together with $C_{Y^{(2)}}(s)$. From these plots it becomes obvious that (in the mean when many simulations are performed) a periodogram obtained from $N = 256$ values is an estimator having a bias that (i) is negligible at the peak of the spectrum and (ii) amounts to approximately 2 dB at the low frequencies. These results are in line with those obtained from diagnostics (6.115), (6.116) and (6.117) (modified for the Fejer kernel in the remarks to Fig. 8.4): (i) $\mathcal{B}_{\triangle}^{\cap}\big(F_{256}(s)\big) \approx 0.0034$, the width at half height of the Fejer kernel for $N = 256$ is much smaller than $\mathcal{B}_{\triangle}\big(C_{Y^{(2)}}(s)\big) = 0.02$, the bandwidth (7.83) of the spectrum as plotted in Fig. 7.13 (b), and (ii) $\mathrm{dyn}\big(C_{Y^{(2)}}(s)\big) = 35$ dB, the dynamic range of the spectrum as defined in (6.96) is larger than the threshold in (6.117), indicating that leakage due to the side lobes of the Fejer kernel is likely to bias the result.

With increasing dynamic range and decreasing bandwidth of the spectrum under analysis the bias in the periodogram becomes more severe. Expanding the observational period (when possible), a number of sufficiently large N observations can be obtained such that the bias due to the main lobe of

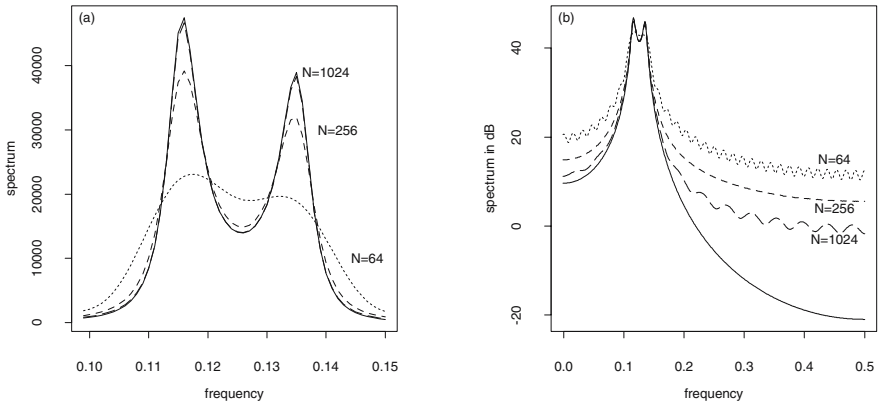


Fig. 8.7. Spectrum $C_X(s)$ of AR[4] model (X_t) in Fig. 7.14 (solid line) together with $\mathrm{E}I_X(s)$ for $N = 64$, $N = 256$ and $N = 1024$ (broken lines), i.e., the expectation functions of periodograms obtained from 64, 256 and 1024 observations, resulting from convolutions of the $C_X(s)$ with $F_{64}(s)$, $F_{256}(s)$ and $F_{1024}(s)$ as required in (8.21).

the Fejer kernel becomes negligible. The bias due to its side lobes, however, persists, since these, although being lower than those of the Dirichlet kernel, do not decay rapidly with increasing distance from the main lobe, and thus do not become small enough ($< 10^{-4}$) to reduce the leakage effectively, as can be seen in Fig. 6.11.

For example, the AR[4] model (X_t) introduced in Fig. 7.14 has a spectrum with bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big) = 0.00474$ obtained in Fig. 7.10 (d) and a dynamic range $\mathrm{dyn}\big(C_X(s)\big) = 65$ dB in $0 \le s \le 0.5$ as can be seen in Fig. 7.14 (b). If realisations of (X_t) having $N = 64$, $N = 256$ and $N = 1024$ values are simulated and used to calculate periodograms then, in the mean over many realisations, these periodograms will be afflicted with the bias shown in Fig. 8.7. From simulation experiments performed in Problem 8.7 using this AR[4] model, it is concluded that the bias due to the side lobes of the Fejer kernel in (8.21) persists with increasing N and increases with decreasing $C_X(s)$, in accordance with (8.23).

Fig. 8.7 demonstrates that, when trying to estimate a spectrum by calculating a periodogram from an observed time series of length N stemming from a stationary stochastic process and thereafter taking the logarithm, a bias in the order of magnitudes is possibly obtained even when N is very large (examples with $N > 10^6$ are given in [134]). Such a large bias is clearly a risk that a gambler can take, not a data analyst! Hence, the periodogram is not a recommended estimator for the spectrum. An exception is the estimation of a spectrum having a small dynamic range on condition that the number of observations is large enough.

Estimators for the spectrum are introduced in Chaps. 9 and 10.

8.4 Its Covariance Function and its Distribution

A linear process (X_t) (2.25) is the output of a stochastic LTI filter (7.84) with (W_t), i.e., a white noise process, being its input and (b_t) its impulse response function. The output spectrum $C_X(s)$ is obtained by multiplying the squared absolute value of the frequency response function $B(s)$ with $C_W(s)$, the spectrum of the input: $C_X(s) = B(s)\overline{B(s)}C_W(s) = |B(s)|^2\sigma_W^2$, as required in (7.90). This equation becomes an approximation when $C_X(s)$ and $C_W(s)$ are substituted with periodograms $I_X(s)$ and $I_W(s)$, both calculated as required in (8.1,4) from (x_t) and (w_t), $t = 0, 1, \ldots, N-1$, on condition that (i) (x_t) is the output generated by the input (w_t) and (ii) $N \to \infty$. This result is derived in Sect. 8.5.2 and summarised in (8.48).

Using (8.48) as well as (1.15) and (1.16), approximations (8.24) for the variance and covariance functions as well as the probability distribution of $I_X(s)$ can be derived from the properties of $I_W(s)$ summarised in (8.7).

Let $I_X(s_k)$ and $I_X(s)$ be the periodogram calculated as required in (8.1,3,4) from a time slice (x_t), $t = 0, \ldots, N-1$, in a realisation of a linear process (X_t) as defined in (2.25) having a continuous spectrum $C_X(s)$. Then, for $N \to \infty$:

1. *$I_X(s_k)$ and $I_X(s_l)$ are approximately independent for $k \neq l$*
2. *the normalised periodogram*
   $\big(2/C_X(s_k)\big)I_X(s_k)$ is approximately $\chi^2_{(2)}$ distributed for $-1/2 < s_k < 0 < s_k < 1/2$, and
   $\big(1/C_X(s_k)\big)I_X(s_k)$ is approximately $\chi^2_{(1)}$ distributed for $s_k = -1/2, 0, 1/2$

(8.24)

3. $\mathrm{E}I_X(s_k) = \int_{-1/2}^{1/2} F_N(s_k - r)C_X(s)\mathrm{d}r \to C_X(s)$
4. $\mathrm{Cov}\big(I_X(s_k), I_X(s_l)\big) \approx 0$ *for $k \neq l$*
5. $\mathrm{Var}I_X(s_k) \approx \begin{cases} C_X^2(s_k) & \text{for } -1/2 < s_k < 0 < s_k < 1/2 \\ 2C_X^2(s_k) & \text{for } s_k = -1/2, 0, 1/2 \end{cases}$
6. $\mathrm{Cov}\big(I_X(s), I_X(s+r)\big) \approx 0$, *$r$ being small*
7. $\mathrm{Var}I_X(s) \approx \begin{cases} C_X^2(s) & \text{for } -1/2 < s < 0 < s < 1/2 \\ 2C_X^2(s) & \text{for } s = -1/2, 0, 1/2 \end{cases}$

Properties (8.24) also apply (duly modified for a spectrum being possibly non-continuous) to periodograms calculated from realisations of almost all discrete-time stochastic processes which are stationary and ergodic. Example stochastic processes having periodograms with properties (8.24) despite their integrated spectra are not differentiable for all $-1/2 \leq s \leq 1/2$ are given in [117] or [153].

(8.24,1) implies that random variables $I_X(s_k)$ and $I_X(s_l)$ have negligible correlations for Fourier frequencies $s_k \neq s_l$. A realisation of a random function having this property fluctuates wildly as demonstrated by (i) the periodograms calculated from a simulation of an AR[2] process in Fig. 8.4

and (ii) the empirical spectra in Fig. 6.5 (which are identical with the periodograms as shown in the remarks to Fig. 8.1). Despite $\Delta s_k = 1/N \to 0$ for $N \to \infty$, $I_X(s_k)$ and $I_X(s_l)$ remain approximately independent for $k \neq l$.

Approximations (8.24,2) can be derived for finite N on condition that (X_t) is a normal process as defined in (2.3), a property induced by a normal input process (W_t), as proposed in (5.56). If (X_t) is normally distributed then $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$, both being linear combinations of the (X_t) as defined in (8.1,1), are normally distributed as well. Furthermore, because $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ are normally distributed, it is concluded (as in the case of a periodogram stemming from a white noise process in the remarks to (8.8)) that a periodogram stemming from the realisation of a normal linear process is $\chi^2_{(n)}$-distributed, $n = 1, 2$, on condition that it is normalised as proposed in (8.24,2). This normalisation is arrived at using (8.48). These approximations also apply to a periodogram calculated from a non-normal linear process on condition that $N \to \infty$, as shown, e.g., in [21]. Instructive examples can be obtained when the simulations in Problem 8.8 are performed.

The first moment function in (8.24,3) is a special case of (8.21). The second moment functions in (8.24,4,5) imply that a periodogram, calculated for Fourier frequencies s_k, fluctuates wildly (neighbour values are either not correlated or have only very small correlations) about the spectrum with a variance that is proportional to the square of the spectrum. These properties are shared with a periodogram that is calculated for all real frequencies $-1/2 \leq s, s + r \leq 1/2$: although the variances $\mathrm{Var}\big(I_X(s)\big)$ and covariances $\mathrm{Cov}\big(I_X(s), I_X(s+r)\big)$ are (among others) a function of N, (8.24,6,7) hold for $N \to \infty$. A more accurate version of (8.24,6) can be obtained from the solid curve in Fig. 9.6, which is arrived at in (9.14) on condition that the process under analysis has the properties enumerated in (9.13).

(8.24,6,7) imply that the variance of a periodogram does not decrease with increasing N: it merely comes closer to the square of the spectrum. Consequently, the variance of the periodogram is not reduced when the number of observations is increased and, therefore, the periodogram is not a consistent estimator as defined in (1.4). Non-consistent estimators are unusual, because the variance of all estimators introduced so far decreases when the number of observations increases. For example, consistent estimators are the sample mean (in the remarks to (1.4)), the mean of an ergodic time series (as is concluded from (2.56)) and also the estimators for the parameters of an AR[p] model (as is demonstrated in Sect. 5.2.4).

(8.24,6,7) are not only unusual but also inconvenient for three reasons: (i) an estimator having a smaller variance than the squared spectrum is required, (ii) the periodogram is not smooth whereas the spectrum is often smooth or contains smooth segments (confined by the frequencies with jump discontinuities in the integrated spectrum), and (iii), a wildly fluctuating periodogram can suggest that the spectrum of the process under analysis

is discrete or mixed, even though it is continuous (7.71). Spectral estimators that have a smaller variance than the periodogram are introduced in Sect. 9.3.

8.5 Supplements

Two derivations of the expectation function of a periodogram are given in Sect. 8.5.1. In Sect. 8.5.2, an approximation relating a periodogram calculated from a linear process with a periodogram calculated from its innovations is arrived at.

8.5.1 Expectation Function of the Periodogram

For the following derivations, it is assumed that (x_t), $t = 0, \ldots, N-1$, is a time-slice in a realisation of a stationary discrete-time stochastic process (X_t) with expectation function $\mu_X = 0$ and Fourier representation $X_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st} \mathrm{d}Z_X(s)$ as required in (7.51).

$I_X(s)$ in (8.25) is the periodogram as defined in (8.1,4). The sums in (8.25) can be interpreted as polynomials. In their squares, therefore, the terms can be arranged in rows and columns using two indices and thus the double-indexed sums in the following two lines are arrived at. In each term in these sums, the product of two trigonometric functions is substituted using the elementary trigonometric identities, and the sum (8.26) readily follows.

$$
\begin{aligned}
N I_X(s) &= \left(\sum_{t=0}^{N-1} X_t \cos(2\pi st) \right)^2 + \left(\sum_{t=0}^{N-1} X_t \sin(2\pi st) \right)^2 && (8.25)\\
&= \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} X_t \cos(2\pi st) X_u \cos(2\pi su) + \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} X_t \sin(2\pi st) X_u \sin(2\pi su)\\
&= \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} X_t X_u \frac{1}{2} \big(\cos(2\pi s(t-u)) + \cos(2\pi s(t+u)) \big) +\\
&\qquad \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} X_t X_u \frac{1}{2} \big(\cos(2\pi s(t-u)) - \cos(2\pi s(t+u)) \big)\\
&= \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} X_t X_u \cos\big(2\pi s(t-u)\big) && (8.26)\\
&= \sum_{\tau=-(N-1)}^{N-1} \left(\sum_{t=0}^{N-1-|\tau|} X_t X_{t+|\tau|} \right) \cos(2\pi s\tau) && (8.27)
\end{aligned}
$$

The argument $2\pi s(t-u)$ of the cosine functions in (8.26) is substituted with $2\pi s\tau$ and thereafter the diagonals in the matrix of the terms are summed, as

demonstrated when deriving (2.73) or in Problem 6.16. The main diagonal contains N terms X_t^2, the first sub-diagonals $N-1$ terms X_tX_{t-1} as well as X_tX_{t+1}, etc., and (8.27) is arrived at. The sum in parenthesis in (8.27) is the autocorrelation (6.103) of the X_t, $t = 0, 1, \ldots, N-1$ (and thus an even function). If $X_tX_{t+\tau} = 0$ for $|\tau| \geq N$ is assumed, then this sum and $NI_X(s)$ become a Fourier transform pair as proposed in (6.49) and (6.50).

(8.27) is written as in (8.28) and, thereafter, the right side is compared with the empirical covariance function as defined in (2.1,3) and written as estimator, i.e., in (2.1,3), the observations (x_t), $t = 0, \ldots, N-1$ (the origin in the time domain is arbitrary), are substituted with the random variables in this segment of the discrete-time stationary process (X_t). From this comparison it becomes obvious that the periodogram and the empirical covariance function, each calculated from the very same time slice in a realisation of a stationary process (X_t) with $\mu_X = 0$, are a Fourier transform pair as proposed in (6.49) and (6.50), on condition that $\hat{c}_X(\tau)$ for $|\tau| \geq N$. This result is given in (8.29).

$$I_X(s) = \sum_{\tau=-(N-1)}^{N-1} \left(\frac{1}{N} \sum_{t=0}^{N-1-|\tau|} X_t X_{t+|\tau|} \right) \cos(2\pi s\tau) \tag{8.28}$$

$$= \sum_{\tau=-(N-1)}^{N-1} \hat{c}_X(\tau) \cos(2\pi s\tau) \tag{8.29}$$

From (8.29), the expectation function of the periodogram is easily arrived at in (8.19), using the expectation function of the empirical covariance function as obtained in (2.57).

The expectation function of the periodogram can also be derived by applying the Fourier representation of a discrete-time stationary stochastic process (X_t). The estimator $J(s)$ as defined in (8.30) is obtained by assuming that the observations at the exterior of the observational period (and thus also the variables in the corresponding segments of (X_t)) are identically zero, as shown by the multiplication with the rectangle sequence $\Pi_N(t)$ which is defined in (6.52). This is a prerequisite for substituting X_t with its Fourier representation $\int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi s'} \mathrm{d}Z_X(s')$, as defined in (7.51), to obtain (8.31). The order of summation and integration can be reversed and the Fourier transform of the rectangle sequence as obtained in (6.54) and (6.55) is substituted in order to arrive at (8.32).

$$J(s) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t \mathrm{e}^{-\mathrm{i}2\pi st} = \frac{1}{\sqrt{N}} \sum_{t=\infty}^{\infty} \Pi_N(t) X_t \mathrm{e}^{-\mathrm{i}2\pi st} \tag{8.30}$$

$$= \frac{1}{\sqrt{N}} \sum_{t=\infty}^{\infty} \Pi_N(t) \left(\int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi s'} \mathrm{d}Z_X(s') \right) \mathrm{e}^{-\mathrm{i}2\pi st} \tag{8.31}$$

$$= \frac{1}{\sqrt{N}} \int_{-1/2}^{1/2} \sum_{t=\infty}^{\infty} \mathit{\Pi}_N(t) \mathrm{e}^{-\mathrm{i}2\pi(s-s')t} \mathrm{d}Z_X(s')$$

$$= \frac{1}{\sqrt{N}} \int_{-1/2}^{1/2} \mathrm{e}^{-\mathrm{i}\pi(N-1)(s-s')} D_N(s-s') \mathrm{d}Z_X(s') \tag{8.32}$$

The integral transform (8.32) (Riemann integral transforms of deterministic functions are defined in (2.33)) is a stochastic Stieltjes integral (7.21) of, in essence, the Dirichlet kernel with respect to the orthogonal increment process $Z_X(s)$. This integral equation relates the Fourier transform $J(s)$ of the random variables $(X_0, X_1, \ldots, X_{N-1})$ in (X_t) with the orthogonal increment process $Z_X(s)$ pertaining to (X_t) as required in (7.51). If this integral equation could be solved given a realisation of $J(s)$ (i.e., the Fourier transform of the observations) then a realisation of $Z_X(s)$ could be obtained, whereafter the spectrum or integrated spectrum of (X_t) could be inferred using (7.51,3) and the stationarity of (X_t). It is, however, rather difficult to solve (8.32) for $Z(s)$, given $J(s)$ [134].

To circumvent these difficulties, $J(s)$ is multiplied with its complex-conjugate $\overline{J(s)}$ as in (8.33), and the periodogram $I_X(s) = J(s)\overline{J(s)} = |J(s)|^2$ as defined in (8.1,4) is obtained in (8.34).

$$\overline{J(s)} = \frac{1}{\sqrt{N}} \int_{-1/2}^{1/2} \overline{\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-s')}} D_N(s-s') \overline{\mathrm{d}Z_X(s')} \tag{8.33}$$

$$\begin{aligned} I_X(s) &= J(s)\overline{J(s)} = |J(s)|^2 \\ &= \frac{1}{N} \int_{-1/2}^{1/2} \mathrm{e}^{-\mathrm{i}\pi(N-1)(s-s')} D_N(s-s') \mathrm{d}Z_X(s') \times \\ &\qquad \int_{-1/2}^{1/2} \overline{\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-s'')}} D_N(s-s'') \overline{\mathrm{d}Z_X(s'')} \end{aligned} \tag{8.34}$$

$$\begin{aligned} \mathrm{E}I_X(s) &= \frac{1}{N} \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \mathrm{e}^{-\mathrm{i}\pi(N-1)(s-s')} \overline{\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-s'')}} \times \\ &\qquad D_N(s-s') D_N(s-s'') \Big(\delta(s'-s'') \mathrm{E}\big(\mathrm{d}Z_X(s') \overline{\mathrm{d}Z_X(s'')}\big) \Big) \\ &= \int_{-1/2}^{1/2} F_N(s-s') \mathrm{E}\big(|\mathrm{d}Z_X(s')|^2\big) \end{aligned} \tag{8.35}$$

Taking the expectations, (8.35) is arrived at by means of integrating with respect to an orthogonal increment process as demonstrated in Sect. 7.5.2, and thereafter using the relationship between $F_N(s)$ and $D_N(s)$ as derived in (6.57).

8.5.2 The Periodogram of a Linear Process

A white noise process (W_t) as defined in (2.10) or (2.11) is the Fourier-Stieltjes transform of an orthogonal increment process $Z_W(s)$. Hence, a realisation (w_t) of (W_t) is associated with a realisation $z_W(s)$ of $Z_W(s)$ and, vice-versa, a realisation $z_W(s)$ of $Z_W(s)$ is associated with a realisation (w_t) of (W_t) as required in (7.51): $w_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}z_W(s)$ for $t = \ldots, -1, 0, 1, \ldots$. A realisation of (W_t) generates in (2.25) a realisation of the linear process (X_t) on condition that the weights (b_t) converge absolutely. For weights $(b_t) = (\ldots, 0, b_0, b_1, \ldots)$, being a sequence that converges absolutely, and (W_t), a white noise process with $\mu_W(t) = 0$ and $\sigma_W^2(t) = \sigma_W^2$, a realisation of a linear process becomes, in (8.36), the output of a stochastic LTI filter as defined in (7.84). The Fourier representation of (w_t) is substituted and, because the weights converge absolutely, the order of summation and integration is interchanged to arrive at (8.37).

$$x_t = \sum_{u=-\infty}^{\infty} b_u w_{t-u}, \quad t = \ldots, -1, 0, 1, \ldots \tag{8.36}$$

$$= \sum_{u=-\infty}^{\infty} b_u \left(\int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi s(t-u)}\mathrm{d}z_W(s) \right)$$

$$= \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st} \sum_{u=-\infty}^{\infty} b_u \mathrm{e}^{-\mathrm{i}2\pi su}\mathrm{d}z_W(s) \tag{8.37}$$

$$\int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}z_X(s) = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st} B(s)\mathrm{d}z_W(s) \tag{8.38}$$

$$\mathrm{d}z_X(s) = B(s)\mathrm{d}z_W(s) \tag{8.39}$$

$$x_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st} B(s)\mathrm{d}z_W(s), \quad t = \ldots, -1, 0, 1, \ldots \tag{8.40}$$

Substituting Fourier representation $x_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}z_X(s)$ on the left side and Fourier transform $B(s) = \sum_{u=-\infty}^{\infty} b_u \mathrm{e}^{-\mathrm{i}2\pi su}$ on the right side, (8.38) is obtained and (8.39) immediately follows. From (8.39) it is concluded that (i) orthogonal increments $\mathrm{d}z_W(s)$ imply orthogonal increments $\mathrm{d}z_X(s)$ and vice-versa, since $B(s)$ is a complex number given a frequency s, and (ii) (x_t) and $B(s)dz_W(s)$ are a Fourier-Stieltjes transform pair, as stipulated in (8.40).

In applications, observations (x_t) of (X_t) are available for $t = 0, 1, \ldots, N-1$, and are assumed to be identically zero at the exterior of this interval. Using this interval, these assumptions can be written $(x_t)\big(\Pi_N(t)\big)$, with $\Pi_N(t)$ being the rectangle sequence as defined in (6.52,1). Consequently, the Fourier transform of the observations can be written as in (8.41).

$$\sum_{t=0}^{N-1} x_t \mathrm{e}^{-\mathrm{i}2\pi st} = \sum_{t=-\infty}^{\infty} \big((x_t)\Pi_N(t)\big)\mathrm{e}^{-\mathrm{i}2\pi st} \tag{8.41}$$

$$= \int_{-1/2}^{1/2} \big(D_N(s-r)\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-r)}\big)B(r)\mathrm{d}z_W(r) \quad (8.42)$$

(8.42) results from (8.41) when convolution theorem (6.48,7) is applied, since (i) $B(s)\mathrm{d}z_W(s)$ is the Fourier representation of (x_t) obtained in (8.39) and (ii) $D_N(s)\mathrm{e}^{-\mathrm{i}\pi(N-1)(s)}$ is the Fourier transform of the rectangle sequence obtained in (6.54) and (6.55).

Borrowing from the derivation of (8.34), the periodogram calculated as required in (8.1) from a realisation of a linear process is written in (8.43) by multiplying the Fourier transform in (8.41) with its complex-conjugate. Thereafter (8.42) is substituted and (8.44) arrived at.

$$\begin{aligned} I_X(s) &= \frac{1}{N}\left(\sum_{t=0}^{N-1} X_t \cos(2\pi st)\right)^2 + \frac{1}{N}\left(\sum_{t=0}^{N-1} X_t \sin(2\pi st)\right)^2 \\ &= \frac{1}{N}\left(\sum_{t=0}^{N-1} x_t \mathrm{e}^{-\mathrm{i}2\pi st} \times \overline{\sum_{t=0}^{N-1} x_t \mathrm{e}^{-\mathrm{i}2\pi st}}\right) \quad (8.43) \\ &= \frac{1}{N}\int_{-1/2}^{1/2} D_N(s-r)\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-r)}B(r)\mathrm{d}z_W(r) \times \\ &\qquad \int_{-1/2}^{1/2} D_N(s-r')\overline{\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-r')}}\,\overline{B(r')dz_W(r')} \quad (8.44) \\ \mathrm{E}I_X(s) &= \frac{1}{N}\int_{-1/2}^{1/2}\int_{-1/2}^{1/2} \mathrm{e}^{-\mathrm{i}\pi(N-1)(s-r)}\overline{\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-r')}} \times \\ &\qquad D_N(s-r)D_N(s-r')B(r)\overline{B(r')}\Big(\delta(r-r')\mathrm{E}\big(\mathrm{d}Z_W(r)\overline{\mathrm{d}Z_W(r')}\big)\Big) \\ &= \frac{1}{N}\int_{-1/2}^{1/2} D_N(s-r)D_N(s-r)B(r)\overline{B(r)}\mathrm{E}\big(|\mathrm{d}Z_W(r)|^2\big) \\ &= \int_{-1/2}^{1/2} F_N(s-r)|B(r)|^2\mathrm{E}\big(|\mathrm{d}Z_W(r)|^2\big) \quad (8.45) \end{aligned}$$

Since $z_W(s)$ stems from the complex-valued random function $Z_W(s)$ having orthogonal increments as required in (7.51), (8.45) follows by taking the expectations and thereafter integrating with respect to an orthogonal increment process as demonstrated (i) in Sect. 7.5.2 and (ii) when deriving (8.35), with the Fejer kernel $F_N(s)$ as in (6.57).

If the identity sequence $(e_t) = (\ldots, 0, e_0 = 1, 0, \ldots)$ as defined in (2.41) is substituted for (b_t) in (8.36), then (i) input and output of the filter become identical, i.e., a white noise process, and (ii), since $\mathcal{F}_{-\mathrm{i}}(e_t) = 1$, $I_W(s)$, the periodogram calculated from a realisation of a white noise process, and its expectation are arrived at in (8.46) and (8.47).

$$x_t = \sum_{u=-\infty}^{\infty} e_u w_{t-u} = w(t) \quad t = \ldots, -1, 0, 1, \ldots$$

$$I_W(s) = \frac{1}{N}\int_{-1/2}^{1/2} D_N(s-r)\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-r)}\mathrm{d}z_W(r) \times \int_{-1/2}^{1/2} D_N(s-r')\overline{\mathrm{e}^{-\mathrm{i}\pi(N-1)(s-r')}dz_W(r')} \qquad (8.46)$$

$$\mathrm{E}I_W(s) = \int_{-1/2}^{1/2} F_N(s-r)\mathrm{E}\big(|\mathrm{d}Z_W(r)|^2\big) \qquad (8.47)$$

The Fejer kernel in (8.45) and (8.47) is shown to be a defining sequence for $\delta(s-n)$, $n = \ldots, -1, 0, 1, \ldots$, in the remarks to (6.62), and, further, $\delta(s-n)$ is the identity function under convolution as defined in (6.144), as concluded from propositions (6.67) and (6.68). For large N, consequently, (8.45) and (8.47) become $\mathrm{E}I_X(s) \approx |B(r)|^2\mathrm{E}\big(|\mathrm{d}Z_W(r)|^2\big)$ and $\mathrm{E}I_W(s) \approx \mathrm{E}\big(|\mathrm{d}Z_W(r)|^2\big)$.

By comparing (8.44) with (8.46) as well as (8.45) with (8.47), the approximation in (8.48) is arrived at. The equation in (8.48) is obtained from (7.90,4) since $B(s)$ is the frequency response function pertaining to (b_t), i.e., the impulse response function in filter (8.36).

$$I_X(s) \approx |B(r)|^2 I_W(s) = \frac{C_X(s)}{\sigma_W^2} I_W(s) \quad \text{for } N \to \infty \qquad (8.48)$$

In (8.48), the periodogram $I_X(s)$ is approximated using the product of (i) $I_W(s)$ and (ii) the squared absolute value of the frequency response function in the linear process $\big(C_X(s) = B(s)\overline{B(s)}C_W(s) = B(s)\overline{B(s)}\sigma_W^2$ is obtained from (7.90,4)$\big)$, on condition that N becomes very large. The results obtained above are summarized in the introduction to Sect. 8.4.

8.6 Problems

8.1. Let $I_X(s_k)$ be a periodogram calculated from a realisation stemming from a normal white noise process (W_t), s_k being the Fourier frequencies. Show that $I_X(s_k)$ is exponentially distributed as defined in (7.4). A hint is found in the remarks to (8.8).

8.2. Simulate a realisation of a white noise process with $N = 4096$ and thereafter calculate the periodogram from the simulated values. Plot the periodogram (i) as obtained and (ii), subsequent to taking the logarithms, in dB scale. Then plot the histograms of both periodograms, prior and subsequent to the logarithmic transformation. Derive the probability density of $J_W(s) = 10\log_{10}(I_W(s))$, on condition that $I_W(s)$ is exponentially distributed as defined in (7.4). Hint: Let X be an exponentially distributed random variable and let $Y = 10\log_{10}(X)$. Then the Stieltjes integral $F(y) = \int_{10\log_{10}(x)\le y} \mathrm{d}F(x)$ will result from (1.5) and, when evaluating this integral, $1 - \mathrm{e}^{-\lambda 10^{x/10}}$, i.e., the distribution of Y, will be arrived at.

8.3. $N\hat{\sigma}_X^2 = \sum_{k=-(m-1)}^{m} I_X(s_k)$, i.e., the sum of all values in a periodogram is equal to the product of N (the number of observations) and $\hat{\sigma}_X^2$ (the empirical variance of the observations) subject to certain conditions.

8.4. Apply R expressions as given in file `/path/problem84.R` to simulate realisations of the AR[2] process having the spectrum as plotted in Fig. 8.2 (b). Select time slices of lengths $N = 512$, $N = 1028$ and $N = 2046$ from the simulations, calculate the periodograms and thereafter the normalised cumulated periodograms using (8.9). Then plot the normalised integrated spectra of the model together with the normalised cumulated periodograms.

8.5. Calculate the differences of the global average temperatures plotted in Fig. 2.23. Thereafter, plot the correlation function of the differences. Do these differences stem from a white noise process? Answers can be obtained (i) by applying (2.59), i.e., comparing the empirical correlation function with the templates in Fig. 2.22, (ii) from the Ljung-Box statistic in (2.60), and (iii) from the periodogram tests introduced in Sect. 8.2.2. Do these answers agree with each other?

8.6. Calculate periodograms from time slices of length $N = 64$, $N = 256$ and $N = 1024$ of realisations of the AR[2] model introduced in Fig. 7.13, then plot these periodograms together with the spectrum of the process, as demonstrated in Fig. 8.4. Does the bias of all periodograms depend on N?

8.7. As in Problem 8.6, but using the AR[4] model in Fig. 7.14 and larger N.

8.8. Simulate realisations `wts` with $N = 4096$ of a normal white noise process (W_t) with $\mu_W = 0$ and $\sigma_W^2 = 1$ and thereafter, using `xts <- arima.sim(...,innov=wts,...)`, generate realisations of linear processes. Calculate

- the spectra of the linear processes, borrowing from the R expressions in the remarks to Fig. 7.14, and
- periodograms from their realisations using the R expressions in the remarks to Fig. 8.1.

Normalise each periodogram with the pertaining spectrum and plot a histogram of the normalised periodogram values. Are the histograms obtained in agreement with the properties stipulated in (8.24,2)? Thereafter repeat the experiments using non-normal white noise processes.

9 Estimators for a Continuous Spectrum

Let $X(t)$ be a discrete-time stationary stochastic process having a continuous spectrum $C_X(s)$ and a covariance function $c_X(\tau)$, such that $c_X(\tau)$ and $C_X(s)$ are a Fourier transform pair as required in (7.65) and (7.66). Then a plot of $C_X(s)$ gives, except for processes (X_t) having very simple spectra, a more detailed picture of the second moment function of (X_t) than does the plot of $c_X(\tau)$, as is demonstrated in Sect. 7.5.5. In this chapter therefore, estimators for $C_X(s)$ are introduced which (i) have a negligible bias and (ii) are consistent. These estimators are, in general, preferred to the periodogram having the disadvantages discussed in Sect. 8.3 and the remarks to (8.24,7).

The model for (X_t) with a continuous $C_X(s)$ is defined in Sect. 9.1 and it is assumed, in the following sections, that a time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of (X_t) has been observed. Can then $C_X(s)$ be estimated without bias (1.3) and consistently (1.4) from (x_t)?

In Sect. 9.2, the observations are tapered (a data taper can reduce the leakage as shown in Sect. 6.7) to obtain an estimator (called direct spectral estimator) which is bias-free and, in Sect. 9.3, a direct spectral estimator is smoothed to arrive at an estimator whose variance decreases when N increases, N being the length of the observational period. The length of the observational period has to be chosen according to the dynamic range and bandwidth of the spectrum to be estimated. If (i) approximations for both are available and (ii) N can be chosen accordingly then the estimation is straightforward and the bias and the variance of the estimator can be controlled; if not, the estimation becomes difficult. These difficulties can be overcome when the procedures introduced in the remarks to Figs. 9.3 and 9.4 as well as in Sect. 9.3.5 are applied. Alternatives to tapering are the parametric spectral estimation and the pre-whitening introduced in Sect. 9.2.5.

In Sect. 9.4, (i) a seismogram (a record of fine oscillations in the solid earth generated by, e.g., a possibly distant earthquake) and (ii) wind speed measurements in a turbulent atmospheric flow are assumed to stem from stationary stochastic processes, both having a continuous spectrum. Under this assumption, methods as introduced in Sects. 9.2 and 9.3 are applied to estimate the spectra of these processes. The last two sections contain, as usual, the supplements and the problems.

The estimation of a continuous spectrum is dealt with in [108].

9.1 The Model

Let $\mathcal{L}$ be a stochastic LTI filter (7.84), having an impulse response sequence that converges absolutely (2.24). Then the output of $\mathcal{L}$ is a stationary linear process with a continuous spectrum provided that its input is a white noise process, as is concluded from (7.90). Obviously, definitions (2.25) (linear process) and (5.38) (MA[∞] representation of the ARMA[p, q] model as defined in (5.36)) are stochastic LTI filters having one-sided impulse response functions. These definitions are slightly modified to capture a two-sided sum in (9.1,1), and, obviously, $C_X(s)$ in (9.1,2) is continuous in $-1/2 \le s \le 1/2$, due to $B(s)$ and $C_W(s)$ being continuous in this interval.

Let (W_t) be a white noise process with $\mu_W = 0$, σ_W^2 and thus spectrum $C_W(s) = \sigma_W^2$, $-1/2 \le s \le 1/2$, and sequence $(b_t) = (\ldots, -b_{-1}, b_0, b_1, \ldots)$ be absolutely convergent (2.24) with Fourier transform $B(s) = \mathcal{F}_{-\mathrm{i}}(b_t)$ as defined in (6.49). Then:

1. $X_t = \sum_{u=-\infty}^{\infty} b_u W_{t-u}$, $t = \ldots, -1, 0, 1, \ldots$, *is a stochastic LTI filter as defined in (7.84), and*
2. $C_X(s) = B(s)\overline{B(s)}C_W(s) = |B(s)|^2\sigma_W^2$, $-1/2 \le s \le 1/2$, *is the continuous spectrum of (X_t) as derived from (7.90).*

(9.1)

An example of an LTI filter with a two-sided impulse response sequence is given in Problem 7.8, where a realisation of a white noise process is smoothed using a two-sided moving average. If, however, $b_t = 0$ for $t < 0$ is assumed in (9.1,1) then this model becomes a linear process as defined in (2.25).

In Sects. 9.2 and 9.3, (i) the stochastic LTI filter as defined in (9.1,1) is used as a model for discrete-time stationary processes (X_t) having a continuous spectrum $C_X(s)$ and (ii) a variety of estimators for $C_X(s)$ are computed using a time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of (X_t). These estimators are demonstrated using realisations simulated from AR[p] models. Very useful is the AR[4] model $X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9838X_{t-4} + W_t$, where (W_t) is a white noise process with $\mu_W = 0$ and variance $\sigma_W^2 = 1$, on account of its large dynamic range (65 dB in Fig.7.14 (b)) and small bandwidth (0.00474 in Fig. 7.10 (d)). A realisation of this model is plotted above in Fig. 7.14.

In this chapter, simulation experiments are didactic tools promoting the teaching of spectral estimators because a first estimate as required in (9.3) using a cosine taper with, e.g., $p = 0.05$, can not only be compared with a second estimate obtained using a cosine taper with, e.g., $p = 0.10$, but also with the model spectrum. Thus, the performance of spectral estimators becomes readily visible. Real data (a seismogram and wind speed measurements) are analysed in Sect. 9.4, both stemming from stationary stochastic processes whose spectra are not completely unknown. This analysis demonstrates how two entities collaborate in spectral estimation, these being (i) preliminary knowledge of the process under analysis available from theory and/or prior experiments and (ii) methods as introduced in Sects. 9.2 and 9.3.

9.2 Direct Spectral Estimator

Sect. 8.3 discusses the bias obtained in (8.19) that has to be taken into account when a spectrum $C_X(s)$ $\big($or an integrated spectrum $C_X^{(I)}(s)\big)$ is estimated using a periodogram $I_X(s)$ $\big($or a cumulated periodogram $\sum_{-1/2<r_l\le s_k} I_X(r_l)\big)$ calculated from a time slice (x_t), $t = 0, \ldots, N-1$, in a realisation of a discrete-time stationary process (X_t) under investigation. This bias is shown to be due to convolving with the Fejer kernel $F_N(s)$ in (8.21), a situation similar to the one encountered in Sect. 6.5.3. There, it is demonstrated that convolving with the Dirichlet kernel $D_N(s)$ in (6.95) generates distortions when a Fourier transform $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ pertaining to a function $f(t)$ in $L^2(-\infty, \infty)$ is calculated from a finite interval in the time domain. From the similarity of these convolution integrals, it is concluded in Sect. 8.3 that the bias of $I_X(s)$,

1. due to the main lobe of the Fejer kernel, becomes negligible within small neighbourhoods of peaks in $C_X(s)$ on condition that the number of observations N is as large as that recommendend by diagnostic (6.116), i.e., large enough to assert that $\mathcal{B}_{\triangle}^{\cap}\big(F_N(s)\big) \le \mathcal{B}_{\triangle}\big(C_X(s)\big)/2$, where $\mathcal{B}_{\triangle}^{\cap}\big(F_N(s) \approx (7/8N)$ is the width at half height of the main lobe of the Fejer kernel as obtained in the remarks to (6.97) and $\mathcal{B}_{\triangle}\big(C_X(s)\big)$ is the bandwidth of the continuous spectrum as defined in (7.83), and
2. due to the side lobes of the Fejer kernel, remains substantially large in greater distances from the peaks of $C_X(s)$ even when N is large enough to trace the narrowest peak, on condition that $C_X(s)$ has a dynamic range which is greater than the threshold proposed in (6.117).

Consequently, a periodogram is subject to a bias due to the convolution integral in (8.21), as demonstrated in Figs. 8.6 and 8.7.

Assuming that the number of observations available is as large as or even larger than recommended in (6.116), an answer to the following question is given in this section: does a data taper as defined in (6.110,2) exist such that the Fejer kernel in (8.21) can be substituted with a function that has smaller side lobes? The probability that such a data taper exists is quite large because a cosine taper $\widehat{h_N^{(e)}}(t)$ as defined in (6.111,2), for example, strikingly suppresses the distortions generated by the side lobes of its kernel $\widehat{H_N^{(e)}}(s)$ in (6.114), as is demonstrated in Fig. 6.28.

In Sect. 9.2.1, the direct spectral estimator is defined using a suitable data taper, and the bias of the estimator thus obtained is shown to be considerably smaller than the bias of the periodogram. The variance and covariance functions of the direct spectral estimator are derived in Sect. 9.2.2, and its probability distribution is given in Sect. 9.2.3. In Sect. 9.2.4, the direct spectral estimator is calculated for frequencies s'_k with $\Delta s'_k < \Delta s_k$, $\Delta s_k = s_{k+1} - s_k = (k+1)/N - k/N = 1/N$ the Fourier frequencies as

defined in (6.21), and in Sect. 9.2.5, alternatives to using a data taper are introduced.

9.2.1 Definition and Expectation

In Sect. 6.7, leakage due to convolution (6.94) is reduced by tapering in (6.114). This motivates to define a data taper in (9.2,1) which is shown to be suitable for spectral estimation in the remarks to (9.8). Using this data taper, an estimator for the continuous spectrum $C_X(s)$ of model (X_t) (9.1) is defined in (9.3) and (9.4).

Let $h_N(t)$ be a data taper as defined in (6.110) such that $h_N(t)$ and its kernel $H_N(s) = \mathcal{F}_{-\mathrm{i}}\big(h_N(t)\big)$ are a Fourier transform pair as defined in (6.49) and (6.50), and let N be a positive integer number. Then:

1. $h_N^{(d)}(t) = h_N(t)\big/\sqrt{\Sigma(h^2)}$, *with* $\Sigma(h^2) = \sum_{t=0}^{N-1}\big(h_N(t)\big)^2$, *is called a variance-normalised data taper or a data taper normalised for spectral estimation (cf. the remarks to (9.8)).*
2. $\mathcal{H}_N^{(d)}(s) = \big(H_N(s)\overline{H_N(s)}\big)\big/\Sigma(h^2) = \mathcal{F}_{-\mathrm{i}}\big(h_N^{(d)} \star h_N^{(d)}(t)\big)$, $-1/2 \le s \le 1/2$, *and being periodic with period 1, is called the spectral window pertaining to* $h_N^{(d)}(t)$.

(9.2)

$h_N^{(d)}(t)$ as defined in (9.2,1) is "smooth" because, in (6.110), w is required to be a smooth function. This implies, for the reasons given in the remarks to (6.110), that the side lobes of $H_N(s)$ and $\mathcal{H}_N^{(d)}(s)$ are smaller than those of $D_N(s)$ and $F_N(s)$. $F_N(s)$ in (6.57) is the Fourier transform of $\Lambda_{2N+1}^{(e)}(t) = (1/N)\big(\Pi_N \star \Pi_N(t)\big)$, i.e., the autocorrelation of $\Pi_N(t)$ as required in (6.103). Consequently, $\Pi_N(t)$ as in (6.52,1) acts as "data taper" when a periodogram is calculated and $F_N(s)$ is the pertaining spectral window. In convolution (8.21) therefore, the leakage due to the side lobes of $F_N(s)$ is reduced when $\Pi_N(t)$ is substituted with $h_N^{(d)}(t)$, acting as data taper when the direct spectral estimator $\widehat{C}_X^{(d)}(s)$ is calculated using definition (9.3).

In (9.3), a time slice $t = 0, \ldots, N-1$ in a realisation (x_t), $t = \ldots, -1, 0, 1, \ldots$, stemming from a stationary discrete-time stochastic process (X_t) with $\mu_X = 0$, is tapered using a variance-normalised data taper $h_N^{(d)}(t)$ as defined in (9.2,1). Thus, $h_N^{(d)}(t)x_t$, $t = \ldots, -1, 0, 1, \ldots$ is arrived at: a sequence that is in $L^2(-\infty, \infty)$, because $\sum_{t=-\infty}^{\infty}\big(h_N^{(d)}x_t\big)^2 = \sum_{t=0}^{N-1}\big(h_N^{(d)}x_t\big)^2 < \infty$, i.e., it has a finite norm on $t = \ldots, -1, 0, 1, \ldots$, as required in (6.3). In product $h_N^{(d)}(t)x_t$, observations x_t are as usual substituted with random variables X_t to obtain an estimator, and thus $\widehat{C}_X^{(d)}(s)$ becomes the squared absolute value of $\mathcal{F}_{-\mathrm{i}}\big(h_N^{(d)}(t)X_t\big)$, i.e., the Fourier transform of $h_N^{(d)}(t)X_t$ as defined in (6.49) and (6.50). Hence, the direct spectral estimator $\widehat{C}_X^{(d)}(s)$ is defined for all real

$-1/2 \leq s \leq 1/2$ and is periodic with period 1, properties shared by the spectrum of a discrete-time stochastic process, as is concluded from (7.51), (7.64), (7.65) and (7.66). In (9.4), the direct spectral estimator is written as squares of the real and imaginary part of $h_N^{(d)}(t)X_t\mathrm{e}^{-\mathrm{i}2\pi st}$ in (9.3), as a direct consequence of (6.2,4) and (9.2,1).

$$\widehat{C}_X^{(d)}(s) = \left| \sum_{t=0}^{N-1} h_N^{(d)}(t) X_t \mathrm{e}^{-\mathrm{i}2\pi st} \right|^2 = \left| \sum_{t=0}^{N-1} \frac{h_N(t)X_t}{\sqrt{\Sigma(h^2)}} \mathrm{e}^{-\mathrm{i}2\pi st} \right|^2 \tag{9.3}$$

$$= \left(\sum_{t=0}^{N-1} \frac{h_N(t)X_t}{\sqrt{\Sigma(h^2)}} \cos(2\pi st) \right)^2 + \left(\sum_{t=0}^{N-1} \frac{h_N(t)X_t}{\sqrt{\Sigma(h^2)}} \sin(2\pi st) \right)^2 \tag{9.4}$$

$$= \sum_{\tau=-(N-1)}^{N-1} \left(\frac{1}{\Sigma(h^2)} \sum_{t=0}^{N-1-|\tau|} h_N(t)X_t h_N(t+|\tau|)X_{t+|\tau|} \right) \cos(2\pi s\tau) \tag{9.5}$$

$$= \sum_{\tau=-\infty}^{\infty} \hat{c}_X^{(d)}(\tau)\mathrm{e}^{-\mathrm{i}2\pi s\tau} \quad \text{where} \tag{9.6}$$

$$\hat{c}_X^{(d)}(\tau) = \begin{cases} \frac{1}{\Sigma(h^2)} \sum_{t=0}^{N-1-|\tau|} h_N(t)X_t h_N(t+|\tau|)X_{t+|\tau|} & \text{for } |\tau| < N \\ 0 & \text{for } |\tau| \geq N \end{cases} \tag{9.7}$$

(9.5), (9.6) and (9.7) summarise the results derived in Sect. 9.5.1: $\widehat{C}_X^{(d)}(s) = \mathcal{F}_{-\mathrm{i}}\big(\hat{c}_X^{(d)}(\tau)\big)$, i.e., $\widehat{C}_X^{(d)}(s)$ and $\hat{c}_X^{(d)}(\tau)$ are a Fourier transform pair as defined in (6.49) and (6.50), where both, $\widehat{C}_X^{(d)}(s)$ and $\hat{c}_X^{(d)}(\tau)$, are calculated from tapered (using a variance-normalised data taper as defined in (9.2,1)) observations of a stationary stochastic process.

The expectation of the empirical covariance function calculated from tapered observations is arrived at in (9.8), using (i) formula (1.15,4) and (ii) the expectation of the empirical covariance function as derived in (2.57):

$$\begin{aligned} \mathrm{E}\big(\hat{c}_X^{(d)}(\tau)\big) &= \frac{1}{\Sigma(h^2)} \sum_{t=0}^{N-1-|\tau|} h_N(t)h_N(t+|\tau|)\mathrm{E}\big(X_t X_{t+|\tau|}\big) \\ &= \frac{\sum_{t=0}^{N-1-|\tau|} h_N(t)h_N(t+|\tau|)}{\sum_{t=0}^{N-1}\big(h_N(t)\big)^2} c_X(\tau). \end{aligned} \tag{9.8}$$

In (9.8), $\mathrm{E}\big(\hat{c}_X^{(d)}(\tau)\big) = c_X(\tau)$ for $\tau = 0$, and, consequently, the empirical covariance function calculated from tapered observations estimates the covariance function without bias at least for lag $\tau = 0$, i.e., $\hat{c}_X^{(d)}(0)$ is an unbiased estimator for the variance σ_X^2 of the stationary process (X_t). This desired property is due to the normalisation of $h_N(t)$ with $\Sigma(h^2)$ in definition (9.2,1), where $h_N^{(d)}(t)$ is called variance-normalised data taper.

When (9.8) is substituted in (9.6), the expectation of the direct spectral estimator is arrived at in (9.9). (9.10) is obtained by applying convolution

theorem (6.48,7). Thereafter, (9.11) and (9.12) follow using definition (9.2,2). Alternatively, the expectation of a direct spectral estimator is derived in Sect. 9.5.2 from the Fourier representation of the process under analysis.

$$\mathrm{E}\left(\widehat{C}_X^{(d)}(s)\right) = \sum_{\tau=-\infty}^{\infty}\left(\frac{\sum_{t=0}^{N-1-|\tau|} h_N(t)h_N(t+|\tau|)}{\Sigma(h^2)}c_X(\tau)\right)\mathrm{e}^{-\mathrm{i}2\pi s\tau} \tag{9.9}$$

$$= \left(\sum_{\tau=-\infty}^{\infty}\frac{\sum_{t=0}^{N-1-|\tau|} h_N(t)h_N(t+|\tau|)}{\Sigma(h^2)}\mathrm{e}^{-\mathrm{i}2\pi s\tau}\right) * \mathrm{d}C_X^{(I)}(s) \tag{9.10}$$

$$= \left(\mathcal{H}_N^{(d)}(s)\right) * \mathrm{d}C_X^{(I)}(s) = \int_{-1/2}^{1/2}\mathcal{H}_N^{(d)}(s-r)\mathrm{d}C_X^{(I)}(s) \quad \begin{array}{l}\text{convolution}\\ \text{as in (6.144)}\end{array} \tag{9.11}$$

$$= \left(\mathcal{H}_N^{(d)}(s)\right) * C_X(s) = \int_{-1/2}^{1/2}\mathcal{H}_N^{(d)}(s-r)C_X(s)\mathrm{d}s \quad \begin{array}{l}\text{when the}\\ \text{spectrum is}\\ \text{continuous}\end{array} \tag{9.12}$$

When the expectation of the periodogram obtained in (8.21) is compared with the one of the direct spectral estimator arrived at in (9.11), it is concluded that both, the periodogram and the direct spectral estimator, are biased for finite N. In both cases, the bias is due to the convolution of the spectrum to be estimated with the spectral window ($F_N(s)$ is the spectral window of the periodogram as obtained in the remarks to (9.3)). Since $F_N(s)$ is a defining sequence for the periodic delta function with period 1, $\delta(s+n)$, $n = \ldots, -1, 0, 1, \ldots$, as defined in the remarks to (6.62), $I_X(s)$ is unbiased for $N \to \infty$ as argued in the remarks to (8.19) and (8.21). Is the spectral window $\mathcal{H}_N^{(d)}(s)$ in (9.2,2) a defining sequence for $\delta(s+n)$ as well?

$h_N^{(d)} \star h_N^{(d)}(0) = \left(1/\Sigma(h^2)\right)\left(h_N \star h_N(0)\right) = 1$, due to the normalisation of the data taper $h_N(t)$ required in (9.2,1), implies $\int_{-1/2}^{1/2}\mathcal{H}_N^{(d)}(s) = 1$, as is concluded from definitions (9.2,2) and (6.103), the properties of the autocorrelation derived in Sect. 6.9.6 and proposition (6.48,1). Alternatively, this desired property is obtained in Problem 9.2 by applying Parseval's identity. A spectral window $\mathcal{H}_N^{(d)}(s)$, $N = 1, 2, 3, \ldots$, is therefore reconcilable with (6.60,2). If it is in agreement with (6.60,1) as well, then it is a defining sequence for $\delta(s)$ in its definition interval $-1/2 \le s \le 1/2$, and the direct spectral estimators in (9.11) and (9.12) become unbiased for $N \to \infty$.

If N is finite, then the direct spectral estimators in both (9.11) and (9.12) are biased due to the convolution of the spectrum with the spectral window; however, the bias is reduced substantially as compared with the periodogram. This is demonstrated in Fig. 9.1. There, the expectations of the direct spectral estimators are calculated, using (9.12), by convolving the spectrum of the AR[4] model in Fig. 7.14 with the spectral windows $\widehat{\mathcal{H}_N^{(d)}}(s)$ pertaining to cosine tapers $\widehat{h_N}(t)$, as defined in (6.111,1) and plotted for $p = 0$, $p = 0.05$ and $p = 0.1$ in Fig. 9.2. Cosine taper $\widehat{h_N}(t)$ for $p = 0$ is identical with $\Pi_N(t)$ in (6.52,1) and, therefore, the expectation function of the periodogram in (8.21)

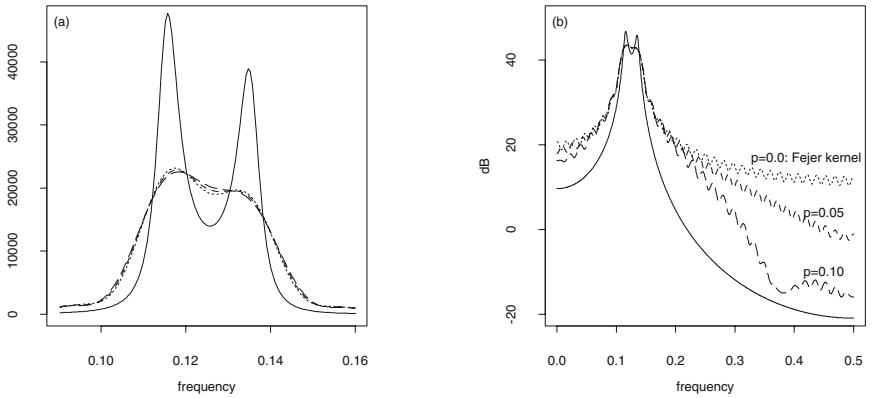


Fig. 9.1. Spectrum (solid line) pertaining to the AR[4] model in Fig. 7.14 together with expectations (9.12) of direct spectral estimators (broken lines) calculated from $N = 64$ observations stemming from a realisation of this model, subsequent to tapering with cosine tapers for $p = 0.0$ (the periodogram), $p = 0.05$ and $p = 0.10$.

is obtained, i.e., $\mathcal{H}_N^{(d)}(s)$ in (9.12) is substituted with $F_N(s)$, the Fejer kernel being the spectral window of the periodogram. This expectation is identical with the one plotted in Fig. 8.7 (b) for this AR[4] model, and $N = 64$. In Fig. 9.1 (b), for higher frequencies at larger distances from the peaks, the bias of the periodogram is reduced when the spectrum is convolved with the spectral windows $\widehat{\mathcal{H}}_N^{(d)}(s)$ pertaining to cosine tapers $\widehat{h_N}(t)$ for $p = 0.05$ and $p = 0.10$. The bias is reduced substantially, i.e., by orders of magnitudes.

To the example cosine tapers $\widehat{h_N}(t)$ (6.111,1) for $N = 64$ and $p = 0.0$, $p = 0.05$ as well as $p = 0.10$ as plotted on the left in Fig. 9.2, pertain the spectral windows $\widehat{\mathcal{H}}_N^{(d)}(s)$ as plotted on the right in this figure. $\widehat{\mathcal{H}}_N^{(d)}(s) = \mathcal{F}_{-\mathrm{i}}\big(\widehat{h}_N^{(d)} \star \widehat{h}_N^{(d)}(t)\big)$ is calculated by Fourier transforming the autocorrelation of the cosine taper, subsequent to normalising with $\sum_{t=0}^{N-1}\big(\widehat{h_N}(t)\big)^2$ as required in (9.2,1). $\widehat{\mathcal{H}}_N^{(d)}(s)$, $N = 1, 2, 3, \ldots$, is a defining sequence for $\delta(s + n)$, $n = \ldots, -1, 0, 1, \ldots$, as is derived in Problem 9.3. A direct spectral estimator calculated from cosine tapered observations is therefore unbiased for $N \to \infty$, a desired property as is concluded from the remarks to (9.12). Cosine data tapers are quite frequently used in spectral estimation. However, alternative tapers are available. For example, DPSSs as introduced in the remarks to (6.127) are used as data tapers in multi-taper spectral estimation [108].

For the case of spectral windows $\widehat{\mathcal{H}}_N^{(d)}(s)$ pertaining to variance-normalised cosine tapers $\widehat{h_N}(t)$, Fig. 9.2 demonstrates that $\widehat{\mathcal{H}}_N^{(d)}(s)$ has

– large side lobes on condition that $\widehat{h_N}(t)$ increases and decreases towards both of its ends with slopes having large absolute values, i.e., when p is small (in this case, the tails of $\widehat{h_N}(t)$ are short), and

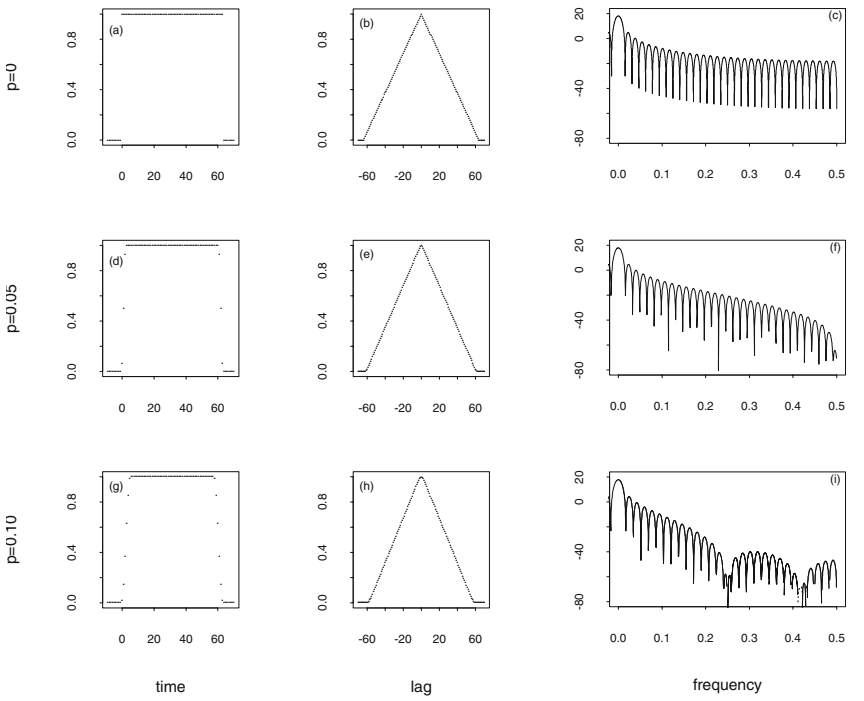


Fig. 9.2. Cosine tapers $\widehat{h_N}(t)$ as defined in (6.111,1) (on the left), their autocorrelations normalised with $\sum_{t=0}^{N-1}\big(\widehat{h_N}(t)\big)^2$ (in the centre) and their spectral windows $\widehat{\mathcal{H}}_N^{(d)}(s)$ as defined in (9.2,2) in dB (on the right, when negative with broken lines) for $N = 64$ as well as $p = 0$, $p = 0.05$ and $p = 0.10$.

– small side lobes on condition that $\widehat{h_N}(t)$ increases and decreases with slopes having small absolute values, i.e., when p is large (in this case, the tails of $\widehat{h_N}(t)$ are long).

This behaviour is shared by all spectral windows $\mathcal{H}_N^{(d)}(s) = \mathcal{F}_{-\mathrm{i}}\big(h_N^{(d)} \star h_N^{(d)}(t)\big)$ as defined in (9.2,2), a conclusion drawn from the discussion in Sect. 6.7.1, because the autocorrelation of $h_N^{(d)}(t)$ is a time-limited sequence, as introduced in (6.108).

The example cosine tapers $\widehat{h_N}(t)$ and spectral windows $\widehat{\mathcal{H}}_N^{(d)}(s)$ in Fig. 9.2 (together with the widths at half height of their main lobes in Table 9.1) also show that the main lobe of $\widehat{\mathcal{H}}_N^{(d)}(s)$ increases in width when the side lobes decrease in height, i.e., when p increases, a behaviour shared by spectral window (9.2,2) as is concluded from (6.48). A spectral window with a wider main lobe will blur a peak in the spectrum to be estimated more than one with a narrower main lobe, i.e., the loss of resolution in the frequency domain

Table 9.1. Height and width at half height $\mathcal{B}_{\triangle}^{\cap}\big(\widehat{\mathcal{H}}_N^{(d)}(s)\big)$ of the main lobes of spectral windows $\widehat{\mathcal{H}}_N^{(d)}(s)$ (9.2,1) pertaining to cosine tapers $\widehat{h_N}(t)$ (6.111) for $N = 64$. $\mathcal{B}_{\triangle}^{\cap}\big(\widehat{\mathcal{H}}_N^{(d)}(s)\big)$ results in Problem 9.3 when definition (6.97) is applied. $\mathcal{B}_{\triangle}^{\cap}\big(F_N(s)\big) \approx 7/(8N)$, i.e., $\mathcal{B}_{\triangle}^{\cap}\big(\widehat{\mathcal{H}}_N^{(d)}(s)\big)$ for $p = 0.0$, is obtained in plots of Fejer kernels for a varietey of N, as is demonstrated when $\mathcal{B}_{\triangle}^{\cap}\big(D_N(s)\big) \approx 6/(5N)$ is approximated in the remarks to (6.97).

p	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
height $\approx$	64.00	62.01	59.54	57.35	54.47	52.36	50.31	46.44	42.64
$\mathcal{B}_{\triangle}^{\cap}\big(\widehat{\mathcal{H}}_N^{(d)}(s)\big) \approx$	.0137	.0146	.0152	.0160	.0172	.0180	.0190	.0208	.0225

increases, as is demonstrated in Sect. 6.5.3. Consequently, an increased bias of a direct spectral estimator in the bands with a peak is the penalty paid for reducing the bias a larger distance away from a peak by tapering. In Fig. 9.1 (a) for example, the bias of the direct spectral estimator increases slightly with increasing p in the peak region of the spectrum; the increase in bias is however small compared to the bias generated by the too small number of observations used for the estimation (i.e., $N = 64$, whereas diagnostic (6.116) recommends $1/N = \Delta s_k < 0.00474/2$, 0.00474 being the bandwidth of the spectrum to be estimated drawn in Fig. 7.10 (d)).

From Figs. 9.1 and 9.2 it becomes evident that a bias in a direct spectral estimator is more efficiently reduced when the side lobes of the spectral window decrease. This result is not surprising for two reasons:

1. the bias in the direct spectral estimator due to the convolution in (9.11) and (9.12) is comparable to the leakage generated in (6.113) and (6.114) by convolution $G(s) * H_N^{(e)}(s)$ when the tapered observations $(g_t)\big(h_N^{(e)}(t)\big)$ are Fourier transformed, and
2. the leakage in (6.113) and (6.114) reduces when the side lobes in $H_N^{(e)}(s)$ become smaller, $H_N^{(e)}(s)$ being the kernel pertaining to $h_N^{(e)}(t)$, as is demonstrated in Problem 6.28.

To what extent is tapering required in order to reduce the bias due to the side lobes of the spectral window when a spectrum of a discrete-time stationary stochastic process is estimated? An answer to this question can be obtained from diagnostics as introduced in Sect. 6.7.2, modified as recommended in the remarks to Figs. 8.4 and 8.5, provided that approximations for the dynamic range (6.96) and the bandwidth (7.83) of the spectrum under analysis are available. If these parameters of the spectrum to be estimated are not available, then a diagnostic for the leakage due to the side lobes of the spectral window can be obtained when the procedure introduced in the remarks to (6.117) is applied: the observations are tapered with a sequence of data

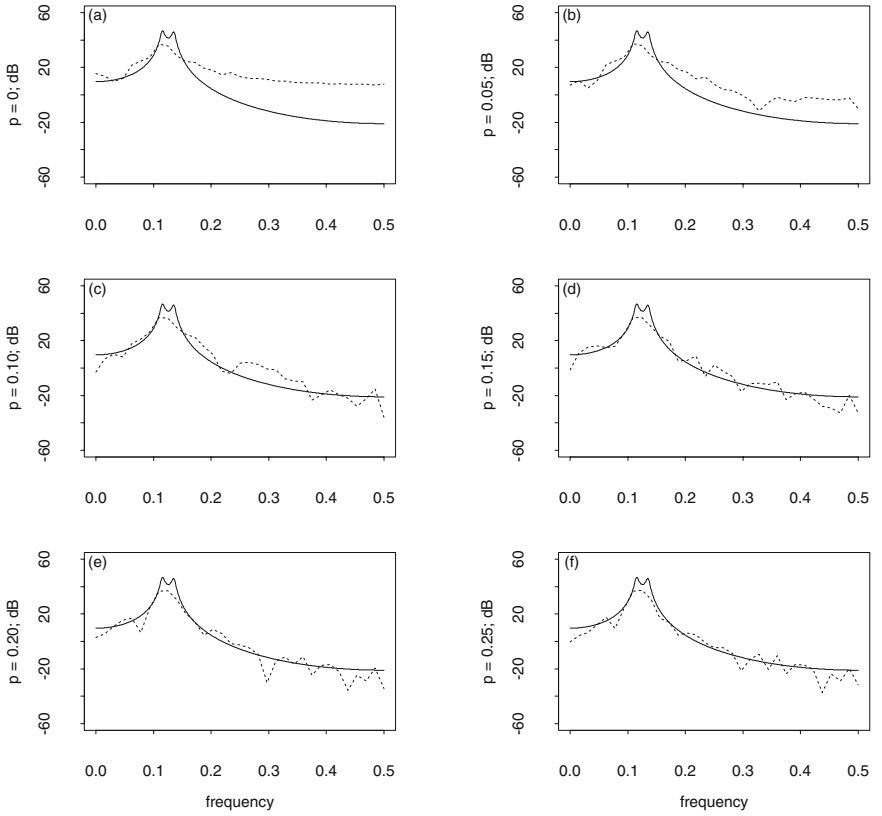


Fig. 9.3. Spectrum $C_X(s)$ of the AR[4] process (X_t) as in Fig. 7.14 (solid lines) together with direct spectral estimators $\widehat{C}_X^{(d)}(s)$ calculated as required in (9.3) from $N = 64$ observations in a realisation of (X_t), which are multiplied with cosine tapers $\widehat{h_N}(t)$ as defined in (6.111,1) for $p = 0.00, 0.05, 0.10, 0.15, 0.20, 0.25$ and subject to a normalisation as required in (9.2), from left to right and top to bottom.

tapers having tails of increasing lengths and, subsequent to tapering, direct spectral estimators are calculated.

In Figs. 9.3 and 9.4, for example, direct spectral estimators are plotted which are calculated from two time slices of lengths $N = 64$ and $N = 256$ from a realisation of the AR[4] model (X_t) with spectrum $C_X(s)$ as plotted in Fig. 7.14, subsequent to multiplying the observations with sequences of cosine tapers. From these figures it becomes obvious that, in order to suppress the bias for frequencies $s \geq 0.20$, a cosine taper with $p \geq 0.15$ is required when the direct spectral estimator is calculated from $N = 64$ observations, whereas this goal can be achieved using a cosine taper with $p \geq 0.05$ when $N = 256$ observations are available.

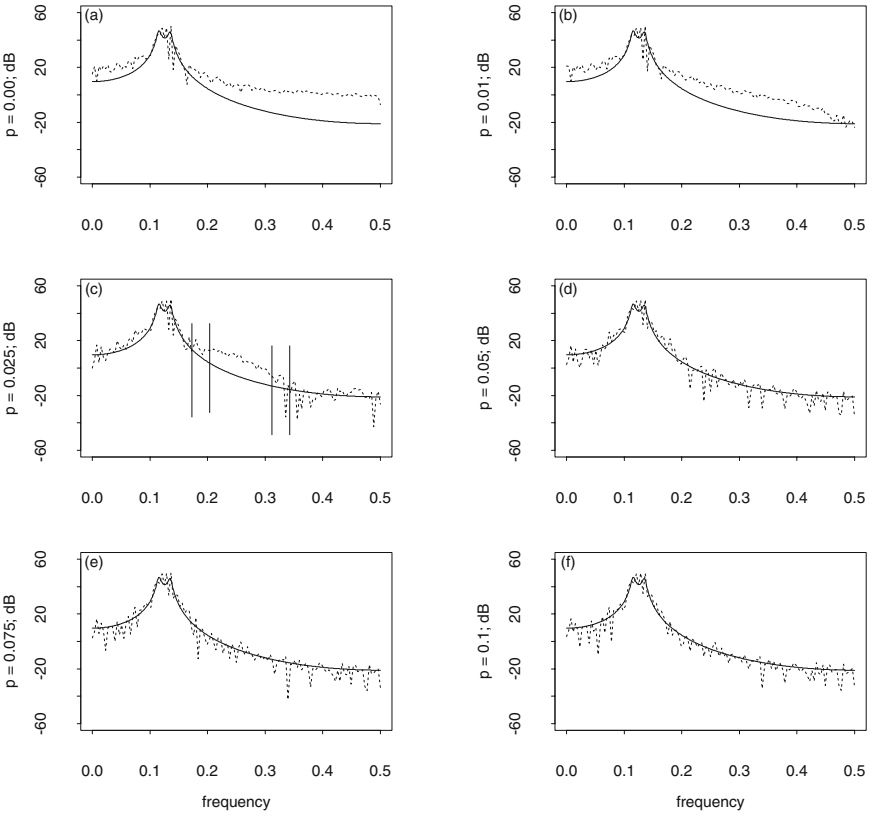


Fig. 9.4. As Fig. 9.3, but here $N = 256$ observations are multiplied with cosine tapers for $p = 0.000, 0.010, 0.025, 0.050, 0.075, 0.100$.

Could these results be obtained if the spectrum of the AR[4] model were not plotted with solid lines in Figs. 9.3 and 9.4? A bias becoming reduced by tapering leaves two kinds of "footprints" in the plots of a sequence of spectral estimators calculated using data tapers with tails increasing in length:

1. At the exterior of the bands with peaks, the periodogram is substantially larger than the spectral estimators because the side lobes of the Fejer kernel (the spectral window of the periodogram) are larger than the side lobes of the spectral windows pertaining to data tapers as defined in (9.2). With increasing length of the tails in the data tapers, i.e., with increasing p when cosine tapers are applied, the spectral estimators become stable, i.e., pairs in the sequence of estimators become approximately identical, as demonstrated by the estimators for $p = 0.20, 0.25$ in Fig. 9.3 and for $p = 0.050, 0.075, 0.100$ in Fig. 9.4. The estimators in the first pair having this property are then selected as estimators for the spectrum, as

argued in the remarks to (6.117). There, a sequence of cosine tapers with increasing p is applied to reduce the leakage due to the side lobes of the kernel in (6.114).

2. In a logarithmic plot, a direct spectral estimator fluctuates within a confidence interval of constant height. This consequence of (9.19) is demonstrated in Figs. 9.7, 9.8 and 9.9. It implies that a direct spectral estimator locally (2.55,2) fluctuates with constant amplitudes provided that it is unbiased, a property possessed by those direct spectral estimators with cosine tapers for $p = 0.20, 0.25$ in Fig. 9.3 and for $p = 0.050, 0.075, 0.100$ in Fig. 9.4, reconcilable with the result obtained above. The other estimators in these figures, in contrast, fluctuate with locally non-constant amplitudes, a behaviour pointing to a biased estimator. For example, the direct spectral estimator obtained using a cosine taper for $p = 0.025$ in Fig. 9.4 (c) is, within the bands plotted with solid vertical lines, subject to changes in the amplitudes of its fluctuations.

Cosine tapers $\widehat{h_N}(t)$ as defined in (6.111) and plotted in Figs. 6.26 and 9.2 are obtained in R using `spec.taper()`, as demonstrated in the remarks to (6.111). The pertaining normalised autocorrelations $\big(1/\Sigma(\widehat{h^2})\big)\big(\widehat{h_N} \star \widehat{h_N}(t)\big)$ with $\Sigma(\widehat{h^2}) = \sum_{t=0}^{N-1}\big(\widehat{h_N}(t)\big)^2$, and spectral windows $\widehat{\mathcal{H}}_N^{(d)}(s)$, as plotted in Fig. 9.2, can be calculated using the R expressions in Sect. 6.9.8 as templates and/or the hint given in Problem 9.3. Despite the plots of the spectral windows being obtained using `fft()` from the normalised autocorrelations, $\widehat{\mathcal{H}}_N^{(d)}(s)$ and $\big(1/\Sigma(\widehat{h^2})\big)\big(\widehat{h_N} \star \widehat{h_N}(t)\big)$ as well as $\widehat{H}_N(s)$ and $\widehat{h_N}(t)$ are Fourier transform pairs as proposed in (6.49) and (6.50). This is implied by definitions (9.2,2) and (6.110,2).

Using a cosine taper and applying definitions (9.2) and (9.3), a direct spectral estimator can be calculated using `spec.univariate()` introduced in Sect. 9.5.6. As an alternative to the normalisation of the data taper in `spec.univariate()`, the squared absolute value of the Fourier transform is normalised with $\Sigma(\widehat{h^2}) = \sum_{t=0}^{N-1}\big(\widehat{h_N}(t)\big)^2$ in `spec.pgram()`, the usual R function for computing a direct spectral estimator. In this form, $\Sigma(\widehat{h^2})$ plays the role of $N = \sum_{t=0}^{N-1}\big(\mathit{\Pi}_N(t)\big)^2$ used for the normalisation of a periodogram as required in (8.1,3), because $\mathit{\Pi}_N(t)$ (6.52,1) is the "data taper" of the periodogram.

If a direct spectral estimator is calculated by applying a data taper $h_N(t)$ to a time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of a stationary stochastic process (X_t), then the observations multiplied with the weights in the tails of $h_N(t)$ are downweighted as compared with the observations multiplied with the centre segment of $h_N(t)$. Consequently, oscillations in the product $h_N(t)(x_t)$ are damped near the end points of the observational period, as is demonstrated in Fig. 9.5.

In statistics, different weights are applied to observations x_i in order to compensate for different variances of the observed random variables X_i. For

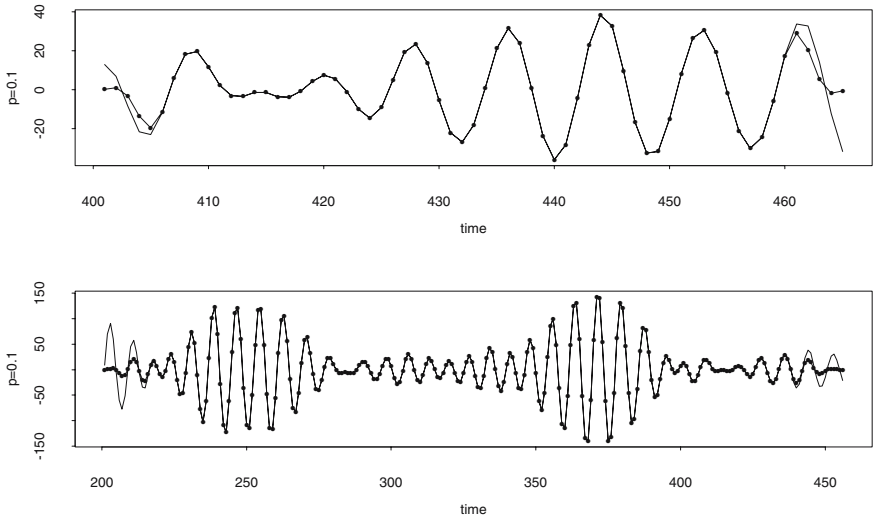


Fig. 9.5. Time slices of lengths $N = 64$ (above) and $N = 256$ (below) in a realisation of the AR[4] process (X_t) as in Fig. 7.14 (solid line), both multiplied with cosine tapers for $p = 0.1$ ($\bullet$ on solid lines).

example, a weighted least square regression is calculated when the variances in the diagonal of the residual covariance matrix (3.16) are not constant. A non-constant variance of the observations is compensated for, in order to obtain estimates with a minimal variance. If, however, observations (x_t), $t = 0, 1 \ldots, N-1$, stemming from a discrete-time stationary stochastic process (X_t) and thus having identical variance as implied by (2.8,2), are multiplied with a data taper $h_N(t)$, $t = 0, 1 \ldots, N-1$, then estimates calculated from the weighted observations do not have minimal variance. Hence, a direct spectral estimator $\widehat{C}_X^{(d)}(s)$ is supposed to have a variance that is larger than that of a periodogram $I_X(s)$, even though both are calculated from identical observations (x_t), $t = 0, 1 \ldots, N-1$, in a realisation (x_t) of a stationary stochastic process (X_t).

9.2.2 Variance and Covariance Functions

The periodogram $I_X(s)$ and the direct spectral estimator $\widehat{C}_X^{(d)}(s)$ are both calculated from a time slice (x_t), $t = 1, 2, \ldots, N-1$, in a realisation (x_t) of a discrete-time stochastic process (X_t) having a continuous, discrete or mixed spectrum, on condition that (X_t) is stationary. The expectation of the periodogram as derived in (8.21) is the convolution of the spectrum to be estimated with the Fejer kernel $F_N(s)$. (8.21) becomes a special case of the expectation of the direct spectral estimator as obtained in (9.11), when $\Pi_N(t)$

is substituted for $h_N^{(d)}(t)$ in (9.3) and, correspondingly, $F_N(s)$ for $\mathcal{H}_N^{(d)}(s)$ in (9.11).

Are the variance and covariance functions of the periodogram as proposed in (8.24) a special case of the variance and covariance functions of the direct spectral estimator? An answer to this question can be obtained from approximation (9.14) which applies when conditions (9.13) are satisfied [108]. (9.13,1) requires that the process under analysis has a continuous spectrum (9.1,2); and (9.13,4) implies that the continuous spectrum to be estimated is locally constant in a frequency band $2\mathcal{B}_{\triangle}^{\cap}\big(H_N(s)\big)$ wide. $\mathcal{B}_{\triangle}^{\cap}\big(H_N(s)\big)$ is the width at half height calculated for the main lobe of the kernel $H_N(s) = \mathcal{F}_{-\mathrm{i}}\big(h_N(t)\big)$, i.e., the taper applied, and thus (6.98) requires that $\mathcal{B}_{\triangle}^{\cap}\big(H_N(s)\big)$ becomes smaller with increasing N. Consequently, (9.13,4) is satisfied on condition that N is sufficiently large. Approximations for the threshold N can be inferred from the dynamic range and bandwidth of the spectrum to be estimated, provided that these parameters are known.

1. *Let* (x_t), $t = 0, 1 \ldots, N-1$, *be an observed time slice in a realisation* (x_t) *of model* (X_t) *as defined in (9.1,1).*
2. *Let* $\widehat{C}_X^{(d)}(s)$ *be a direct spectral estimator calculated as required in (9.3) from* (x_t), $t = 0, 1 \ldots, N-1$, *using a variance-normalised data taper as defined in (9.2):* $h_N^{(d)}(t) = h_N(t)\big/\sqrt{\Sigma(h^2)}$, *with* $\Sigma(h^2) = \sum_{t=0}^{N-1}\big(h_N(t)\big)^2$. (9.13)
3. *Let* (X_t) *be normal as defined in (2.3).*
4. *Let* $C_X(s)$ *be locally constant in a frequency band* $[s-\varsigma, s+\varsigma]$ *of width* $2\varsigma = 2\mathcal{B}_{\triangle}^{\cap}\big(H_N(s)\big)$ *about* s.

From assumptions (9.13) the covariance function of the direct spectral estimator is derived in Sect. 9.5.3 with the results in (9.14), (9.15), (9.16) and (9.17), which hold for $s \neq -1/2, 0, 1/2$ and $s + r \neq -1/2, 0, 1/2$.

$$\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s+r)\big) \approx \big(C_X^2(s)\big)\big(H{\oplus}H(r)\big) \qquad \text{with} \tag{9.14}$$

$$H{\oplus}H(r) = \frac{1}{\Big(\sum_{t=0}^{N-1}\big(h_N(t)\big)^2\Big)^2}\left|\int_{-1/2}^{1/2} H_N(r')H_N(r-r')\mathrm{d}r'\right|^2 \tag{9.15}$$

$$\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s)\big) = \mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s) \tag{9.16}$$

$$\begin{aligned}\mathrm{Cor}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s+r)\big) &= \mathrm{Cov}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s+r)\big)/\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big)\\ &\approx \big(C_X^2(s)\big)\big(H{\oplus}H(r)\big)/C_X^2(s) = H{\oplus}H(r)\end{aligned} \tag{9.17}$$

The covariance function of the direct spectral estimator in (9.14) for lag r, r being a displacement in the frequency domain, is approximated by the product of $C_X^2(s)$ and $H{\oplus}H(r)$, i.e., the square of the spectrum to be estimated and the squared absolute value of convolution $\int_{-1/2}^{1/2} H_N(r')H_N(r-r')dr'$, subsequent to a normalisation. This normalisation causes $H{\oplus}H(0) = 1$ for $r =$

0, as obtained in (9.124), which implies approximation (9.16) for the variance of the direct spectral estimator. Approximation (9.17) for its correlation is a direct consequence of (9.14) and (9.16).

$H{\oplus}H(r)$ results from a self-convolution (introduced in the remarks to (6.103)) of kernel $H_N(s)$ pertaining to data taper $h_N(t)$. $H_N(s)$, being the Fourier transform of a time-limited sequence as required in (6.110), has main and side lobes in its real and imaginary parts, similar to those of the examples given in Sect. 6.6.4. $H{\oplus}H(0) = 1$ for $r = 0$ is obtained in (9.124), i.e., the maximal value of $H{\oplus}H(r)$ is arrived at when both, $H_N(r')$ and its displaced duplicate $H_N(r - r')$, become identical. For increasing r, $H{\oplus}H(r)$ decreases and becomes small for $r \geq \mathcal{B}_{\triangle}^{\cap}\big(|H_N(s)|\big)$, $\mathcal{B}_{\triangle}^{\cap}\big(|H_N(s)|\big)$ the width at half height of the main lobe of $|H_N(s)|$ as defined in the remarks to (6.97), because, for displacements $r \geq \mathcal{B}_{\triangle}^{\cap}\big(|H_N(s)|\big)$, the main lobes of $|H_N(r')|$ and $|H_N(r - r')|$ only overlap in a small interval. This is the reason why, under assumption (9.13,4), the approximation in (9.14) can be derived from (9.114), as commented on in the remarks to (9.117) and (9.118), and that consequently approximation (9.18) is arrived at.

Let $\widehat{C}_X^{(d)}(s)$ *be a direct spectral estimator being calculated under the preconditions stipulated in (9.13). Then*
$\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s+r)\big) \approx 0$ *for* $r \geq 2\mathcal{B}_{\triangle}^{\cap}\big(|H_N(s)|\big)$. (9.18)

For example, $H{\oplus}H(r)$ pertaining to cosine tapers $\widehat{h_N}(t)$ (6.111) for $p = 0.0$, $p = 0.05$, $p = 0.1$, $p = 0.3$ and $p = 0.5$ as well as $N = 64$ are plotted in Fig. 9.6. $H{\oplus}H(r)$ as plotted in Fig. 9.6 holds for arbitrary N on condition that lag r is measured in units of $\Delta s_k = 1/N$ as defined in (6.21) (Problem 9.6). If N is an even number (as in the example given in Fig. 9.6) then $\mathcal{B}_{\triangle}^{\cap}\big(|\widehat{H}_N(s)|\big) \approx \mathcal{B}_{\triangle}^{\cap}\big(\widehat{H}_{N+1}^{(e)}(s)\big)$, with $\widehat{H}_{N+1}^{(e)}(s)$ being the kernel pertaining to the even cosine taper $\widehat{h}_{N+1}^{(e)}(t)$ (6.111,2). Some $\mathcal{B}_{\triangle}^{\cap}\big(\widehat{H}_N^{(e)}(s)\big)$ for $N = 65$, i.e., some widths at half height of main lobes in even cosine tapers with $N = 65$, are given in Table 6.2, those for $p = 0.0$, $p = 0.05$, $p = 0.1$, $p = 0.3$ and $p = 0.5$ are plotted in Fig. 9.6 using vertical lines.

A cosine taper $\widehat{h_N}(t)$ (6.111) with $p = 0.0$ is identical with $\Pi_N(t)$ (6.52,1), with $D_N(s)\mathrm{e}^{-\mathrm{i}\pi(N-1)s} = \mathcal{F}_{-i}\big(\Pi_N(t)\big)$ (6.54) being its kernel and $F_N(s)$ (6.57) is spectral window. The properties of $D_N(s)$ and $F_N(s)$ are enumerated in (6.58) and (6.59), from which it is concluded that $F_N(s)$ is a defining sequence for $\delta(s)$ (6.62). In addition, $\mathcal{B}_{\triangle}^{\cap}\big(|D_N(s)\mathrm{e}^{-\mathrm{i}\pi(N-1)s}|\big) \approx \mathcal{B}_{\triangle}^{\cap}\big(D_{N+1}(s)\big) \approx 5/(6N)$ is arrived at in the remarks to Fig. 6.24. For $N = 65$, $\mathcal{B}_{\triangle}^{\cap}\big(|D_N(s)\mathrm{e}^{-\mathrm{i}\pi(N-1)s}|\big) \approx 0.01188$ is drawn as a vertical solid line in Fig. 9.6.

The correlation of the periodogram (solid line in Fig. 9.6) rapidly decreases with increasing lag r: $H{\oplus}H(r) < 0.05$ for $r > \mathcal{B}_{\triangle}^{\cap}\big(|D_N(s)\mathrm{e}^{-\mathrm{i}\pi(N-1)s}|\big)$ and $H{\oplus}H(r) \approx 0.0$ for $r > 2\mathcal{B}_{\triangle}^{\cap}\big(|D_N(s)\mathrm{e}^{-\mathrm{i}\pi(N-1)s}|\big) \approx 12/(5N)$ as approximated in the remarks to Fig. 6.24. Consequently, also the covariance of the

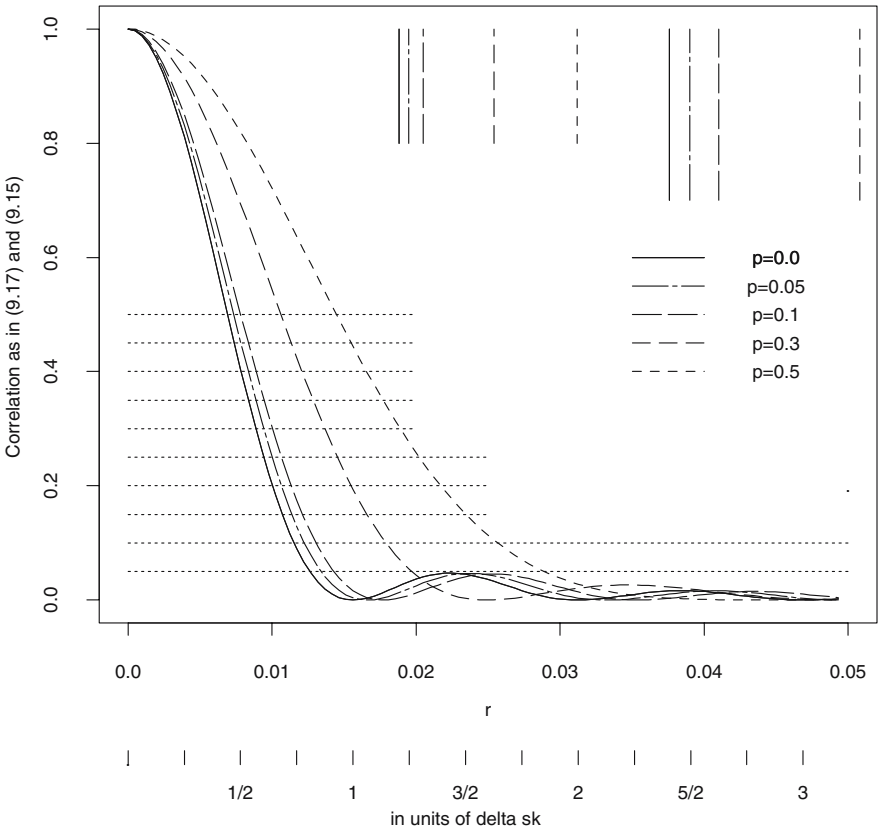


Fig. 9.6. Correlation of direct spectral estimators calculated using cosine tapers $\widehat{h_N}(t)$ as defined in (6.111,1) with parameters $N = 64$ and $p = 0.0, 0.05, 0.1, 0.3, 0.5$: $H^{\oplus}H(r)$ as defined in (9.15) as well as (vertical lines) approximations for $\mathcal{B}_{\triangle}^{\cap}\big(|\widehat{H_N}(s)|\big)$ and $2\mathcal{B}_{\triangle}^{\cap}\big(|\widehat{H_N}(s)|\big)$ as in Table 6.2. r is the lag or displacement in the frequency domain and can be read in units of $\Delta s_k = 1/N$ as well.

periodogram becomes negligible for $r > 2\mathcal{B}_{\triangle}^{\cap}\big(|D_N(s)\mathrm{e}^{-\mathrm{i}\pi(N-1)s}|\big)$. This specifies (8.24,6) more precisely: (i) $\mathrm{Cov}\big(I_X(s), I_X(s+r)\big) \approx 0$, for $r > 12/(5N)$ and s and $s + r$ not being identical with Fourier frequencies s_k, and (ii) $\mathrm{Cov}\big(I_X(s), I_X(s+r)\big) = 0$ for s and $s + r$ being identical with s_k.

The correlations of spectral estimators calculated subsequent to multiplying the observations with cosine tapers $\widehat{h_N}(t)$ for $p = 0.05$, $p = 0.1$, $p = 0.3$ and $p = 0.5$ are drawn using broken lines in Fig. 9.6, together with the pertaining $\mathcal{B}_{\triangle}^{\cap}\big(|\widehat{H_N}(s)|\big)$. For moderate ($p \le 0.1$) cosine tapers, the correlations are not far from the correlation of the periodogram (solid line), i.e., $H^{\oplus}H(r) \le 0.05$ for $r > \mathcal{B}_{\triangle}^{\cap}\big(|\widehat{H_N}(s)|\big)$ and $H^{\oplus}H(r) \approx 0$ for $r > 2\mathcal{B}_{\triangle}^{\cap}\big(|\widehat{H_N}(s)|\big)$,

since the main lobes of $|\widehat{H}_N(s)|$ are not much wider than the main lobe of $|D_N(s)\mathrm{e}^{-\mathrm{i}\pi(N-1)s}|$ obtained in (6.54) as the kernel pertaining to $\Pi_N(t)$ in (6.52,1), with $D_N(s)$ being the Dirichlet kernel in (6.55).

Fig. 9.6 allows for determining r_0 such that $\mathrm{Cor}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s+r)\big) = H{\oplus}H(r) \approx 0$ for $r \geq r_0 = n\Delta s_k$, with $\mathrm{Cor}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s+r)\big)$ as approximated in (9.17) and n an integer number as small as possible. For all examples in Fig. 9.6, $n = 2$ is arrived at, because $H{\oplus}H(r) < 0.05$ for $r \geq 2\Delta s_k$. n is the minimal lag (in units of Δs_k) for which the correlations in a direct spectral estimator become negligible.

In Fig. 9.6 it is evident that correlations $H{\oplus}H(r)$ of a direct spectral estimator calculated from observations "tapered" with a rectangle sequence (the periodogram) differ only slightly from those correlations of a direct spectral estimator calculated using a moderate ($p \leq 0.1$) cosine taper. However, (8.24,4) implies that the covariances of the periodogram become approximately zero for Fourier frequencies s_k, whereas the covariances of a direct spectral estimator are, with $H{\oplus}H(r)$, proportional to the square of the spectrum to be estimated, as is concluded from (9.14). In applications therefore, periodogram values $I_X(s_k)$ and $I_X(s_{k-1})$, in lag $\Delta s_k = s_k - s_{k-1} = 1/N$, are approximately non-correlated; in contrast, values $\widehat{C}_X^{(d)}(s_k)$ in a direct spectral estimator are approximately uncorrelated for lags $r \geq 2\Delta s_k$.

The correlations in a direct spectral estimator can be demonstrated using simulated realisations obtained from a first process having a spectrum with a small dynamic range (15 dB, AR[2] model in Fig. 7.12) and a second process having a spectrum with a large dynamic range (65 dB, AR[4] model in Fig. 7.14). These spectra are plotted with solid lines on the left in Figs. 9.7 and 9.8, together with periodograms and direct spectral estimators (using cosine tapers) obtained from time slices of length $N = 512$ in simulated realisations of these models. From, for example, 1000 simulated realisations of these models, 1000 periodograms and 1000 direct spectral estimators could, in fact, be calculated using both cosine tapers and, thereafter, histograms and scatterplots of the spectral estimators could be obtained as demonstrated in Figs. 2.5 and 2.6, and empirical correlations calculated. It is, however, more expedient to calculate the empirical correlations of the estimators from only one realisation, provided that the spectral estimator is not biased.

The estimators plotted in Figs. 9.7 and 9.8 (except for the periodogram shown in plot (a) in Fig. 9.8) are not biased, i.e., $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X(s)$, as is (i) obvious from the plots and (ii) in line with the remarks to Figs. 8.6 and 8.7 as well as Figs. 9.3 and 9.4. Further, $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s)$ in (8.3.8), because both models are reconcilable with assumptions (9.13). Under these preconditions and because the spectra of the simulated processes are known, estimators $\widehat{C}_X^{(ds)}(s_k) = \big(\widehat{C}_X^{(d)}(s_k) - C_X(s_k)\big)/C_X(s_k)$ are arrived at. These estimators are approximately stationary . Thereafter, the empirical correlation functions $\mathrm{Cor}\big(\widehat{C}_X^{(ds)}(s_k), \widehat{C}_X^{(ds)}(s_{k+l})\big)$ are calculated by applying definition

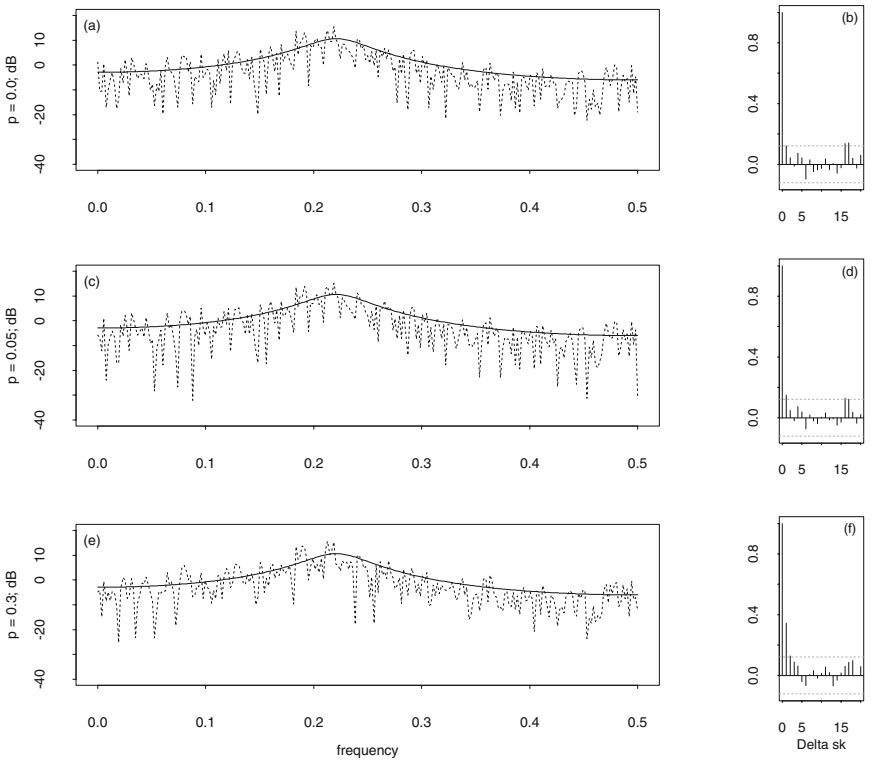


Fig. 9.7. On the left, periodogram and direct spectral estimators calculated using cosine tapers with $p = 0.05$ and $p = 0.3$ (from top to bottom) from a time slice of length $N = 512$ in a realisation of the AR[2] model in Fig. 7.12; on the right, the empirical correlation functions of the estimators, calculated subsequent to a normalisation with the model spectrum (solid line).

(2.1,4) for lags $l\Delta s_k$, $l = 1, 2, \ldots$, and plotted on the right in Figs. 9.7 and 9.8. The bias of the periodogram in Fig. 9.8 (a) is large, and thus it is not possible to construct a stationary estimator from this periodogram, which further implies that no empirical correlations are available.

Do the empirical correlation functions plotted on the right in Figs. 9.7 and 9.8 stem from realisations of white noise processes? When these plots are compared with those on the right in Fig. 2.22, it is then concluded from (2.59) that

1. the direct spectral estimators calculated using cosine tapers with $p = 0.05$ and $p = 0.30$ in plot (c) and (e) in Fig. 9.7 have correlations for lag $1\Delta s_k = 1/N = 1/512$ that are small but not negligible, at least in the

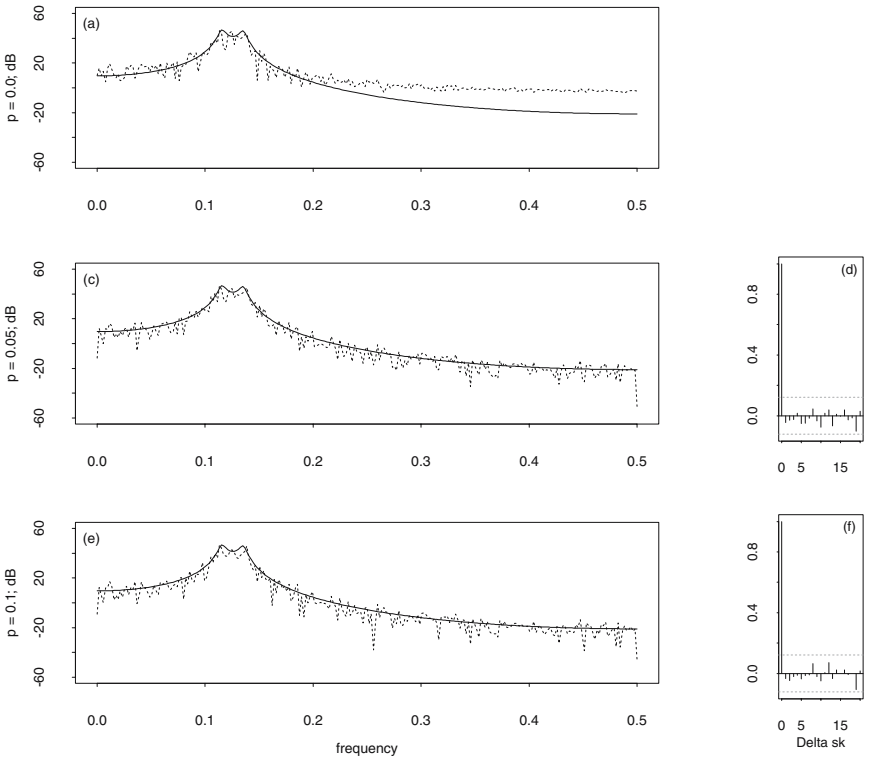


Fig. 9.8. As Fig. 9.8, with the difference that the estimators are calculated from a realisation of the AR[4] model in Fig. 7.14 using cosine tapers with $p = 0.05$ and $p = 0.1$

case of the estimator obtained using a cosine taper with $p = 0.30$, as is seen in plot (f)

2. the direct spectral estimators calculated using cosine tapers with $p = 0.05$ and $p = 0.10$ in plot (c) and (e) in Fig. 9.8 have correlation functions that remain safely in the interior of their .95 confidence intervals and thus are non-correlated, and
3. the periodogram in plot (a) in Fig. 9.7 is non-correlated for lags $1\Delta s_k$, $2\Delta s_k$, $3\Delta s_k$, ..., in agreement with (8.24,4).

Consequently, when the spectrum with a small dynamic range of approximately 15 dB pertaining to the AR[2] model in Fig. 7.12 is estimated in Fig. 9.7, then a cosine taper does not reduce the bias of the periodogram (the periodogram in plot (a) is approximately unbiased) but merely generates, in the estimator, correlations that are not wanted. In Fig. 9.8 however, a cosine taper does reduce the bias in the estimators for a spectrum with a

large dynamic range of approximately 45 dB pertaining to the AR[4] model in Fig. 7.14, and the correlations in the estimator for lags Δs_k, $2\Delta s_k$, ... remain safely in the interior of their .95 interval. These results are in line with the correlations of a direct spectral estimator calculated using cosine tapers that can be inferred from Fig. 9.6.

As a general conclusion drawn from the simulation experiments performed in Figs. 9.7 and 9.8, it is recommended to use moderate data tapers (e.g., cosine tapers with $p \leq .10$) for spectral estimation (on condition that a sufficient reduction of the bias will be arrived at using such a taper).

9.2.3 Probability Distribution

The asymptotic distribution for $N \to \infty$ is derived, e.g., in [20], with the result in (9.19).

Let $\widehat{C}_X^{(d)}(s)$ be a direct spectral estimator (9.3) for the continuous spectrum $C_X(s)$ of model (X_t) (9.1). Then, for $N \to \infty$:

1. $\big(2/C_X(s)\big)\widehat{C}_X^{(d)}(s)$ *is* $\chi^2_{(2)}$ *distributed for* $-1/2 < s < 0$ *and* $0 < s < 1/2$, *and*
2. $\big(1/C_X(s)\big)\widehat{C}_X^{(d)}(s)$ *is* $\chi^2_{(1)}$ *distributed for* $s = -1/2, 0, 1/2$.

(9.19,1,2) become approximations for large finite N. (9.19)

Using the properties of the $\chi^2_{(n)}$ distribution as enumerated in the remarks to (8.8), it can be inferred from (9.19,1,2) that, in the limiting case for $N \to \infty$, the variance of the direct spectral estimator is proportional to the square of the spectrum to be estimated. This result is in line with the one proposed for finite N in (9.16) under the assumptions in (9.13). (9.16) and (9.19) imply that the variance of the direct spectral estimator does not decrease with with increasing N: $\widehat{C}_X^{(d)}(s)$ is not a consistent estimator for $C_X(s)$, a property shared by the periodogram, as is concluded from (8.24,6,7).

Consequently, a direct spectral estimator $\widehat{C}_X^{(d)}(s)$ and a periodogram $I_X(s)$, both calculated from identical observations (x_t), $t = 0, 1 \ldots, N-1$, in a realisation (x_t) of model (X_t) as defined in (9.1), do have identical variance, a result in contradiction to the one supposed in the remarks concluding Sect. 9.2.1. However, these estimators are distinct in their covariance functions: $\mathrm{Cov}\big(I_X(s_k), I_X(s_l)\big) \approx 0$ for $k \neq l$, s_k and s_l Fourier frequencies as defined in (6.21,1), whereas $\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s_k), \widehat{C}_X^{(d)}(s_l)\big) \approx 0$ for $(l-k)\Delta s_k \geq 2\mathcal{B}_{\triangle}^{\cap}\big(|H_N(s)|\big)$ as required in (9.18). Thus, the reduction of the bias arrived at by tapering is compensated by covariances in the direct spectral estimator thus obtained that are not negligible for lag Δs_k (and possibly neither for $2\Delta s_k$ when the data taper is not moderate).

The probability distribution of the direct spectral estimator as proposed in (9.19) is demonstrated in Fig. 9.9 using direct spectral estimators calculated from time slices of lengths $N = 256$, $N = 512$ and $N = 1024$ in a realisation of

the AR[4] model in Fig. 7.14. These time slices are multiplied with a cosine taper $p = 0.075$, chosen such that estimators are obtained which are not biased. What can be seen in Fig. 9.9?

Firstly, $\widehat{C}_X^{(d)}(s_k)$, i.e., the direct spectral estimators calculated for Fourier frequencies s_k, fluctuate in the logarithmic plots on the left in Fig. 9.9 within a confidence interval of constant height, reconcilable with the $\chi^2_{(2)}$ approximated distribution of $\big(2/C_X(s_k)\big)\widehat{C}_X^{(d)}(s_k)$ for $-1/2 < s < 0 < s < 1/2$ in (9.19,1,2). This approximation implies $\Pr\big(Q_2(p) \le \big(2\widehat{C}_X^{(d)}(s_k)\big)/C_X(s_k) \le Q_2(1-p) \approx 1-2p$, with $Q_2(p)$ being the p quantile of the $\chi^2_{(2)}$ distribution, and, consequently, $\big(2\widehat{C}_X^{(d)}(s_k)\big)/Q_2(1-p) \le C_X^2(s_k) \le \big(2\widehat{C}_X^{(d)}(s_k)\big)/Q_2(p)$ is the approximate $2p$ confidence interval for $\widehat{C}_X^{(d)}(s_k)$. Both bounds of this confidence interval are proportional to $\widehat{C}_X^{(d)}(s_k)$, implying a constant ratio bound/estimator which becomes, subsequent to a logarithmic transform, the constant difference between bound and estimator. More on confidence intervals of direct spectral estimators can be read in Sect. 9.3.4, which introduces confidence intervals for smoothed direct spectral estimators.

Secondly, all three direct spectral estimators fluctuate with approximately identical amplitudes, irrespective of N. This result is in line with the histograms of $\widehat{C}_X^{(d)}(s_k)/C_X(s_k)$, i.e., the direct spectral estimators normalised with the spectrum (feasible in simulation experiments), as plotted on the right in Fig. 9.9. These histograms do not change when N changes, and, consequently, the distribution of a direct spectral estimator does not depend on N, as required in (9.19).

The same two results also apply to the periodogram (Problem 8.8). As a third result, applying both to the periodogram and the direct spectral estimator, use the hints given in Problem 8.2 to show that a direct spectral estimator is no longer exponentially distributed subsequent to taking the logarithms, as is obvious from the plots on the left in Fig. 9.9.

It is demonstrated in Fig. 9.9 that the variance of a direct spectral estimator does not depend on the length of the observational period N and thus does not decrease when N increases, reconcilable with the approximations obtained in (9.19) and (9.16). Consequently, a direct spectral estimator is not a consistent estimator as defined in (1.4). This property is shared by the periodogram as implied by (8.24,5,7) and is inconvenient for the reasons already mentioned in the remarks to (8.24):

1. an estimator having a smaller variance than $\mathrm{Var}\big(I_X(s)\big) \approx C_X^2(s)$ or $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s)$, both for $-1/2 < s_k < 0 < s_k < 1/2$, as obtained in (8.24,7) and (9.16), is required
2. both estimators are not smooth whereas the spectrum is often smooth or contains smooth segments (confined by frequencies with jump discontinuities in the integrated spectrum), and

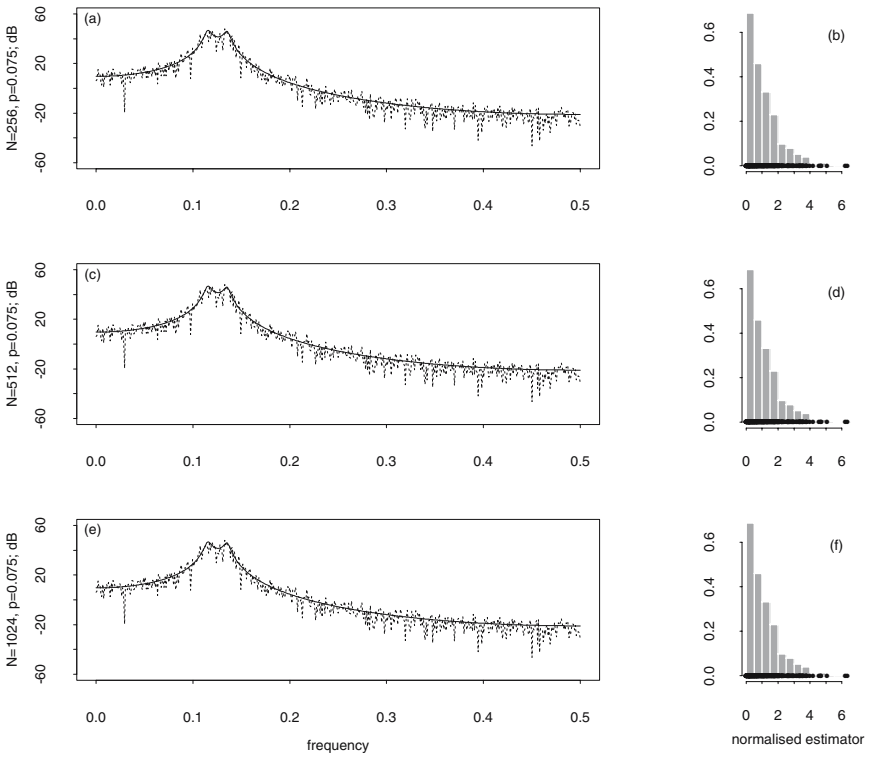


Fig. 9.9. On the left, direct spectral estimators $\widehat{C}_X^{(d)}(s)$ calculated as required in (9.3) using cosine tapers with $p = 0.075$ from time slices of lengths $N = 256$ (a), $N = 512$ (c) and $N = 1024$ (e) in a realisation of the AR[4] model in Fig. 7.14, together with, on the right, histograms of $\widehat{C}_X^{(d)}(s)(s_k)/C_X(s_k)$.

3. a wildly fluctuating estimator can suggest that the spectrum of the process under analysis is discrete or mixed, even though it is continuous as defined in (7.71,1).

For example, when the spectrum of the AR[4] model in Fig. 7.14 is estimated, the variance of a direct spectral estimator is approximately 40000^2 for the peak frequencies, and $(10^{-2})^2$ for the high frequencies, i.e., intolerably large.

An increase in the length of the observational period does not reduce the variance of either the periodogram or of the direct spectral estimator. However, the estimators are calculated on a finer grid of frequencies as implied by (6.21,1). For example, the direct spectral estimators plotted on the left in Fig. 9.9 have been calculated for 2, 5 and 10 Fourier frequencies in band $0.11 \leq s \leq 0.12$, corresponding to $N = 256$, $N = 512$ and $N = 1024$ observations, as shown in Table 9.2.

In Sect. 9.3, a direct spectral estimator is smoothed to reduce its variance.

Table 9.2. Direct spectral estimators $\widehat{C}_X^{(d)}(s_k)$ in Fig. 9.9 in band $0.11 \le s \le 0.12$.

$N \setminus s_k$	.1103	.1113	.1123	.1133	.1142	.1152	.1162	.1172	.1182	.1191
256				19474				11877		
512		19699		7819		7042		19031		28491
1024	7882	20611	20606	27109	25716	13248	5604	50106	9713	31727

9.2.4 Calculated for Non-Fourier Frequencies

A direct spectral estimator as defined in (9.3) can be calculated for discrete frequencies s'_k which are not required to be the Fourier frequencies pertaining to N as defined in (6.21,1): Frequencies s'_k resulting in (9.20,2) from zero-padding the tapered observations are on a finer grid than the Fourier frequencies s_k, since $\Delta s'_k = s'_k - s'_{k-1} = k/N' - (k-1)/N' = 1/N' < \Delta s_k = 1/N$.

Let $\big(h_N^{(d)}(t)(x_t)\big)$, $t = 0, 1, \ldots, N-1$, *be observations stemming from a stationary stochastic process, multiplied with a variance-normalised data taper as defined in (9.2). This sequence is zero-padded as defined in the remarks to (6.26) to arrive at* $N' > N$ *values:*

1. $\big(h_N^{(d)}(t)(x_t)\big)' = \big(h_N^{(d)}(0)\times x_0, h_N^{(d)}(1)\times x_1, \ldots, h_N^{(d)}(N-2)\times x_{N-2}, h_N^{(d)}(N-1)\times x_{N-1}, 0_N, \ldots, 0_{N'-1}\big)$.

   Then $\mathcal{F}_{-i}\big(h_N^{(d)}(t)(x_t)\big)'$, *the discrete Fourier transform of the tapered and zero-padded observations, is calculated as required in (6.26) for a finer grid of*
2. *frequencies* $s'_k = k/N'$, $k = 0, 1, \ldots, m'$, *with* m' *being the largest integer number* $\le N'/2$, *and,*
3. $\big(\mathcal{F}_{-i}\big(h_N^{(d)}(t)(x_t)\big)'\big)(s'_k) = \big(\mathcal{F}_{-i}\big(h_N^{(d)}(t)(x_t)\big)\big)(s_k)$, *for* $s'_k = s_k$, *is obtained using the rationale in the remarks to (6.26).*

(9.20)

For example, direct spectral estimators $\widehat{C}_X^{(d)}(s'_k)$ are calculated from a time slice of length $N = 64$ in a realisation of the AR[4] model in Fig. 7.14 for frequencies s'_k as defined in (9.20,2), by zero-padding the tapered observations as required in (9.20,1). Thereafter, their empirical correlation functions are calculated in the same manner as those in Figs. 9.7 and 9.8. The estimators for $N' = 96$ and $N' = 128$ are shown in plots (c) and (e) in Fig. 9.10, their correlation functions in plots (d) and (f). These show correlations at the exterior of the .95 confidence intervals (2.59) for lag $1\Delta s'_k$, $\Delta s'_k = 1/96$, in plot (d), and lags $\le 5\Delta s'_k$, $\Delta s'_k = 1/128$, in plot (f). These correlation plots also demonstrate that the confidence intervals become narrower with increasing m' (and not m), with m' being the number of values $\widehat{C}_X^{(d)}(s'_k)$ as required in (9.20,2).

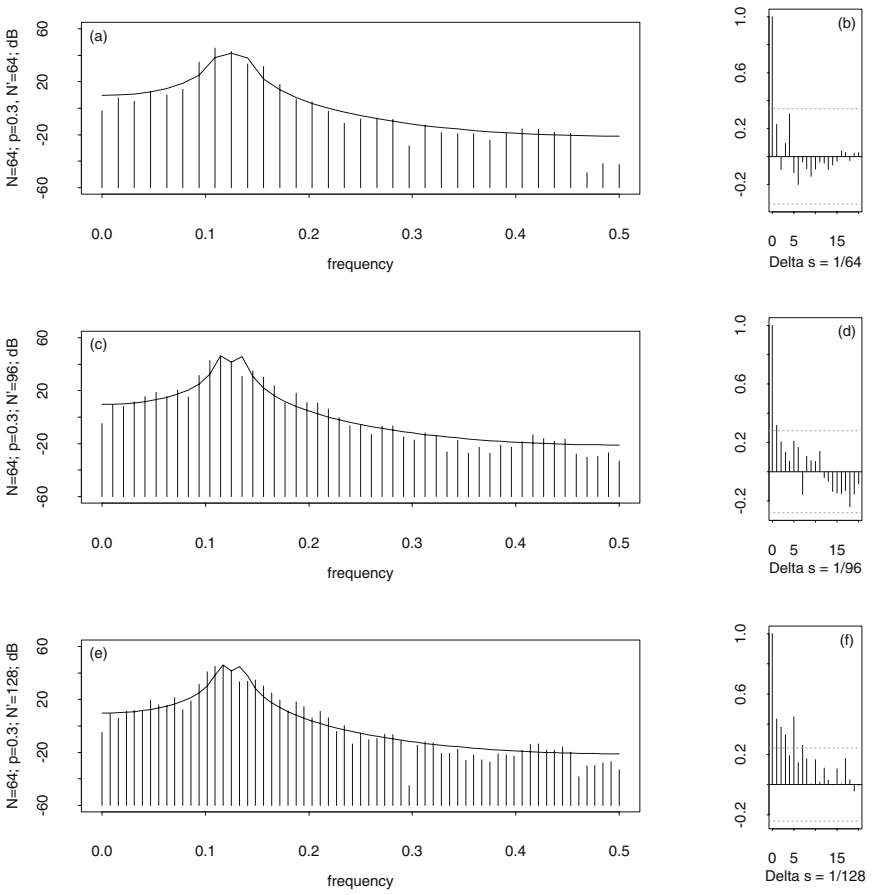


Fig. 9.10. On the left, spectrum of the AR[4] model in Fig. 7.14 together with direct spectral estimators (vertical lines) calculated using a cosine taper for $N = 64$ and $p = 0.3$ for frequencies s'_k as defined in (9.20) ($N' = 64, 96, 128$, from top to bottom). On the right, the empirical correlation functions of the estimators, calculated as in Fig. 9.7.

In the empirical correlation functions plotted on the right in Fig. 9.10, the maximal lag $n\Delta s'_k$ with a non-zero correlation increases with decreasing $\Delta s'_k$, i.e., increasing N'. This behaviour is induced by $H \oplus H(r)$ as defined in (9.15) (examples are given in Fig. 9.6) which does not depend on N'. $H \oplus H(r)$ is not dependent on N' because $h_N^{(d)}(t)$ does not depend on N' in (9.20), which implies that $h_N(t)$ does not depend on N' either.

$\widehat{C}_X^{(d)}(s'_k)$, with $\Delta s'_k = 1/N' = 1/(2N)$, $N' = 2N$, and $\hat{c}_X^{(d)}(\tau)$ defined in (9.7) are a discrete Fourier transform pair as defined in (6.22,4) and (6.26), a

result obtained in (9.23) and (9.24). There, $\widehat{C}_X^{(d)}(s'_k)$ is the direct spectral estimator calculated as required in (9.3) for frequencies s'_k with $\Delta s'_k = (1/2)\Delta s_k$ as defined in (6.21); and $\hat{c}_X^{(d)}(\tau)$ is the empirical covariance function calculated from tapered (using a variance-normalised data taper as defined in (9.2,1)) observations of a stationary stochastic process.

Frequencies s'_k with $\Delta s'_k = 1/(2N)$ are defined in (9.21). In (9.22), the discrete Fourier transform $G_{2N}(s'_k)$ of $\hat{c}_X^{(d)}(\tau)$, s'_k frequencies as in (9.21), is calculated as required in (6.22,4) and (6.26). From this transform, (9.23) results because $\hat{c}_X^{(d)}(\tau) = 0$, for $|\tau| \geq N$, as defined in (9.7). This Fourier sum contains $(N-1)+1+N = 2N$ terms and the s'_k defined in (9.21) are its frequencies. The inverse transform is arrived at in (9.24) by applying (6.22,4) and (6.27).

$$s'_k = k/N' = k/(2N) \quad \text{with} \quad \begin{array}{l} k = -(N-1), \ldots, -1, 0, 1, \ldots, N \\ \text{and therefore } -1/2 \leq s'_k \leq 1/2 \end{array} \tag{9.21}$$

$$G_{2N}(s'_k) = \sum_{\tau=-(N-1)}^{N-1} \hat{c}_X^{(d)}(\tau) \mathrm{e}^{-\mathrm{i}2\pi s'_k \tau} \tag{9.22}$$

$$= \sum_{\tau=-(N-1)}^{N} \hat{c}_X^{(d)}(\tau) \mathrm{e}^{-\mathrm{i}2\pi s'_k \tau} = \widehat{C}_X^{(d)}(s'_k) \tag{9.23}$$

$$\hat{c}_X^{(d)}(\tau) = \frac{1}{2N} \sum_{k=-(N-1)}^{N} G_{2N}(k/(2N)) \mathrm{e}^{+\mathrm{i}2\pi(\tau/(2N))k} \quad \text{with} \tag{9.24}$$

$$\tau = -(N-1), \ldots, -1, 0, 1, \ldots, N \tag{9.25}$$

Using (9.23), a direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$, s'_k frequencies with $\Delta s'_k = 1/(2N)$, is calculated as the discrete Fourier $-$i-transform of $\hat{c}_X^{(d)}(\tau)$ defined in (9.7), and, vice-versa, $\hat{c}_X^{(d)}(\tau)$ is obtained as the discrete Fourier $+$i-transform of $\widehat{C}_X^{(d)}(s'_k)$ using (9.24). This discrete Fourier transform pair favours the derivations in Sect. 9.3.2.

When (9.23) is applied, one arrives at a direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$, s'_k frequencies with $\Delta s'_k = 1/(2N)$, on the second path, as demonstrated in Problem 9.5. On the first path, however, $\widehat{C}_X^{(d)}(s'_k)$ is obtained more readily: the observations are tapered, the sequence thus obtained is zero-padded and then (9.3) is applied. For example, the direct spectral estimator as plotted in Fig. 9.10 (e) is calculated from a time slice of length $N = 64$ in a realisation of the model in Fig. 7.14 using `spec.univariate(...,taper=0.30,pad=1,...)` introduced in Sect. 9.5.6.

Alternatively, a direct spectral estimator that (i) applies a cosine taper as defined in (6.111,1) and (ii) pads as required in (9.20,1) with N zeros to arrive at $N' = 2N$, is calculated using `spec.pgram(...,taper=0.30,pad=1,...)`.

9.2.5 Alternatives: Parametric Estimation, Pre-Whitening

The bias that has to be taken into account when a spectrum with a large dynamic range is estimated using the periodogram can be reduced on condition that the observations are tapered. The bias of the direct spectral estimator thus obtained in Sect. 9.2.1 is considerably smaller than the one of the periodogram. The favourable reduction of the bias is compensated by correlations in the direct spectral estimators which are brought about by the data taper and therefore are not present in the periodogram, as can be seen in Sect. 9.2.2. The unfavourable properties (biased and/or correlated) of both, the periodogram and direct spectral estimator, motivate the introduction of two alternatives to using a data taper in this section: (i) parametric spectral estimation and (ii) pre-whitening.

A parametric estimation for the spectrum of the model in (9.1,1) is arrived at by fitting a linear process to the observations using the procedures introduced in Sect. 5.4 ([83], [89]). Often, autoregressive models are estimated ([26], [60], [10]). Thereafter, estimates for the model spectrum are calculated as demonstrated in Sect. 7.4.2, with estimates substituted for model parameters.

For example, the AR[4] model $X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9838X_{t-4} + W_t$, with $\mu_W = 0$ and $\sigma_W^2 = 1$, is estimated using $N = 512$ simulated values as plotted above in Fig. 7.14. From this simulated time series, Burg's algorithm (5.21) produces $\tilde{a}_1 = 2.7483216$, $\tilde{a}_2 = -3.7817308$, $\tilde{a}_3 = 2.6122472$, $\tilde{a}_4 = -0.9038562$ and $\tilde{\sigma}_W^2 = 0.949931$, having only small deviations from the maximum likelihood estimates $\breve{a}_1 = 2.7504098$, $\breve{a}_2 = -3.7834073$, $\breve{a}_3 = 2.6127735$, $\breve{a}_4 = -0.9027626$ and $\breve{\sigma}_W^2 = 0.9474084$. With these values acting as substitutes for the model parameters in the R expressions given in the remarks to Fig. 7.14, those empirical spectra are obtained which are plotted with broken lines in Fig. 9.11 (a) in the band $0.1 \leq s \leq 0.15$. From this plot, it becomes evident that (i) the maximum likelihood estimate, despite being too low, estimates the frequencies of the peaks satisfactorily and (ii) Burg's estimate underestimates both the height and frequencies of the peaks. At the exterior of $0.1 \leq s \leq 0.15$, the differences between the estimates and the spectrum are small.

The approximations (5.26) imply that both estimates, the maximum likelihood and Burg's, become more accurate on condition that the number of observations increases. This property is inherited by the empirical spectra computed from the parameter estimates, and therefore the empirical spectra come closer to the spectrum of the model when the length N of the simulated time series increases, as demonstrated in Problem 9.7. For very large N, however, Burg's estimate is preferred to the maximum likelihood one, since the time needed for the computation of the maximum likelihood estimates can be considerable.

As demonstrated in Fig. 9.11 (a), parametric spectral estimates can be biased owing to errors in the empirical parameters; these errors, however,

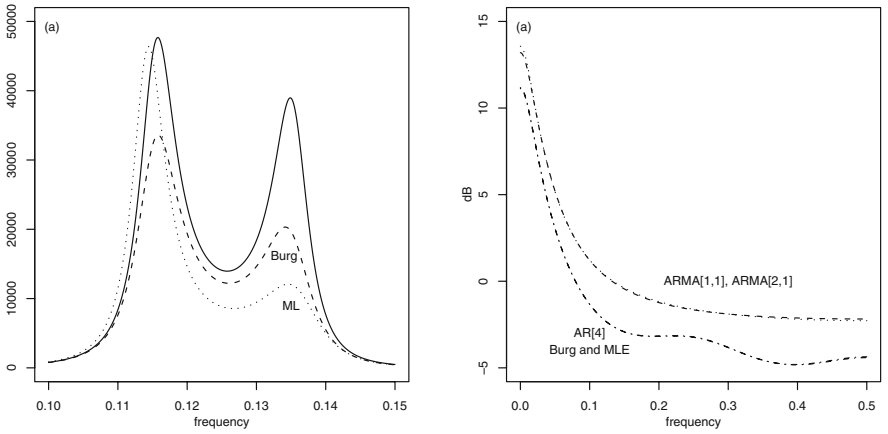


Fig. 9.11. On the left, spectrum of the AR[4] model in Fig. 7.14 in the interval with the peaks (solid line) together with two parametric estimates; on the right, empirical spectra of the SO index obtained from the estimates in Table 5.3.

decrease with increasing number of observations available. In addition, parametric spectral estimates can be biased due to uncertainties in the empirical parameters stemming from models ambiguously identified since contradictory results can be obtained from the procedures introduced in Sects. 5.4.1 and 5.4.2. In Table 5.3, for example, estimates for AR[4], ARMA[1, 1] and ARMA[2, 1] models for the SO index are given, all of which are considered to fit the data when the procedures introduced in Sect. 5.4.2 are applied. The empirical spectra of the SO index obtained from these estimates are plotted in Fig. 9.11 (b). The empirical AR[4] spectra (Burg's and maximum likelihood estimates) are practically identical, i.e., they cannot be distinguished when plotted. Slightly larger are the differences between the maximum likelihood estimates of the ARMA[1, 1] and ARMA[1, 2] spectra. However, both ARMA spectra are approximately 2 dB larger than their AR[4] counterpart. The empirical spectra of the SO index obtained from the estimates of the AR[4], ARMA[1, 1] and ARMA[1, 2] models in Table 5.3 demonstrate that the uncertainties due to a difficult model identification are inherited by the empirical spectra.

Pre-whitening is the second alternative to tapering an observed time series (x_t), $t = 0, 1, \ldots, N-1$. As a prerequisite for pre-whitening, it is assumed that the continuous spectrum $C_X(s)$ of the process under analysis (X_t) has a large dynamic range and is roughly known, with details to be estimated. This knowledge stems from previous experiments and/or is derived from theory. Then the observations (x_t), $t = 0, 1, \ldots, N-1$, are modified by applying a stochastic filter as defined in (7.84) such that they appear, subsequent to filtering, as a realisation of a process having a spectrum with a very small dynamic range, i.e., a spectrum that is approximately flat: $(y_t) = (g_t) * (x_t)$,

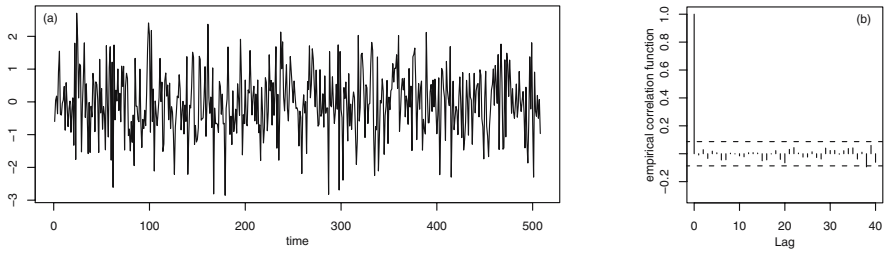


Fig. 9.12. Output (y_t) of a pre-whitening filter as defined in (8.3.13) having the realisation (x_t) of the AR[4] model in Fig. 7.14 (a) as its input, together with its empirical correlation function $\hat{c}_Y(\tau)$.

with (g_t) being the impulse response sequence and (y_t) a realisation of the output process (Y_t) whose spectrum $C_Y(s)$ has a small dynamic range.

If (Y_t) is the convolution of (g_t) and (X_t) then $C_Y(s) = |G(s)|^2 C_X(s)$ is obtained from (7.90,4), with $G(s)$ the frequency response function as defined in (7.84). The output spectrum $C_Y(s)$ has a small dynamic range on condition that (9.27) holds, and, under this condition, can be readily estimated by calculating the output periodogram $I_Y(s)$ from (y_t), as demonstrated in Fig. 8.4. Subsequent to estimating $C_Y(s)$, $C_X(s)$ is arrived at using (9.26), and the favourable properties of $I_Y(s)$ as proposed in (8.7) are inherited by the *pre-whitened estimator* $\widehat{C}_X^{(pw)}(s)$ for $C_X(s)$, on condition that $C_Y(s)$ is approximately flat.

$$(y_t) = (g_t) * (x_t) \qquad C_Y(s) = |G(s)|^2 C_X(s) \tag{9.26}$$

$$\left(|G(s)|^2\right) \frac{1}{C_X(s)} \approx \text{constant} \tag{9.27}$$

In order to arrive at a pre-whitening estimate therefore, a filter (g_t) in (9.26) is required such that (9.27) holds. Very often, (g_t) is obtained from an AR$[p]$ model fitted to (x_t) using the procedures introduced in Sect. 5.4. Assuming $\mu_X = 0$ and using only one index in (5.2), where the AR$[p]$ model is defined, $(w_t) = (1, -a_1, \ldots, -a_p) * (x_t)$ is obtained, with (w_t) stemming from a white noise process and $(a_1, \ldots, a_p)$ being the model coefficients. Then an MA$[\infty]$ representation $(x_t) = (1, -a_1, \ldots, -a_p)^{-1} * (w_t)$ of this realisation will exist. If both sides in this MA$[\infty]$ representation are convolved with $(1, -a_1, \ldots, -a_p)$ then (w_t) is regained. Since (w_t) has a flat spectrum, a pre-whitening filter as required in (9.26) and (9.27) has been found: (x_t) is the input, $(1, -a_1, \ldots, -a_p)$ the impulse response sequence, (w_t) the output, and $G(s) = \mathcal{F}_{-i}(1, -a_1, \ldots, -a_p)$ the frequency response function. $G(s)$ and $|G(s)|^2$ can be calculated using the R expressions in the remarks to Fig. 7.14, once estimates (Burg's or ML, as introduced in Sect. 5.2) and for $(1, -a_1, \ldots, -a_p)$ have been obtained.

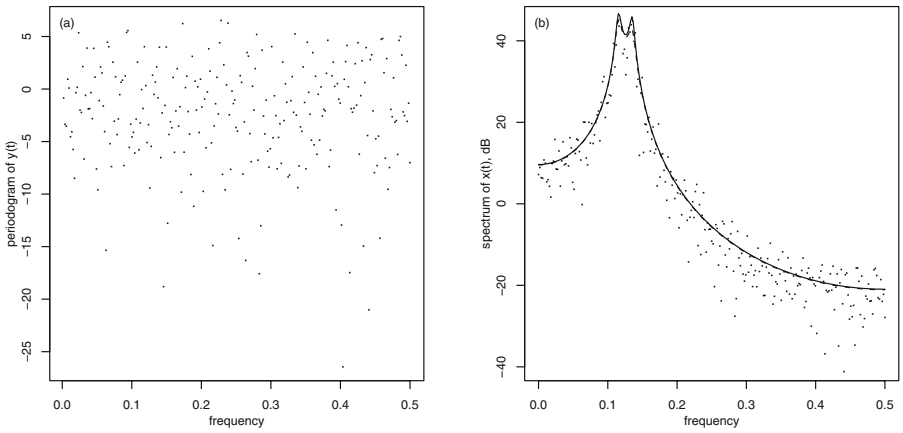


Fig. 9.13. Periodogram $I_Y(s_k)$ in dB obtained from $y(t)$ as plotted in Fig. 9.12 (on the left) and $\widehat{C}_X^{(pw)}(s_k)$ (on the right), both plotted with small dots. On the right, in addition, the spectrum of the AR[4] model (solid line) together with its parametric estimation obtained using Burg's algorithm (broken line), details of Burg's estimation in Fig. 9.11.

For example, the spectrum of the AR[4] model in Fig. 7.14 is estimated in the following paragraphs. From the time slice in a simulated realisation in Fig. 7.14 (a), Burg's estimates $\tilde{a}_1 = 2.7678735$, $\tilde{a}_2 = -3.8178623$, $\tilde{a}_3 = 2.6453416$, $\tilde{a}_4 = -0.9142806$ as well as $\tilde{\sigma}_W^2 = 1.098001$ are calculated in the remarks to Fig. 9.11 (a). From these estimates, the impulse response function (g_t) in a pre-whitening filter as derived above readily follows, and

```
#R time series object x: simulation in Fig. 7.14 (a)
#R vector ar: Burg's estimates obtained from x
ar <- c(2.7678735, -3.8178623, 2.6453416, -0.9142806)
g <- concat(1, -ar)
y <- filter(x,g,method="convolution",sides=1)
#R time series object y: output of the pre-whitening filter
```

$(y_t) = (1, -\tilde{a}_1, -\tilde{a}_2, -\tilde{a}_3, -\tilde{a}4) * (x_t)$ is obtained as output. (y_t) is plotted in Fig. 9.12 together with its empirical correlation function $\hat{c}_Y(\tau)$. When the plots in Fig. 9.12 are compared with those in Fig. 2.22 it then becomes clear that (y_t) is approximately a realisation of a white noise process. Consequently, $I_Y(s_k)$, i.e., the periodogram of (y_t), as plotted in Fig. 9.13 (a), possesses the properties summarised in (8.7). Using

```
specy <- spec.pgram(y,taper=0.0,pad=0,...)
specydb <- 10*log10(specy$spec)
```

this periodogram is calculated as demonstrated in Sect. 8.1.

It remains to calculate the frequency response function $G(s_k)$ of the pre-whitening filter, with s_k being the Fourier frequencies for $N = 512$, i.e.,

those of $I_Y(s_k)$ obtained above. $G(s_k)$ is computed borrowing from the R expressions in the remarks to Fig. 7.14

```
s1 <- (0:256)/512   #257 frequencies with delta sk = 1/512
sw <- 1.098001      #variance of innovations (Burg's estimate)
sw2 <- rep(sw,257)
q <- as.complex(rep(1+0i,257))
for(i in 1:4) {q <- q - (ar[i]*exp(-1i*(2*i)*pi*s1)) }
#squared absolute value of frequency response function
Gsq <- sw2/((Mod(q))*(Mod(q)))
Gsqdb <- 10*log10(Gsq)            #in decibel
```

and its squared absolute value $|G(s_k)|^2$ easily follows.

$C_Y(s) = |G(s)|^2 C_X(s)$ in (9.26) implies $\widehat{C}_X^{(pw)}(s_k) = I_Y(s_k)/|G(s_k)|^2$ and, since both $I_Y(s_k)$ and $|G(s_k)|^2$ have been made available above in dB,

```
specxpwdb <- specydb - Gsqdb
```

computes the pre-whitened estimator $\widehat{C}_X^{(pw)}(s_k)$ for the spectrum of the AR[4] model as plotted in Fig. 9.13 (b). $\widehat{C}_X^{(pw)}(s_k)$ is an unbiased estimator for the spectrum of the AR[4] model, comparable to the direct spectral estimator $\widehat{C}_X^{(d)}(s)$ in Fig. 9.9 (c), except for band $0.10 \leq s_k \leq 0.15$ with its peaks. There, $\widehat{C}_X^{(pw)}(s_k)$ inherits the bias of Burg's estimate in Fig. 9.11 (a).

9.2.6 Summary

A direct spectral estimator is calculated as defined in (9.3) by means of Fourier transforming the observations subsequent to a multiplication with a variance-normalised data taper as defined in (9.2,1). Tapering plays a key role in spectrum estimation [19]. When the observations are tapered, the expectation of the periodogram in (8.21) becomes the expectation of the direct spectral estimator in (9.11), with the spectral window $\mathcal{H}_N^{(d)}(s)$ substituted for the Fejer kernel $F_N(s)$. $\mathcal{H}_N^{(d)}(s)$ has smaller side lobes than does $F_N(s)$ and, consequently, the bias due to leakage is reduced when a spectrum with a large dynamic range is estimated.

However, when a spectrum with a small dynamic range (in the extreme case, the constant spectrum of a white noise process) is estimated, the bias of the periodogram is negligible in (8.21). Since the bias cannot be reduced in this case, the non-negligible covariances of a direct spectral estimator for small lags (9.14) induced by the data taper are not compensated for. Consequently, the periodogram is the preferred estimator for (almost) flat spectra.

A sequence of data tapers with progressively increasing tail lengths is applied to arrive at the diagnostics in the remarks to Figs. 9.7 and 9.8. These diagnostics allow for deciding whether, and to what extent, tapering is required.

Alternative to tapering, a spectrum having a large dynamic range can be estimated by (i) fitting a linear process to the observations, using the

techniques in Sect. 5.4 and thereafter calculating an empirical spectrum as demonstrated in Sect. 7.4.2 (parametric spectral estimation), and (ii) by prewhitening, i.e., estimating a stochastic LTI filter (7.84) such that the spectrum of the output process has a small dynamic range.

Tapering efficiently reduces the bias in a direct spectral estimator; its variance as approximated in (9.16), however, remains proportional to the square of the spectrum to be estimated, a property shared by the periodogram, as implied by (8.24,7). This variance can be reduced by smoothing, as will be shown in Sect. 9.3.

9.3 Smoothed Direct Spectral Estimator

The variance of a direct spectral estimator for the continuous spectrum $C_X(s)$ in (9.1) is approximately the square of the spectrum to be estimated, i.e., $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s)$ as proposed in (9.16). Consequently, the variance of a direct spectral estimator does not depend on the length of the observational period N and thus does not decrease when N increases, as is demonstrated in Fig. 9.9. Obviously, the variance of a direct spectral estimator can be reduced by applying a moving average to a direct spectral estimator calculated for discrete frequencies, as is concluded from the discussion in the remarks to Fig. 9.9 and Table 9.2.

Let $\widehat{C}_X^{(d)}(s'_k)$ be a direct spectral estimator calculated for discrete frequencies s'_k as defined in (9.20,2), with $\Delta s'_k = s'_k - s'_{k-1} = 1/N'$. Then the number of s'_k in an arbitrary band having a constant width will increase on condition that $\Delta s'_k$ decreases, and, consequently, the variance of a mean calculated from those $\widehat{C}_X^{(d)}(s'_k)$ contained in the band decreases for increasing N', i.e., an increasing N when N'/N is kept constant. Thus an estimator for the spectrum is obtained which has a variance depending on the number of observations. Such an estimator can be defined and discussed in both domains:

1. in the frequency domain, local means of a direct spectral estimator calculated for discrete frequencies are obtained in Sect. 9.3.1 by smoothing with a moving average (2.39,1), a convolution corresponding,
2. in the time domain, to a multiplication of the empirical covariance function calculated from tapered observations (9.7) with the Fourier transform of the moving average, as is discussed in Sect. 9.3.2,

and therefore advantage can be taken of both forms, once the moment functions and the distribution of the smoothed direct spectral estimator have been derived in Sects. 9.3.3 and 9.3.4.

Obviously, the moving average applied for smoothing in the frequency domain depends on the bandwidth of the spectrum to be estimated: narrow and sharp forms (peaks and/or valleys) are not captured when local means of a direct spectral estimator are calculated over frequency bands being wider

than the forms of the spectrum. Decisions regarding the width of the band for smoothing become difficult when the bandwidth of the spectrum to be estimated is not known, as is discussed in Sect. 9.3.5.

9.3.1 Discretely Smoothed Direct Spectral Estimator

Discretely smoothed direct spectral estimators are defined in (9.28) using a moving average, with the properties discussed in the remarks to Fig. 6.31 and also in Problem 7.8. Unlike the example in Fig. 6.31, no values are missing at the start of the smoothed sequence or at its end, because the direct spectral estimator is periodic with period 1 and thus is available at the exterior of its definition interval $-1/2 \le s \le 1/2$, when the convolution sum for its endpoints is computed by this so-called *circular* convolution, with circular borrowed from `filter(..., sides=2, circular=T, ...)` as introduced below in the remarks to (9.29).

Let $\widehat{C}_X^{(d)}(s)$ be a direct spectral estimator (9.3) calculated for discrete frequencies $-1/2 \le s_k \le 1/2$ in (6.21) (the Fourier frequencies) or for $-1/2 \le s'_k \le 1/2$ in (9.20,2), and being periodic with period 1; and let $V_m(s_k)$ and $V_m(s'_k)$ be even (6.32,1) sequences defined on frequencies s_k or s'_k with $\sum_{j=-M}^{M} V_m(s_j) = 1$ and $\sum_{j=-M}^{M} V_m(s'_j) = 1$. Then:

1. *$V_m(s_k)$ and $V_m(s'_k)$ are called smoothing kernels, and*
2. $$\widehat{C}_X^{(m)}(s_k) = \sum_{j=-M}^{M} V_m(s_j)\widehat{C}_X^{(d)}(s_{k-j}) \quad \text{for } -1/2 \le s_k \le 1/2 \text{ by circular convolution}$$
3. $$\widehat{C}_X^{(m)}(s'_k) = \sum_{j=-M}^{M} V_m(s'_j)\widehat{C}_X^{(d)}(s'_{k-j}) \quad \text{for } -1/2 \le s'_k \le 1/2 \text{ by circular convolution}$$

are called discretely smoothed direct spectral estimators, with M and/or m being their parameters. (9.28)

The convolutions in (9.28) are linear and time-invariant transformations as defined in (2.30) and (2.32); however, they are not stochastic filters (7.84) for the reason to be given in Problem 9.9.

Both, $\widehat{C}_X^{(m)}(s_k)$ and $\widehat{C}_X^{(m)}(s'_k)$, are obtained as moving averages of direct spectral estimators calculated for discrete frequencies, and thus are calculated for discrete frequencies solely. The spectrum to be estimated, however, is continuous. Consequently, it could be more advantageous to smooth a direct spectral estimator $\widehat{C}_X^{(d)}(s)$ as defined in (9.5) for real s by making use of a smoothing kernel defined for real s in a convolution integral. This alternative is discussed in Sect. 9.3.2.

Often, a direct spectral estimator calculated for discrete frequencies is smoothed using the modified discrete Daniell kernel as defined in (9.29). If

$V_m(s_j)$ or $V_m(s'_j)$ in (9.28) are substituted with modified discrete Daniell kernels $V_{dD,M}(s_k)$ or $V_{dD,M}(s'_k)$ then (i) $\widehat{C}_X^{(d)}(s_k)$ is averaged in (9.28,2) over frequencies $s_{k-M}, \ldots, s_k, \ldots, s_{k+M}$ or (ii) $\widehat{C}_X^{(d)}(s'_k)$ is averaged in (9.28,3) over frequencies $s'_{k-M}, \ldots, s'_k, \ldots, s'_{k+M}$; i.e., within bands of widths $2M\Delta s_k$ or $2M\Delta s'_k$. These are the widths of a modified discrete Daniell kernel as obtained by applying definition (9.125).

Let $-1/2 \le s_k \le 1/2$ and $-1/2 \le s'_k \le 1/2$ be frequencies as used in (9.28) and let $M \ll N$ and $M \ll N'$, with $\ll$ smaller by orders of magnitude. Then

$$\left.\begin{matrix} V_{dD,M}(s_k) \\ V_{dD,M}(s'_k) \end{matrix}\right\} = \begin{cases} 1/(4M) & \text{for } k = -M, M \\ 1/(2M) & \text{for } -M < k < M \\ 0 & \text{for } |k| > M \end{cases} \tag{9.29}$$

is called a modified discrete Daniell kernel.

In R, a direct spectral estimator is discretely smoothed using R function `filter(..., sides=2, circular=T, ...)` as introduced in the remarks to Fig. 2.16, where a moving average and differences as defined in (2.39) are calculated. When used with these arguments, `filter()` computes a convolution assuming that the direct spectral estimator is periodic and, therefore, at the start of the smoothed sequence, and again at its end, no values are missing. In R function `spec.univariate()` as introduced in Sect. 9.5.6, a direct spectral estimator in R vector `spec` is smoothed with a modified discrete Daniell kernel using `filter()`.

For example, the direct spectral estimators in Fig. 9.10 calculated using a cosine taper with $p = 0.3$ from $N = 64$ observations in a realisation of the AR[4] model in Fig. 7.14 for (i) frequencies s_k with $\Delta s_k = 1/64$ (above, plot a) and (ii) s'_k with $\Delta s'_k = 1/128 = 1/N'$, i.e., subsequent to zero-padding as required in (9.20) (below, plot e), are plotted once more in Fig. 9.14, plots (a) and (c). These direct spectral estimators are smoothed as required in (9.28,2,3) using $V_{dD,M}(s_k)$ with $M = 1$ and $V_{dD,M}(s'_k)$ with $M = 3$, $V_{dD,M}(s_k)$ and $V_{dD,M}(s'_k)$ being modified discrete Daniell kernels as defined in (9.29). This convolution corresponds to computing local weighted means of the direct spectral estimators in moving frequency bands of widths $2M\Delta s_k = 2\Delta s_k$ and $2M\Delta s'_k = 6\Delta s'_k$. The smoothed direct spectral estimators are plotted with $\bullet$ in Fig. 9.14, plots (a) and (c).

Both estimators are smoothed in bands which are too wide, i.e., $2\Delta s_k = 2 \times (1/64) \approx 0.031$ and $6\Delta s'_k = 6 \times (1/128) \approx 0.047$, as compared to the bandwidth of the spectrum of the AR[4] model in Fig. 7.14, having a bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big) = 0.00474$ and dynamic ranges of approximately 20 dB in $0.10 \le s \le 0.15$ as well as 65 dB in $0.0 \le s \le 0.5$. Consequently, the smoothed direct spectral estimators in Fig. 9.14 have a bias which is quite large in the band with the peaks, and which decreases with increasing distance from the peaks.

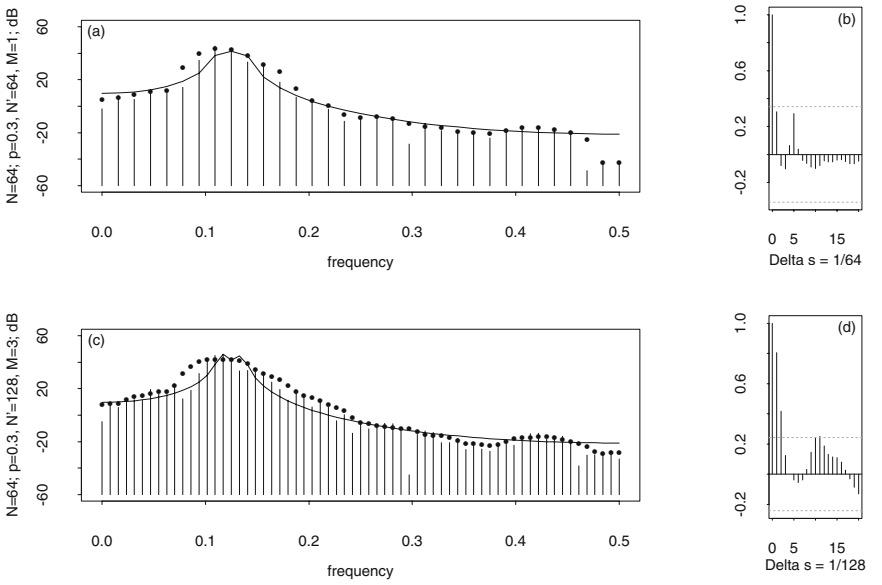


Fig. 9.14. On the left, direct spectral estimators in Fig. 9.10 (a) and (e) smoothed with modified discrete Daniell kernels with $M = 1$ and $M = 3$ (plotted with •). On the right, empirical correlation functions of the smoothed direct spectral estimators.

Both smoothed direct spectral estimators in Fig. 9.14, $\widehat{C}_X^{(m)}(s_k)$ and $\widehat{C}_X^{(m)}(s'_k)$, become approximately stationary when $\big(\widehat{C}_X^{(m)}(s_k) - C_X(s_k)\big)/C_X(s_k)$ and $\big(\widehat{C}_X^{(m)}(s'_k) - C_X(s'_k)\big)/C_X(s'_k)$ are computed as demonstrated in the remarks to Figs. 9.7 and 9.8. Thereafter, the empirical correlation functions of the transformed estimators are plotted on the right in Fig. 9.14. The empirical correlations thus obtained are larger that those of the direct spectral estimators, as becomes obvious when plots (b) and (d) in Fig. 9.14 are compared with plots (b) and (f) in Fig. 9.10, because the correlations in a smoothed direct spectral estimator are generated by two sources, i.e., tapering and smoothing. The correlation function of a discretely smoothed direct spectral estimator depends on the weight function, i.e., the discrete kernel applied when smoothing; the maximal lag with non-zero correlations, however, depends on the width of the moving frequency band over which a weighted mean of the direct spectral estimator is computed; both of these approximations result from the examples in Problem 7.8.

The moving average $(1/4)\widehat{C}_Y^{(d)}(s_{k-1}) + (1/2)\widehat{C}_X^{(d)}(s_k) + (1/4)\widehat{C}_X^{(d)}(s_{k+1})$, as computed for $\widehat{C}_X^{(m)}(s_k)$ in Fig. 9.14 (a), therefore contains at least two approximately non-correlated terms and thus $\mathrm{Var}\big(\widehat{C}_X^{(m)}(s_k)\big) < \mathrm{Var}\big(\widehat{C}_X^{(d)}(s_k)\big)$. From the same argument follows that $\mathrm{Var}\big(\widehat{C}_X^{(m)}(s'_k)\big) < \mathrm{Var}\big(\widehat{C}_X^{(d)}(s'_k)\big)$.

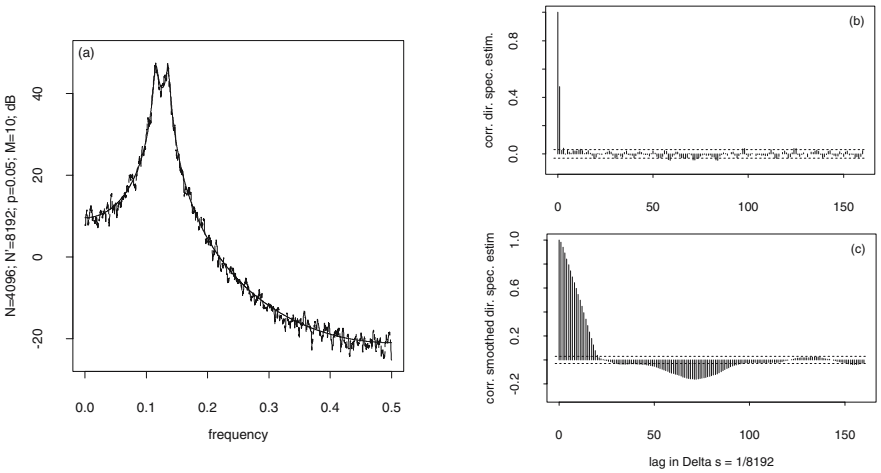


Fig. 9.15. On the left, spectrum (solid line) of the AR[4] model in Fig. 7.14 and smoothed direct spectral estimator (broken line) calculated from $N = 4096$ observations in a realisation of this model. The observations are multiplied with a cosine taper for $p = 0.05$ and then zero-padded as required in (9.20) and thus a direct spectral estimator for $N' = 8192$ and $\Delta s'_k = 1/8192$ is arrived at. Thereafter, the direct spectral estimator is smoothed with a modified discrete Daniell kernel having parameter $M = 10$.

The estimators in Fig. 9.14 are thus biased and their variances still too large. As a second example therefore, the spectrum of the AR[4] model in Fig. 7.14 is estimated using a time slice of length $N = 4096$ in a realisation of this model. Using $N = 4096$ observations, can an estimator be calculated that is not biased for all frequencies and which has a variance much smaller than the squared spectrum, i.e., the variance of the direct spectral estimator approximated in (9.16)?

An answer to this question can be found in a second simulation experiment. A direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ is calculated from $N = 4096$ observations in a realisation of the AR[4] model (X_t) in Fig. 7.14, subsequent to tapering with a $p = 0.05$ cosine taper and zero-padding to arrive at an estimator for frequencies $\Delta s'_k = 1/N' = 1/8192 = 0.000122$ as defined in (9.20). Thereafter, $\widehat{C}_X^{(d)}(s'_k)$ is smoothed using a modified discrete Daniell kernel with $M = 10$, i.e., over a band having width $20 \times \Delta s'_k = 0.002441$, to obtain $\widehat{C}_X^{(m)}(s'_k)$ as defined in (9.28). $\widehat{C}_X^{(m)}(s'_k)$ is plotted together with $C_X(s'_k)$ in Fig. 9.15 (a), and together with both, $C_X(s'_k)$ as well as $\widehat{C}_X^{(d)}(s'_k)$, in Fig. 9.16 for $0.11 \le s'_k \le 0.14$ and $0.43 \le s'_k \le 0.46$.

Distinctly, the smoothed direct spectral estimator is approximately unbiased for all frequencies, in the band with the peaks as well as for the higher

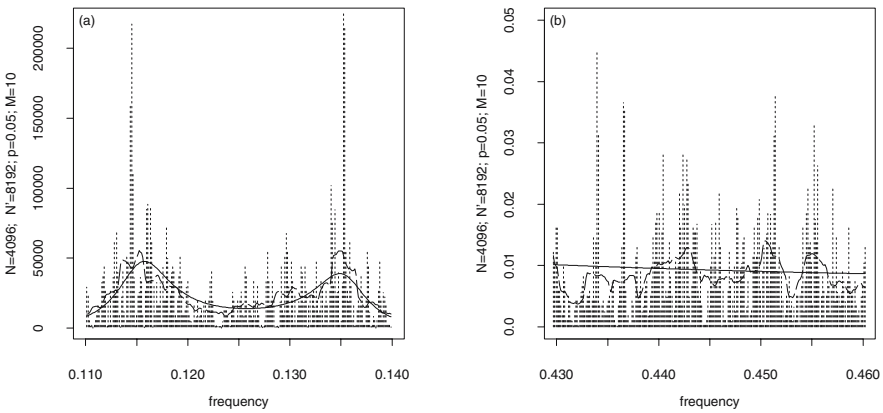


Fig. 9.16. Details from Fig. 9.15 in bands $0.11 \leq s'_k \leq 0.14$ (on the left) and $0.43 \leq s'_k \leq 0.46$ (on the right): spectrum (solid line), direct spectral estimator (vertical broken lines), smoothed direct spectral estimator (broken line fluctuating about the solid line).

frequencies a larger distance away from the peaks. This property is due to (i) the unbiased direct spectral estimator and (ii) $2M\Delta s'_k = 0.002441 < 0.00474 = \mathcal{B}_{\triangle}\big(C_X(s)\big)$, i.e., both bandwidths, the one of the smoothing kernel and the one of the spectrum to be estimated (in the remarks to Fig. 7.14), being approximately reconcilable with diagnostics (6.115) and (6.116).

Both spectral estimators, $\widehat{C}_X^{(d)}(s'_k)$ and $\widehat{C}_X^{(m)}(s'_k)$, are transformed to obtain realisations of stationary random sequences $\big(\widehat{C}_X^{(d)}(s'_k) - C_X(s'_k)\big)/C_X(s'_k)$ and $\big(\widehat{C}_X^{(m)}(s'_k) - C_X(s'_k)\big)/C_X(s'_k)$, whose empirical correlation functions are plotted in Fig. 9.15 (b) and (c). This transformation is introduced in the remarks to Figs. 9.7 and 9.8. In Fig. 9.15 (b), the empirical correlations are small for lags $\geq 2\Delta s'_k = 2(1/8192) = (1/4096)$ in accordance with the correlation of a direct spectral estimator for a cosine taper with $p = 0.05$ as plotted in Fig. 9.6. Consequently, 10 terms in $(1/40)\widehat{C}_X^{(d)}(s'_{k-10}) + (1/20)\widehat{C}_X^{(d)}(s'_{k-9}) + \ldots + (1/20)\widehat{C}_X^{(d)}(s'_k) + \ldots + (1/20)\widehat{C}_X^{(d)}(s'_{k+9}) + (1/40)\widehat{C}_X^{(d)}(s'_{k+10})$ are approximately non-correlated, and thus $\widehat{C}_X^{(m)}(s'_k)$ is the mean of at least 10 independent random variables, provided that the AR[4] model is normal. From this rationale it is concluded that the variance of direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ is reduced by at least one order of magnitude when it is smoothed using a modified discrete Daniell kernel for $M = 10$. In Fig. 9.15 (c), the empirical correlations decrease with constant slope and approximate zero-correlation (2.59) is arrived at for lag $20\Delta s'_k$. For larger lags, the empirical correlation function damps out, reconcilable with the examples given in Problem 7.8.

It can be seen in Fig. 9.16 that $\widehat{C}_X^{(m)}(s'_k)$ (i) fluctuates with relatively large amplitudes about $C_X(s'_k)$, i.e., the spectrum to be estimated, but (ii) is not systematically larger or smaller than $C_X(s'_k)$ — neither in the band with the peaks nor for the higher frequencies — in accordance with the result obtained above, i.e., that $\widehat{C}_X^{(m)}(s'_k)$ is unbiased. Thus, the fluctuations in $\widehat{C}_X^{(m)}(s'_k)$ are assumed to be due to its variance being still relatively large, despite the reduction of variance attained when $\widehat{C}_X^{(d)}(s'_k)$ is convolved as required in (9.28) to arrive at $\widehat{C}_X^{(m)}(s'_k)$. Clearly, the variance of $\widehat{C}_X^{(m)}(s'_k)$ is further reduced when the number of $\widehat{C}_X^{(d)}(s'_k)$ available in a frequency band of width $(1/2)\mathcal{B}_{\triangle}\big(C_X(s)\big) = 0.00474/2$ increases, which is easily arrived at in Problem 9.10 when N increases, on condition that $N' = 2N$.

The simulation experiments performed above, with the results plotted in Figs. 9.14, 9.15 and 9.16, demonstrate that the moment functions of a discretely smoothed direct spectral estimator calculated from N observations depend on

1. a data taper being applied because a bias in a direct spectral estimator is not reduced by smoothing, and
2. the bandwidth of the discrete smoothing kernel: with a kernel which is wider than half the bandwidth of the spectrum to be estimated (the maximal width recommended in (6.115) and (6.116)), one runs the risk of obtaining a biased smoothed estimator, whereas with a kernel narrower than recommended, the variance of the smoothed estimator is not as small as would be desired.

These dependencies are further analysed in the following paragraphs where the expectation and the variance functions of a smoothed direct spectral estimator are derived.

The expectation function $\mathrm{E}\big(\widehat{C}_X^{(m)}(s'_k)\big)$ of a discretely smoothed direct spectral estimator as defined in (9.28) is approximated in (9.30)

$$\begin{aligned}\mathrm{E}\big(\widehat{C}_X^{(m)}(s'_k)\big) &= \mathrm{E}\left(\sum_{j=-M}^{M} V_m(s'_j)\widehat{C}_X^{(d)}(s'_{k-j})\right)\\ &= \sum_{j=-M}^{M} V_m(s'_j)\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_{k-j})\big) \approx \sum_{j=-M}^{M} V_m(s'_j)C_X(s'_{k-j}) \qquad (9.30)\end{aligned}$$

under the assumption that $\widehat{C}_X^{(d)}(s'_k)$ is approximately unbiased, i.e., that $\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_k)\big) \approx C_X(s'_k)$, which is arrived at on condition that the observations have been tapered as discussed in Sect. 9.2.1.

In the simulation experiment in Figs. 9.15 and 9.16, for example, $\widehat{C}_X^{(d)}(s'_k)$ is calculated subsequent to tapering a time slice of length $N = 4096$ in a realisation of the AR[4] model in Fig. 7.14 and zero-padding such that $\Delta s'_k = 1/(2N) = 1/8192$. Due to the multiplication with a cosine taper for

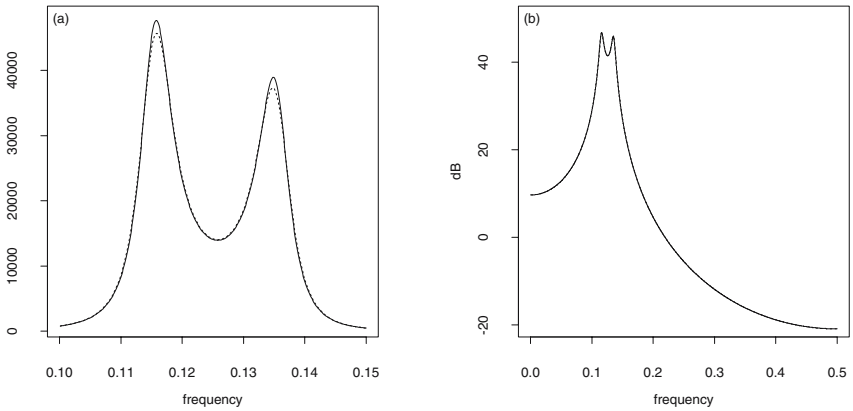


Fig. 9.17. Spectrum of the AR[4] model in Fig. 7.14 (solid line) and $\mathrm{E}\big(\widehat{C}_X^{(m)}(s'_k)\big)$ of the smoothed direct spectral estimator in Figs. 9.15 and 9.16 (broken line) approximated using (9.30).

$p = 0.05$, the leakage can be neglected, and thus $\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_k)\big) \approx C_X(s'_k)$ as is concluded from the examples in Figs. 9.3 and 9.4. Thereafter, $\widehat{C}_X^{(d)}(s'_k)$ is smoothed using a modified discrete Daniell kernel with $M = 10$ to obtain $\widehat{C}_X^{(m)}(s'_k)$. Hence, $\mathrm{E}\big(\widehat{C}_X^{(m)}(s'_k)\big)$ can be computed using (9.30) since $C_X(s'_k)$ is known in a simulation experiment, and the approximation plotted with a broken line in Fig. 9.17 is obtained.

From the logarithmic plot (b) in Fig. 9.17, $\widehat{C}_X^{(m)}(s'_k)$ is concluded to be unbiased for $0 \le s'_k \le 1/2$; from plot (a) in this figure it becomes, however, obvious that $\widehat{C}_X^{(m)}(s'_k)$ underestimates $C_X(s'_k)$ at the peaks. For example, $C_X(s'_k) = 38971$ for $s'_k = 0.1348877$ whereas $\mathrm{E}\big(\widehat{C}_X^{(m)}(0.1348877)\big)$ is the weighted mean of $\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_l)\big) \approx C_X(s'_l)$, with $0.1336670 \le s'_l \le 0.1361084$ and $C_X(s'_l) \le C_X(0.1348877)$ and thus $\mathrm{E}\big(\widehat{C}_X^{(d)}(0.1348877)\big) < C_X(0.1348877)$.

The example in Fig. 9.17 and the results obtained in Problem 9.10 demonstrate that a smoothed direct spectral estimator is approximately unbiased provided that (i) the direct spectral estimator is also approximately unbiased and (ii) the bandwidth of the spectrum to be estimated is considerably larger (at least twice as large as recommended in (6.115)) than the bandwidth of the smoothing kernel.

The variance function of a direct spectral estimator as defined in (9.3) is approximated in (9.16) under assumptions (9.13): $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s)$ for frequencies $s \neq -1/2, 0, 1/2$, i.e., (i) the variance is approximately the square of the spectrum to be estimated and (ii) does not decrease when the number of observations increases. These properties, which are shared by the periodogram, are inconvenient for the reasons discussed in the remarks to (8.24)

and Fig. 9.9 and thus motivate to search for an estimator having a smaller variance depending on the number of observations. Such an estimator is arrived at when a direct spectral estimator, calculated for discrete frequencies s_k or s'_k, is smoothed as required in (9.28). The variance function of the estimator defined in (9.28) can be derived in three steps.

In a first step, the variance function of $\widehat{C}_X^{(m)}(s'_k)$, i.e., a discretely smoothed direct spectral estimator as defined in (9.28), is obtained using (1.15,10)

$$\begin{aligned}\mathrm{Var}\big(\widehat{C}_X^{(m)}(s'_k)\big) &= \mathrm{Var}\left(\sum_{l=-M}^{M} V_m(s'_l)\widehat{C}_X^{(d)}(s'_{k-l})\right)\\ &= \sum_{i=-M}^{M}\sum_{j=-M}^{M} V_m(s'_i)V_m(s'_j)\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s'_{k-i}),\widehat{C}_X^{(d)}(s'_{k-j})\big)\end{aligned} \tag{9.31}$$

without any approximation, i.e., (9.31) is exact. The terms with two indices summed in (9.31) are arranged in a matrix with weighted variances of $\widehat{C}_X^{(d)}(s'_k)$ in the main and weighted covariances in the subdiagonals. Covariances of a direct spectral estimator calculated for discrete frequencies s_k and $s_k + n\Delta s_k$ or s'_k and $s'_k + n\Delta s'_k$, n a small positive integer number, are not identically zero, as implied by (9.14). Under assumptions (9.13), however, these covariances decrease rapidly for $n = 1, 2, 3, ...$, depending on the data taper applied. For a variety of cosine tapers, the decay of the correlations in a direct spectral estimator with increasing displacement in the frequency domain is demonstrated in Fig. 9.6.

This property favours, in a second step, the approximation of the variance of a discretely smoothed direct spectral estimator on condition that the assumptions stipulated in (9.32) are taken into account.

Let $\widehat{C}_X^{(m)}(s'_k)$ be a discretely smoothed direct spectral estimator as defined in (9.28). Then its variance can be approximated on condition that:

1. *(X_t) is normal*
2. *s'_k is more than $|M\Delta s'_k|$ away from $-1/2$, 0, $1/2$*
3. *$\widehat{C}_X^{(d)}(s'_k)$ is approximatively unbiased*
4. *$C_X(s)$ is approximatively constant in a band $2M\Delta s'_k$ wide.*

(9.32)

(9.32,1) is inherited from (9.13,3) and (9.32,2) excludes neighbour frequencies of $-1/2$, 0, and $1/2$, thus asserting that $\widehat{C}_X^{(d)}(-1/2)$, $\widehat{C}_X^{(d)}(0)$ or $\widehat{C}_X^{(d)}(1/2)$ are not included in the moving average, since approximations (9.14) and (9.15) do not apply there. If $\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_k)\big) \approx C_X(s'_k)$ as required in (9.32,3), then (9.32,4) also propagates to $\widehat{C}_X^{(d)}(s'_k)$. This directly implies approximation (9.33) and,

$$\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_{|k-l|})\big) \approx C_X(s'_k) \tag{9.33}$$

$$\mathrm{Var}\big(\widehat{C}_X^{(d)}(s'_{|k-l|})\big) \approx C_X^2(s'_k) \tag{9.34}$$

$$\frac{\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s'_k), \widehat{C}_X^{(d)}(s'_{|k-l|})\big)}{\mathrm{Var}\big(\widehat{C}_X^{(d)}(s'_k)\big)} \approx H^{\oplus}H(|k-l|\Delta s'_k) \tag{9.35}$$

further, approximations (9.34) and (9.35) can be derived from (9.14), (9.16) and (9.17) for lags $s'_k - s'_{k-l} = l\Delta s'_k$, with $\Delta s'_k = 1/N'$ and $l = -M, \ldots, -1$, $0, 1, \ldots, M$, on condition that a value for M can be found such that (9.32,4) applies. An appropriate value for $2M$, i.e., the width of the band used for computing the moving average as required in (9.28), is readily obtained in simulation experiments provided that the spectrum to be estimated is known. If the spectrum is completely unknown, then it is difficult to arrive at a value for $2M$. Possible solutions to this problem are discussed in Sect. 9.3.5.

In the third step, approximations (9.34) and (9.35) are substituted in (9.31), which is repeated in the first line of the following derivation, and thus (9.36) is arrived at. The terms with indices i and j in the sum on the right side in (9.36), i.e., the correlations of the direct spectral estimator as proposed in (9.17), are arranged in a matrix. They depend, under assumptions (9.32), on differences $l = i - j$ solely, and are summed over the diagonals in (9.37).

$$\mathrm{Var}\big(\widehat{C}_X^{(m)}(s'_k)\big) = \sum_{i=-M}^{M}\sum_{j=-M}^{M} V_m(s'_i)V_m(s'_j)\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s'_{k-i}), \widehat{C}_X^{(d)}(s'_{k-j})\big)$$

$$\approx C_X^2(s'_k)\sum_{i=-M}^{M}\sum_{j=-M}^{M}\big(H^{\oplus}H(|i-j|\Delta s'_k)\big)\big(V_m(s'_i)\big)\big(V_m(s'_j)\big) \tag{9.36}$$

$$\approx C_X^2(s'_k)\sum_{l=-M}^{M}\big(H^{\oplus}H(|l|\Delta s'_k)\big)\sum_{\ell=-M}^{M-|l|}\big(V_m(s'_\ell)\big)\big(V_m(s'_{\ell+|l|})\big) \tag{9.37}$$

$$\approx C_X^2(s'_k)\sum_{l=-M^{\oplus}}^{M^{\oplus}}\big(H^{\oplus}H(|l|\Delta s'_k)\big)\sum_{\ell=-M}^{M-|l|}\big(V_m(s'_\ell)\big)\big(V_m(s'_{\ell+|l|})\big) \tag{9.38}$$

The second sum in (9.37) is the autocorrelation (6.103) $V_m \star V_m(l\Delta s'_k)$ of the discrete smoothing kernel $V_m(s'_k)$ for lag $l\Delta s'_k$, because $V_m(s'_k)$ becomes identically zero for $k > M$ as implied by definition (9.28,3). An approximation having a sufficient accuracy is arrived at when the sums over the first $M^{\oplus}$, $M^{\oplus} < M$, diagonals are computed in (9.38) because $H^{\oplus}H(|l|\Delta s'_k)$ becomes negligibly small for $M^{\oplus}\Delta s_k < |l|\Delta s'_k < M\Delta s_k$. $M^{\oplus}$ can be inferred from Fig. 9.6 provided that the direct spectral estimator is calculated using a cosine taper. From (9.38), a further approximation will be derived in Problem 9.12.

For example, subsequent to multiplicating a cosine taper for $p = 0.05$ with $N = 4096$ observations in a realisation of the AR[4] model in Fig. 7.14 and zero-padding, direct spectral estimators are computed for frequencies s'_k with $\Delta s'_k = 1/N'$, $N' = 2N$, and then smoothed with a modified discrete

Daniell kernel $V_{dD,M}(s'_k)$ for $M = 10$ to obtain the estimator as plotted in Figs. 9.15 and 9.16. The variance function of this smoothed direct spectral estimator can be approximated using (9.38) in a straightforward calculation. Since $\Delta s'_k = 1/(2N) = (1/2)\Delta s_k$ in this case, the following correlations are read in Fig. 9.6: $H^{\oplus}H(0\Delta s'_k) = 1.0$, $H^{\oplus}H(1\Delta s'_k) \approx 0.5$, $H^{\oplus}H(2\Delta s'_k) \approx 0.0$, $H^{\oplus}H(3\Delta s'_k) \approx 0.05$, $H^{\oplus}H(4\Delta s'_k) \approx 0.0$, and $H^{\oplus}H(l\Delta s'_k) \approx 0.0$, for $l > 4$, and thus $M^{\oplus} = 3$ in (9.38).

$H^{\oplus}H(l\Delta s'_k)$, $l = 0, 1, 2, 3$, thus approximated from Fig. 9.6 are used as weights for the autocorrelations of $V_{dD,M}(s'_k)$, which are readily computed subsequent to a multiplication of $V_{dD,M}(s'_k)$ with $2M = 20$: (i) for lag $l = 0$, the squared weights in $V_{dD,M}(s'_k)$ sum to 19.5, (ii) for lag $l = 1$, $(0.5\times1.0+18\times(1.0\times1.0)+0.5\times1.0) = 19$, (iii) for lag $l = 2$, $(0.5\times1.0+17\times(1.0\times1.0)+0.5\times1.0) = 18$, and (iv) for lag $l = 3$, $(0.5 \times 1.0 + 16 \times (1.0 \times 1.0) + 0.5 \times 1.0) = 17$ are obtained. Thereafter, the first sum in (9.38): $19.5 \times 1.0 + 2 \times 19.0 \times 0.5 + 2 \times 18.0 \times 0.0 + 2 \times 17.0 \times 0.05 = 40.2$ is computed and compensated for the above multiplication with $2M = 20$: $40.2/400 = 0.1005$ is arrived at. Thus $\mathrm{Var}\big(\widehat{C}_X^{(m)}(s'_k) \approx 0.01005 \times C_X^2(s'_k) \approx 0.10 \times C_X^2(s'_k)$ is obtained for the variance of the smoothed direct spectral estimator in Figs. 9.15 and 9.16 under assumptions (9.32), a result in favourable agreement with the one obtained in the remarks to Fig. 9.16. There it was found that the variance of the direct spectral estimator is reduced by an order of magnitude subsequent to smoothing with the modified discrete Daniell kernel to obtain the smoothed estimator as plotted in Figs. 9.15 and 9.16.

The examples in Figs. 9.14, 9.15 and 9.16 demonstrate that the variance of a direct spectral estimator $\widehat{C}_X^{(d)}(s_k)$ or $\widehat{C}_X^{(d)}(s'_k)$, calculated for discrete frequencies s_k with $\Delta s_k = 1/N$ (or s'_k with $\Delta s'_k = 1/N'$), can be reduced by smoothing using a moving average. Thus, an estimator is obtained that (i) is unbiased, a property inherited from the direct spectral estimator, and (ii) has a variance that depends on N and M, with N being the length of the observational period and $2M\Delta s_k$ (or $2M\Delta s'_k$) the width of the band in which the direct spectral estimator is averaged in definition (9.28). Consequently, the goals formulated in the remarks concluding Sects. 8.3 and 8.4 can be arrived at on condition that, in a band being $2M\Delta s_k$ or $2M\Delta s'_k$ wide, (i) the number of $\widehat{C}_X^{(d)}(s_k)$ or $\widehat{C}_X^{(d)}(s'_k)$ having approximately zero-correlations is large enough to arrive at a substantial reduction of their variance and (ii) the spectrum to be estimated is approximately constant. Both conditions are easily met in simulation experiments where the spectrum is known. If the spectrum to be estimated is not known, approximations for its bandwidth can be obtained in Sect. 9.3.5 and N can be chosen accordingly.

A continuous direct spectral estimator $\widehat{C}_X^{(d)}(s)$, $-1/2 \le s \le 1/2$, is obtained in (9.6) as Fourier transform of $\hat{c}_X^{(d)}(\tau)$, i.e., the empirical correlation function calculated from tapered observations as defined in (9.7). $\widehat{C}_X^{(d)}(s)$ appears superior $\big($compared with its discrete versions $\widehat{C}_X^{(d)}(s_k)$ or $\widehat{C}_X^{(d)}(s'_k)\big)$

when a continuous spectrum $C_X(s)$ is to be estimated. Under which conditions can the variance of $\widehat{C}_X^{(d)}(s)$ be reduced by means of smoothing with a continuous kernel in a convolution integral?

9.3.2 Lag Window Spectral Estimator

Let (X_t) be a discrete-time stationary stochastic process having a continuous spectrum $C_X(s)$ and a covariance function $c_X(\tau)$, and let (x_t) be a realisation of (X_t). Then a direct spectral estimator $\widehat{C}_X^{(d)}(s)$ and its corresponding empirical covariance function $\hat{c}_X^{(d)}(\tau)$, both calculated as required in (9.6) and (9.7) from tapered observations in (x_t), are a Fourier transform pair as defined in (6.49) and (6.50). This property is shared by $C_X(s)$ and $c_X(\tau)$, a result which is obtained in (7.65) and (7.66). The first moment function of $\widehat{C}_X^{(d)}(s)$ is derived in (9.12), and, under assumptions (9.13), the second moment functions of $\widehat{C}_X^{(d)}(s)$ are obtained in (9.14), (9.15) and (9.16). An estimator having a smaller variance than $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s)$ in (9.16) is hopefully arrived at when $\widehat{C}_X^{(d)}(s)$ is smoothed, below in (9.39), by convolving with a function $V_m(s)$ having the properties required in (9.44).

$$\widehat{C}_X^{\vdash\tau\dashv}(s) = \int_{-1/2}^{1/2} V_m(s-r)\widehat{C}_X^{(d)}(r)\mathrm{d}r \quad \begin{array}{l}\text{convolution}\\ \text{as in (6.144)}\end{array} \tag{9.39}$$

$$= \sum_{\tau=-\infty}^{\infty} \Big(v_m(\tau)\hat{c}_X^{(d)}(\tau)\Big)\mathrm{e}^{-\mathrm{i}2\pi s\tau} \quad \begin{array}{l}\text{with } v_m(\tau) \text{ and } V_m(s)\\ \text{a Fourier transform pair}\\ \text{as in (6.49) and (6.50)}\end{array} \tag{9.40}$$

$$= \sum_{\tau=-\infty}^{\infty} \Big(v_m^{\vdash\tau\dashv}(\tau)\hat{c}_X^{(d)}(\tau)\Big)\mathrm{e}^{-\mathrm{i}2\pi s\tau} \quad \begin{array}{l}\text{with } v_m^{\vdash\tau\dashv}(\tau) \text{ as}\\ \text{defined in (9.43)}\end{array} \tag{9.41}$$

$$= \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)\widehat{C}_X^{(d)}(r)\mathrm{d}r \quad \begin{array}{l}v_m^{\vdash\tau\dashv}(\tau) \text{ and } V_m^{\vdash\tau\dashv}(s)\\ \text{a Fourier transform pair}\\ \text{as in (6.49) and (6.50)}\end{array} \tag{9.42}$$

$$v_m^{\vdash\tau\dashv}(\tau) = \begin{cases} v_m(\tau) & \text{for } \tau = -(N-1), \ldots, -1, 0, 1, \ldots, N-1 \\ 0 & \text{for } |\tau| \geq N \end{cases} \tag{9.43}$$

$v_m^{\vdash\tau\dashv}(\tau)$, being a time-limited sequence (6.108), is a lag window and $\widehat{C}_X^{\vdash\tau\dashv}(s)$ is a lag window spectral estimator.
$V_m(s)$ and $V_m^{\vdash\tau\dashv}(s)$ are (continuous) smoothing kernels provided that they are defined in $-1/2 \leq s \leq 1/2$, periodic with period 1, even (6.32), defining sequences for $\delta(s+n)$, $n = \ldots, -1, 0, 1, \ldots$, i.e., the periodic delta function introduced in the remarks to (6.62), and $V_m(s) \geq 0$ for all m and s. (9.44)

(9.39) implies (9.40) as is concluded from convolution theorem (6.48,7), which is obtained in (6.145) for Fourier transform pairs as defined in (6.49) and (6.50). The Fourier sums in (9.40) and (9.41) are equal, as is concluded from

(9.7) where the empirical covariance function calculated from tapered observations is defined. Consequently, convolution integrals (9.39) and (9.42) are equal despite $V_m(s) \neq V_m^{\vdash\tau\dashv}(s)$. This implies that $\widehat{C}_X^{\vdash\tau\dashv}(s) \geq 0$ since both $V_m(s) \geq 0$ and $\widehat{C}_X^{(d)}(r) \geq 0$ in (9.39). Further properties of the smoothing kernels and their Fourier transforms are discussed in the following two paragraphs.

Sequence $v_m(\tau)$ is possibly band-limited as defined in (6.108,2), whereas $v_m^{\vdash\tau\dashv}(\tau)$ in (9.43) is a time-limited sequence (6.108,1). Thus a lag window $v_m^{\vdash\tau\dashv}(\tau)$ depends on the parameter m in $V_m(s)$ and on the number N of the observed values. Both weight sequences $v_m(\tau)$ and $v_m^{\vdash\tau\dashv}(\tau)$ modify $\hat{c}_X^{(d)}(\tau)$, i.e., the empirical covariances calculated from tapered observations, in a manner similar to a data taper or data window modifying the observations in e.g., (6.113), (9.3) or (9.7). $v_m^{\vdash\tau\dashv}(\tau)$ is therefore called lag window.

The continuous smoothing kernels $V_m(s)$ and $V_m^{\vdash\tau\dashv}(s)$ depend on parameter m in the sense that both kernels become narrower in $-1/2 \leq s \leq 1/2$ with increasing m, a property shared by the discrete kernels $V_m(s_k)$ and $V_m(s'_k)$ used in (9.28) for smoothing $\widehat{C}_X^{(d)}(s_k)$ or $\widehat{C}_X^{(d)}(s'_k)$ calculated for discrete frequencies. Distinct from the discrete case, however, a sequence of continuous smoothing kernels $V_m(s)$ $\big($or $V_m^{\vdash\tau\dashv}(s)\big)$ is a defining sequence for $\delta(s+n)$, $n = \ldots, -1, 0, 1, \ldots$, as required in (9.44). This implies that (adapted from (6.60) for interval $-1/2 \leq s \leq 1/2$): (i) $\int_{-1/2}^{1/2} V_m(s)\mathrm{d}s = \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s)\mathrm{d}s = 1$ for $m = 1, 2, 3, \ldots$, with further implications to be derived in Problem 9.11, and (ii), in the limiting case for $m \to \infty$, $V_m(s) \to \infty$ $\big($or $V_m^{\vdash\tau\dashv}(s) \to \infty\big)$ for $s = 0$ as well as $V_m(s) = 0$ $\big($or $V_m^{\vdash\tau\dashv}(s) = 0\big)$ for $-1/2 \leq s < 0$ and $0 < s \leq 1/2$.

Using (9.41), a lag window spectral estimator $\widehat{C}_X^{\vdash\tau\dashv}(s)$ for the continuous spectrum $C_X(s)$ of model (X_t) in (9.1) is calculated in the following steps:

1. an observed time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of (X_t) is multiplied with a variance-normalised data taper $h_N^{(d)}(t)$ as defined in (9.2), to reduce the leakage in a subsequent Fourier transform
2. from the tapered observations, $\hat{c}_X^{(d)}(\tau)$ is calculated using (9.7) as demonstrated in Problem 9.5
3. $\hat{c}_X^{(d)}(\tau)$ is multiplied with the lag window as required in (9.41), and
4. $\widehat{C}_X^{\vdash\tau\dashv}(s)$, i.e., the lag window spectral estimator, is obtained by means of Fourier transforming the product $v_m^{\vdash\tau\dashv}(\tau) \times \hat{c}_X^{(d)}(\tau)$ as required in (9.41)

on condition that a lag window is available which has (whose pertaining smoothing kernel(s) have, resp.) the properties required in (9.44). Moreover, the bandwidth of the smoothing kernel (e.g., the width at half height of its main lobe as introduced in Sect. 9.5.4) and the bandwidth of the spectrum to be estimated should be approximately reconcilable with diagnostics (6.115) and (6.116), as is recommended in the case of a discrete smoothing kernel in the remarks concluding the discussion of Figs. 9.14, 9.15 and 9.16. If the

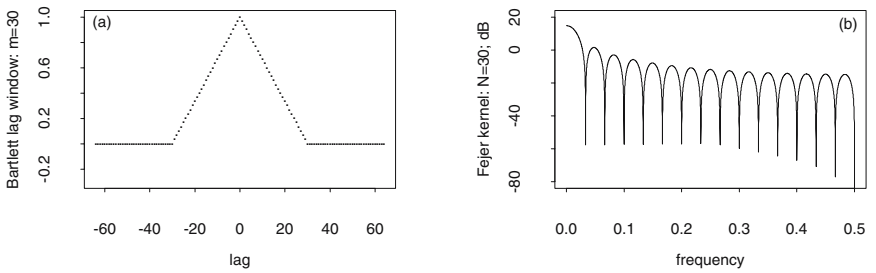


Fig. 9.18. The Bartlett lag window for $m = 30$ and the pertaining smoothing kernel are identical with the even triangle sequence $\Lambda^{(e)}_{2N+1}(t)$ in (6.56) and the Fejer kernel $F_N(s)$ in (6.57) for $N = 30$.

bandwidth of the spectrum to be estimated is not known, the bandwidth of the smoothing kernel can be assessed using the procedures introduced in Sect. 9.3.5.

Despite the width of the smoothing kernel (and not its form) is the important question concerning spectral estimation as argued in the remarks to (9.80): a variety of lag windows having distinct forms have been proposed (a summary is given in [113]), those introduced by Daniell [42] and Bartlett [9] are given here as examples.

A first example for a smoothing kernel $V_m(s)$ is the periodic function plotted in Fig. 6.23 (b): this function is defined in $-1/2 \le s \le 1/2$, periodic with period 1 and, with parameters chosen accordingly, becomes a defining sequence for $\delta(s+n)$, $n = \ldots, -1, 0, 1, \ldots$, as is concluded from definition (6.61) and Fig. 6.12 (a). Its Fourier transform $v_m(\tau)$, i.e., the sequence plotted in Fig. 6.23 (a), is not time-limited as implied by (6.109). In contrast, lag window $v_m^{\vdash\tau\dashv}(\tau)$ obtained from $v_m(\tau)$ by applying definition (9.43) with $N = 64$ as plotted in Fig. 6.23 (c), is a time-limited sequence and, therefore, its Fourier transform $V_m^{\vdash\tau\dashv}(s)$ in Fig. 6.23 (d) is not band-limited. A rectangle function as plotted in Fig. 6.23 (b) can be defined using only one parameter m in (9.45), whereafter, from the smoothing kernel thus defined, the corresponding lag window for parameters m and N can be derived.

$$V_{Da,m}(s+n) = V_{Da,m}(s) = \begin{cases} m & \textit{for } -1/(2m) \le s \le 1/(2m) \\ 0 & \textit{for } |s| > 1/(2m) \end{cases}$$
$$\textit{with } -1/2 \le s \le 1/2,\ 0 < 1/(2m) \ll 1/2 \textit{ and } n = \ldots, -1, 0, 1, \ldots, \qquad (9.45)$$
is called Daniell kernel. From its Fourier transform $v_{Da,m}(\tau)$, *the Daniell lag window* $v_{Da,m}^{\vdash\tau\dashv}(\tau)$ *is obtained using (9.43).*

As a second example, a Bartlett lag window for $m = 30$ is plotted in Fig. 9.18 (a). A Bartlett lag window has only one parameter m and is identical, for $m = N$, with the even triangle sequence $\Lambda^{(e)}_{2N+1}(t)$ in (6.56). The smoothing kernel pertaining to the Bartlett lag window is the Fejer kernel

$F_N(s)$, with $m = N$, as obtained in (6.57). $F_N(s)$ has the properties of a continuous smoothing kernel as is concluded from (6.58), (6.59) and the remarks to (6.62).

A lag window can also be obtained when a discrete smoothing kernel is used in (9.28) to smooth a direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ calculated for discrete frequencies s'_k with $\Delta s'_k = 1/N'$. In this case, the Fourier sum in (9.6) becomes $\widehat{C}_X^{(d)}(s'_k) = \sum_{\tau=-(N-1)}^{N-1} \hat{c}_X^{(d)}(\tau)\mathrm{e}^{-\mathrm{i}2\pi s'_k \tau}$, with $\hat{c}_X^{(d)}(\tau)$ as defined in (9.7), and thus a discretely smoothed direct spectral estimator as in (9.28,3) becomes — since $s'_{k-j} = s'_k - s'_j$ as implied by (6.26) and (6.27) — the weighted Fourier sum in (9.46). Thereafter, (9.47) and (9.48) follow immediately.

$$\widehat{C}_X^{(m)}(s'_k) = \sum_{j=-M}^{M} V_m(s'_j) \left(\sum_{\tau=-(N-1)}^{N-1} \hat{c}_X^{(d)}(\tau)\mathrm{e}^{-\mathrm{i}2\pi(s'_k - s'_j)\tau} \right) \tag{9.46}$$

$$= \sum_{\tau=-(N-1)}^{(N-1)} \left(\sum_{j=-M}^{M} V_m(s'_j)\mathrm{e}^{\mathrm{i}2\pi s'_j \tau} \right) \hat{c}_X^{(d)}(\tau)\mathrm{e}^{-\mathrm{i}2\pi s'_k \tau} \tag{9.47}$$

$$= \sum_{\tau=-(N-1)}^{(N-1)} \Big(v_{m,M}(\tau)\hat{c}_X^{(d)}(\tau) \Big)\mathrm{e}^{-\mathrm{i}2\pi s'_k \tau} \quad \begin{array}{l} v_{m,M}(\tau) \text{ and } V_m(s'_j) \\ \text{a Fourier transform} \\ \text{pair as in (6.22,1,2)} \end{array} \tag{9.48}$$

$$= \widehat{C}_X^{\vdash\tau\dashv}(s) \quad \text{with } s = s'_k \text{ provided that} \tag{9.49}$$

$$\widehat{C}_X^{\vdash\tau\dashv}(s) = \sum_{\tau=-\infty}^{\infty} \Big(v_{m,M}^{\vdash\tau\dashv}(\tau)\hat{c}_X^{(d)}(\tau) \Big)\mathrm{e}^{-\mathrm{i}2\pi s\tau} \quad \begin{array}{l} \text{with } v_{m,M}^{\vdash\tau\dashv}(\tau) \text{ as} \\ \text{defined in (9.52)} \end{array} \tag{9.50}$$

$$= \int_{-1/2}^{1/2} V_{m,M}^{\vdash\tau\dashv}(s-r)\widehat{C}_X^{(d)}(r)\mathrm{d}r \quad \begin{array}{l} v_{m,M}^{\vdash\tau\dashv}(\tau) \text{ and } V_{m,M}^{\vdash\tau\dashv}(s) \\ \text{a Fourier transform pair} \\ \text{as in (6.49) and (6.50)} \end{array} \tag{9.51}$$

$$v_{m,M}^{\vdash\tau\dashv}(\tau) = \begin{cases} v_{m,M}(\tau) & \text{for } \tau = -(N-1), \ldots, -1, 0, 1, \ldots, N-1 \\ 0 & \text{for } |\tau| \geq N \end{cases} \tag{9.52}$$

From definition (9.7), it is concluded that the Fourier sum in (9.48) equals the one in (9.50) provided that s'_k is substituted for s and therefore (9.49) is obtained.

This result implies that any discretely smoothed direct spectral estimator $\widehat{C}_X^{(m)}(s'_k)$ can be written as a lag window estimator $\widehat{C}_X^{\vdash\tau\dashv}(s)$ using (9.50), (9.51) and (9.52), on condition that the lag window $v_{m,M}^{\vdash\tau\dashv}(\tau)$ and its continuous smoothing kernel $V_{m,M}^{\vdash\tau\dashv}(s)$ have the properties required in (9.44). Both, $v_{m,M}^{\vdash\tau\dashv}(\tau)$ and $V_{m,M}^{\vdash\tau\dashv}(s)$, depend on (i) m and M (in the subscripts) as inherited from the moving average in (9.46), and (ii) N as required in (9.52), symbolised by the superscript.

For example, $V_{dD,M}(s'_k)$, for $M = 3$ and $\Delta s'_k = 1/N'$, $N' = 2N$ and $N = 64$, as defined in (9.29), i.e., a modified discrete Daniell kernel with bandwidth

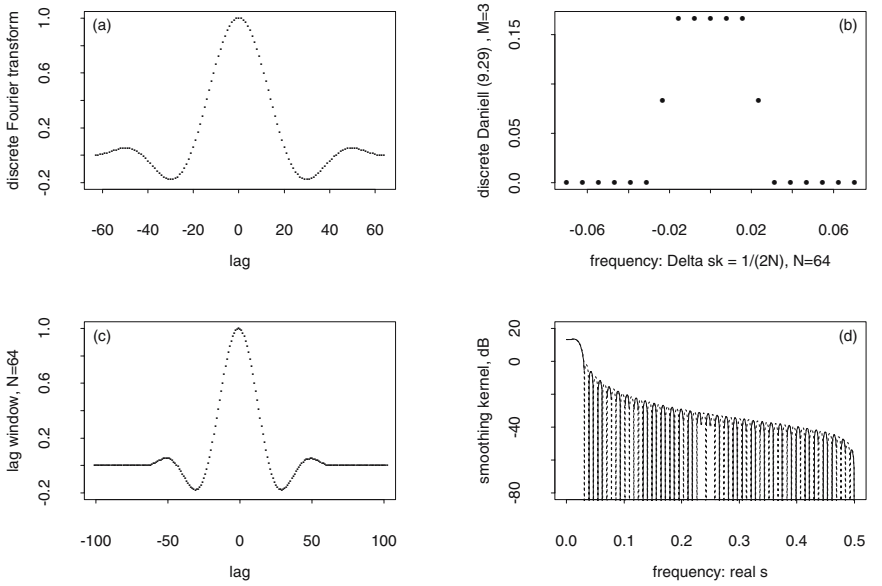


Fig. 9.19. The sequence in plot (a) is the discrete Fourier transform of the modified discrete Daniell kernel (9.29) for $M = 3$ and $\Delta s'_k = 1/128$ in plot (b). From the sequence in (a), the lag window in plot (c) is obtained using definition (9.52). The lag window and its continuous smoothing kernel in plot (d) are a Fourier transform pair as defined in (6.49) and (6.50).

$6\Delta s'_k = 6/128$ is plotted in Fig. 9.19 (b). Its discrete Fourier +i-transform $v_{dD,M}(\tau)$ is calculated using (6.22,1) for lags $\tau = -63, \ldots, -1,\ 0, 1, \ldots, 64$ and thereafter plotted in Fig. 9.19 (a). Thus $v_{dD,M}(\tau)$ and $V_{dD,M}(s'_k)$ are a discrete Fourier transform pair with the normalisation in the $-$i-transform as required in (6.22,1) and $v_{dD,M}(\tau)$ is periodic with period $2N = 128$.

Thereafter, when definition (9.52) is applied to $v_{dD,M}(\tau)$, the lag window $v^{\vdash\tau\dashv}_{dD,M}(\tau)$ pertaining to the modified discrete Daniell kernel is arrived at. $v^{\vdash\tau\dashv}_{dD,M}(\tau)$ as plotted in Fig. 9.19 (c) and the pertaining continuous smoothing kernel $V^{\vdash\tau\dashv}_{dD,M}(s)$ in Fig. 9.19 (d) are a Fourier transform pair as defined in (6.49) and (6.50). $\sum_{k=-M}^{M} V_{dD,M}(s'_k) = 1$ in (9.29) implies $v_{dD,M}(\tau = 0) = 1$ as is concluded from (6.48,1). This property is inherited by $v^{\vdash\tau\dashv}_{dD,M}(\tau)$ and, therefore, $\int_{-1/2}^{1/2} V^{\vdash\tau\dashv}_{dD,M}(s)\mathrm{d}s = 1$. If N increases and M remains constant then (i) $v_{dD,M}(\tau)$ becomes wider and (ii) $V_{dD,M}(s'_k)$ becomes narrower since $s'_k = 1/(2N)$ decreases, and further, (iii) $v^{\vdash\tau\dashv}_{dD,M}(\tau)$ becomes wider and (iv) $V^{\vdash\tau\dashv}_{dD,M}(s)$ decreases in its width. It thus becomes plausible that $V^{\vdash\tau\dashv}_{dD,M}(s)$ for $N = 1, 2, \ldots$ is a defining sequence for $\delta(s + n)$, $n = \ldots, -1, 0, 1, \ldots$, as required in (9.44).

Consequently, any direct spectral estimator smoothed discretely with a modified Daniell kernel can be written as a lag window spectral estimator, as proposed in (9.48), and therefore, properties of the lag window spectral estimator that will be derived in Sects. 9.3.3 and 9.3.4 are shared by a direct spectral estimator smoothed discretely with a modified Daniell kernel.

The above example demonstrates (i) the conclusion in the remarks to (9.49), (9.50), (9.51) and (9.52), i.e., that any discretely smoothed direct spectral estimator $\widehat{C}_X^{(m)}(s'_k)$ can be written as a lag window estimator under certain conditions (also being demonstrated) as well as (ii) its implication, i.e., that the properties of the lag window spectral estimator, e.g., its moment functions to be derived in Sects. 9.3.3, are shared by the direct spectral estimator.

The inverse, however, is not true, i.e., (9.49) does not imply that any lag window spectral estimator $\widehat{C}_X^{\vdash\tau\dashv}(s)$, being defined in (9.41) for all real frequencies s in $-1/2 \le s \le 1/2$, can be written as a discretely smoothed direct spectral estimator $\widehat{C}_X^{(m)}(s'_k)$ or $\widehat{C}_X^{(m)}(s_k)$, being defined for discrete frequencies $-1/2 \le s'_k \le 1/2$ or $-1/2 \le s_k \le 1/2$ in (9.28). For that set of frequencies s'_k with $\Delta s'_k = 1/(2N)$ defined in (9.53), as an exception, any lag window spectral estimator can be written as a discretely smoothed direct spectral estimator.

A prerequisite, however, for the derivation of this exception are definitions (9.53), (9.54) and (9.55) as well as the Fourier transforms in (9.56) and (9.57). (9.56) is a repetition of (9.23) and (9.24), where $\hat{c}_X^{(d)}(\tau)$ and $\widehat{C}_X^{(d)}(s'_k)$ are obtained as a discrete Fourier transform pair normalised in the +i-transform as required in (6.22,4). In (9.57), a smoothing kernel as defined in (9.42) is obtained as Fourier −i-transform of the lag window for discrete frequencies as in (9.53), bearing in mind that a smoothing kernel and a lag window are a Fourier transform pair as defined in (6.49) and (6.50).

$$s'_k = k/N' = k/(2N), \quad \text{with} \quad \begin{array}{l} k \in \mathrm{N}_1,\ \mathrm{N}_1 = (-N, \ldots, \\ \qquad -1, 0, 1, \ldots, N) \end{array} \tag{9.53}$$

$$\mathrm{N}_2 = (-(N-1), \ldots, -1, 0, 1, \ldots, N) \tag{9.54}$$

$$\mathrm{N}_3 = (-(N-1), \ldots, -1, 0, 1, \ldots, (N-1)) \tag{9.55}$$

$$\hat{c}_X^{(d)}(\tau) = \frac{1}{2N} \sum_{j=-(N-1)}^{N} \widehat{C}_X^{(d)}(s'_j) \mathrm{e}^{+\mathrm{i}2\pi s'_j \tau}, \quad \text{with } \tau \in \mathrm{N}_2 \tag{9.56}$$

$$\begin{aligned} V_m^{\vdash\tau\dashv}(s'_k) &= \sum_{\tau=-\infty}^{\infty} v_m^{\vdash\tau\dashv}(\tau) \mathrm{e}^{-\mathrm{i}2\pi s'_k \tau} \quad \begin{array}{l} \text{as required in (9.42),} \\ \text{i.e., using (6.49) and (6.50)} \end{array} \\ &= \sum_{\tau=-(N-1)}^{N-1} v_m^{\vdash\tau\dashv}(\tau) \mathrm{e}^{-\mathrm{i}2\pi s'_k \tau} \quad \text{definition (9.43)} \end{aligned} \tag{9.57}$$

For the frequencies defined in (9.53), a lag window spectral estimator is written as in (9.58) by applying definitions (9.41) and (9.7). These definitions

imply $\widehat{C}_X^{\vdash\tau\dashv}(s) = \sum_{\tau=-(N-1)}^{(N-1)} \big(v_m^{\vdash\tau\dashv}(\tau)\hat{c}_X^{(d)}(\tau)\big)\mathrm{e}^{-\mathrm{i}2\pi s\tau}$, for $-1/2 \le s \le 1/2$, i.e., that a lag window spectral estimator is completely determined when the lag window and the empirical covariance function calculated from tapered observations are known for lags $\tau = -(N-1), \ldots, -1, 0, 1, \ldots, (N-1)$.

$$\widehat{C}_X^{\vdash\tau\dashv}(s'_k) = \sum_{\tau=-(N-1)}^{(N-1)} \big(v_m^{\vdash\tau\dashv}(\tau)\hat{c}_X^{(d)}(\tau)\big)\mathrm{e}^{-\mathrm{i}2\pi s'_k\tau} \quad s'_k \text{ in (9.53)} \tag{9.58}$$

$$= \sum_{\tau=-(N-1)}^{(N-1)} v_m^{\vdash\tau\dashv}(\tau) \left(\frac{1}{2N} \sum_{j=-(N-1)}^{N} \widehat{C}_X^{(d)}(s'_j)\mathrm{e}^{+\mathrm{i}2\pi s'_j\tau} \right) \mathrm{e}^{-\mathrm{i}2\pi s'_k\tau}$$

$$= \sum_{\tau=-(N-1)}^{(N-1)} v_m^{\vdash\tau\dashv}(\tau) \left(\frac{1}{2N} \sum_{j=-(N-1)}^{N} \widehat{C}_X^{(d)}(s'_{k-j})\mathrm{e}^{+\mathrm{i}2\pi s'_{k-j}\tau} \right) \mathrm{e}^{-\mathrm{i}2\pi s'_k\tau}$$

$$= \sum_{j=-(N-1)}^{N} \left(\frac{1}{2N} \sum_{\tau=-(N-1)}^{(N-1)} v_m^{\vdash\tau\dashv}(\tau)\mathrm{e}^{-\mathrm{i}2\pi s'_j\tau} \right) \widehat{C}_X^{(d)}(s'_{k-j})$$

$$= \frac{1}{2N} \sum_{j=-(N-1)}^{N} V_m^{\vdash\tau\dashv}(s'_j)\widehat{C}_X^{(d)}(s'_{k-j}) \tag{9.59}$$

The above lines are obtained (i) by substituting from (9.56), (ii) because $\widehat{C}_X^{(d)}(s'_j)$ is periodic with period $2N$ and thus can be summed within arbitrary intervals of length $2N$ as demonstrated in (6.23), (iii) re-arranging the terms in the two-fold sum, and finally, the moving average in (9.59), i.e., a discretely smoothed direct spectral estimator as defined in (9.28,3), is arrived at by substituting from (9.57). This derivation is borrowed from [108].

From (9.59) it is concluded that the approximations obtained in (9.30), (9.34) and (9.35) for the moment functions of a discretely smoothed direct spectral estimator also apply to a lag window estimator on condition that both estimators are calculated for frequencies s'_k as defined in (9.53). This favours the derivation of the moment functions of the lag window spectral estimator in the following Sect. 9.3.3. Vice-versa, results obtained in Sect. 9.3.3 for a lag window estimator as defined in (9.41), (9.42), (9.43) and (9.44) also apply to any discretely smoothed direct spectral estimator, as is concluded from equation (9.49) and the example in Fig. 9.19.

9.3.3 Moment Functions of the Lag Window Estimator

Let a lag window estimator $\widehat{C}_X^{\vdash\tau\dashv}(s)$ for a continuous spectrum $C_X(s)$ be calculated from a finite number of observations. Then $\widehat{C}_X^{\vdash\tau\dashv}(s)$ is biased because (i) a bias is inherited from the direct spectral estimator and (ii) an additional bias is generated when the direct spectral estimator is smoothed. Can the effects of both these bias-generating sources be separated in the expectation

function of $\widehat{C}_X^{\vdash\tau\dashv}(s)$? An answer is given below in the first part of this subsection. In the second part, the variance and covariance functions of $\widehat{C}_X^{\vdash\tau\dashv}(s)$ are derived under assumptions which include those in (9.13) as a subset.

The expectation function of a lag window spectral estimator is readily obtained because both, integrating and calculating an expectation, are linear operators and therefore, their order can be reversed to obtain (9.61) from (9.60). Thereafter, $\mathrm{E}\big((\widehat{C}_X^{(d)}(s)\big) = \big(\mathcal{H}_N^{(d)}(s)\big) * C_X(s)$ as obtained in (9.12) is substituted, and (9.62) follows because both, the smoothing kernel $V_m^{\vdash\tau\dashv}(s)$ and the spectral window $\mathcal{H}_N^{(d)}(s)$ are defined in $-1/2 \le s \le 1/2$ and are periodic with period 1, properties shared by $C_X(s)$, i.e., the spectrum to be estimated.

$$\mathrm{E}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) = \left(\mathrm{E}\int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)\widehat{C}_X^{(d)}(r)\mathrm{d}r\right) \quad \begin{array}{l}\text{convolution}\\ \text{as in (6.144)}\end{array} \tag{9.60}$$

$$= \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)\mathrm{E}\big(\widehat{C}_X^{(d)}(r)\big)\mathrm{d}r \tag{9.61}$$

$$= \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)\left(\int_{-1/2}^{1/2} \mathcal{H}_N^{(d)}(r-r')C_X(r')\mathrm{d}r'\right)\mathrm{d}r$$

$$= \int_{-1/2}^{1/2}\left(\int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)\mathcal{H}_N^{(d)}(r-r')\mathrm{d}r\right)C_X(r')\mathrm{d}r' \tag{9.62}$$

$$= \int_{-1/2}^{1/2} \mathcal{V}_m(s-r')C_X(r')\mathrm{d}r', \quad \text{with} \tag{9.63}$$

$$\mathcal{V}_m(s) = \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)\mathcal{H}_N^{(d)}(r)\mathrm{d}r \quad \begin{array}{l}\text{being the spectral}\\ \text{window of the lag}\\ \text{window estimator}\end{array} \tag{9.64}$$

From (9.62), the expectation function of the lag window estimator is arrived at in (9.63), with $\mathcal{V}_m(s)$ as defined in (9.64).

If $\mathcal{H}_N^{(d)}(s)$ in (9.12) is the spectral window pertaining to the direct spectral estimator $\widehat{C}_X^{(d)}(s)$ calculated using a data taper $h_N^{(d)}(t)$ as defined in (9.2,1), then $\mathcal{V}_m(s)$ is the spectral window pertaining to $\widehat{C}_X^{\vdash\tau\dashv}(s)$ calculated using a lag window $v_m^{\vdash\tau\dashv}(\tau)$ as defined in (9.42). Consequently, $\mathcal{V}_m(s)$ depends on $h_N^{(d)}(t)$ and $v_m^{\vdash\tau\dashv}(\tau)$.

For $N \to \infty$, $\mathcal{H}_N^{(d)}(s) \to \delta(s-n)$, $n = \ldots, -1, 0, 1, \ldots$, i.e., a delta function being periodic with period 1 as introduced in the remarks to (6.63), due to the normalisation in (9.2,1) and under the additional condition in the remarks to (9.12). For $N \to \infty$ therefore, the expectation of the lag window estimator in (9.62) becomes $\int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)C_X(r)\mathrm{d}r$ as is concluded from (6.67) and (6.68), $\delta(x)$ being the identity element under convolution (2.28,1) for functions with a real argument. This convolution integral then becomes, for $N \to \infty$, $\int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)C_X(r)\mathrm{d}r = C_X(s)$ for the reason that

$V_m^{\vdash\tau\dashv}(s) \to \delta(s-n)$, $n = \ldots, -1, 0, 1, \ldots$, as required in (9.44). Thus a lag window spectral estimator is asymptotically unbiased for $N \to \infty$, on condition that the direct spectral estimator being smoothed is asymptotically unbiased for $N \to \infty$.

The convolution in (9.64) becomes the Fourier transform of the product in (9.65), as is concluded from convolution theorem in (6.144) and (6.145) and definition (9.2,2).

$$\begin{aligned}\mathcal{V}_m(s) &= \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}((s-s')\mathcal{H}_N^{(d)}(s')\mathrm{d}s' \\ &= \sum_{\tau=-(N-1)}^{N-1} \Big(\big(v_m^{\vdash\tau\dashv}(\tau)\big)\big(h_N^{(d)} \star h_N^{(d)}(\tau)\big)\Big)\mathrm{e}^{-\mathrm{i}2\pi s\tau} \end{aligned} \tag{9.65}$$

(9.65) lends itself for the calculation of a spectral window pertaining to a lag window spectral estimator on condition that both, the variance-normalised data taper $h_N^{(d)}(t)$ pertaining to the direct spectral estimator and the lag window $v_m^{\vdash\tau\dashv}(\tau)$, are known.

For example, the spectral window $\mathcal{V}_m(s)$ of the discretely smoothed direct spectral estimator $\widehat{C}_X^{(m)}(s'_k)$ in Figs. 9.15 and 9.16 can be computed as required in (9.65), since $\widehat{C}_X^{(m)}(s'_k)$ is a lag window estimator as is concluded from (9.49). The result thus obtained is plotted in Fig. 9.21 (b), (c) and (d), with the steps in this computation demonstrated in Fig. 9.20 (a) and (c), as well as in Fig. 9.21 (a) and (b): (i) the autocorrelation of the variance-normalised cosine taper in Fig. 9.20 (a) is computed as demonstrated in Fig. 9.2, (ii) the lag window in Fig. 9.20 (c) is obtained as demonstrated in Fig. 9.19 and (iii), the product of the autocorrelation and the lag window as plotted in Fig. 9.21 (a) is Fourier transformed.

$\mathcal{B}_\triangle^\cap\big(\widehat{\mathcal{H}}_N^{(d)}(s)\big) \approx 0.000214 \approx 0.000244 = 2\Delta s'_k = 2/(2N)$, $N' = 2N$, $N = 4096$, i.e., the width at half height of the main lobe in the spectral window of the direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ as plotted in Fig. 9.20 (b) is approximately twice the distance of the discrete frequencies s'_k, for which $\widehat{C}_X^{(d)}(s'_k)$ is smoothed using the modified discrete Daniell kernel $V_{dD,M}(s'_k)$ with $M = 10$ in a band of width $20\Delta s'_k$. Consequently, the main lobe in the continuous smoothing kernel pertaining to $V_{dD,M}(s'_k)$ in Fig. 9.20 (d) is approximately one order of magnitude wider than the main lobe of the spectral window, i.e., $\mathcal{B}_\triangle^\cap\big(V_{dD,M}^{\vdash\tau\dashv}(s)\big) \approx 0.002440$. These widths at half height are, however, assessed in non-logarithmic plots which are not shown.

$\mathcal{B}_\triangle^\cap\big(V_{dD,M}^{\vdash\tau\dashv}(s)\big) \approx 10\mathcal{B}_\triangle^\cap\big(\widehat{\mathcal{H}}_N^{(d)}(s)\big)$ implies $\mathcal{B}_\triangle^\cap\big(V_{dD,M}^{\vdash\tau\dashv}(s)\big) \approx \mathcal{B}_\triangle^\cap\big(\mathcal{V}_m(s)\big) \approx 0.00240$, i.e., the width at half height of the main lobe in the spectral window of the smoothed estimator as plotted in Fig. 9.21 (b). An identical value is obtained when definition (6.97) is appplied in Fig. 9.21 (c).

The side lobes of $\widehat{\mathcal{H}}_N^{(d)}(s)$ in Fig. 9.20 (b) are, for $|s| > 0.007$, smaller by two or three orders of magnitude than those of $F_N(s)$ for $N = 4096$, the

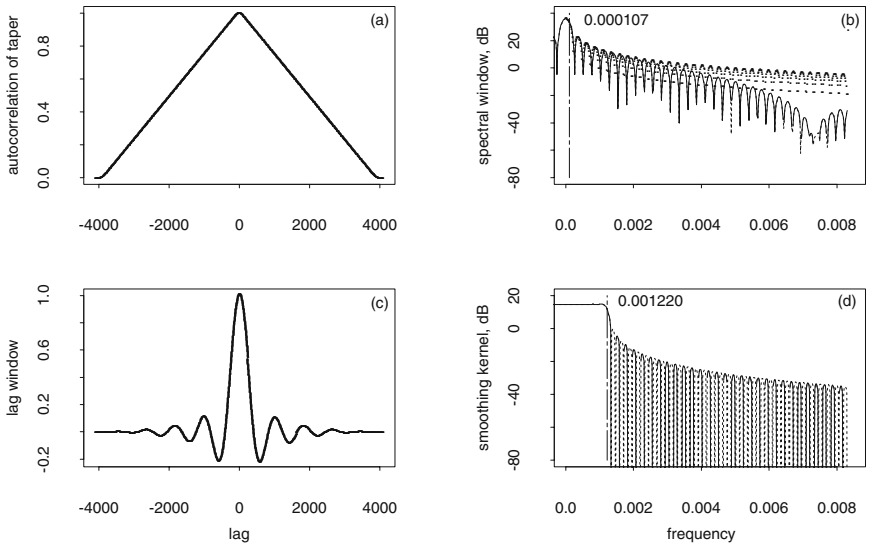


Fig. 9.20. In plot (a), autocorrelation of the variance-normalised data taper $h_N^{(d)}(t)$ (a cosine taper for $N = 4096$ and $p = 0.05$) applied to compute the direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$, $\Delta s'_k = 1/(2N)$, which is discretely smoothed using a modified discrete Daniell kernel $V_{dD,M}(s'_k)$ with $M = 10$ to obtain the discretely smoothed direct spectral estimator in Figs. 9.15 and 9.16. In plot (c), lag window $v_{dD,M}^{\vdash\tau\dashv}(\tau)$ pertaining to $V_{dD,M}(s'_k)$, for $M = 10$, $N = 4096$ and $N' = 2N$. On the right in plots (b) and (d), Fourier transforms of $h_N^{(d)} \star h_N^{(d)}(\tau)$ $\big($the spectral window of $\widehat{C}_X^{(d)}(s'_k)\big)$ and $V_{dD,M}^{\vdash\tau\dashv}(\tau)$: negative side lobes with broken lines, a long vertical line for assessing the width at half height of main lobes in the spectral window and the smoothing kernel. The Fejer kernel for $N = 4096$ is added in plot (b) using a broken line.

Fejer kernel being the spectral window of the periodogram in (8.21), which is plotted in Fig. 9.20 (b) as well. Due to these small side lobes, the leakage in (9.12) becomes negligible and the direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ is approximately unbiased for frequencies a larger distance away from the peaks in the spectrum to be estimated.

From plots (b) and (d) in Fig. 9.21, it becomes obvious that $\mathcal{V}_m(s)$ has side lobes smaller by two or three orders of magnitude than those of $F_N(s)$ for $|s| > 0.008$ (a behaviour similar to $\widehat{\mathcal{H}}_N^{(d)}(s)$ in Fig. 9.20 (b) for $|s| > 0.007$). This spectral window, therefore, reduces the leakage in convolution (9.63) to the extent that $\widehat{C}_X^{\vdash\tau\dashv}(s)$ and thus $\widehat{C}_X^{(m)}(s'_k)$ in Figs. 9.15 and 9.16 become approximately unbiased for frequencies a larger distance away from the peaks in the spectrum to be estimated. This conclusion is reconcilable with plot (b) in Fig. 9.17, showing the expectation of $\widehat{C}_X^{(m)}(s'_k)$ on the logarithmic scale as

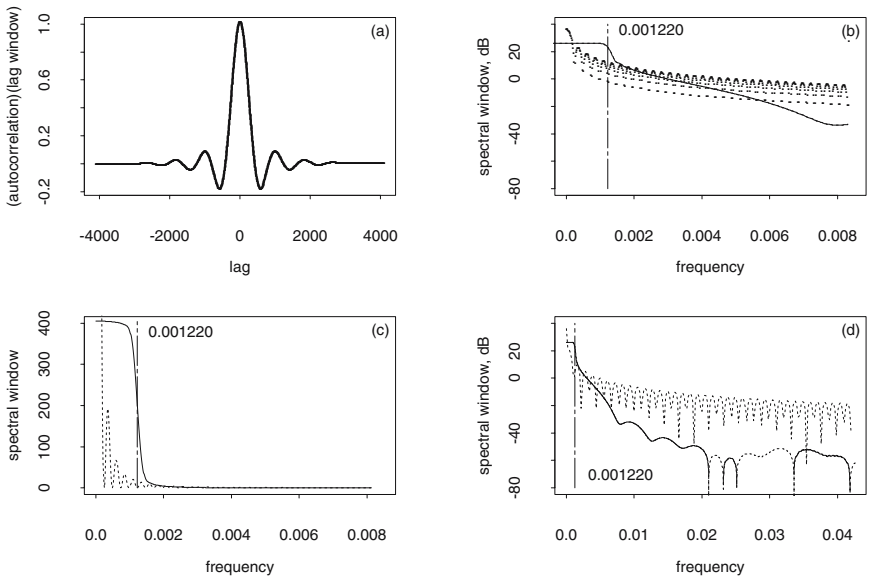


Fig. 9.21. In plot (a), the product of the autocorrelation of the variance-normalised data taper and the lag window as plotted in Fig. 9.20 (a) and (c), together with its Fourier transform in plot (b), i.e., the spectral window of the discretely smoothed direct spectral estimator as plotted in Figs. 9.15 and 9.16. Details of the spectral window in plots (c) and (d). In plots (b), (c) and (d), negative side lobes and the Fejer kernel for $N = 4096$ are plotted with broken lines, together with a long vertical line for assessing the width at half height of the spectral window.

approximated using (9.30) under the assumption that $\widehat{C}_X^{(d)}(s'_k)$ is unbiased due to tapering.

At the peaks of the spectrum to be estimated, however, $\widehat{C}_X^{(m)}(s'_k)$ is biased as is obvious in Fig. 9.17 (b). This bias originates from smoothing $\widehat{C}_X^{(d)}(s'_k)$ in a band of width $0.00244 = 20\Delta s'_k$, $\Delta s'_k = 1/(2N)$, $N' = 2N$, $N = 4096$, to obtain $\widehat{C}_X^{(m)}(s'_k)$, on condition that $\widehat{C}_X^{(d)}(s'_k)$ is unbiased. This bandwidth is recommended by diagnostic (6.115) since $\mathcal{B}_\triangle\big(C_X(s)\big) \approx 0.005$, $\mathcal{B}_\triangle\big(C_X(s)\big)$ being the bandwidth of the spectrum to be estimated obtained by applying (7.83) (the widths at half height of both peaks being approximately 0.005) in Fig. 7.10 (d) for the spectrum of the AR[4] model (X_t) in Fig. 7.14. Is the bandwidth of the estimator too large as compared to the bandwidth of the spectrum, despite diagnostic (6.115)? As a prerequisite for giving an answer to this question in (9.72), a bias in a lag window spectral estimator is assumed to be due to (i) leakage in (9.12) and (ii) smoothing in (9.42), with the contribution of both sources added as in (9.67).

If a biased direct spectral estimator $\widehat{C}_X^{(d)}(s)$ (e.g., the estimators in Fig. 9.4 (a), (b) and (c)) is smoothed to obtain a lag window estimator $\widehat{C}_X^{\vdash\tau\dashv}(s)$, then $\widehat{C}_X^{\vdash\tau\dashv}(s)$ inherits the bias in $\widehat{C}_X^{(d)}(s)$ and is subject to an additional bias due to smoothing as required in (9.42). Often a bias generated by smoothing will afflict the estimator in the peak region of the spectrum to be estimated, as is demonstrated in Fig. 9.17 (b). Consequently, the overall bias of $\widehat{C}_X^{\vdash\tau\dashv}(s)$ as defined in (9.66) is the sum of two distinct terms: $\mathrm{bias}_{\mathcal{H}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big)$ originates from leakage due to the convolution with the spectral window $\mathcal{H}_N^{(d)}(s)$ in (9.12), and $\mathrm{bias}_{\mathcal{V}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big)$ is generated by smoothing the direct spectral estimator with a smoothing kernel.

$$\mathrm{bias}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) = -\mathrm{E}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) + C_X(s) \tag{9.66}$$

$$= \mathrm{bias}_{\mathcal{H}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) + \mathrm{bias}_{\mathcal{V}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big), \quad \text{where} \tag{9.67}$$

$$\mathrm{bias}_{\mathcal{H}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) = -\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big) + C_X(s), \quad \begin{matrix}\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big) \\ \text{as in (9.12)}\end{matrix} \tag{9.68}$$

From (9.67) it is concluded that $\widehat{C}_X^{\vdash\tau\dashv}(s)$ is approximately unbiased for all $-1/2 \le s \le 1/2$ on condition that (i) $\widehat{C}_X^{(d)}(s)$ is approximately unbiased, i.e., $-\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big) + C_X(s) \approx 0$, and (ii) this property is preserved despite smoothing, i.e., $\mathrm{bias}_{\mathcal{V}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) \approx 0$.

$\mathrm{bias}_{\mathcal{V}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big)$ is approximated in (9.72) under the assumption that $\mathrm{bias}_{\mathcal{H}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) \approx 0$, and thus $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X(s)$ in (9.12). Substituting this approximation in (9.61), the expectation of the lag window estimator is approximated in (9.69) under the assumption that the bias of the direct spectral estimator is negligible. $\mathrm{bias}_{\mathcal{V}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big)$ follows in (9.70).

$$\mathrm{E}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) \approx \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r) C_X(r) \mathrm{d}r \tag{9.69}$$

$$\mathrm{bias}_{\mathcal{V}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) \approx -\int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r) C_X(r) \mathrm{d}r + C_X(s) \tag{9.70}$$

$$\approx \frac{{C_X}''(s)}{2} \int_{-1/2}^{1/2} r^2 V_m^{\vdash\tau\dashv}(r) \mathrm{d}r \tag{9.71}$$

$$\approx \frac{{C_X}''(s)}{24} \Big(\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)\Big)^2 \tag{9.72}$$

(9.71) is derived in [113] using a Taylor series expansion of $C_X(s)$. From this approximation, (9.72) readily follows by substituting the σ-width of the smoothing kernel as defined in (9.128).

A bias due to smoothing as in (9.72) is proportional to ${C_X}''(s)$, i.e., the second derivative of the spectrum to be estimated. $|{C_X}''(s)|$ is large in bands where $C_X(s)$ has spiky peaks and narrow valleys, and small in bands where

$C_X(s)$ is moderately curved. Consequently, the product of the second derivative and the squared σ-width of the smoothing kernel remains lower than a pre-selected threshold for the bias due to smoothing, on condition that the maximal $|C_X{}''(s)|$ is known and the direct spectral estimator is smoothed by means of an adequate kernel.

In applications, a lag window is applied such that the pertaining smoothing kernel $V_m^{\vdash\tau\dashv}(s)$ has a σ-width not larger than half the bandwidth of the spectrum to be estimated, i.e., $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big) \leq \mathcal{B}_\triangle^\cap\big(C_X(s)\big)/2$, as is recommended in (6.115). If $\mathcal{B}_\triangle^\cap\big(C_X(s)\big)$ is not known then the procedures introduced in Sect. 9.3.5 can be helpful to arrive at an adequate width of the smoothing kernel.

In Figs. 9.15 and 9.16 for example, the discrete direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$, with $\Delta s'_k = 1/N' = 1/8192 = 0.000122$, is smoothed using a modified discrete Daniell kernel $V_{dD,M}(s'_k)$ with $M = 10$. The lag window pertaining to $V_{dD,M}(s'_k)$ is plotted, together with its continuous smoothing kernel $V_{dD,M}^{\vdash\tau\dashv}(\tau)$, in Fig. 9.20 (c) and (d). In plot (d), $\mathcal{B}_\triangle^\cap\big(V_{dD,M}(s'_k)\big) \approx .0024$, i.e., the width at half height of the main lobe in the continuous smoothing kernel, is approximated, as demonstrated in Fig. 6.24. Is this width in agreement with diagnostic (6.115)? Using this diagnostic, $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big) \leq 0.0047/2$ is obtained, 0.0047 being the bandwidth of the spectrum to be estimated as obtained in Figs. 7.10 and 7.14. Since $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big) \approx\approx 2M\Delta s'_k$ in Table 9.4, $2M \times 0.000122 \leq 0.00234$ and $M \leq 9.713$ follow, $\approx\approx$ a rough approximation. The smoothing kernel thus arrived at is narrow enough to estimate $C_X(s)$ without bias for almost all frequencies, except those in the bands with the peaks as is obvious in Fig. 9.17 (a). In these bands, $|C_X{}''(s)|$ becomes too large for the width of the smoothing kernel in (9.72), resulting in $\mathrm{bias}_\mathcal{V}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) > 0$. Above, $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$ is substituted with $2M\Delta s'_k$, i.e., the bandwidth of the modified discrete Daniell kernel as defined in (9.125), because

1. its equivalent width, as well as its σ-width as in Table 9.4, are calculated despite the remarks to (9.126) and (9.128)
2. the bandwidth of a discrete smoothing kernel as defined in (9.125) is intuitively appealing since it is the width of a moving average as defined in the remarks to (2.39), and
3. parameter M readily follows when this bandwidth is used in diagnostic (6.115).

This example demonstrates that (9.63), (9.64), (9.67) and (9.72) imply, in condensed form, that a lag window spectral estimator is unbiased on condition that (i) the observations are tapered adequately, i.e., taking into account the dynamic range of the spectrum to be estimated as demonstrated in Figs. 9.3 and 9.4, and (ii) the direct spectral estimator calculated from the tapered observations is smoothed taking into account the bandwith of the spectrum as recommended in (6.115).

(9.63), (9.64) and (9.65) suggest that a lag window spectral estimator could be calculated in the following steps: (i) computing the empirical covariance function from observations that have not been tapered, (ii) multiplying this covariance function with a lag window whose smoothing kernel has, in addition to properties (6.108), side lobes that are small enough to suppress leakage even when a spectrum having a large dynamic range is estimated, and (iii), calculating a Fourier transform of the product resulting from step (ii). The smoothing kernel is required to have small side lobes because it is, in this case, the spectral window: the lag window is applied to the empirical covariance function in order to arrive at an estimator with both a negligible bias and a small variance.

The construction of a lag window possessing all the above properties is difficult because an associated smoothing kernel, possibly having negative side lobes with absolute values much larger than the examples given in Figs. 9.2 and 9.21, induces the risk of obtaining a negative expectation of the spectrum, as is demonstrated in Fig. 6.22. There, the convolution becomes negative for displacements such that the dominating peak becomes coincident with a negative side lobe in the Dirichlet kernel. In spectral estimation, it is much easier (i) to keep the bias to the minimum by means of tapering the observations and (ii) to reduce the variance by means of applying a lag window to the empirical covariance function (calculated from tapered observations), an operation which corresponds, in the frequency domain, to smoothing a direct spectral estimator (calculated from tapered observations).

Above, in the first part of Sect. 9.3.3, the lag window spectral estimator is analysed with respect to its expectation function. From this analysis it follows that $\mathrm{bias}_{\mathcal{V}}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big)$, i.e., the bias due to smoothing as approximated in (9.72), is the price to be paid for obtaining a lag window estimator that has a smaller variance than that of the direct spectral estimator being smoothed. Is the reduction in variance worth the cost? An answer to this question is obtained when the variance function of the lag window spectral estimator is derived in the second part of Sect. 9.3.3, which follows.

Any lag window spectral estimator $\widehat{C}_X^{\vdash\tau\dashv}(s'_k)$, s'_k discrete frequencies with $\Delta s'_k = 1/(2N)$ as defined in (9.53), can be written as a discretely smoothed direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$, as is concluded from (9.59). Consequently, the variance of a lag window estimator $\widehat{C}_X^{\vdash\tau\dashv}(s'_k)$ can be approximated using (9.38), i.e., the approximation for the variance of a discretely smoothed direct spectral estimator. This approximation is obtained under assumptions (9.32). Under these assumptions, for example, the variance function of the discretely smoothed direct spectral estimator in Figs. 9.15 and 9.16 is approximated in the remarks to (9.59).

Approximations for the variance function of a lag window spectral estimator $\widehat{C}_X^{\vdash\tau\dashv}(s)$ calculated for real frequencies $-1/2 \leq s \leq 1/2$ can be derived under the assumptions stipulated in (9.73), a version of (9.32) adapted for continuous frequencies. (9.73,1) requires that the direct spectral estimator be-

ing smoothed has a negligible bias and a variance very close to the square of the spectrum to be estimated. The first property is the result of tapering, and the second can be attributed to a direct spectral estimator reconcilable with assumptions (9.13), as is concluded from (9.16). (9.73,3) implies (9.13,3), and (9.73,2), stipulating that the width of the band in which $C_X(s)$ is required to be constant is larger than the autocorrelation width of the smoothing kernel, implies (9.13,4) on condition that $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) \geq 2\mathcal{B}_\triangle^\cap\big(H_N(s)\big)$.

Let $\widehat{C}_X^{\vdash\tau\dashv}(s) = \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)\widehat{C}_X^{(d)}(r)\mathrm{d}r$ *be a lag window estimator for a continuous spectrum* $C_X(s)(s)$ *as defined in (9.41), (9.42), (9.43) and (9.44). Then* $\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big)$ *can be approximated under the following assumptions:*

1. $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X(s)$ *and* $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s)$
2. $C_X(s)$ *is approximately constant in band* $s-\zeta \leq s \leq s+\zeta$ *having width* 2ζ, $2\zeta \geq \mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$, $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ *the autocorrelation width of the smoothing kernel as defined in (9.129)*
3. (X_t) *is normal as required in (9.13,3), and*
4. *the distance between* s *and* $-1/2$, 0, *or* $1/2$ *is larger than* $|\zeta|$.

(9.73)

These assumptions imply those in (9.32) on condition that $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) \geq 2M\Delta s'_k$, and therefore, approximation (9.38) for the variance of a discretely smoothed direct spectral estimator (9.28,3) also applies to a lag window spectral estimator calculated for frequencies s'_k as defined in (9.53), since, for these frequencies, $\widehat{C}_X^{\vdash\tau\dashv}(s'_k) = \frac{1}{2N}\sum_{j=-(N-1)}^{N} V_m^{\vdash\tau\dashv}(s'_j)\widehat{C}_X^{(d)}(s'_{k-j})$ as obtained in (9.59). When $\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s'_k)\big)$ is approximated using (9.38), then the autocorrelations of the smoothing kernel for discrete displacements $l \times \Delta s'_k$ are computed, which are then weighted with the correlations of the direct spectral estimator as obtained in (9.17) and added in order to arrive at the result. These computations are feasible, as is demonstrated in the remarks to (9.38).

However, approximation (9.138) derived in Sect. 9.5.5 is (i) easier to compute and (ii) related to simple properties (i.e., properties that can be described using a few numbers) of the data taper and the smoothing kernel applied for computing $\widehat{C}_X^{\vdash\tau\dashv}(s'_k)$. These are the reasons why (9.138) became the usual approximation for the variance of a lag window spectral estimator computed for discrete frequencies s'_k with $\Delta s'_k = 1/(2N)$ as obtained in (9.59) under the assumption that $C_X(s)$ is locally (2.55) constant in a band approximately 2ζ wide. 2ζ is the width of band $s-\zeta \leq s \leq s+\zeta$ in (9.73,2). This approximation is reformulated in (9.75) and (9.76), using $\Xi(h)$ as defined in (9.74) and $\Delta s_k = 1/N$ as in (6.21).

$$\Xi(h) = \frac{N\Sigma(h^4)}{\big(\Sigma(h^2)\big)^2} = \frac{N\sum_{t=0}^{N-1}(h_N(t))^4}{\Big(\sum_{t=0}^{N-1}(h_N(t))^2\Big)^2} \tag{9.74}$$

$$\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s'_k)\big) \approx \big(C_X^2(s'_k)\big)\frac{\Xi(h)}{N}\frac{1}{\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)} \tag{9.75}$$

Table 9.3. $\Xi(h)$ for cosine tapers calculated in Problem 9.15 using definition (9.74).

p	0	0.025	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.5
$\Xi(h)$	1.000	1.027	1.055	1.116	1.184	1.261	1.347	1.444	1.553	1.675	1.944

$$\approx \frac{(\Delta s_k)\big(\Xi(h)\big)}{\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)} \big(C_X^2(s'_k)\big) \quad \text{with } \Delta s_k = 1/N \tag{9.76}$$

In (9.75) and (9.76), $\Delta s_k = 1/N$, $\Xi(h)$ and $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ are determined by the number of observations available, the data taper applied to the observations, and the kernel used for smoothing the direct spectral estimator. Consequently, this approximation for $\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s'_k)\big)$ does not depend on s'_k and thus applies for continuous frequencies $-1/2 \le s \le 1/2$, (except for those excluded in (9.73,4)), as formulated in (9.77) and (9.78).

$$\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) \approx \Xi(h)\frac{1}{N}\frac{1}{\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)}\big(C_X^2(s)\big) \tag{9.77}$$

$$\approx \frac{(\Delta s_k)\big(\Xi(h)\big)}{\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)}\big(C_X^2(s)\big) \quad \text{with } \Delta s_k = 1/N \tag{9.78}$$

Clearly, these approximations apply under the assumptions in (9.73), which are equivalent to those made in Sect. 9.5.5.

(9.77) (or (9.78)) is preferred to (9.38) for approximating the variance of a smoothed direct spectral estimator because (i) it applies for continuous frequencies $-1/2 \le s \le 1/2$, except for those excluded in (9.73,4), (ii) $\Xi(h)$ and $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ can be easily computed and are related to properties of the data taper and the smoothing kernel that can be easily seen in a plot, e.g., $\Xi(h)$ increases when the tails of the data taper become longer as demonstrated in Table 9.3 for the case of cosine tapers, and (iii) it allows for approximating the variance of a lag window estimator for large N in the following paragraph.

If, in (9.77) (or (9.78)), $\Xi(h)$ and $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ remain fixed (due to a data taper and a smoothing kernel not being changed), then $\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big)$ decreases with $1/N = \Delta s_k$ when N increases. Consequently, $\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) \to 0$ on condition that $\Xi(h)/\big(N\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)\big) \to 0$ for $N \to \infty$. This is the property required in definition (1.4), and thus a lag window spectral estimator attains a second moment function as desired in the introduction to Sect. 9.3.

The variance function of a lag window spectral estimator as approximated in (9.77) (or (9.78)) is proportional to the squared spectrum to be estimated, the factors being $\Xi(h)$, $1/N$ and $1/\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$. Of these, N and $\Xi(h)$ are determined by the observations available and the data taper applied, and $1/\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ by the kernel used for smoothing the direct spectral estimator having a variance as approximated in (9.16). Consequently,

when (9.16) is compared with (9.77) and (9.78), it becomes evident that $\big(\Xi(h)\big) \times (1/N) \times \big(1/\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)\big)$ is the variance reduction attained by smoothing a direct spectral estimator. This variance reduction is discussed in the following paragraphs for two cases: i.e., for smoothing a periodogram and for smoothing a direct spectral estimator.

Assume that a spectrum with a small dynamic range is estimated using a periodogram $I_X(s_k)$ (8.1,3), a possibility conceded in the concluding remarks of Sect. 8.3. Thereafter, $I_X(s_k)$ can be smoothed to arrive at an estimator that is not subject to the disadvantages enumerated in the concluding remarks of Sect. 8.4. The estimator thus obtained, i.e., the smoothed periodogram $I_X^{(m)}(s_k)$, will have a variance that can be approximated using (9.77) (or (9.78)) under assumptions (9.73) because (i) $\mathit{\Pi}_N(t)$ in (6.52,1) becomes the "data taper" of the periodogram as argued in the remarks to definitions (9.2) and (ii) any discretely smoothed direct spectral estimator is a lag window estimator (as is concluded in the remarks to (9.52)). $\Xi(h) = 1$ then follows from definition (9.74), and thus (9.78) becomes

$$\mathrm{Var}\big(I_X^{(m)}(s_k)\big) \approx \frac{\mathit{\Delta} s_k}{\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)}\big(C_X^2(s)\big) \tag{9.79}$$

$$\approx (1/d)\big(C_X^2(s)\big) \quad \text{with} \quad d \times \mathit{\Delta} s_k = \mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big). \tag{9.80}$$

When approximation (9.80) is compared with the one in (8.24,7), it becomes obvious that the variance of $I_X^{(m)}(s_k)$ is $(1/d) \approx (1/d_1)$ times smaller than the variance of $I_X(s_k)$, d_1 being the largest positive integer smaller than d. This variance reduction is reconcilable with the one obtained when $I_X^{(m)}(s_k)$ is calculated as the mean of d_1 random variables $I_X(s_k)$ under iid. conditions since the $I_X(s_k)$ are (i) identically distributed as required by (9.73,1,2) and (ii) independent as implied by (8.24,1). Consequently, the variance of a periodogram $I_X(s_k)$ (8.1,3) is reduced by approximately $(1/d)$ when it is discretely smoothed, with d being the autocorrelation width of the smoothing kernel in units of $\mathit{\Delta} s_k$. For example, the spectrum of a turbulent flow in the atmospheric boundary layer is estimated in Sect. 9.4.2 by smoothing a periodogram calculated for the Fourier frequencies.

A spectrum with a large dynamic range is estimated using a direct spectral estimator $\widehat{C}_X^{(d)}(s)$ (9.3) having a variance $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s)$ as obtained in (9.16). $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big)$ is reduced when $\widehat{C}_X^{(d)}(s)$ is smoothed and the variance of the smoothed direct spectral estimator is approximated in (9.77) and (9.78) under assumptions (9.73). For example, the variance of the smoothed direct spectral estimator in Figs. 9.15 and 9.16, which is a lag window estimator in accordance with the remarks to (9.52), is approximated using the following factors in (9.77) (or (9.78)): $\Xi(h) = 1.055$ for a cosine taper with $p = 0.05$ from Table 9.3, and $\mathcal{B}_\star\big(V_{dD,M}^{\vdash\tau\dashv}(s)\big) = 0.0026$ for a continuous smoothing kernel associated with a modified discrete Daniell kernel for $M = 10$, $N = 4096$ and $\mathit{\Delta} s'_k = 1/(2N)$ from Table 9.4. With these factors, $\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) \approx$

$(1/4096) \times 1.055 \times (1/0.0026) \times \left(C_X^2(s)\right) \approx 0.0991 \times \left(C_X^2(s)\right) \approx 0.10 \times \left(C_X^2(s)\right)$ is arrived at, reconcilable with the variance reduction obtained for this example in the remarks to (9.38) under assumptions (9.32). Identical variance reductions are obtained from both approximations, i.e., the one in (9.38) as well as its counterpart in (9.77) and (9.78), because the assumptions in (9.32) and (9.73) are equivalent.

(9.73,2) requires that a continuous (9.1,2) spectrum to be estimated be constant in a band having a width equal to the autocorrelation width of the smoothing kernel applied. This specification has to be adhered to in applications since it is counterproductive to average realised values which stem from random variables having large differences in their expectations.

The variance of a direct spectral estimator is reduced when it is smoothed in accordance with (9.73), however, this reduction in variance is only attained at the cost of an increase in bias, as is concluded from (9.71) and (9.72): in (9.72) the bias due to smoothing is proportional to the squared σ-width of the smoothing kernel and the second derivative of the spectrum, whereas, in (9.77) and (9.78), the variance is proportional to the inverse of the autocorrelation width and the square of the spectrum. For example, when the direct spectral estimator in Fig. 9.16 is discretely smoothed, its variance is reduced by a factor 10, as calculated above, however, the smoothed estimator is slightly biased at the peaks of the spectrum, as is demonstrated in Fig. 9.17 (a). An estimator having (i) a smaller variance than the one plotted in Figs. 9.15 and 9.16 and (ii) a bias not larger than the one shown in Fig. 9.17 (a) can be computed on condition that N increases, because, in (9.77) and (9.78), the autocorrelation width of the smoothing kernel cannot be increased without increasing its σ-width in (9.72). Thus (9.77) as well as (9.78) and (9.72) imply a bias-variance tradeoff, on condition that N and $\Xi(h)$ remain constant.

Consequently, the statistical properties of a lag window spectral estimator depend on the σ- and autocorrelation widths, and not on a particular form, of the smoothing kernel applied [74]. This is in agreement with the results obtained in Problems 6.31 and 7.8. There, an input sequence in filters (6.120) and (7.84) can be smoothed to approximately identical output sequences when it is convolved with a variety of impulse response sequences having different forms but identical width. In contrast, when the input sequence is convolved with impulse response sequences having identical forms but different widths, the output sequences become distinct from each other. From these experiments it is concluded that, when a direct spectral estimator is computed, a smoothing kernel having form A performs as satisfactorily as another one having form B, A and B arbitrary, on condition that the smoothing kernels have (i) identical width [74] and (ii) properties as required in (9.28) and/or (9.44).

In applications, either a lag window or a discrete smoothing kernel are applied such that the pertaining smoothing kernel $V_m^{\vdash\tau\dashv}(s)$ has a σ-width

not larger than half the bandwidth of the spectrum to be estimated, i.e., $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big) \leq \mathcal{B}_\triangle^\cap\big(C_X(s)\big)/2$, as is recommended in (6.115). Obviously, the estimator thus obtained (i) is biased at its peaks as is concluded from (9.72) and, assuming $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) \approx \mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$, (ii) does not satisfy (9.73,2) or (9.32,4). This is a minor dilemma compared to the one you will find yourself in if you do not know the bandwidth of the spectrum to be estimated. In this difficult situation, the procedures introduced in Sect. 9.3.5 can be helpful to arrive at an adequate width of the smoothing kernel. It is, however, clear that both the variance and bias of the smoothed direct spectral estimator become smaller when N increases and, simultaneously, the width of the smoothing kernel increases with $b \times \Delta s_k$, where $b < 1$ and $\Delta s_k = 1/N$.

9.3.4 Distribution of the Lag Window Estimator

A periodogram $I_W(s_k)$ calculated as required in (8.7) for the Fourier frequencies s_k using a time slice in a realisation of a normal (2.3) white noise process is $\chi^2_{(n)}$ distributed, $n = 1$ or $n = 2$, as proposed in (8.7,3) for the following reasons: (i) the periodogram is, for each Fourier frequency s_k, the sum of the squares of $\widehat{A}_k$ and $\widehat{B}_k$, which are stochastically independent from each other and normally distributed, and (ii) the sum of n independent random variables having a standard normal distribution is $\chi^2_{(n)}$ distributed. When both $\widehat{A}_k^2$ and $\widehat{B}_k^2$ are added to obtain the periodogram $I_W(s_k)$ then $n = 2$, and when $I_W(s_k) = A_k^2$ (for frequencies $s_k = 0$ and possibly $-1/2$ and $1/2$) then $n = 1$, as distinguished in (8.7,3).

For large finite N, a periodogram $I_X(s)$, calculated from observations x_t, $t = 0, 1, \ldots, N-1$, in a realisation of model (9.1) and normalised as in (8.24,2), i.e., with the continuous spectrum of the model, is approximately $\chi^2_{(n)}$ distributed, $n = 1$ or $n = 2$. This approximation also applies to a normalised direct spectral estimator calculated using observations stemming from model (9.1), as proposed in (9.19).

Under the same assumptions, i.e., for large finite N, a chi-square approximation also applies to the distribution of the discretely smoothed direct spectral estimators $\widehat{C}_X^{(m)}(s_k)$ and $\widehat{C}_X^{(m)}(s'_k)$ as defined in (9.28). A $\chi^2_{(n)}$ distribution, with n to be approximated, becomes perspicuous from the following rationale: $\widehat{C}_X^{(m)}(s_k)$ or $\widehat{C}_X^{(m)}(s'_k)$ are the weighted sums of $2M+1$ random variables in estimators $\widehat{C}_X^{(d)}(s_k)$ or $\widehat{C}_X^{(d)}(s'_k)$, which are normally distributed, however not independent, on condition that the estimators being smoothed have the properties required in (9.28) and (9.32).

From this plausible result, it is concluded that a chi-square approximation also applies to the distribution of a lag window spectral estimator $\widehat{C}_X^{\vdash\tau\dashv}(s'_k)$, s'_k being discrete frequencies with $\Delta s'_k = 1/(2N)$, as defined in (9.53), for the reason that, for this set of discrete frequencies, any lag window spectral estimator can be written as a discretely smoothed direct spectral estimator, as derived in (9.59). For large finite N it is therefore reasonable to assume

that a lag window spectral estimator $\widehat{C}_X^{\vdash\tau\dashv}(s)$, as defined in (9.41), (9.42), (9.43) as well as (9.44) and having properties as required in (9.73), is $\chi^2_{(n)}$ distributed, with n to be approximated.

Under this assumption, $\mathrm{E}\big((1/a)\widehat{C}_X^{\vdash\tau\dashv}(s)\big) = (1/a)\big(\mathrm{E}\widehat{C}_X^{\vdash\tau\dashv}(s)\big) = n$ implying (9.81) and $\mathrm{Var}\big((1/a)\widehat{C}_X^{\vdash\tau\dashv}(s)\big) = (1/a^2)\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s)\big) = 2n$ are obtained, due to the fact that the expectation of a random variable having a $\chi^2_{(n)}$ distribution is n, i.e., the number of degrees of freedom, its variance twice this number. Using these results,

$$a = \big(\mathrm{E}\widehat{C}_X^{\vdash\tau\dashv}(s)\big)/n \tag{9.81}$$

$$n = 2\Big(\big(\mathrm{E}\widehat{C}_X^{\vdash\tau\dashv}(s)\big)^2\Big)\Big/\big(\mathrm{Var}\widehat{C}_X^{\vdash\tau\dashv}(s)\big) \tag{9.82}$$

the normalisation $(1/a)$ and the number n of degrees of freedom in the $\chi^2_{(n)}$ approximation for the distribution of $(1/a)\widehat{C}_X^{\vdash\tau\dashv}(s)$ are arrived at above, in (9.81) and (9.82).

If $\mathrm{E}\widehat{C}_X^{\vdash\tau\dashv}(s) \approx C_X(s)$ due to an adequate tapering and a number of observations N being large (under these preconditions, a lag window spectral estimator becomes approximately unbiased, as argued in the remarks to (9.64)), then (i) (9.81) becomes approximation (9.83) and (ii) approximation (9.84) is obtained by substituting approximation (9.77) for $\mathrm{Var}\widehat{C}_X^{\vdash\tau\dashv}(s)$ in (9.82). Substituting for a and n, the *distribution of the* $\big((n/C_X(s)\big)$*-normalised lag window spectral estimator* is approximated in (9.85). (9.86) follows because the variance of a random variable having a $\chi^2_{(n)}$ distribution is $2n$.

$$a \approx \big(C_X(s)\big)/n \tag{9.83}$$

$$n \approx \frac{2\big(C_X(s)\big)^2 N\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)}{\Xi(h)\big(C_X(s)\big)^2} = \frac{2N\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)}{\Xi(h)} \tag{9.84}$$

$$\frac{1}{a}\widehat{C}_X^{\vdash\tau\dashv}(s) = \frac{n}{C_X(s)}\widehat{C}_X^{\vdash\tau\dashv}(s) \textit{ is approximately } \chi^2_{(n)} \textit{ distributed} \tag{9.85}$$

$$\mathrm{Var}\widehat{C}_X^{\vdash\tau\dashv}(s) \approx \frac{2n}{n^2}C_X^2(s) = \frac{2}{n}C_X^2(s) \tag{9.86}$$

n is the number of degrees of freedom in the $\chi^2_{(n)}$ approximation for the distribution of random function $\big((n/C_X(s)\big)\widehat{C}_X^{\vdash\tau\dashv}(s)$, i.e., the *number of degrees of freedom associated with the* $\big((n/C_X(s)\big)$*-normalised lag window spectral estimator.*

When approximation (9.86) is compared with the one obtained in (9.16), it is concluded that smoothing as applied in (9.28) or (9.42) reduces the variance of the direct spectral estimator: the variance of the smoothed estimator is $(2/n)$ times smaller than the variance of the estimator prior to smoothing. Thus the variance of the lag window spectral estimator as approximated in (9.77) and (9.86) decreases when the autocorrelation width of

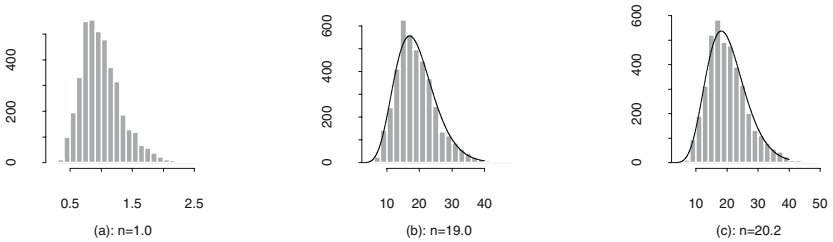


Fig. 9.22. Smoothed direct spectral estimator as in Fig. 9.15: histograms obtained subsequent to a normalisation (9.85), with $C_X(s)$ being the spectrum of the AR[4] model in Fig. 7.14 and $n = 1.0$ (a), $n = 19.0$ (b) and $n = 20.2$ (c).

the smoothing kernel increases, on condition that N remains constant. However, a smoothing kernel which is too wide can generate (i) an undesired bias as is concluded from approximation (9.72), and (ii) a poor approximation of the variance when the width of kernel is not reconcilable with (9.73,2) or (9.32,4). Thus approximation (9.72) as well as those in (9.77) and (9.86) imply a bias-variance tradeoff, which is discussed in the remarks to (9.80).

Usually, (9.84) is applied to compute the degrees of freedom associated with a lag window spectral estimator. When this approximation is applied to the discretely smoothed direct spectral estimator in Fig. 9.15 (which is also a lag window spectral estimator as concluded from the remarks to (9.52)), the following results are obtained.

In order to arrive at the estimator in Fig. 9.15, $N = 4096$ observations of the AR[4] model in Fig. 7.14 are multiplied with a cosine taper for $p = 0.05$ in a first step. In a second step, the tapered observations are zero-padded such that a direct spectral estimator is calculated for discrete frequencies s'_k with $\Delta s'_k = 1/8192$. In a third and final step, the direct spectral estimator is smoothed with a modified discrete Daniell kernel (9.29) having parameter $M = 10$. To the cosine taper applied pertains a correction factor $\Xi(h) = 1.055$ (in Table 9.3), and the variance reduction due to smoothing can be described using either $\mathcal{B}_{||}\big(V_{dD,M}(s'_k)\big) = 2M\Delta s'_k = 0.0024$ as defined in (9.125) or $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) = 0.0026$ in (9.84) and (9.129), both widths as in Table 9.4. Using $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) = 0.0026$, $n \approx 20.2$ is obtained, whereas $n \approx 19.0$ results when $2M\Delta s'_k = 0.0024$ is used.

When these approximations and the spectrum of the AR[4] model (as plotted in Fig. 7.14) are substituted in (9.85), two normalised estimators are obtained whose histograms are plotted in Fig. 9.22 (b) and (c).

The histograms in Fig. 9.22 (b) and (c) are plotted together with the $\chi^2_{(n)}$ approximations (9.85), $n = 19.0$ and $n = 20.2$, for the densities of the normalised estimators. The fit of both approximations is fairly satisfactory, because, in this example, the width of the smoothing kernel is chosen such that (i) the bias due to smoothing (9.72) does not become too large at the

peaks of the spectrum and (ii) assumption (9.73) is still adequately satisfied. It is, however, difficult to satisfy both assumptions and, at the same time, to reduce the variance of the direct spectral estimator by an order of magnitude, when only $N = 4096$ observations are available for the estimation of such a narrow spectrum having bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big) = 0.00474$ as shown in Figs. 7.10 (c) and 7.14 (b).

The $\chi^2_{(n)}$ approximation (9.85) for the distribution of $\big((n/C_X(s)\big)\widehat{C}_X^{\vdash\tau\dashv}(s)$ allows for approximating the probability in (9.87), with $Q_n(p)$ being the p quantile of the $\chi^2_{(n)}$ distribution. (9.87) implies the *approximative* $2p$ *confidence interval for the lag window spectral estimator* in (9.88),

$$\Pr\left(Q_n(p) \le \frac{n\widehat{C}_X^{\vdash\tau\dashv}(s)}{C_X(s)} \le Q_n(1-p)\right) \approx 1 - 2p \tag{9.87}$$

$$\Pr\left(\frac{n\widehat{C}_X^{\vdash\tau\dashv}(s)}{Q_n(1-p)} \le C_X(s) \le \frac{n\widehat{C}_X^{\vdash\tau\dashv}(s)}{Q_n(p)}\right) \approx 1 - 2p \tag{9.88}$$

provided that assumptions (9.73) are satisfied. Under these assumptions, approximation (9.77) for the variance of the lag window spectral estimator is obtained, which is thereafter substituted in (9.82) to obtain approximation (9.84) for the degrees of freedom associated with the $\big((n/C_X(s)\big)$-normalised lag window spectral estimator. The approximative $2p$ confidence interval in (9.88) depends, on condition that n is computed as required in (9.84), on (i) $\Xi(h)$ which compensates for correletions (9.17) induced by the cosine taper in the direct spectral estimator, (ii) $\mathcal{B}_{\star}\big(V_m^{\vdash\tau\dashv}(s)\big)$, i.e., the autocorrelation width of the smoothing kernel, (iii) the length N of the time slice in the realisation observed, and (iv) the spectrum $C_X(s)$ to be estimated.

In Fig. 9.23 for example, .95 confidence intervals (9.88) for the discretely smoothed direct spectral estimator as in Fig. 9.15 are plotted in bands $0.110 \le s \le 0.140$ and $0.430 \le s \le 0.460$, together with both the model spectrum to be estimated and the estimator. These confidence intervals are computed using $n = 19.0$ in the $\chi^2_{(n)}$ approximation (9.85). This approximation for the density of the normalised estimator (any discretely smoothed direct spectral estimator is a lag window estimator as stipulated in the remarks to (9.52)) is plotted in Fig. 9.22 (b).

A confidence interval as obtained in (9.88) applies for each frequency s individually, a property shared by the prediction intervals as defined in (5.71). Hence, such a confidence interval contains the spectrum for 95 out of 100 frequencies in the mean over many realisations. For the example in Fig. 9.23, $4077 = 4096 - 2 \times 10 + 1$ confidence intervals are computed for discrete frequencies $0 + 10\Delta s'_k \le s'_k \le 0.5 - 10\Delta s'_k$, since the approximations for the variance of the smoothed estimator, and consequently for $n = 19.0$ in the $\chi^2_{(n)}$ approximation, do not apply for frequencies close to 0 and 0.5, in accordance with (9.32,2). Of these, 3869 (or 94.89 %) contain the spectrum

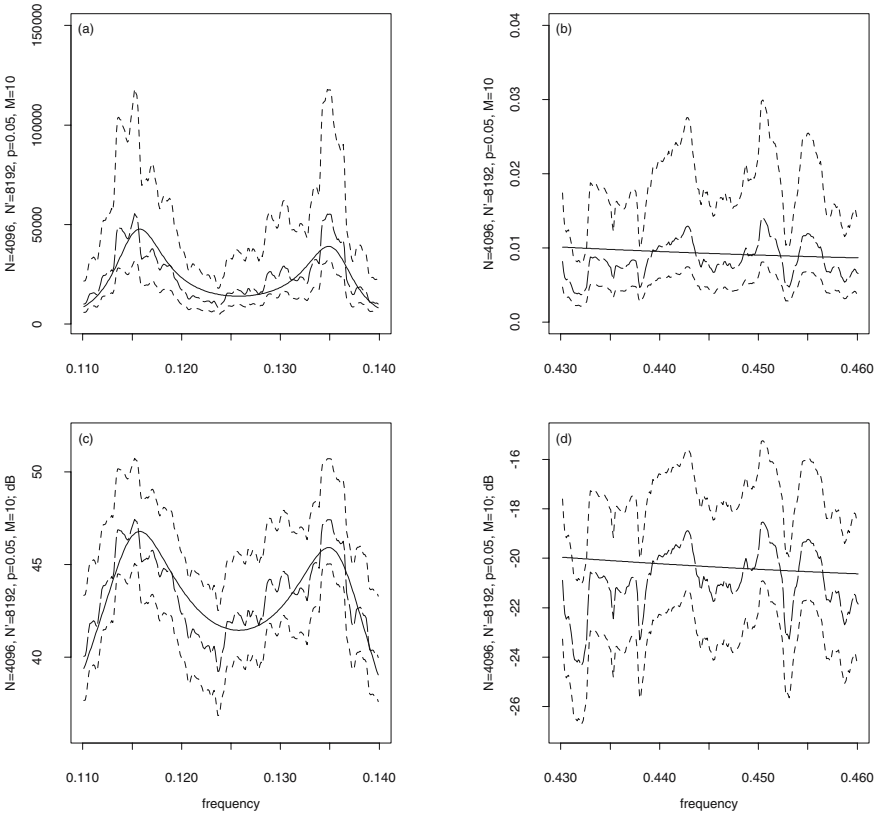


Fig. 9.23. Smoothed direct spectral estimator as in Figs. 9.15 and 9.16 for the spectrum of the AR[4] model in Fig. 7.14 (solid line), together with .95 confidence intervals (9.88) with $n = 19.0$ in the $\chi^2_{(n)}$ approximation (9.85).

of the AR[4] model. If $n = 20.2$ is used in the $\chi^2_{(n)}$ approximation, 3845 out of 4077 confidence intervals (or 94.06 %) contain the spectrum. Consequently, a $\chi^2_{(n)}$ distribution with $n \approx 20.0$ is a fairly good approximation (9.85) for the distribution of the smoothed direct spectral estimator in Figs. 9.15, 9.16 and 9.23.

The approximative confidence interval obtained in (9.88) has a lower and an upper bound which are both proportional to the estimator. Consequently, the quotient of each bound and estimator is constant for all frequencies and thus becomes, subsequent to a logarithmic transformation, the constant difference between each bound and estimator. In a logarithmic plot therefore, the bounds of the confidence intervals are at a constant distance from the estimator and the height of the intervals remains constant, as demonstrated below in Fig. 9.23, plots (c) and (d). In a non-logarithmic plot, however, the

height of the confidence interval is proportional to both the estimator and spectrum, on condition that the estimator is approximatively unbiased, as demonstrated above in Fig. 9.23, plots (a) and (b).

The above approximations were computed using `spec.univariate()` introduced in Sect. 9.5.6. In `spec.univariate()`, (i) a direct spectral estimator is computed using a cosine taper, (ii) the estimator thus obtained is smoothed with a modified discrete Daniell kernel (9.29), and (iii) the degrees of freedom in the $\chi^2_{(n)}$ approximation for the distribution of the estimator are computed as required in (9.84) and (9.129). For the approximations in Fig. 9.22 (b) and in Fig. 9.23, $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) = 0.0026$ was substituted with $\mathcal{B}_{||}\big(V_{dD,M}(s'_k)\big) = 2M\Delta s'_k = 0.0024$, $\mathcal{B}_{||}\big(V_{dD,M}(s'_k)\big)$ as in (9.125).

9.3.5 Estimating a Spectrum with Unknown Bandwidth

The statistical properties of a lag window spectral estimator (and therefore also of a discretely smoothed direct spectral estimator as is concluded from (9.50), (9.51) and (9.52)) depend on $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$ and $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$, i.e., the σ- and autocorrelation widths of its smoothing kernel $V_m^{\vdash\tau\dashv}(s)$, as argued in the remarks to (9.80). Subsequent to this discussion it is recommended, in the last paragraph of Sect. 9.3.3, to obtain $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$ using diagnostic (6.115), on condition that the bandwidth of the spectrum to be estimated is known, and then to assume $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) \approx \mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$. Under this assumption, both properties of the smoothed direct spectral estimator, i.e., the bias due to smoothing as proposed in (9.72) and the variance as proposed in (9.75) or (9.76), can be approximated. Both $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ or $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$ can also be used to define the bandwidth of a lag window spectral estimator. However, such a definition is controversial [108] since it ties the bandwidth of the estimator to either $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ or $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$. Moreover, neither a periodogram nor a direct spectral estimator have a smoothing kernel and thus their bandwidths can not be defined using $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ or $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$.

An alternative definition can be justified using the following rationale which reconsiders the bandwidths proposed in [112]. The expectation of an estimator for a continuous spectrum $C_X(s)$ is the convolution of the spectrum to be estimated with the spectral window of the estimator: $\big(F_N(s)\big) * C_X(s)$ in (8.21) is the expectation of a periodogram, $\big(\mathcal{H}_N^{(d)}(s)\big) * C_X(s)$ in (9.12) is the one of a direct spectral estimator, and $\big(\mathcal{V}_m(s)\big) * C_X(s)$ in (9.63) the one of a lag window spectral estimator. $\big(\mathcal{V}_m(s)\big) * C_X(s)$ is also the expectation of a discretely smoothed direct spectral estimator as is concluded from the remarks to (9.52). These spectral windows are required to be defining sequences for $\delta(s+n)$, $n = \ldots, -1, 0, 1, \ldots$, i.e., the periodic version of the delta function as introduced in the remarks to (6.62), such that the estimators become unbiased in the limiting case for $N \to \infty$. For finite N, however, these spectral windows have main and side lobes since they are the Fourier transforms of the sequences in (8.19), (9.9) and (9.65), all of which are time-

limited (6.108,1) and thus not band-limited as implied by (6.109). The side lobes of the spectral window are neglected in the following definition.

$$
\begin{aligned}
&B\times\mathcal{B}_{\triangle}^{\cap}\big(\big) \text{ is the bandwidth of a spectral estimator provided that:}\\
&1.\ \mathcal{B}_{\triangle}^{\cap}\big(\big) \text{ is the width at half height (6.97) of the main lobe in the}\\
&\quad\text{associated spectral window, i.e., of } F_N(s) \text{ in (8.21), } \mathcal{H}_N^{(d)}(s)\\
&\quad\text{in (9.12) and } \mathcal{V}_m(s) \text{ in (9.63), and}\\
&2.\ B \text{ is constant.}
\end{aligned}
\tag{9.89}
$$

Constant B in (9.89) accounts for choosing arbitrarily the half-power height as defined in (6.97): when this height is substituted with another height (the e-folding height for example, i.e., the height at which the power equals 1/e of the maximal height) another bandwidth is obtained. Clearly, identical heights should be used in definitions (9.89) and (7.83,1) when diagnostic (6.115) is used to assess the bandwidth of a spectral estimator provided that the bandwidth of the spectrum to be estimated is known. Definition (9.89) also applies to a periodogram or a direct spectral estimator calculated in order to estimate a spectrum having a discrete part, as is concluded from the remarks to (10.18).

Definition (9.89) can be applied to calculate the bandwidth of a spectral estimator on condition that (i) a value for B is known, and (ii) the spectral window has been computed as demonstrated in Figs. 9.20 and 9.21. In practical applications however, diagnostic (6.115) is often applied with any one of the approximations (9.90), (9.91) or (9.92) substituted for $\mathcal{B}_{\triangle}^{\cap}\big(\big)$ in (9.89) and assuming that $B = 1$. For example, approximation (9.90) is compared with the bandwidth of the spectrum to be estimated in the remarks to Figs. 9.14, 9.15 and 9.16.

Approximation (9.90) is motivated by the width of a moving average as proposed in the remarks to definition (2.39),

$$
\begin{aligned}
&\text{Let } \widehat{C}_X^{(m)}(s_k) \text{ and } \widehat{C}_X^{(m)}(s'_k) \text{ be discretely smoothed direct spectral}\\
&\text{estimators, as defined in (9.28,2,3), whose discrete smoothing}\\
&\text{kernels have approximately rectangular shape, i.e., small tails.}\\
&\text{Then } \widehat{C}_X^{(m)}(s_k) \text{ and } \widehat{C}_X^{(m)}(s'_k) \text{ have approximative bandwidths}\\
&\mathcal{B}_{||}\big(V_{m,M}(s_k)\big) \approx 2M\Delta s_k \text{ and } \mathcal{B}_{||}\big(V_{m,M}(s'_k)\big) \approx 2M\Delta s'_k.
\end{aligned}
\tag{9.90}
$$

and is thus reconcilable with (9.125) where the width of a discrete smoothing kernel is defined. The computational advantages of the autocorrelation width (as compared to the σ- and equivalent widths in Sects. 6.6 and 9.5.4) favour approximation (9.91):

$$
\begin{aligned}
&\text{Let } \widehat{C}_X^{\vdash\tau\dashv}(s) = \textstyle\int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s-r)\widehat{C}_X^{(d)}(r)\mathrm{d}r \text{ be a lag window esti-}\\
&\text{mator as defined in (9.42), (9.43) and (9.44). Then } \widehat{C}_X^{\vdash\tau\dashv}(s) \text{ has}\\
&\text{approximate bandwidth } \mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big),\ \mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) \text{ as in (9.129).}
\end{aligned}
\tag{9.91}
$$

If this approximation is applied to calculate the bandwidth of a discretely smoothed direct spectral estimator $\widehat{C}_X^{(m)}(s'_k)$ as defined in (9.28,2,3) (any

$\widehat{C}_X^{(m)}(s'_k)$ can be written as a lag window estimator $\widehat{C}_X^{\vdash\tau\dashv}(s)$ using (9.50), (9.51) and (9.52)), then $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) \approx 2M\Delta s'_k = \mathcal{B}_{||}\big(V_{m,M}(s'_k)\big)$ as discussed in Sect. 9.5.4 for discrete smoothing kernels having approximately rectangular shape. Approximation (9.92) accounts for the differences in the widths at half height of the main lobes in the spectral windows as given in Table 9.1:

Let $\widehat{C}_X^{(d)}(s)$ a direct spectral estimator as defined in (9.4) and $I_X(s)$ be a periodogram as defined in (8.1,4) and Then:
1. *$\Xi(h)/N = \Xi(h)\Delta s_k$ is an approximation for the bandwidth of $\widehat{C}_X^{(d)}(s)$, and* (9.92)
2. *$\Delta s_k = 1/N$ is an approximation for the bandwidth of $I_X(s)$.*

If approximation (9.92) is substituted for bandwidth (9.89) in diagnostic (6.115), then diagnostic (6.116) is obtained.

Caution is called for when diagnostic (6.115) is applied together with any one of the approximations (9.90), (9.91) or (9.92) to determine the bandwidth of a spectral estimator: this procedure does not guarantee that an estimator has (i) a negligible bias for all frequencies and (ii) a variance as small as desired. This restriction is imposed for the reason that (i) the bandwidth and thus the statistical properties of a spectral estimator depend not only on the bandwidth of the spectrum to be estimated but also on the length N of the observational period, and (ii) a bias-variance tradeoff is implied by (9.77) as well as (9.78) and (9.72), as is demonstrated in the remarks to (9.80).

However, diagnostics (6.115) or (6.116) can not be applied when the bandwidth of the spectrum to be estimated is not known. Possible solutions to this problem are proposed in (9.93) and the following remarks.

An unknown bandwidth $\mathcal{B}_\triangle\big(C_X(s)\big)$ as defined in (7.83,1) of a continuous spectrum to be estimated can be approximated by:
1. *using geophysical knowledge of the observed phenomenon*
2. *window closing, i.e., smoothing a direct spectral estimator using kernels with gradually decreasing widths*
3. *cutting off a lag window for lags where the empirical covariance function becomes negligible*
4. *estimating the bandwidth from an empirical covariance function as proposed in (9.95) on condition that the spectrum to be estimated has only one peak* (9.93)
5. *determining the width at half height of the narrowest form found in a plot of an empirical spectrum obtained from parametric spectral estimation as introduced in Sect. 9.2.5*
6. *approximating the bandwidth in a plot of a direct spectral estimator.*

In applications, procedures as proposed in (9.93) are used to obtain an approximation for the unknown bandwidth of the spectrum to be estimated,

as demonstrated in Sect. 9.4.1, where the spectrum of microseisms is estimated. However, simulation experiments using linear processes with known spectra are better suited to demonstrate the performance of the procedures as proposed in (9.93) in the following paragraphs.

As suggested in (9.93,1), the bandwidth of an unknown continuous spectrum can be obtained from geophysical knowledge on condition that (i) spectra of similar stationary processes have already been estimated, and/or (ii) the bandwidth of the unknown spectrum can be inferred from geophysical theory. If neither empirical nor theoretical knowledge is available then spectral estimation becomes an exploratory tool and, usually, a combination of the other procedures enumerated in (9.93) is applied to obtain an approximation of the bandwidth of the spectrum to be estimated.

When the procedure called window closing (9.93,2) is applied, a direct spectral estimator is smoothed using a variety of smoothing kernels which have decreasing widths [74]. For example, a direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ for the spectrum of the AR[4] model in Fig. 7.14 is calculated as required in (9.3) for discrete frequencies s'_k with $\Delta s'_k = 1/(2N) = 0.000122$, $N = 4096$, using a cosine taper with $p = 0.05$, and thereafter plotted in Figs. 9.15 and 9.16 as well as, with vertical broken lines, in Fig. 9.24 for band $0.10 \leq s'_k \leq 0.15$. $\widehat{C}_X^{(d)}(s'_k)$ is smoothed using modified discrete Daniell kernels $V_{dD,M}(s'_k)$ as defined in (9.29) having parameters $M = 80$, $M = 40$, $M = 20$, and $M = 10$. Jenkins and Watts in [74] recommend to begin with a wide kernel and thus to calculate an estimator which is too smooth and in which, presumably, narrow forms of the spectrum are not resolved. The estimator for $M = 80$ in Fig. 9.24 for example, does not resolve the peaks in the spectrum to be estimated, since its bandwidth $160\Delta s'_k \approx 0.01953$ as approximated using (9.90) is larger than 0.01364, i.e., the width at half height of the valley between the peaks as obtained in Fig. 7.10 (d).

Subsequent to plotting the estimator having a bandwidth too large, the smoothing is repeated using a sequence of kernels decreasing in width until, in the smoothed estimators thus obtained, forms become visible that are considered to be too narrow, i.e., generated by the random fluctuations of the direct spectral estimator having variance $\mathrm{Var}\big(\widehat{C}_X^{(d)}(s)\big) \approx C_X^2(s)$ as obtained in (9.16). When, for example, the direct spectral estimator in Fig. 9.24 is smoothed using a modified discrete Daniell kernel with $M = 5$ then an estimator with four peaks will result (not plotted) all of which seem too narrow. Since the estimators for $M = 40$, $M = 20$ and $M = 10$ in Fig. 9.24 have two peaks, the estimator for $M = 20$ is selected as estimator for the unknown spectrum. This estimator has bandwidth $40\Delta s'_k \approx 0.0049 \approx 0.00474 = \mathcal{B}_{\triangle}\big(C_X(s)\big)$, the bandwidth of the spectrum to be estimated as obtained in Fig. 7.10 (d). Consequently, the bandwidth of the estimator thus obtained is too large as compared to the bandwidth obtained when diagnostic (6.115) is applied.

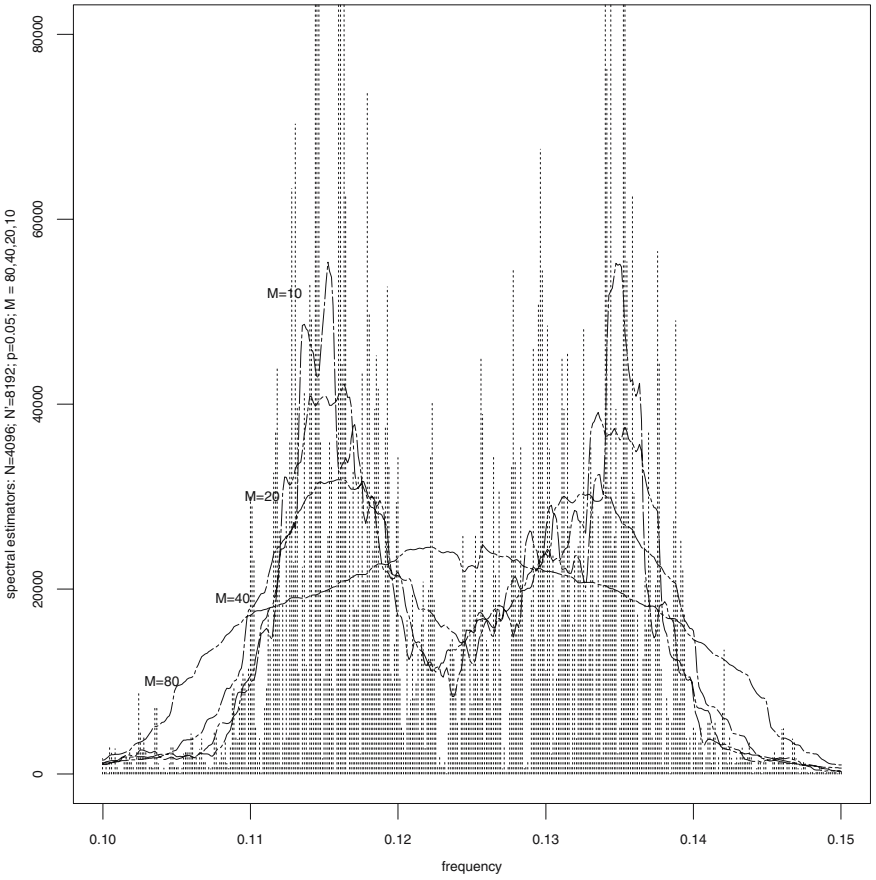


Fig. 9.24. Window closing: the direct spectral estimator as in Figs. 9.15 and 9.16 (vertical broken lines) for the spectrum of the AR[4] model in Fig. 7.14 (not plotted) is smoothed using modified discrete Daniell kernels as defined in (9.29) for $M = 80, 40, 20, 10$.

As demonstrated above, the analyst performing window closing has some knowledge about the smoothness of the spectrum to be estimated, and will select the smoothed estimator that comes closest to his idea of what the spectrum should look like. Consequently, window closing is a subjective method.

(9.93,3) recommends to use a lag window $v_m^{\vdash\tau\dashv}(\tau)$ such that $v_m^{\vdash\tau\dashv}(\tau) = 0$ for lags $\tau > \tau_0$, with τ_0 being the lag for which $|\hat{c}_X(\tau)|$ comes close to zero, i.e., $|\hat{c}_X(\tau)| > \tau_0) \approx 0$. $\hat{c}_X(\tau)$ is the empirical covariance function as defined in (2.1,3). When (9.93,3) is applied, a lag window $v_m^{\vdash\tau\dashv}(\tau)$ is obtained that allows for calculating the width of the pertaining smoothing kernel using definitions (9.126), (9.128) or (9.129), since $V_m^{\vdash\tau\dashv}(s)$ is the Fourier transform

of $v_m^{\vdash\tau\dashv}(\tau)$ in (9.42). The width of the smoothing kernel thus obtained is then considered to be an approximation for the bandwidth of the spectrum to be estimated.

In Fig. 9.25 (b) for example, an empirical covariance function $\hat{c}_X(\tau)$ calculated from $N = 4096$ observations stemming from a realisation of the AR[4] model (X_t) in Fig. 7.14 becomes negligible for lags $\tau > \tau_0 = 3800$. If a lag window spectral estimator is calculated from $\hat{c}_X(\tau)$ then, following (9.93,3), $\tau_0 = 3800$ becomes the cut-off lag: a Bartlett lag window (as defined in Fig. 9.18), for example, becomes the even triangle sequence $\Lambda_{2m+1}^{(e)}(\tau)$ with parameter $m = 3800$. The Fourier transform of this Bartlett lag window is the Fejer kernel $F_{3800}(s)$ having a main lobe being 0.000301 wide as approximated using $\mathcal{B}_{\triangle}^{\cap}\big(F_N(s)\big) \approx 7/(8N)$ given in Table 9.1. Obviously, the approximation thus obtained, i.e., $\mathcal{B}_{\triangle}^{\cap}\big(F_{3800}(s)\big) \approx 0.000301$, is an order of magnitude smaller than the bandwidth 0.00474 of the spectrum to be estimated as obtained in Fig. 7.10 (d). What generates this too large difference?

(9.93,3) assumes that (i) there is a general relationship between the bandwidth of a continuous spectrum and the width of the covariance function, i.e., a relationship that also applies to spectra having more than one peak, and (ii) the empirical covariance function is not biased for large lags.

Unfortunately, neither the first nor the second assumption applies, in general, and therefore, procedure (9.93,3) is as subjective as the window closing in (9.93,2). The first assumption is invalidated by the counter examples given in Figs. 7.15 and 7.17. In these examples, it is not possible to see the properties of the spectrum (number and frequencies of its peaks, its bandwidth or its dynamic range) in the plots of the covariance functions. The second assumption is falsified by the remark to (2.58) pointing to the possibility that an empirical covariance function often overestimates the theoretical covariances for large lags. For example, the oscillations in the covariance function of the AR[4] model in Fig. 7.14 damp out very rapidly with absolute values becoming small for lags $\tau > 150$, as can be seen in Fig. 9.25 (a). In contrast, an empirical covariance function calculated from a time slice of length $N = 4096$ of this model shows quite large covariances for lags $400 \leq \tau \leq 3800$, as demonstrated in Fig. 9.25 (a) and (b).

(9.93,4) proposes approximation (9.95) which is derived in [139] for the bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big)$ of an unknown continuous spectrum $C_X(s)$. Approximation (9.95) is easily computed from an empirical covariance function $\hat{c}_X(\tau)$ obtained from a time slice (x_t), $t = 0, 1, \ldots, N$, in a realisation of the stationary process (X_t) under analysis, on condition that $C_X(s)$ has only one peak.

$$\mathcal{B}_{\triangle}\big(C_X(s)\big) \approx \mathcal{B}_{\tau}\big(C_X(s)\big) = \frac{\big(c_X(0)\big)^2}{2\sum_{\tau=-(N-1)}^{N-1}(1-\tau/N)\big(c_X(\tau)\big)^2} \tag{9.94}$$

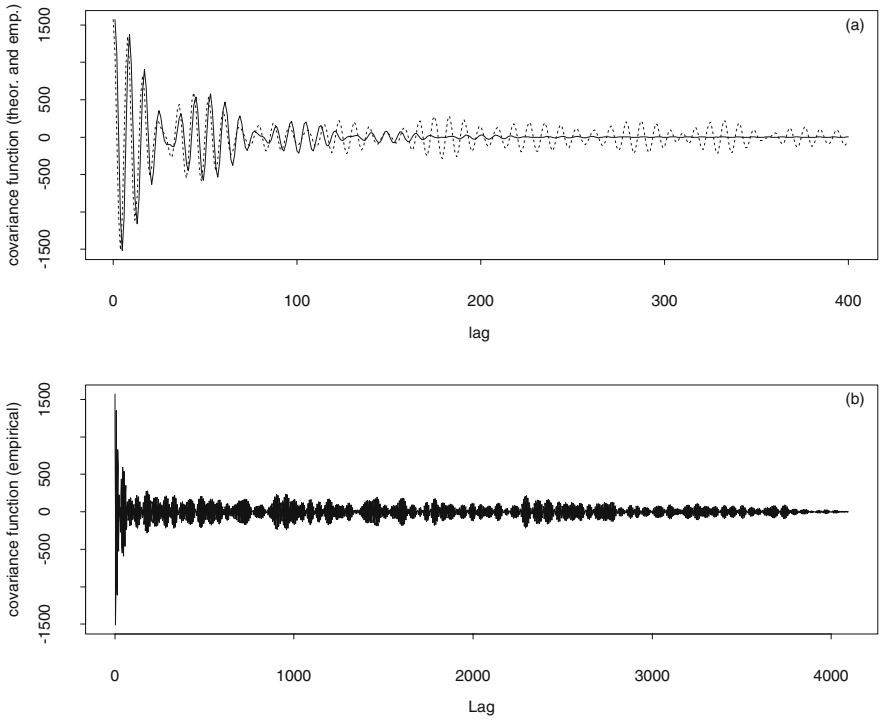


Fig. 9.25. Above in plot (a), the covariance function $c_X(\tau)$ of the AR[4] model (X_t) in Fig. 7.14 is plotted (solid line) for lags $\tau = 1, \ldots, 400$, together with an empirical covariance function $\hat{c}_X(\tau)$ (broken line). $\hat{c}_X(\tau)$ is also plotted below in (b) for lags $\tau = 1, \ldots, 4095$. $c_X(\tau)$ is estimated from those $N = 4096$ observations used to calculate the spectral estimators as plotted in Figs. 9.15, 9.16 and 9.24.

$$\approx \widehat{\mathcal{B}}_\tau\big(C_X(s)\big) = \frac{5\big(\hat{c}_X(0)\big)^2}{6\sum_{\tau=-(N-1)}^{N-1}(1-\tau/N)\big(\hat{c}_X(\tau)\big)^2} - \frac{1}{N} \qquad (9.95)$$

This restriction is the price paid for a procedure using only explicitly formulated assumptions and estimators. Thus, approximation (9.95) is considered to be objective, in contrast to approximations obtained using the subjective methods (9.93,2,3) discussed above. If, however, the number of peaks in the spectrum to be estimated is unknown, then an estimate (9.95) has to be compared with results obtained from other methods in (9.93). For example, bandwidth $\mathcal{B}_\triangle\big(C_X(s)\big) = 0.00474$ obtained in Fig. 7.10 (d) for the spectrum of the AR[4] model (X_t) defined in Fig. 7.14 is approximated as follows: (i) $\mathcal{B}_\tau\big(C_X(s)\big) = 0.015$ is computed using (9.94) from the theoretical covariance function in Fig. 9.25 (a), and (ii) $\widehat{\mathcal{B}}_\tau\big(C_X(s)\big) = 0.018$ is computed using (9.95) from the empirical covariance function in Fig. 9.25 (b). Both results clearly

overestimate the bandwidth of this spectrum, however, both are not far from 0.01364, i.e., the width at half height of the valley between the peaks in this spectrum as obtained in Fig. 7.10 (d). This overestimation results because approximations (9.94) and (9.95) only apply when the unknown spectrum has only one peak.

(9.93,5) recommends to (i) perform a parametric spectral estimation as introduced in Sect. 9.2.5, (ii) plot the empirical spectrum thus obtained, and (iii) apply definition (7.83,1) to arrive at the bandwidth of the parametric estimate which then (iv) is assumed to be an approximation for the bandwidth of the theoretical spectrum. In Fig. 9.11 (a), for example, the left peak of both estimates is narrower than the right one and, consequently, the widths at half height of the left peaks become the bandwidths of these empirical spectra: 0.005 is obtained for the Burg estimate and 0.007 for the maximum likelihood estimate. Both estimates are obtained from a time slice of length $N = 512$ in a realisation of the AR[4] model defined in Fig. 7.14, and both are, despite the relatively small N, useful approximations for the bandwidth of this model spectrum which is arrived at, in Fig. 7.10 (d), as width at half height of its right peak.

The procedure proposed in (9.93,6) can be applied when peaks and valleys can be clearly distinguished in a direct spectral estimator calculated from a time slice in a realisation of the process under analysis. In Fig. 9.31 (b) for example, 0.08 s^{-1} is obtained as approximation for the width at half height of the dominating (and narrowest) peak in the spectrum of the microseisms.

9.3.6 Summary and Alternatives

An approximately unbiased direct spectral estimator can be obtained subsequent to adequately tapering the observations, as summarised in Sect. 9.2.6. The variance of an estimator thus obtained is thereafter reduced by convolving with a smoothing kernel as proposed in Sects. 9.3.1 and 9.3.2.

The smoothing reduces the variance of the direct spectral estimator but can induce an additional bias in the smoothed direct spectral estimator, as is concluded from approximations (9.77) and (9.72), since both, bias and variance of the smoothed estimator, depend on the width of the smoothing kernel used. It is therefore important to apply diagnostic (6.115) in order to obtain a smoothing kernel whose width is reconcilable with the known bandwidth of the spectrum to be estimated, as proposed in the remarks to (9.92). If, however, the bandwidth of the spectrum to be estimated is not known, an approximation is substituted in (6.115) which can be obtained from a combination of the procedures proposed in (9.93).

In R, a discretely smoothed direct spectral estimator as defined in (9.28) is computed using `spec.univariate()` or `spec.pgram()`, both functions being discussed in Sect. 9.5.6. It is, however, also possible to compute a lag window estimator as proposed in the remarks to (9.44) by: (i) calculating an empirical covariance function from tapered observations using R expressions as in

Problem 9.5, (ii) multiplying the estimator thus obtained with a lag window having properties (9.44) (this is simply a multiplication of two R vectors), and (iii) using `fft()` in order to Fourier transform the product thus obtained: the result is the lag window spectral estimator as defined in (9.41).

As an alternative to smoothing a direct spectral estimator, a spectrum can be estimated by averaging direct spectral estimators calculated from segments in a time series, an idea first proposed by Bartlett in [9]. Bartlett splits the time series observed into N/N_S contiguous segments such that N_S observations are in each segment, and then computes the periodogram from each segment. He assumes that the N/N_S periodograms thus obtained are independent of each other despite the correlations in the stationary time series observed and, under this assumption, arrives at a mean periodogram. The mean periodogram thus computed has a variance that is approximately N/N_S smaller than the variances of the periodograms calculated from the segments. The estimator thus obtained is approximately a Bartlett lag window estimator (Fig. 9.18 shows the Bartlett lag window for $m = 30$).

Clearly, it is also possible to multiply the observations in each segment with a data taper and to average the direct spectral estimators computed from the tapered segments. In addition, the segments can be chosen to overlap thus producing an overall estimator having better properties of the second moment function than the one obtained from non-overlapped segments [144]. Weighted overlapped-segments spectral estimators can be very efficiently implemented and thus can be computed for extremely long time series ([29], [102]) and can also be used to compute robust spectral estimates [34].

9.4 Examples

In this section, the spectra of (i) the background noise in a seismogram ("microseisms") and (ii) a turbulent atmospheric flow are estimated. These spectra are known to be continuous, and further knowledge stemming from earlier observations and/or theory is available. The spectra to be estimated are, however, not completely known prior to the estimation and, therefore, the situation is more realistic than the one encountered in simulation experiments as performed in Sects. 9.2 and 9.3 for didactical reasons.

9.4.1 Microseisms

Microseisms are fine vibrations of the earth (i.e., the background noise) seen in seismograms, which have not been dominated by the much larger vibrations generated by a substantial earthquake having occurred possibly a large distance away from the recording station, e.g., on another continent. Earthquakes are not evenly distributed around the globe. They predominantly occur at the plate boundaries where oceanic and continental plates grind

against each other, triggered by their relative motions. Oceanic and continental plates are sections of the earth's crust which are transported at a speed of a few centimeters per year by convection in the mantle of the earth (the layer below its crust). The convection in the mantle is maintained by heat which is generated mainly from radioactive decay in the interior of the earth, as discussed in Sect. 1.1.

This process is called plate-tectonics. Plate-tectonics shapes the earth: (i) spreading of the crust generates oceanic ridges, (ii) island arcs associated with volcanism and deep trenches appear in the oceans when one oceanic plate moves beneath another one in the subduction zones, (iii) such island arcs can later become part of a continental margin (with an oceanic plate being subduced beneath a continental one), and (iv) mountain belts are created by the compression of two continental plates.

In regions with plate boundaries, faults appear. A fault is a fracture (or a variety of related fractures) in the rock, which allows for relative displacements of the adjacent sides. Small faults can be a few millimeters in length, large faults possibly reaching a length of thousands of kilometers. Often, faults are active over long (geologic) periods, involving repeated displacements. In a fault, shear stress is generated by plate-tectonic motions due to friction between both sides of the rock. If the shear stress becomes larger than the frictional forces, the rock on one side of the fault will then suddenly slip with respect to the other and thus release the stored stress energy. The segment of the fault where this abrupt movement occurs is known as the focus, the point on the surface above the focus is called the epicentre of the earthquake.

Most of the energy released in the focus dissipates in the form of heat, a small fraction, however, is emitted in the form of elastic waves in the rock which disperse in all directions (seismic waves). The rock normally damps all elastic waves, however, the lower frequency waves ($< 10\ \mathrm{s}^{-1} = 10\ \mathrm{seconds}^{-1}$) are less damped than those having higher frequencies. Consequently, low-frequency seismic waves generated by a substantial earthquake propagate over large distances around the earth, and, since they disperse in all directions, may deeply penetrate the earth on their journey towards, e.g., another continent.

The seismic waves emitted by a substantial earthquake will cause the ground to tremble so violently that, at small distances from the epicentre, buildings, bridges, dams, etc. can be damaged. At large and very large distances from the epicentre, e.g., on another continent, the seismic waves will cause fine vibrations of the ground which, although not being discernible to the human senses, can be measured using an appropriate device: a heavy weight is suspended such that it remains immobile due to its large mass when the ground vibrates. In this manner, the movements of the ground relative to the immobile weight are recorded. A record thus obtained is called a seismogram, and the device, a seismograph. A substantial earthquake generates

a seismogram that shows "footprints" of (i) the focal point, (ii) the structure of the rock in which the seismic waves have travelled from the focal point to the seismograph, (iii) the ground beneath the seismograph, and (iv) the seismograph itself.

The solid surface of the earth never stands completely still. Fine vibratory disturbances recorded by a seismograph can be due to either natural phenomena or to human activity. Ground vibrations stemming from human activities (e.g., traffic, construction works) are undesirable in a seismogram and, therefore, seismographs are mounted in remote areas, wherever practicable deep inside a natural cave or an abandoned mine.

If a seismograph is installed as required above, it then will record fine vibrations with periods between 4 s and 12 s, i.e., with frequencies between $0.08\ \mathrm{s}^{-1}$ and $0.25\ \mathrm{s}^{-1}$. These fine vibrations are the dominant natural noise in seismograms. They are aroused by ocean waves breaking along the coastline of a continent, thus generating elastic waves in the rock which propagate into the interior of the continent. These tiny vibrations are therefore called marine or sea microseisms. The amplitudes of marine microseisms depend strongly on the weather conditions prevailing at the coast and near ocean regions. For lower and higher frequencies than those pertaining to the marine microseisms, the natural noise in a seismogram is much smaller and often vanishes in relation to the noise stemming from human activities.

In Fig. 9.26, for example, a seismogram is plotted that was recorded in Nilore, Pakistan, (33.65 N, 73.26 E) on September 7, 1997, 10:15:45.0 (hours:minutes:seconds in UTC). This seismogram was generated by an earthquake having magnitude 5.6 that occurred at a remote place in Pakistan (30^0 N, 67.82 E), its focal point being at a depth of 33 km. The distance between the earthquake and the seismometer is approximately 650 km. A description of the Nilore seismometer is available from [49]. This seismogram is a plot of the speed of the earth in vertical direction (in ms^{-1} (meters/second)) against time (in s (seconds)), the sampling interval is $\Delta t = 0.025$ s. This seismogram is made available to the public by the IRIS data centre [72].

An earthquake generates both compression waves and shear waves when a large amount of energy stored in the rock under stress is abruptly released. In a compression wave (or P-wave), the particles in the rock vibrate forward and backward in the same direction the wave travels, in a shear wave (or S-wave), the particles vibrate perpendicularly or transversely to the direction of propagation. Compression waves travel swiftly through the rock at speeds of between 2 and 8 kms^{-1}, whereas shear waves are slower, attaining speeds ranging between 60% and 70% of the compression waves. Although speeds of seismic waves undergo substantial changes on their passage through the earth, the ratio between the average speeds of compression waves and shear waves remains approximately constant. Consequently, there is a time delay (the $S-P$ time) between the arrival of the compression waves and the arrival of the shear waves generated by a distant (thousands of km, often on another

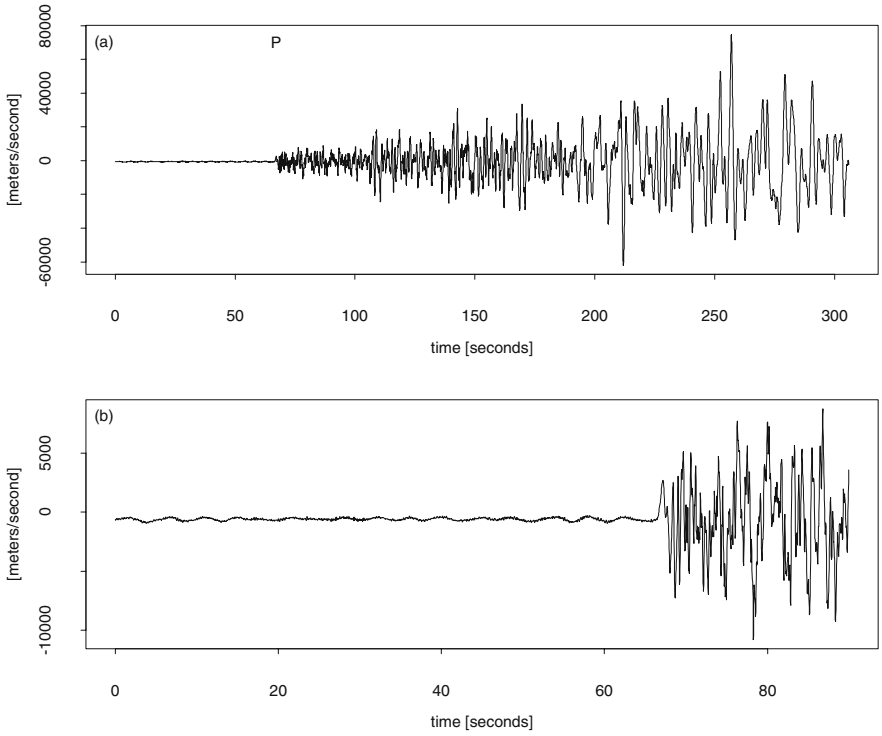


Fig. 9.26. Earthquake on September 7, 1997 in Pakistan, seismogram recorded at Nilore, Pakistan ($33^{6}5$ N, 73.26 E). The record begins 65 s before the waves generated by the earthquake arrive at the seismograph (point P in plot a).

continent) earthquake. From those $S - P$ times of an earthquake observed in seismograms recorded at a number of stations, the coordinates of its epicentre and the depth of its focal point can be calculated. This is the principal purpose of seismograph networks such as those reporting to the IRIS data centre.

With an additional delay (i.e., later than the arrival of the S-waves), those seismic waves that propagate in the earth's crust (called surface waves) are recorded in a seismogram, their route to the seismograph being longer than the one of waves travelling through the earth's interior.

In Fig. 9.26 (a) for example, the compression waves arrive at time point P at the Nilore seismograph, the S-waves at a later but unknown time point (in a seismogram, it is difficult to read the arrival of the S-waves stemming from an earthquake occurring a small distance away). The time slice from 0 through to 65 s in the seismogram shown in Fig. 9.26 is plotted in Fig. 9.27. Immediately following this period, the compression waves of the earthquake arrive at the seismograph inducing the vibrations of the ground to increase by an order of magnitude, as is obvious from a comparison with plot (b) in Fig. 9.26.

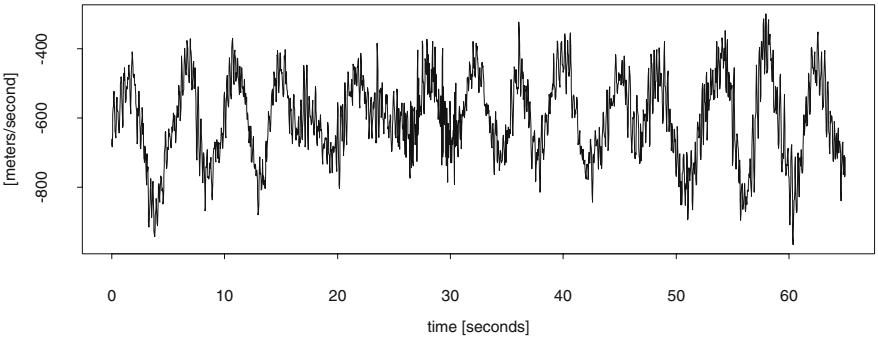


Fig. 9.27. Microseisms: time slice from 0 through to 65 seconds in the seismogram plotted in Fig. 9.26.

In Fig. 9.27, the recorded natural seismic noise becomes clearly visible: the ground vibrates with approximately 12 or 13 oscillations per minute which corresponds to a period smaller than but very close to 5 s, yet falling well within the range of the periods of marine microseisms as introduced above. Is it possible to calculate a better estimate for the period of these marine miscroseisms?

It is assumed that the microseisms in Fig. 9.27 are a time slice from a realisation of a discrete-time stationary stochastic process having a spectrum as asserted by the spectral representation theorem (7.51). Further, it is assumed that the microseisms are reconcilable with model (9.1,2), i.e., have a continuous spectrum due to the mechanical properties of the rocks through which the seismic waves have travelled.

Following the recommendations in the remarks to Figs. 9.3 and 9.4, periodograms and direct spectral estimators of the microseisms are calculated using R function `spec.univariate()` introduced in Sect. 9.5.6

```
earthquake <- ts(scan("/path/earthquake.dat"),
                                start=0,frequency=40,)
microseisms <- window(earthquake,0,65)
microseismsdem <- microseisms - mean(microseisms)
ms.spec <- spec.univariate(microseismsdem,taper=0,  #periodogram in
            spans=1,pad=1,confint=0.95,fast=F)      #Fig. 9.28 (a)
plot(ms.spec$freq,10*log10(ms.spec$spec),type="n",
                                xlab="frequency [1/s]",ylab="dB")
points(ms.spec$freq,10*log10(ms.spec$spec),pch=".")
```

and thereafter plotted in Figs. 9.28 and 9.29. For frequencies higher than 17.5 s^{-1}, both periodograms in Fig. 9.28 obviously no longer fluctuate locally within a confidence interval of constant height, as required (in a logarithmic plot) by (9.19). This points to a possible bias due to the side lobes of the Fejer kernel in convolution (8.21). This bias is reduced such that it becomes

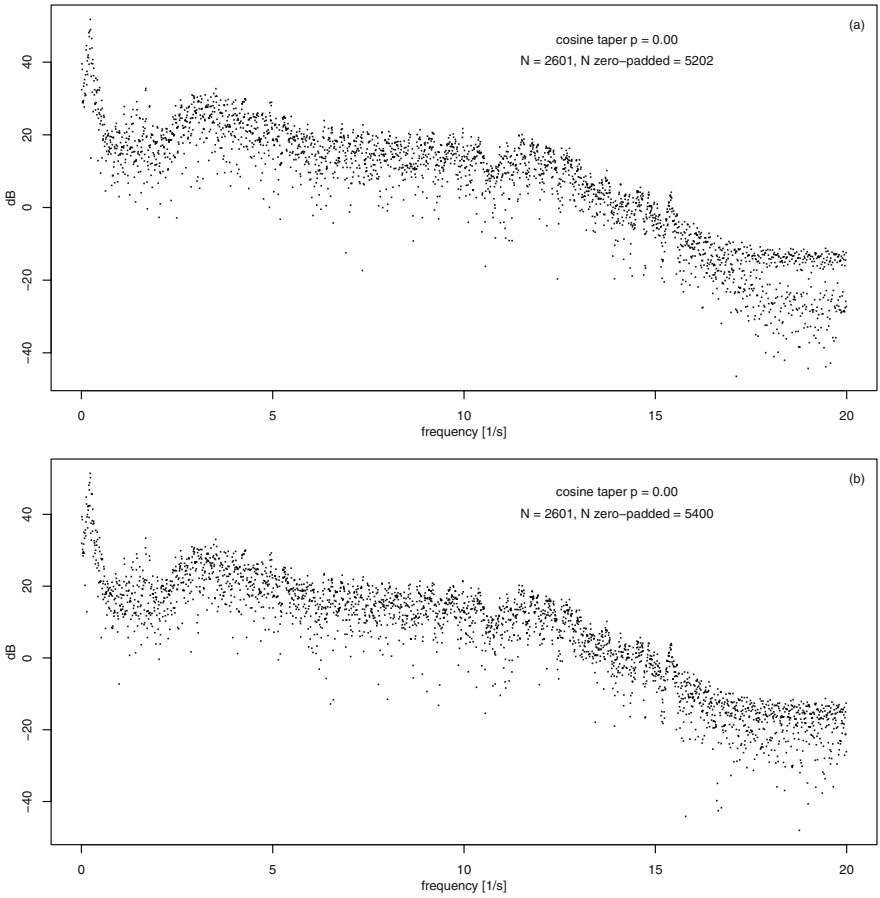


Fig. 9.28. Periodograms of the microseisms as plotted in Fig. 9.27. The $N = 2601$ observations are zero-padded to arrive at $N' = 5202$ values (above, plot a) and $N' = 5400$ values (below, plot b).

negligible once a spectral window having smaller side lobes than those of a Fejer kernel is convolved with the spectrum in (9.12), as is demonstrated in Fig. 9.29. There, direct spectral estimators calculated using cosine tapers for $p = 0.05$ and $p = 0.10$ become stable at approximately -20 dB for the high frequencies. It is therefore assumed, in agreement with the recommendations in the remarks to Figs. 9.3 and 9.4 in Sect. 9.2.1 as well as to (6.117) in Sect. 6.7.2, that the bias due to the side lobes of the spectral window becomes negligible when a cosine taper for $p = 0.05$ is applied.

The direct spectral estimators in Fig. 9.29 decrease from approximately 20 dB for frequency 10 s^{-1} to approximately -20 dB for frequencies higher than 17.5 s^{-1}. This large decrease (approximately four orders of magnitude)

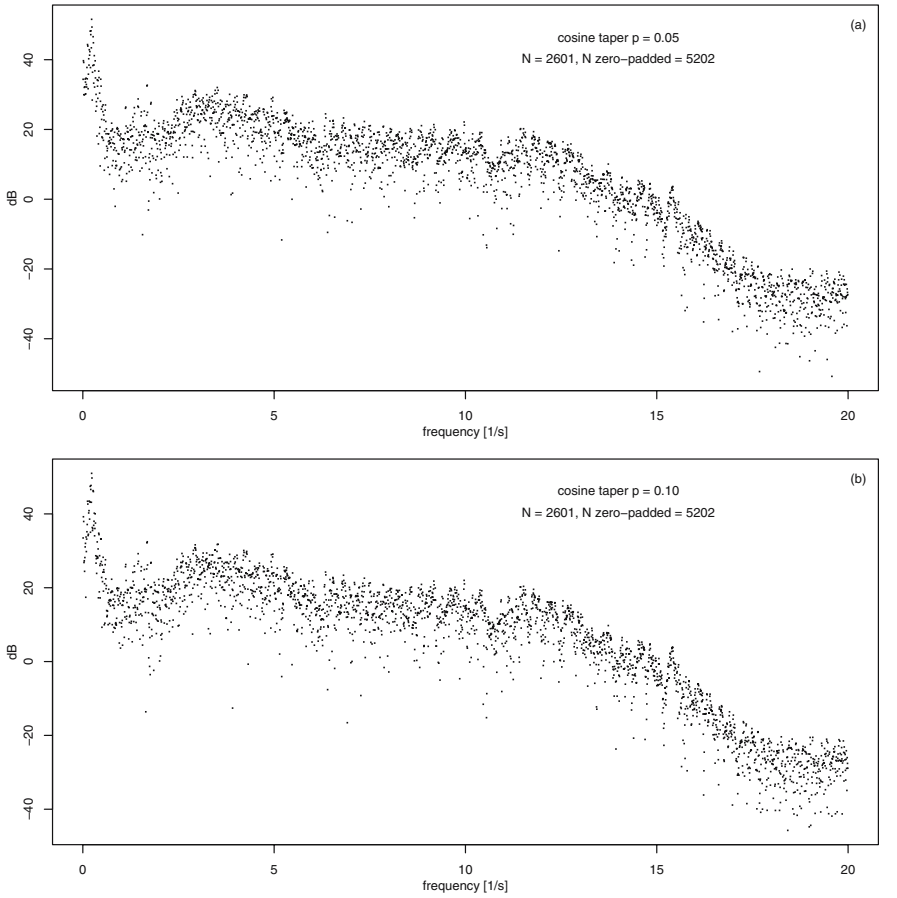


Fig. 9.29. Direct spectral estimators of the microseisms as plotted in Fig. 9.27. The $N = 2601$ observations are multiplied with variance-normalised cosine tapers for $p = 0.05$ (above, plot a) and $p = 0.10$ (below, plot b) and thereafter zero-padded to arrive at $N' = 5202$ values in both cases.

is due to the absorption of vibrations in the rock which increases substantially for those having frequencies higher than 10 s^{-1}. Thus the Fourier transform of a seismogram becomes small for frequencies higher than 17.5 $s^{-1} < 20$ s^{-1}, and an aliasing can be neglected. Obviously, the sampling interval $\Delta t = 0.025$ s with $1/(2\Delta t) = 20$ s^{-1}, used for recording the seismogram in Fig. 9.26 at Nilore station, was chosen in agreement with the recommendations in the remarks to the definition of the Nyquist frequency in (6.92).

Since the rock absorbs seismic waves with increasing degree for frequencies higher than 10 s^{-1}, seismograms are usually analysed for frequencies lower than 10 s^{-1}. For this band, the direct spectral estimator as in Fig. 9.29 (a)

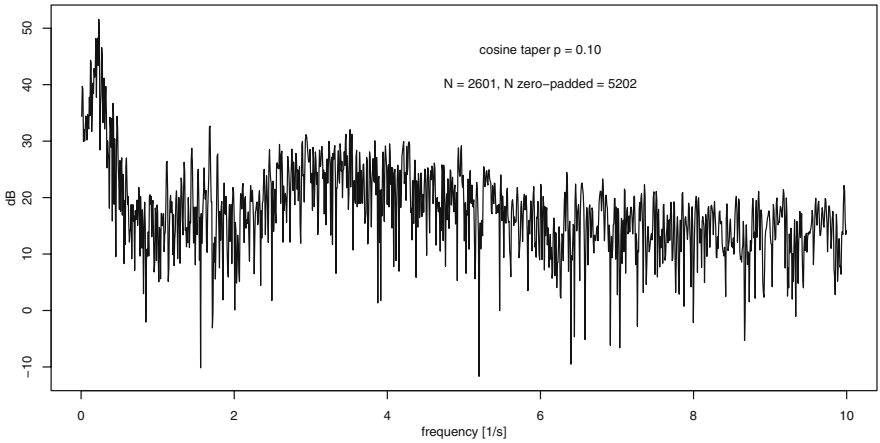


Fig. 9.30. Spectrum of microseisms: direct spectral estimator as in Fig. 9.29 (a) for band $0 \leq s \leq 10\ \mathrm{s}^{-1}$.

is plotted in Fig. 9.30. In this plot, a narrow peak approximately 50 dB high for frequencies between 0.2 and 0.25 s^{-1}, and also a broad peak attaining 30 dB for 3.0 s^{-1} and then slowly decreasing to 20 dB for 6.0 s^{-1}, become clearly visible. Less clearly visible is a third, relatively flat peak for frequencies between 1.0 and 2.0 s^{-1}.

The dominant oscillations visible in the microseisms as in Fig. 9.26 have periods between 4 and 5 s, i.e., frequencies between 0.2 and 0.25 s^{-1}, and therefore are likely to produce the narrow peak which is prominent in Figs. 9.29 and 9.30. This motivates to take a closer look at the direct estimator for the spectrum of the microseism in the band with the peak. Hence, this estimator is re-plotted with higher resolution for frequencies between 0 and 1.5 s^{-1} in Fig. 9.31, subsequent to taking its logarithms and also prior to the logarithmic transformation. In the non-logarithmic plot (b), all details of the estimator are hidden, except the dominating peak between frequencies 0.0 and 0.5 s^{-1} which remains visible. This peak is approximately 50000 $(\mathrm{m/s})^2$ high, an approximation arrived at by accounting for the extreme values due to the long-tailed chi-square distribution (9.19) of the direct-spectral estimator. Further, it is assumed that this peak arises from a base height of 0 $(\mathrm{m/s})^2$, and consequently, at half of its height, it is $10\Delta s'_k(1/\Delta t) = 0.0768935 \approx 0.08\ \mathrm{s}^{-1}$ wide, calculated using $\Delta s'_k = 1/N' = 0.0001922$ with $N' = 2N$ and $N = 2601$ as well as $\Delta t = 0.025$. This is an approximation in agreement with (9.93,6) for the bandwidth (7.83,1) of the spectrum pertaining to the microseisms in Fig. 9.27, which is assumed to be continuous above, in the remarks to this figure.

Given this approximation for the bandwidth of the spectrum to be estimated, the remark in the last paragraph of Sect. 9.3.3 recommends to smooth

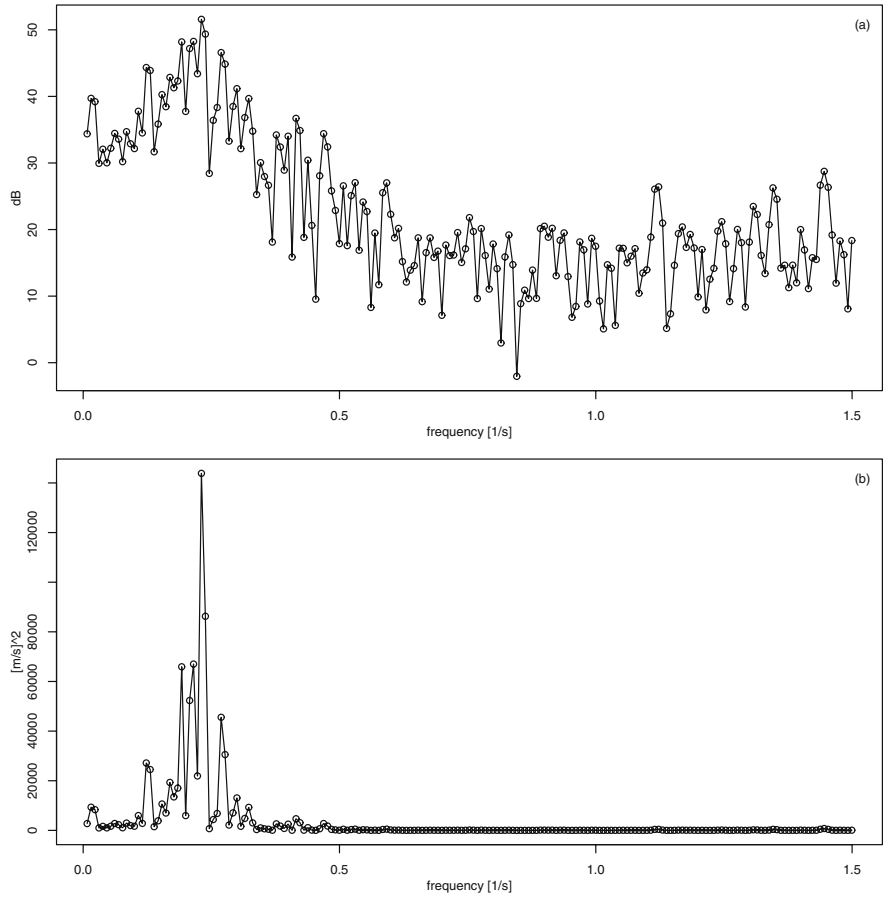


Fig. 9.31. Spectrum of microseisms: direct spectral estimator as in Fig. 9.29 (a) for band $0 \le s \le 1.5\ \mathrm{s}^{-1}$ on logarithmic (a) and non-logarithmic (b) scales.

the direct spectral estimator with a kernel having a σ-width $\mathcal{B}_\sigma\big(V_m^{\vdash\tau\dashv}(s)\big)$ and also an autocorrelation width $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$ not larger than $0.04\ \mathrm{s}^{-1}$. These widths are defined in (9.128) and in (9.129) and, in case of smoothing with a modified discrete Daniell kernel, $\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) \approx \mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) \approx \mathcal{B}_{||}\big(V_{m,M}(s_k)\big) = 2M\Delta s'_k$ as is concluded from the results obtained in Sect. 9.5.4, with $\mathcal{B}_{||}\big(V_m(s_k)\big)$ as defined in (9.125). Thus, when smoothing with a modified discrete Daniell kernel as defined in (9.29), it is recommended to select $M = 2$ or $M = 3$, resulting in bandwidths (9.90) of $0.031\ \mathrm{s}^{-1}$ or $0.046\ \mathrm{s}^{-1}$, bearing in mind that $\mathcal{B}_{||}\big(V_{m,M}(s_k)\big)$ in (9.90) and (9.125) is multiplied with $(1/\Delta t)$ in case of $\Delta t \neq 1$. These are the bandwidths pertaining to the smoothed direct spectral estimators plotted in Fig. 9.32.

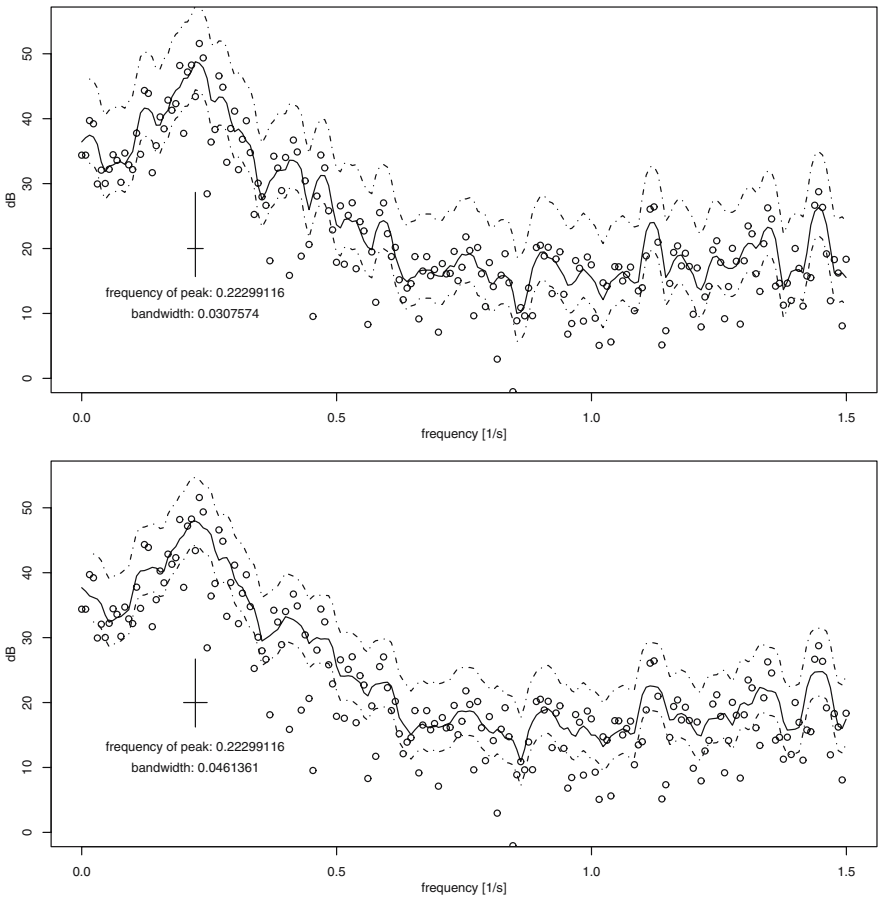


Fig. 9.32. Spectrum of microseisms: a direct spectral estimator as in Fig. 9.31 (a) (here plotted using symbol ∘), calculated for discrete frequencies in distance $\Delta s'_k(1/\Delta t) = 0.00768935\text{s}^{-1}$ wide, $\Delta s'_k = 1/N'$ with $N' = 2N$ and $N = 2601$ as well as $\Delta t = 0.025$, is smoothed over five (above, a) and seven discrete frequencies (below, b), in both cases using a modified discrete Daniell kernel (9.29) with parameter $M = 2$, $M = 3$ resp.

The smoothed estimators in Fig. 9.32 (a) and (b) remain safely in the region described by the scattering points of the direct spectral estimator. This behaviour suggests that the bias of both estimators is not large, even in the peak region where $|C_X{}''(s)|$ attains its maximum, as implied by (9.72) ($C_X{}''(s)$ is the second derivative of the spectrum to be estimated). Hence, both smoothing kernels applied to arrive at the estimators in Fig. 9.32 are not too wide. The bandwidths of the estimators as required in (9.90) are represented by the horizontal bar in the cross plotted underneath the peak

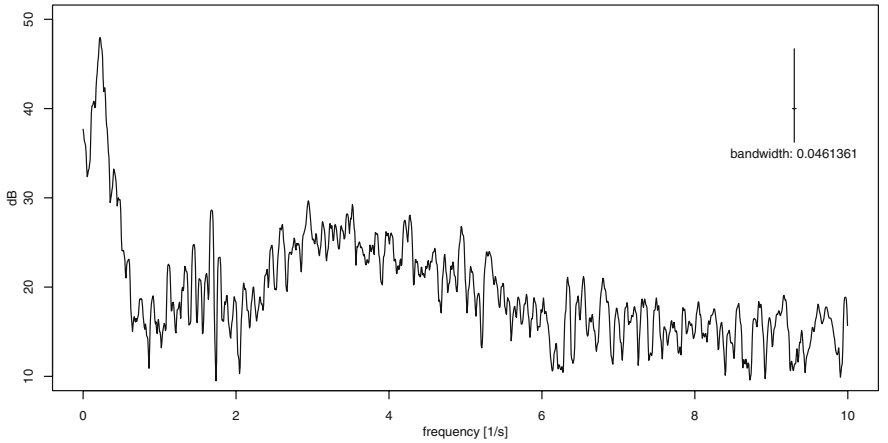


Fig. 9.33. Smoothed direct spectral estimator as in Fig. 9.32 (b) for frequencies between 0.0 and 10s^{-1}.

in each estimator. The vertical bar in each cross is plotted for the frequency where the smoothed estimators attain their maxima. This frequency is identical in both plots, namely 0.22299116 s^{-1}.

The same peak frequency is obtained when the direct spectral estimator is smoothed using a modified discrete Daniell kernel for $M = 4$ (not shown in Fig. 9.32). With increasing bandwidth, however, the estimates for the peaks decrease in their heights, as demonstrated in Fig. 9.32 for kernels with $M = 2$ and $M = 3$, in agreement with the smoothing properties of a moving average as demonstrated in Figs. 2.16 and 7.8. Consequently, the oscillations in the microseisms in Fig. 9.27 increase in their amplitudes for frequencies between 0.20 and 0.25 s^{-1} where they attain their maximal power for approximately frequency 0.22299116 s^{-1}. To these frequencies correspond periods between 4 and 5 s, obviously the periods of the oscillations dominating the microseisms.

Confidence intervals at the 0.95 level are plotted 4.32 dB (3.76 dB, resp.) below and 8.64 dB (6.70 dB, resp.) above the smoothed direct spectral estimators in Fig. 9.32 (a) and (b). The constant (in a logarithmic plot, reconcilable with (9.88)) heights of these confidence intervals is given by the vertical bar in the cross plotted beneath the peak in each estimator.

The estimator obtained using a modified discrete Daniell kernel for $M = 3$, as shown in Fig. 9.32 (b), is also plotted for frequencies between 0.0 and 10.0 s^{-1} in Fig. 9.33, together with the cross, yet without confidence intervals. When this cross is plotted, using a ruler and pencil, both in the primary peak and the highest secondary peak (approximately frequency 3.0 s^{-1}), it becomes obvious that, in a small band about the frequency of the main peak, the spectrum of the microseisms is approximately 20 dB larger than for the other frequencies. From this difference of about two orders of magnitude, it is concluded that the seismic noise in Fig. 9.27 is dominated by oscillations

with frequencies between 0.2 and 0.25 s^{-1}, i.e., by frequencies which are reconcilable with those of marine microseisms introduced in the remarks to Figs. 9.26 and 9.27.

Marine microseisms having frequencies between 0.1 and 1.0 s^{-1} can be seen in every seismogram, on condition that no seismic waves generated from an earthquake are recorded by the seismograph. The amplitudes of marine microseisms increase with increasing height of the ocean waves and decrease as the waves become weaker [23]. Marine microseisms observed in northern and central Europe often have periods of approximately 7 s. Their amplitudes are related to the weather conditions in the North Atlantic via the following (simplified) mechanism: (i) high wind speeds associated with a storm system generate ocean waves with large heights, (ii) when these waves arrive at the coast of Norway, the large-scale pressure variations in the ocean increase in their amplitudes and thus (iii) generate microseisms with larger amplitudes ([47], [48]).

9.4.2 Turbulent Atmospheric Flow

A *turbulent flow in a liquid or gas* has the following properties:

1. A particle moves irregularly in three dimensions on eddy-like trajectories which are changing constantly.
2. Mass, momentum and heat are mixed very efficiently by the eddies emerging and vanishing. For example, the smoke blown out of a chimney is mixed with the ambient air and thus is no longer visible once it has been transported over a short distance.
3. The kinetic energy of the flow is constantly changed into internal energy or heat due to the viscosity (the internal friction) of the liquid or gas, which requires that energy be permanently fed into the flow. If the energy input stops then the turbulent motion in the flow vanishes after a delay corresponding to the mean lifetime of the eddies.
4. The energy is transferred from the larger to the smaller eddies. The largest ones are loaded with energy from the mean kinetic energy of the flow, in the smallest, energy is dissipated into heat or internal energy. The smallest eddies are far smaller (by many orders of magnitude) than the largest and, between these extremes, there are eddies of any size: a cascade of eddies can be observed in a turbulent flow.
5. Simple examples of turbulent flows can be simulated using a super computer [96].
6. Turbulent flows in the atmospheric boundary layers are described statistically in most cases.

A turbulent flow is often described using random functions having a four-dimensional parameter (e.g., [51], [93]). When, for example, a *turbulent flow in the atmospheric boundary layer* is analysed, a coordinate system is chosen such that (i) the positive direction of the first coordinate x_1 is identical with

the direction of the mean velocity of the flow, i.e., the wind direction, (ii) the positive direction of x_3 is the positive vertical direction, and (iii) x_2 is perpendicular to x_1 and x_3. For example, $\boldsymbol{U}(\boldsymbol{x}, t)$ is written for the velocity of the turbulent flow in these three directions in a given region and period, slightly different from the symbols used for a multivariate stochastic process with a multi-dimensional parameter in the remarks to definition (2.2). Using this definition, $U_1(x_1, t)$ becomes the velocity in wind direction as function of x_1 and t, $U_3(x_1, t)$ the velocity in vertical direction as function of x_1 and t, $U_3(x_3, t)$ the velocity in vertical direction as function of x_3 and t, etc.

If time is kept constant, then three spatial coordinates remain: this is the system used to formulate hypotheses on a turbulent flow in the atmospheric boundary layer, and, in this system, the velocity $\boldsymbol{U}(\boldsymbol{x}, t)$ becomes $\boldsymbol{U}(\boldsymbol{x})$, Often, a turbulent flow is in agreement with the assumption that $\boldsymbol{U}(\boldsymbol{x})$ is stationary. Under this assumption, the covariance function $c_{\boldsymbol{U}}(\boldsymbol{h})$ is a function of the three-dimensional displacement $\boldsymbol{h} = \boldsymbol{x} - \boldsymbol{y}$ solely. For example, the covariance of the vertical velocity at two points having coordinates x_1 and y_1 for the first dimension (i.e., in the direction of the wind) depends exclusively on displacement $h_1 = x_1 - y_1$ on condition that the flow is stationary, as formulated in (9.96).

$$c_{U_3}(x_1, y_1) = c_{U_3}(x_1 - y_1, 0) = c_{U_3}(h_1) \tag{9.96}$$

$$c_{U_3}(h_1) = \int_{-\infty}^{\infty} C_{U_3}(k_1) \mathrm{e}^{\mathrm{i}h_1 k_1} \mathrm{d}k_1 \tag{9.97}$$

$$C_{U_3}(k_1) = \frac{1}{2\pi} \int_{-\infty}^{\infty} c_{U_3}(h_1) \mathrm{e}^{-\mathrm{i}k_1 h_1} \mathrm{d}h_1 \tag{9.98}$$

A three-dimensional velocity $\boldsymbol{U}(\boldsymbol{x})$ has a three-dimensional spectrum $C_{\boldsymbol{U}}(\boldsymbol{k})$, $\boldsymbol{k}$ being the three-dimensional wave number corresponding to $\boldsymbol{x}$, on condition that $\boldsymbol{U}(\boldsymbol{x})$ has the properties required in (7.51). In (9.97) for example, covariance function $c_{U_3}(h_1)$ as in (9.96) and $C_{U_3}(k_1)$, i.e., the spectrum of the vertical velocity in the turbulent flow, are a Fourier transform pair, on condition that $C_{U_3}(k_1)$ is continuous. k_1 is the wave number associated with coordinate x_1. Distinct from the Fourier transforms in (7.62) and (7.63) (there, the covariance function and the continuous spectrum are a Fourier transform pair as defined in (6.35)), the Fourier transforms in (9.97) and (9.98) do not have the factor 2π in their exponential functions and, therefore, a normalisation with this factor is required in one of the transforms [18]. In the case of the stationary turbulent flow, the $-$i-transform in (9.98) is usually normalised.

The properties of a stationary turbulent flow are often described using coordinate system (k_1, k_2, k_3), i.e., the coordinate system of the wave numbers. Numerous experiments have shown that three intervals can be distinguished in plots of velocity spectra obtained from stationary turbulent atmospheric flows. In the first interval, $k_1 \leq p_1$ (written for the first dimension) are the wave numbers of the large eddies, which are fed with turbulent kinetic energy stemming from the kinetic energy of the mean air flow. In the second interval

having wave numbers $p_1 \leq k_1 \leq q_1$, the turbulent energy remains constant but is shed from larger to smaller eddies thus producing an energy cascade. In the third interval with $k_1 \geq q_1$, the eddies become small enough such that viscous forces are no more negligible: turbulent kinetic energy is dissipated and becomes heat. In the second interval $p_1 \leq k_1 \leq q_1$, other forces than inertial ones are negligibly small, and therefore, this interval is called *inertial subrange*. For example, lower bound p_1 confining the *inertial subrange* is plotted in Fig. 9.36. In this example, $p_1 \approx 4.5\ \mathrm{m}^{-1}$, and the inertial subrange stretches as far as approximately $q_1 \approx 20.0\ \mathrm{m}^{-1}$ (not plotted).

This partition of the turbulent velocity spectrum is due to Kolmogorov who in 1941 proposed two hypotheses for locally isotropic turbulent flows having large Reynolds numbers [97]. A turbulent flow is locally isotropic on condition that $\boldsymbol{U}(\boldsymbol{x})$ remains invariant for small $\boldsymbol{h}$ under certain (rotational and axial) symmetry conditions. Reynolds number is the ratio of the inertial forces and the viscosity in a flow and thus is proportional to the typical geometry of a flow (diameter of a tube, length of an aerofoil, i.e., wing of a bird or aeroplane, distance in direction of the mean flow between two obstacles in a plane, etc.) but inversely proportional to the inner frictional forces of the liquid or the gas.

Kolmogorov's hypotheses imply that, in the inertial subrange, the spectrum $C_{\boldsymbol{U}}(\boldsymbol{k})$ is proportional to $\epsilon^{2/3}\boldsymbol{k}^{-5/3}$, ϵ being the rate of dissipation of the turbulent kinetic energy, i.e., the energy transported from large to small eddies. Consequently, if a turbulent flow is locally isotropic and has a large Reynolds number, then

$$C_{U_1}(k_1) = \alpha_1 \epsilon^{2/3} {k_1}^{-5/3} \qquad \text{for} \quad p_1 \leq k_1 \leq q_1 \tag{9.99}$$

is obtained for its velocity spectrum in direction of the mean flow, with $p_1 \leq k_1 \leq q_1$ being its inertial subrange and $0.5 \leq \alpha_1 \leq 0.6$ Kolmogorov's constant ($\boldsymbol{\alpha}$ has no dimension). If a spectrum similar to the one in (9.99) is obtained from wind speed measurements in a turbulent atmospheric flow then this similarity is a point in favour of a flow obeying Kolmogorov's hypotheses, i.e., a flow having locally isotropic eddies with wave numbers $p_1 \leq k_1 \leq q_1$ such that the energy remains constant in this cascade of eddies. A second point is a spectrum with $C_{U_3}(k_1) = C_{U_2}(k_1) = (4/3)C_{U_1}(k_1)$ in the inertial subrange [97].

Estimators for the covariance function $c_{\boldsymbol{U}}(\boldsymbol{h})$ and the spectrum $C_{\boldsymbol{U}}(\boldsymbol{k})$ can be obtained from measurements of $\boldsymbol{U}(\boldsymbol{x}, t)$. The following (and similar experiments [67]) can only be perfomed under favourable circumstances; these experiments have, however, implications concerning the usual practical wind speed measurements. In this experimental design, N instruments measuring the wind speed in three dimensions, e.g., sonic anemometers, are mounted on axis x_1 in the range $a_1 \leq x_1 \leq b_1$ with identical distances Δx_1. Then, for each time point t, realisations $u_d(x_1, t)$ of random variables $U_d(x_1, t)$, $d = 1, 2, 3$, are obtained for N locations between a_1 and b_1. If all exterior conditions

determining the turbulent flow (mean advection of energy which depends on, e.g., the mean wind speed as well as mean temperature and humidity, the geometry and the roughness of the land surface, etc.) remain constant in time, then the flow is assumed to be stationary in time and its moment functions $\mathrm{E}U_d(x_1,t)$, $\mathrm{Var}U_d(x_1,t)$ and $\mathrm{Cov}\big(U_d(x_1,t),U_d(y_1,t)\big)$, $d=1,2,3$, can be estimated by averaging the measurements over all time points t in the observational period. Using the empirical moment functions thus computed, an answer to the question can be obtained, whether the random functions $U_d(x_1,t)$, $d=1,2,3$, are also stationary along axis x_1. If the flow is thus found to be stationary in time and in direction of x_1, then

1. the covariance function and spectrum as in (9.96), (9.97) and (9.98) do exist and
2. $\hat{c}_{U_d}(x_1,y_1)=c_{U_d}(h_1)$, $d=1,2,3$, i.e., the empirical covariances calculated as time averages, are an estimate for $c_{U_d}(h_1)$, which
3. allows for calculating lag window spectral estimators (9.41) for $C_{U_d}(k_1)$, $d=1,2,3$.

This experimental design remains hypothetical for most practical applications, since (i) it is too expensive to implement and (ii) the anemometers mounted on their supporting structures increase the roughness of the surface, and thus can generate non-negligible additional turbulence.

However, the velocities of a turbulent flow in the atmospheric boundary layer can be measured without disturbing the flow noticeably on condition that only one instrument is mounted to perform measurements for only one location: a three-dimensional time series $\boldsymbol{u}(t)$ is obtained with $u_1(t)$ being the velocity in direction of the mean flow, $u_3(t)$ the one in vertical direction and $u_2(t)$ the one in a direction perpendicular to the directions of $u_1(t)$ and $u_2(t)$. From $\boldsymbol{u}(t)$, $\boldsymbol{u}(\boldsymbol{x},t)$ can be obtained on condition that the observed flow satisfies Taylor's hypothesis [111], which is defined in the following paragraph.

Assume a coordinate system that travels in the mean direction of the turbulent flow with a speed that is identical with the mean velocity of the flow in this direction, i.e., those of axis x_1. Then, for exactly one time point the origin of this moving coordinate system is identical with the location of the instrument that measures $\boldsymbol{u}(t)$ for one location on x_1. *Taylor's hypothesis* then assumes that, in this moving coordinate system, the three-dimensional velocities of the turbulent flow remain constant during a period being identical with the mean lifetime of the largest eddies. Thus, Taylor's hypothesis is sometimes called hypothesis of "frozen turbulence". Under Taylor's hypothesis,

$$x_1=\mu_{U_1}t \qquad k_1=(2\pi s)/\mu_{U_1} \qquad k_1C_{U_1}(k_1)=sC_{U_1}(s) \tag{9.100}$$

is arrived at, with μ_{U_1} being the mean wind speed of the flow, assumed to be constant during the mean life time of the largest eddies. Clearly, a turbulent flow is never "frozen" because it is subject to permanent change: in

the restless flow, new eddies are generated and existing ones disappear. Often however, turbulence develops slowly and its velocities undergo only minor changes during the lifetime of the largest eddies and thus, under Taylors's hypothesis, $\boldsymbol{u}(t)$ is an approximation for $\boldsymbol{u}(\boldsymbol{x}, t)$.

Assuming stationarity of the turbulent flow in time, an estimate for $\mu_{U_1}(t)$ can be calculated as time average of observations $u_1(t)$: in (9.101), $u_1(t)$ is convolved with the weights $g(t)$. $\hat{\mu}_{U_1}(t)$ thus obtained is an estimator for μ_{U_1} having properties as required in (2.56). Using this estimate, the observations $u_1(t)$ are written in (9.102) as sum of their mean $\hat{\mu}_{U_1}(t)$ and the deviations from the mean $u_1'(t)$. $u_1'(t)$ are called turbulent fluctuations and the sum in (9.102) is called *Reynold's decomposition.*

$$\hat{\mu}_{U_1}(t) = \int_{-\infty}^{\infty} u_1(t-\tau) g(\tau) \mathrm{d}\tau \tag{9.101}$$

$$u_1(t) = \hat{\mu}_{U_1}(t) + u_1'(t) \quad \text{or} \quad u_1(t) = \bar{u}_1(t) + u_1'(t) \tag{9.102}$$

The width of the moving time slice over which $u_1(t)$ is averaged in (9.101), is chosen (i) large enough to include the average lifetime of the largest eddies in the turbulent flow, and (ii) also small enough such that $\mu_{U_1}(t)$ remains constant in this period, as required in (2.8,1). In the period thus obtained, random function $U_1(t)$ is stationary in its first moment function, and, if $U_1(t)$ is also stationary in its second moment function, then (i) its variance function $\mathrm{Var}\big(U_1(t)\big)$ is constant and (ii) the empirical variance of the turbulent fluctuations is an estimate for the variance of both $U_1'(t)$ and $U_1(t)$.

In addition, it is assumed that $U_1'(t)$ is stationary with an expectation function being identically zero in the interval confined by time points t_1 and t_2, with $t_2 - t_1$ being much larger than the width of the moving average in (9.101). Under this assumption, $\hat{\sigma}^2_{U_1'} = (1/(t_2 - t_1)) \int_{t_1}^{t_2} (u_1'(t))^2 \mathrm{d}t$ becomes an estimate for the variance $\sigma^2_{U_1'}$ of $U_1'(t)$ in interval $t_1 \le t \le t_2$. In this interval, $u_1'(t)$ is a realisation of $U_1'(t)$ which fluctuates about the constant or only slowly changing (i.e., being constant in the period used for averaging in (9.101)) expectation function $\mu_{U_1}(t)$ of $U_1(t)$.

When a turbulent flow in the atmospheric boundary layer is measured and, thereafter, a Reynold's decomposition is performed, then $\boldsymbol{U}(t)$ is assumed to be stationary in periods of a few minutes' duration, because a few minutes is the average lifetime of eddies in atmospheric turbulent flows. Consequently, the moving average in (9.101) is a few minutes wide, whereas the turbulent fluctuations $\boldsymbol{U}'(t)$ are assumed to be stationary over much longer periods, usually 30 or 60 minutes. Using the realisation $\boldsymbol{u}'(t)$ of $\boldsymbol{U}'(t)$ obtained under these assumptions, both covariance function $\boldsymbol{c}'_{\boldsymbol{U}}(t)$ and spectrum $\boldsymbol{C}'_{\boldsymbol{U}}(s)$ can be estimated.

In Fig. 9.34 for example, a time series $u_3(t)$ of vertical velocities in a turbulent atmospheric flow is decomposed as required in (9.101) and (9.102) into (i) its moving average $\bar{u}_3(t)$ having width 5 minutes and (ii) its fluctuations $u_3'(t)$ about $\bar{u}_3(t)$. $u_3(t)$ was measured using a sonic anemometer near

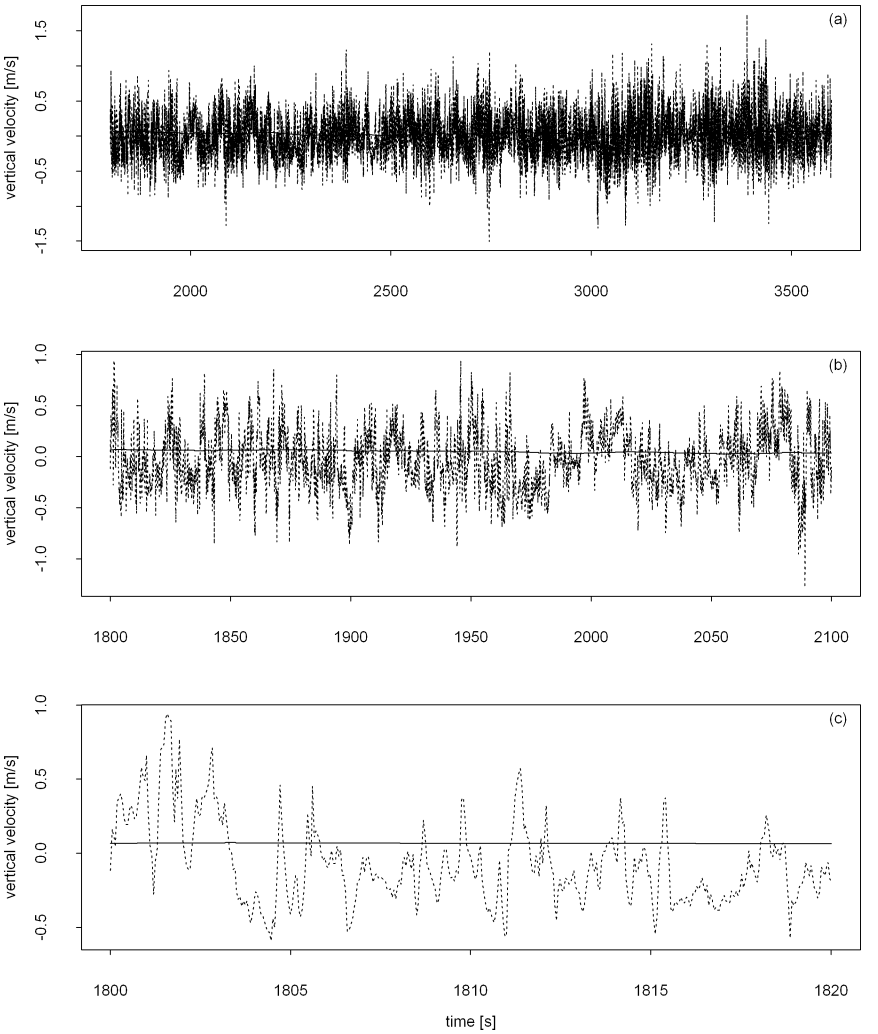


Fig. 9.34. Reynold's decomposition (9.102) of anemometer measurements in a turbulent atmospheric flow: vertical velocity $u_3(t)$ (ms^{-1}) and its 5 minutes' averages for periods having durations of 30 minutes (a), 5 minutes (b) and 20 seconds (c).

San Vittore (Tessin, Switzerland, $46^0 14' 24.6''$ N, $9^0 06' 0.6''$ E) on September 13, 1999 from 12:00:00 through to 14:00:00 UTC (hours:minutes:seconds). The sampling interval was $\Delta t = 48 \times 10^{-3}$ s. The wind speed measurements in San Vittore were performed in order to study the turbulence structure in an Alpine valley ([37], [120]). The results of the Reynold's decomposition

of the San Vittore three-dimensional velocities are made available as files `/path/turwind2.dat` and `/path/turwind3.dat`.

The three-dimensional spatial velocity $\boldsymbol{U}(\boldsymbol{x})$ of a turbulent flow as introduced in the remarks to (9.96) has a continuous three-dimensional spatial spectrum $C_{\boldsymbol{U}}(\boldsymbol{k})$ as defined in the remarks to (9.98). If the spatial spectrum $C_{\boldsymbol{U}}(\boldsymbol{k})$ is continuous then a temporal spectrum $C_{\boldsymbol{U}}(s)$ is also continuous. This is an implication of Taylor's hypothesis (9.100). Consequently, a three-dimensional spectrum $C_{\boldsymbol{U}}(s)$ pertaining to a three-dimensional velocity $\boldsymbol{U}(t)$ measured for one location on x_1 in a turbulent flow is reconcilable with model (9.1,1). In the atmospheric boundary layer, a locally isotropic turbulent flow has a velocity spectrum (9.99) in direction of the mean flow with inertial subrange $p_1 \leq k_1 \leq q_1$, where $q_1 \approx 36$ m^{-1}. Under Taylor's hypothesis, this upper bound is in agreement with the Nyquist frequency (6.92) $s_1^{(Ny)} \approx 10$ s^{-1} of the San Vittore measurements as plotted in (9.34), since $2\pi/1.7654$ [s^{-1}/(ms^{-1})] $= 35.59$ m$^{-1} = k_1^{(Ny)}$, with $10\ s^{-1} \approx 1/(2\Delta t)$, $\Delta t = 48/1000$ s, and $\hat{\mu}_{U_1} = 1.7654$ ms^{-1}. For $k_1 > q_1$, the velocity spectrum decreases at a higher rate than it does in the inertial subrange due to the dissipation of turbulent kinetic energy. Consequently, aliasing is negligible in the San Vittore measurements.

Under these assumptions, i.e., (i) that the $N = 37501$ vertical velocities plotted in Fig. 9.34 (a) are reconcilable with model (9.1) and thus have a continuous temporal spectrum and (ii) that aliasing is negligible, the periodogram and direct spectral estimators using cosine tapers with $p = 0.05$ and $p = 0.10$ are calculated, plotted and compared with each other, as recommended in the remarks to Figs. 9.3 and 9.4. Since the periodogram and the direct spectral estimators cannot be distinguished from each other, this periodogram is not subject to leakage, despite the dynamic range of the spectrum (approximately 25 dB) being larger than the threshold in diagnostic (6.117). Consequently, this spectrum can be estimated by smoothing, in Fig. 9.35, the periodogram calculated from the turbulent fluctuations resulting from the Reynold's decomposition in Fig. 9.34 in the following R expressions:

```
#velocity: u: in direction of flow, v: perpendicular, w: vertical
#temp: temperature
tur <- scan("/path/turwind3.dat",list(u=0,v=0,w=0,temp=0))
#sampling interval: Delta t = 48/1000 seconds
turts <- ts(tur$w,start=0,frequency=1000/48,)
x <- window(turts,1800.0,3600.0) #between 12:30 and 13:00
#spec.univariate() is introduced in Sect. 9.5.6
turper <- spec.univariate(x,taper=0,spans=1,pad=0,
                                        confint=0.95,fast=F)
plot(c(0,10.5),c(-60,0),type="n",xlab="frequency [1/s]",
                ylab="periodogram, M=400, 10*log10[ (m/s)^2 ]")
points(turper$freq,10*log10(turper$spec),pch=".")
turpersm <- spec.univariate(x,taper=0,spans=801,pad=0,
```

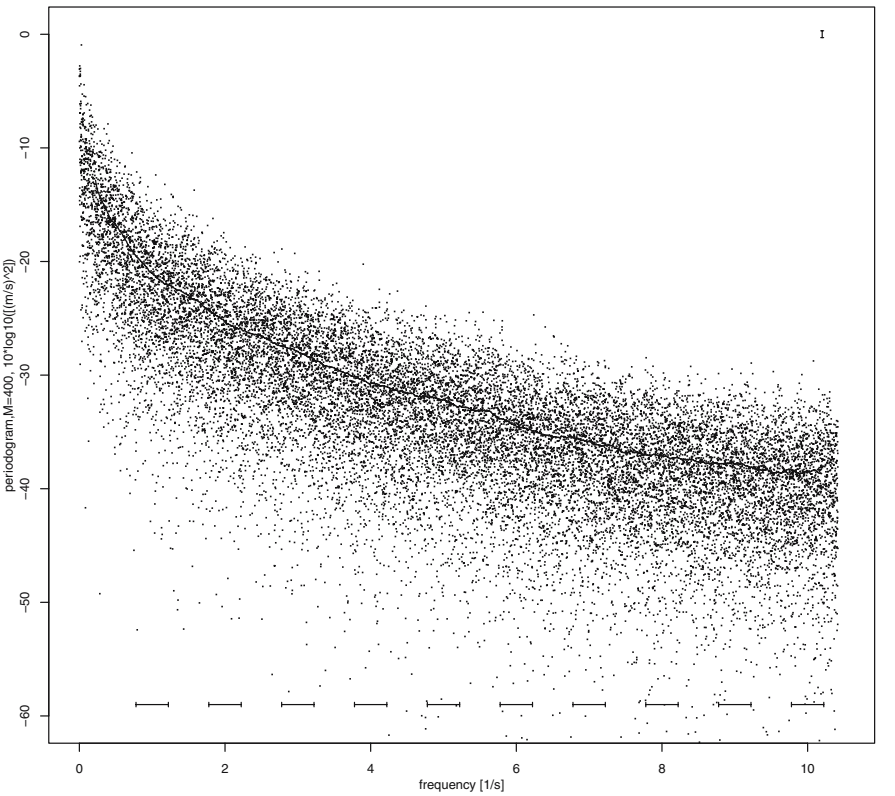


Fig. 9.35. Turbulent fluctuations as plotted in Fig. 9.34 (a): temporal periodogram of the vertical velocity (fine points) discretely smoothed using a modified discrete Daniell kernel having parameter $M = 400$ (solid line). The width of the smoothing kernel is plotted with ten horizontal bars, the height of a .95 confidence interval using a vertical bar (in the upper right corner).

```
                                        confint=0.95,fast=F)
lines(turpersm$freq,10*log10(turpersm$spec),lty=1)
#bars for bandwidth and height of confidence interval:
low <- turpersm$confintdblow
upp <- turpersm$confintdbupp
lines(c(10.2,10.2),c(0+low,0+upp))
lines(c(10.18,10.22),c(0+low,0+low))
lines(c(10.18,10.22),c(0+upp,0+upp))
freqpos <- c(0.0,1.0,2.0,3.0,4.0,5.0,6.0,7.0,8.0,9.0,10.0)
smwwh <- turpersm$bandwidth/2
for(i in 2:11) {
  lines(c(freqpos[i]-smwwh, freqpos[i]+smwwh),c(-59,-59))
```

```
  lines(c(freqpos[i]-smwwh, freqpos[i]-smwwh),c(-59.2,-58.8))
  lines(c(freqpos[i]+smwwh, freqpos[i]+smwwh),c(-59.2,-58.8))
}
```

The smoothed periodogram has bandwidth $0.44 = 800 \times (1/N)(1/\Delta t)$ s^{-1}, computed as required in (9.90) using $\Delta t = 48/1000$ s and $N = 37501$, as well as .95 confidence intervals with height 0.60 dB. Smoothing the periodogram over the band being 0.44 s^{-1} wide produces an estimator which is quite smooth. This is an asset when analysing the spectrum of the vertical velocities in the turbulent flow for frequencies in the inertial subrange (9.99).

Under Taylors's hypothesis (9.100), the approximative .95 confidence intervals of the smoothed periodogram are transformed using the following R expressions

```
#mean velocity of flow in direction x1, calculated as mean of
#the u-values in file /path/turwind2.dat introduced in the
#remarks to Fig. 9.34
umu = 1.7654
clow <- (umu/(2*pi))*turpersm$confintvectorlow
cupp <- (umu/(2*pi))*turpersm$confintvectorupp
k1 <- (2*pi/umu)*turpersm$freq #wave number
```

in approximations for the .95 confidence intervals of estimator $\widehat{C}_{U_3}(k_1)$ for $C_{U_3}(k_1)$. $C_{U_3}(k_1)$ is the vertical component (9.98) in the spatial velocity spectrum of the turbulent flow in direction x_1, i.e., in direction of the mean flow. The logarithms of these approximations for the .95 intervals of $\widehat{C}_{U_3}(k_1)$ are then plotted versus the logarithms of the wave numbers in Fig. 9.36. This plot is generated by the following R expressions

```
plot(c(-1,4),c(-10,-4),type="n",
     ylab="spatial periodogramm, M=400, log[ m*(m/s)^2 ]",
     xlab="wave number k1, log[ 1/m ]")
points(log(k1),log(clow),pch=".")
points(log(k1),log(cupp),pch=".")
lines(c(0.73,3.73),c(-5.1,-10.1))
points(c(0.73,3.73),c(-5.1,-10.1),pch="0")
text(-0.2,-5.1,"(0.73,-5.1)")
text(2.5,-10.0,"(3.73,-10.1)")
lines(c(1.5,1.5),c(-9.0,-6.0))
text(1.5,-9.5, "p1")
#transform symbols for bandwidth
fpwz<- (2*pi/umu)*freqpos
smwwhwz <- (2*pi/umu)*smwwh
for(i in 2:11) {
lines(c(log(fpwz[i]-smwwh), log(fpwz[i]+smwwh)),c(-4.0,-4.0))
lines(c(log(fpwz[i]-smwwh), log(fpwz[i]-smwwh)),c(-4.1,-3.9))
lines(c(log(fpwz[i]+smwwh), log(fpwz[i]+smwwh)),c(-4.1,-3.9)) }
```

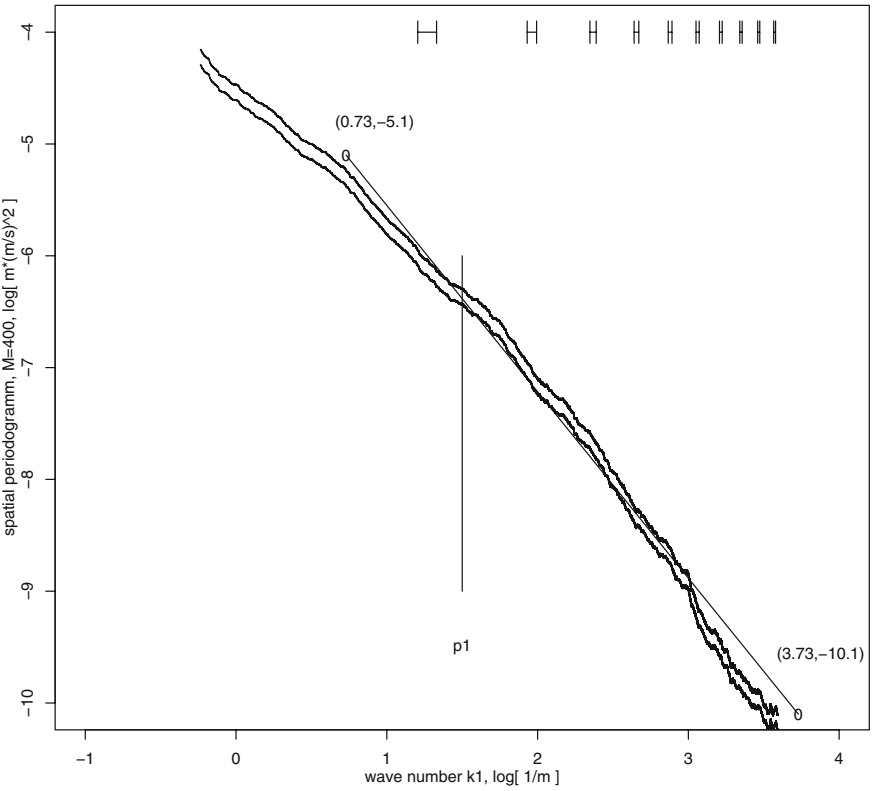


Fig. 9.36. Log-log plot of the .95 confidence intervals pertaining to $\widehat{C}_{U_3}(k_1)$, i.e. the estimator for the vertical component (9.98) in the spatial velocity spectrum of the turbulent flow in Fig. 9.34 (a). The height of the confidence interval is constant whereas the bandwidth of the estimator decreases subsequent to a logarithmic transformation of the wave number. The straight line has slope $-5/3$.

and allows for testing the hypothesis that the turbulent flow in Fig. 9.34 is locally isotropic, i.e., has a spectrum (9.99) being reconcilable with Kolmogorov's hypotheses.

A Kolmogorov spectrum (9.99) decreases in a log-log representation with slope $-5/3$: a straight line having this slope is drawn in Fig. 9.36. This line remains approximately within the .95 confidence intervals for wave numbers $1.5 \leq \log(k_1) \leq 3.0\ \log(\mathrm{m}^{-1})$, i.e., $4.5 \leq k_1 \leq 20.0\ \mathrm{m}^{-1}$. Consequently, (i) the inertial subrange stretches from approximately $4.5\ \mathrm{m}^{-1}$ up to approximately $20\ \mathrm{m}^{-1}$, and (ii) it is not excluded that the turbulent flow is locally isotropic (cf. the remark to (9.99)). The analysis of this turbulent flow is continued in Problem 9.19.

9.5 Supplements

In Sect. 9.5.1, (9.6) is proved, i.e., it is shown that a direct spectral estimator is the Fourier transform of the autocorrelation of the tapered observations, provided that a variance-normalised data taper (9.2,1) is applied. Using this result, the moment functions of a direct spectral estimator are derived in Sects. 9.5.2 and 9.5.3. In Sect. 9.5.4, some definitions introduced in Sect. 6.6 are applied to discuss the width of a smoothing kernel and examples are given in Table 9.4. The usual approximation for the variance of a smoothed direct spectral estimator is derived in Sect. 9.5.5. In Sect. 9.5.6, R function `spec.pgram()` is discussed, and, as an alternative, `spec.univariate()` is proposed.

9.5.1 Direct Spectral Estimator and Autocorrelation of Tapered Observations

A direct spectral estimator $\widehat{C}_X^{(d)}(s)$ as defined in (9.3) and (9.4) can be written as in (9.103). The sums in (9.103) can be interpreted as polynomials. In their squares, therefore, the terms can be arranged in rows and columns using two indices and thus the double-indexed sums in the following two lines are arrived at. In each term in these sums, the product of two trigonometric functions is substituted using the elementary trigonometric identities and the sum (9.104) readily follows.

$$\widehat{C}_X^{(d)}(s) = \left(\sum_{t=0}^{N-1} h_N^{(d)}(t) X_t \cos(2\pi st)\right)^2 + \left(\sum_{t=0}^{N-1} h_N^{(d)}(t) X_t \sin(2\pi st)\right)^2 \tag{9.103}$$

$$= \frac{1}{N} \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} h_N^{(d)}(t) X_t \cos(2\pi st) h_N^{(d)}(u) X_u \cos(2\pi su) +$$
$$\sum_{t=0}^{N-1} \sum_{u=0}^{N-1} h_N^{(d)}(t) X_t \sin(2\pi st) h_N^{(d)}(u) X_u \sin(2\pi su)$$

$$= \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} h_N^{(d)}(t) X_t h_N^{(d)}(u) X_u \frac{1}{2}\big(\cos(2\pi s(t-u)) + \cos(2\pi s(t+u))\big) +$$
$$\sum_{t=0}^{N-1} \sum_{u=0}^{N-1} h_N^{(d)}(t) X_t h_N^{(d)}(u) X_u \frac{1}{2}\big(\cos(2\pi s(t-u)) - \cos(2\pi s(t+u))\big)$$

$$= \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} h_N^{(d)}(t) X_t h_N^{(d)}(u) X_u \cos\big(2\pi s(t-u)\big) \tag{9.104}$$

$$= \sum_{\tau=-(N-1)}^{N-1} \left(\frac{1}{\Sigma(h^2)} \sum_{t=0}^{N-1-|\tau|} h_N(t) X_t h_N(t+|\tau|) X_{t+|\tau|}\right) \cos(2\pi s\tau) \tag{9.105}$$

$$= \sum_{\tau=\infty}^{\infty} \hat{c}_X^{(d)}(\tau)\cos(2\pi s\tau) \qquad \text{where} \tag{9.106}$$

$$\hat{c}_X^{(d)}(\tau) = \begin{cases} \frac{1}{\Sigma(h^2)} \sum_{t=0}^{N-1-|\tau|} h_N(t)X_t h_N(t+|\tau|)X_{t+|\tau|} & \text{for } |\tau| < N \\ 0 & \text{for } |\tau| \geq N \end{cases} \tag{9.107}$$

The argument $2\pi s(t-u)$ of the cosine functions in (9.104) is substituted with $2\pi s\tau$ and thereafter the diagonals in the matrix of the terms are summed, as demonstrated when deriving (2.73) or in Problem 6.16. The main diagonal contains N terms $\big(h_N(t)/\sqrt{\Sigma(h^2)}\big)^2(X_t)^2$, the first subdiagonals $N-1$ terms $\big(h_N(t)/\sqrt{\Sigma(h^2)}\big)\big(h_N(t-1)/\sqrt{\Sigma(h^2)}\big)X_tX_{t-1}$ as well as $\big(h_N(t)/\sqrt{\Sigma(h^2)}\big)\big(h_N(t+1)/\sqrt{\Sigma(h^2)}\big)X_tX_{t+1}$, etc., and (9.105) is arrived at. The sum in parenthesis in (9.105) is the autocorrelation (6.103) of the tapered observations $h_N(t)X_t$ (and thus an even function), normalised with $\Sigma(h^2)$. It is identically zero for $\tau \leq -N$ and $\tau \geq N$ because $h_N(t)h_N(t+|\tau|) = 0$ for $|\tau| \geq N$, as is concluded from definition (6.103), which implies that $\hat{c}_X^{(d)}(\tau)$ in (9.107) is identically zero for $|\tau| \geq N$.

9.5.2 Expectation Function of a Direct Spectral Estimator

Assumptions such as those required for the definition of the direct spectral estimator in (9.3) are made: $J(s)$ in (9.108) is the estimator resulting from a Fourier transform of the observations (x_t), $t = 0, 1, \ldots, N$, stemming from a stationary stochastic process (X_t), calculated subsequent to tapering using a variance-normed data taper as defined in (9.2). The Fourier representation (7.51) of X_t is substituted and thus (9.109) is arrived at, using (i) $\mathrm{d}Z(s) = \overline{\mathrm{d}Z(-s)}$ (as implied by (7.70,3,4), reconcilable with $C_X^{(I)}(s)$ being an even (6.32,1) function, an example is given in (7.43)), (ii) the integration rules, and (iii) $H_N(s)$, i.e., the kernel pertaining to $h_N(t)$ as defined in (6.110).

$$J(s) = \sum_{t=0}^{N-1} h_N^{(d)}(t)X_t\mathrm{e}^{-\mathrm{i}2\pi st}, \qquad X_t = \int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi s'}\mathrm{d}Z_X(s') \tag{9.108}$$

$$\begin{aligned} &= \sum_{t=0}^{N-1} h_N^{(d)}(t)\left(\int_{-1/2}^{1/2} \mathrm{e}^{\mathrm{i}2\pi s'}\mathrm{d}Z_X(s')\right)\mathrm{e}^{-\mathrm{i}2\pi st} \\ &= \int_{-1/2}^{1/2} \sum_{t=0}^{N-1} h_N^{(d)}(t)\mathrm{e}^{-\mathrm{i}2\pi(s-s')t}\mathrm{d}Z_X(s') \\ &= \int_{-1/2}^{1/2} \frac{H_N(s-s')}{\sqrt{\Sigma(h^2)}}\mathrm{d}Z_X(s') = -\int_{-1/2}^{1/2} \frac{H_N(s+s')}{\sqrt{\Sigma(h^2)}}\overline{\mathrm{d}Z_X(s')} \end{aligned} \tag{9.109}$$

Since the complex-valued random function $Z_X(s)$ is required to have orthogonal increments in (7.51), the expectation of the squared absolute value of $J(s)$

is obtained in (9.110) by integrating with respect to an orthogonal increment process as demonstrated in Sect. 7.5.2.

$$\overline{J(s)} = \int_{-1/2}^{1/2} \frac{\overline{H_N(s-s')}}{\sqrt{\Sigma(h^2)}} \overline{\mathrm{d}Z_X(s')}$$

$$\begin{aligned}
\mathrm{E}\big(J(s)\overline{J(s)}\big) &= \mathrm{E}\left(\int_{-1/2}^{1/2} \frac{H_N(s-s')}{\sqrt{\Sigma(h^2)}} \mathrm{d}Z_X(s') \int_{-1/2}^{1/2} \frac{\overline{H_N(s-s'')}}{\sqrt{\Sigma(h^2)}} \overline{\mathrm{d}Z_X(s'')} \right) \\
&= \int_{-1/2}^{1/2} \int_{-1/2}^{1/2} \frac{H_N(s-s')}{\sqrt{\Sigma(h^2)}} \frac{\overline{H_N(s-s'')}}{\sqrt{\Sigma(h^2)}} \Big(\delta(s'-s'') \mathrm{E}\big(\mathrm{d}Z_X(s') \overline{\mathrm{d}Z_X(s'')}\big) \Big) \\
&= \int_{-1/2}^{1/2} \mathcal{H}_N^{(d)}(s-s') \mathrm{E}|\mathrm{d}Z_X(s')|^2 \qquad (9.110)
\end{aligned}$$

In (9.110), $\mathcal{H}_N^{(d)}(s) = \big(H_N(s)\overline{H_N(s)}\big)/\Sigma(h^2) = \mathcal{F}_{-\mathrm{i}}\big(h_N^{(d)}(t)\big) * h_N^{(d)}(t)\big)$ is obtained from the autocorrelation theorem in Sect. 6.9.6, i.e., $\mathcal{H}_N^{(d)}(s)$ is the Fourier transform of the normalised autocorrelation of the data taper applied to the observations. Normalising the autocorrelation of the data taper with $\Sigma(h^2)$ $\big($or the data taper with $\sqrt{\Sigma(h^2)}\big)$ asserts that an empirical covariance function calculated from the tapered observations using (2.1,3) is not biased (at least for lag $\tau = 0$, the estimate for the variance) by the data taper, as argued in the remarks to (9.8).

9.5.3 Covariance Function of a Direct Spectral Estimator

Covariance $\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s+r)\big)$ for lag r (a displacement in the frequency domain) of a direct spectral estimator is approximated in (9.14). This approximation is derived in [108] under the assumptions stipulated in (9.13). An abbreviated version of this derivation is given in this section.

It is assumed that a direct spectral estimator $\widehat{C}_X^{(d)}(s)$ is calculated using (9.3) by multiplying a variance-normalised data taper $h_N^{(d)}(t)$ (9.2,1) with a segment in a realisation of the discrete-time stationary model (X_t) in (9.1,1) having a continuous spectrum $C_X(s)$ and an expectation $\mu_X = 0$ being identically zero. Then the product $h_N^{(d)}(t)(X_t)$ has the Fourier representation obtained in (9.109) which is used to derive the covariance of the direct spectral estimator in (9.111).

$J(s)$ in (9.111) is the Fourier representation of segment $t = 0, 1, 2, \ldots, N-1$ in a normal process (X_t) as required in (9.13,3) and therefore, being a linear combination of multivariate normally distributed (X_t), is complex-valued and normal, as is concluded from the remarks to (1.34). Complex-valued stochastic processes and their moment functions are defined in (7.27). Using these definitions and Isserlis' theorem (which proposes that moments of random variables Y_i, $i = 1, 2, \ldots, n$, with higher order than 2 can all be expressed in terms of the second order moments, on condition that the Y_i are

multivariate normally distributed as required in (2.3) [73]), the covariance of $|J(s)|^2$ for frequencies s and s' is derived borrowing from [108] with the result in (9.111): it is the sum of two squared absolute values of two complex valued expectation functions as defined in (7.27).

The expectation in the first term in (9.111) is evaluated by substituting from (9.109). Thereafter, the expectation of the twofold integral is arrived at in (9.112) by integrating with respect to an orthogonal increment process exploiting the properties of the Fourier representation (7.51) of (X_t) arrived at in (9.109), as demonstrated in Sect. 7.5.2 and also above, in Sect. 9.5.2. From a similar derivation, the expectation in the second term results in 9.113. Substituting s' with $s+r$ and s'' with r' in these expectations, the covariance of the direct spectral estimator in (9.114) is arrived at.

$$\mathrm{Cov}\big(|J(s)|^2,|J(s')|^2\big) = \big|\mathrm{E}\big(\overline{J(s)}J(s')\big)\big|^2 + \big|\mathrm{E}\big(J(s)J(s')\big)\big|^2 \tag{9.111}$$

$$\mathrm{E}\big(\overline{J(s)}J(s')\big) = \mathrm{E}\left(\int_{-1/2}^{1/2}\frac{\overline{H_N(s-s'')}}{\sqrt{\Sigma(h^2)}}\overline{\mathrm{d}Z_X(s'')}\int_{-1/2}^{1/2}\frac{H_N(s'-s''')}{\sqrt{\Sigma(h^2)}}\mathrm{d}Z_X(s''')\right)$$

$$= \int_{-1/2}^{1/2}\int_{-1/2}^{1/2}\frac{\overline{H_N(s-s'')}}{\sqrt{\Sigma(h^2)}}\frac{H_N(s'-s''')}{\sqrt{\Sigma(h^2)}}\Big(\delta(s''-s''')\mathrm{E}\big(\overline{\mathrm{d}Z_X(s'')}\mathrm{d}Z_X(s''')\big)\Big)$$

$$= \frac{1}{\Sigma(h^2)}\int_{-1/2}^{1/2}\overline{H_N(s-s'')}H_N(s'-s'')C_X(s'')\mathrm{d}s'' \tag{9.112}$$

$$\mathrm{E}\big(J(s)J(s')\big) = -\frac{1}{\Sigma(h^2)}\int_{-1/2}^{1/2}H_N(s+s'')H_N(s'-s'')C_X(s'')\mathrm{d}s'' \tag{9.113}$$

$$\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s),\widehat{C}_X^{(d)}(s+r)\big) = \mathrm{Cov}\big(|J(s)|^2,|J(s+r)|^2\big) \tag{9.114}$$

$$= \frac{1}{(\Sigma(h^2))^2}\Bigg(\Bigg|\int_{-1/2}^{1/2}\overline{H_N(s-r')}H_N(s+r-r')C_X(r')\mathrm{d}r'\Bigg|^2 \tag{9.115}$$

$$-\Bigg|\int_{-1/2}^{1/2}H_N(s+r')H_N(s+r-r')C_X(r')\mathrm{d}r'\Bigg|^2\Bigg) \tag{9.116}$$

$$\approx \frac{1}{(\Sigma(h^2))^2}\Bigg|\int_{-1/2}^{1/2}\overline{H_N(s-r')}H_N(s+r-r')C_X(r')\mathrm{d}r'\Bigg|^2 \tag{9.117}$$

for $s \neq -1/2, 0, 1/2$ and $s+r \neq -1/2, 0, 1/2$

The covariance in (9.114) is the difference of the squared absolute values of the integrals in (9.115) and (9.116). The integral in (9.116) becomes negligible as compared to the first one in (9.115) on condition that frequencies s and $s+r$ are not identical with $-1/2$, 0 or $1/2$, as argued in [108]: approximation (9.117) is arrived at.

In approximation (9.117), $\overline{H_N(s-r')}H_N(s+r-r')$ is the product of the kernel of the data taper with its duplicate shifted by displacement r. The main lobe of a kernel $H_N(s)$ is often inside the band $s_0-\varsigma, s_0+\varsigma,]$, s_0

being the frequency where $H_N(s)$ attains its maximum and $\varsigma = \mathcal{B}_{\triangle}^{\cap}\big(H_N(s)$, calculated as demonstrated in Fig. 6.24. For $|r| > 2\varsigma = 2\mathcal{B}_{\triangle}^{\cap}\big(H_N(s)\big)$, the main lobes of the kernel and its duplicate being displaced by r no longer overlap and their product, as required in (9.117), becomes small which implies that the integral is dominated by $C_X(s)$ (the spectrum to be estimated) in frequency bands $[s-\varsigma, s+r+\varsigma,]$, for $r > 0$, and $[s+r-\varsigma, s+\varsigma,]$, for $r < 0$. Consequently, assuming in (9.13,4) that $C_X(s)$ is locally constant in a frequency band $[s-\varsigma, s+\varsigma]$ of width $2\varsigma = 2\mathcal{B}_{\triangle}^{\cap}\big(H_N(s)\big)$ about s, $C_X(s)$ needs no longer to be included in the integral and (9.118) is arrived at, the approximation proposed in (9.14).

$H{\circledast}H(r)$ in (9.119) is defined in $-1/2 \le s \le -1/2$ and periodic with period 1, properties inherited from $H_N(s)$. $H_N(s) = \mathcal{F}_{-\mathrm{i}}\big(h_N(t)\big)$ as required in (6.110,2) implies that convolution theorem (6.48) can be applied to obtain (9.120). (9.121) then follows from definition (9.2,1).

$$\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s), \widehat{C}_X^{(d)}(s+r)\big) \approx \big(C_X^2(s)\big)\big(H{\circledast}H(r)\big) \qquad \text{with} \tag{9.118}$$

$$H{\circledast}H(r) = \frac{1}{(\Sigma(h^2))^2}\left|\int_{-1/2}^{1/2} H_N(r')H_N(r-r')\mathrm{d}r'\right|^2 \tag{9.119}$$

$$= \frac{1}{(\Sigma(h^2))^2}\left|\sum_{\upsilon=0}^{N-1}\Big(\big(h_N(\upsilon)\big)\big(h_N(\upsilon)\big)\Big)\mathrm{e}^{-\mathrm{i}2\pi r\upsilon}\right|^2 \tag{9.120}$$

$$= \left|\sum_{\upsilon=0}^{N-1}\Big(\big(h_N^{(d)}(\upsilon)\big)\big(h_N^{(d)}(\upsilon)\big)\Big)\mathrm{e}^{-\mathrm{i}2\pi r\upsilon}\right|^2 \tag{9.121}$$

$$= \sum_{u=-(N-1)}^{N-1} h{\circledast}h(u)\mathrm{e}^{-\mathrm{i}2\pi ru} = \sum_{u=-\infty}^{\infty} h{\circledast}h(u)\mathrm{e}^{-\mathrm{i}2\pi ru} \tag{9.122}$$

$$h{\circledast}h(u) = \begin{cases} 1/(\Sigma(h^2))^2 \sum_{\upsilon=0}^{N-1-|u|} h_N^2(\upsilon)h_N^2(\upsilon+|u|) & \text{for } |u| < N \\ 0 & \text{for } |u| \ge N \end{cases} \tag{9.123}$$

The sum in (9.121) becomes the first Fourier sum in (9.3) on condition that, in (9.3), t is substituted with υ and X with $h_N^{(d)}$. When these substitutions are applied to (9.6) and (9.7), (9.122) and (9.123) are arrived at, which imply that $h{\circledast}h(u)$ and $H{\circledast}H(r)$ are a Fourier transform pair as defined in (6.49) and (6.50). When the kernel and its shifted duplicate become identical in (9.119), i.e., for $r = 0$,

$$H{\circledast}H(0) = \frac{1}{(\Sigma(h^2))^2}\left|\int_{-1/2}^{1/2} H_N(r')H_N(-r')\mathrm{d}r'\right|^2$$

$$= \frac{1}{(\Sigma(h^2))^2}\left|\sum_{\upsilon=0}^{N-1}\Big(\big(h_N(\upsilon)\big)\big(h_N(\upsilon)\big)\Big)\right|^2 = 1 = \sum_{u=-\infty}^{\infty} h{\circledast}h(u) \tag{9.124}$$

is obtained from (9.120), owing to the normalisation in (9.2,1).

9.5.4 Widths of Smoothing Kernels

In (7.83), the bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big)$ of a continuous spectrum $C_X(s)$ is defined as the width at half height of $C_X(s)$, because a spectrum does not generally have the properties required to calculate an equivalent width (6.99,1) or a σ-width (6.101) which are geometrically meaningful. For example, the spectrum of the AR[4] model $X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9838X_{t-4} + W_t$, with $\mu_W = 0$ and $\sigma_W^2 = 1$, is plotted in Fig. 7.14 (b) and (c) together with its bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big) = 0.00474$ obtained in Fig. 7.10 (d). $\mathcal{B}_{\triangle}\big(C_X(s)\big)$ is compared, in the remarks to Figs. 9.14, 9.16 and 9.16, with $\mathcal{B}_{||}\big(V_{dD,M}(s'_k)\big) = 2M\Delta s'_k$, i.e., the widths as defined (9.125) of the modified discrete Daniell kernels applied to calculate the discretely smoothed direct estimators for $C_X(s)$ plotted there.

Let $V_m(s_k)$ and $V_m(s'_k)$ be discrete smoothing kernels as required in (9.28) and having approximately rectangular shape, i.e., small tails. Then $V_m(s_k)$ and $V_m(s'_k)$ have widths $\mathcal{B}_{||}\big(V_m(s_k)\big) = 2M\Delta s_k$ and $\mathcal{B}_{||}\big(V_m(s'_k)\big) = 2M\Delta s'_k$. (9.125)

Definition (9.125) is motivated by the width of a moving average as proposed in the remarks to definition (2.39).

In the case of a lag window spectral estimator as defined in (9.41), (9.42), (9.43), and (9.44), some of the widths introduced in Sect. 6.6 can be applied to obtain the following definitions for the width of a continuous smoothing kernel.

The equivalent width of a smoothing kernel $V_m^{\vdash\tau\dashv}(s)$ in (9.42) is geometrically meaningful on condition that $V_m^{\vdash\tau\dashv}(s) \geq 0$ for $-1/2 \leq s \leq 1/2$, because a continuous smoothing kernel has its maximum usually in $s = 0$. In addition, if the associated lag window $v_m^{\vdash\tau\dashv}(\tau) \geq 0$ for $\tau = -(N-1), \ldots, -1, 0, 1, \ldots, N-1$, then it is concluded from (9.42) and (6.99,2) that the equivalent width of $V_m^{\vdash\tau\dashv}(s)$ is the inverse of the equivalent width of $v_m^{\vdash\tau\dashv}(\tau)$. Using $v_m^{\vdash\tau\dashv}(0) = \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s)\mathrm{d}s = 1$ as obtained in the remarks to (9.44), the *equivalent width of the smoothing kernel* is obtained in (9.126).

$$\mathcal{B}_{\sqcap}\big(V_m^{\vdash\tau\dashv}(s)\big) = \frac{1}{\mathcal{B}_{\sqcap}\big(v_m^{\vdash\tau\dashv}(\tau)\big)} = \frac{1}{\sum_{\tau=-(N-1)}^{N-1} v_m^{\vdash\tau\dashv}(\tau)} \tag{9.126}$$

If $V_m^{\vdash\tau\dashv}(s) \geq 0$ as required above for definition (9.126), then $V_m^{\vdash\tau\dashv}(s)$ does have the properties of a probability density function, for the reason that $v_m^{\vdash\tau\dashv}(0) = \int_{-1/2}^{1/2} V_m^{\vdash\tau\dashv}(s)\mathrm{d}s = 1$ as concluded in the remarks to (9.44), and thus the calculation (in Problem 9.14) of the *σ-width of the smoothing kernel* (9.128) by applying definition (6.101) becomes geometrically meaningful.

$$\Big(\mathcal{B}_{\sigma}\big(V_m^{\vdash\tau\dashv}(s)\big)\Big)^2 = 12\int_{-1/2}^{1/2} s^2 V_m^{\vdash\tau\dashv}(s)\mathrm{d}s \tag{9.127}$$

$$= 1 + \frac{12}{\pi^2} \sum_{\tau=1}^{N-1} \frac{(-1)^\tau}{\tau^2} v_m^{\vdash\tau\dashv}(\tau) \tag{9.128}$$

In the case of a Daniell lag window (9.45), the integral in (9.127) (and thus the sum in (9.128)), which is identical with the second moment as defined in (6.100), becomes the square of the width of the rectangle function when multiplied by 12.

Definitions (9.126 and (9.128) are geometrically meaningful provided that $v_m^{\vdash\tau\dashv}(\tau) \geq 0$ for $\tau = -(N-1), \ldots, -1, 0, 1, \ldots, N-1$ and $V_m^{\vdash\tau\dashv}(s) \geq 0$ for $-1/2 \leq s \leq 1/2$. However, there are lag windows with negative weights (e.g., the lag window pertaining to the modified discrete Daniell kernel as plotted for $M = 3$ in Fig. 9.19 (c)) and smoothing kernels with negative side lobes (e.g., the continuous smoothing kernel pertaining to the modified discrete Daniell kernel as plotted for $M = 3$ in Fig. 9.19 (d)). If a lag window has negative weights then an equivalent width can become negative. If a smoothing kernel has negative side lobes then the integral in (9.128) can become negative and therefore a σ-width can become imaginary.

These shortcomings are, however, not shared by the autocorrelation width of the smoothing kernel. Using definitions (6.102), (6.103) and (6.104), $\int_{-1/2}^{1/2} \big(V_m^{\vdash\tau\dashv}(s)\big) \star \big(V_m^{\vdash\tau\dashv}(s)\big)\, \mathrm{d}s = \big(v_m^{\vdash\tau\dashv}(0)\big)\big(v_m^{\vdash\tau\dashv}(0)\big) = 1 \times 1 = 1$ is obtained from assumptions (9.42), (9.43) and (9.44), and therefore,

$$\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big) = \frac{1}{\int_{-1/2}^{1/2} \big(V_m^{\vdash\tau\dashv}(s)\big)^2 \mathrm{d}s} = \frac{1}{\sum_{\tau=-(N-1)}^{N} \big(v_m^{\vdash\tau\dashv}(\tau)\big)^2} \tag{9.129}$$

is the *autocorrelation width of the smoothing kernel*. The denominators in (9.129) are equal as is concluded from (6.74,2, Parseval's identity).

For example, some widths of the continuous smoothing kernels plotted in Fig. 6.23 (Daniell), Fig. 9.18 (Bartlett) and Fig. 9.19 (modified discrete Daniell) are given in Table 9.4.

$\mathcal{B}_\triangle^\cap\big(V_m^{\vdash\tau\dashv}(s)\big)$, i.e., the widths at half height of the main lobes in the smoothing kernels, are determined in plots with a high resolution of the frequency domain, except for $\mathcal{B}_\triangle^\cap\big(F_N(s)\big)$ which is calculated using the approximation proposed in Table 9.1. The equivalent, σ- and autocorrelation widths of the Daniell kernel (9.45) are calculated as required in (6.99), (6.101) and (6.104); these computations are favoured by the properties of the rectangle function and its Fourier transform in (6.38), (6.39) and (6.40). In contrast, definitions (9.126), (9.128) and (9.129) are used to compute the widths of smoothing kernels pertaining to $\Lambda_{2m+1}^{(e)}$ (the Bartlett lag window) and $V_{dD,M}^{\vdash\tau\dashv}(\tau)$ (obtained as discrete Fourier transform of the modified discrete Daniell kernel $V_{dD,M}(s'_j)$ (9.29), $N' = 2N$ and $\Delta s'_k = 1/(2N)$). The widths $\mathcal{B}_{||}\big(V_{dD,M}(s'_k)\big)$ of the modified discrete Daniell kernels are obtained by applying definition (9.125).

Table 9.4. Widths $\mathcal{B}_{||}\big(V_{dD,M}(s'_k)\big) = 2M\Delta s'_k$, $\Delta s'_k = 1/(2N)$ as defined in (9.125) of modified discrete Daniell kernels (9.29), equivalent widths $\mathcal{B}_{\sqcap}(\cdot)$, σ-widths $\mathcal{B}_{\sigma}(\cdot)$, autocorrelation widths $\mathcal{B}_{\star}(\cdot)$ of smoothing kernels $V_m^{\vdash\tau\dashv}(s)$ as plotted in Figs. 6.23 (d), 9.18 (b) and 9.19 (d), as well as width at half height $\mathcal{B}^{\cap}_{\triangle}(\cdot)$ of the main lobe of $V_m^{\vdash\tau\dashv}(s)$ (marked * when calculated despite the remarks to (9.126) and (9.128)).

lag window		N	m	M	$2M\Delta s'_k$	$\mathcal{B}_{\sqcap}(\cdot)$	$\mathcal{B}_{\sigma}(\cdot)$	$\mathcal{B}_{\star}(\cdot)$	$\mathcal{B}^{\cap}_{\triangle}(\cdot)$
Daniell (9.45)	$v^{\vdash\tau\dashv}_{Da,m}(\tau)$ Fig. 6.23	64	10			0.1	0.1	0.1	
Bartlett (6.56)	$\Lambda^{(g)}_{2m+1}(\tau)$ Fig. 9.18	64	30			.0333	.0281	.0499	.0300
m. Daniell (9.29)	$v^{\vdash\tau\dashv}_{dD,M}(\tau)$ Fig. 9.19	64		3	.0470			.0535	.0500
		256		4	.0156	.0156*	.0226*	.0167	.0156
		4096		10	.0024	.0024*	.0035*	.0026	.0024

As an aside: computing the σ-width by applying (9.128) is delicate: the results in Table 9.4 were computed using a 64-bit processor and cannot be reproduced using a 32-bit processor and single-precision variables.

From Table 9.4 it is concluded that $\mathcal{B}_{\star}\big(V^{\vdash\tau\dashv}_{dD,M}(s)\big) \approx \mathcal{B}^{\cap}_{\triangle}\big(V^{\vdash\tau\dashv}_{dD,M}(s)\big) \approx \mathcal{B}_{||}\big(V_{dD,M}(s'_k)\big) = 2M\Delta s'_k$, as is obvious for $N = 64$ in Fig. 9.19.

9.5.5 Variance Function of a Lag Window Estimator

Let $\widehat{C}^{\vdash\tau\dashv}_X(s'_k) = (1/(2N))\sum_{j=-(N-1)}^{N} V_m^{\vdash\tau\dashv}(s'_j)\widehat{C}^{(d)}_X(s'_{k-j})$ be a lag window estimator computed for discrete frequencies s'_k with $\Delta s'_k = 1/(2N)$ as obtained in (9.59). For large N — and thus small $\Delta s'_k$ — it is likely that (i) $V_m^{\vdash\tau\dashv}(s'_k) \approx 0$ at the exterior of band $[s'_k - J\Delta s'_k, s'_k + J\Delta s'_k] = [s'_{k-J}, s'_{k+J}]$, J being a small integer number, and (ii) $C_X(s)$ is locally (2.55) constant in $[s'_{k-J}, s'_{k+J}]$. Band $[s'_{k-J}, s'_{k+J}]$ is assumed to have approximately the same width as $s - \zeta \le s \le s + \zeta$ having width 2ζ as defined in (9.73), i.e., $2J\Delta s'_k \approx 2\zeta$, and thus is, in most cases, wider than $[s - \varsigma, s + \varsigma]$ having width $2\varsigma = 2\mathcal{B}^{\cap}_{\triangle}\big(|H_N(s)|\big)$ as required in (9.13,4) for approximating the second moment function of a direct spectral estimator in (9.14), (9.15), (9.16) and (9.17).

For example, the direct spectral estimator discretely smoothed in Figs. 9.15 and 9.16, is the result of tapering the observations with a cosine taper for $p = 0.05$ such that $2\mathcal{B}^{\cap}_{\triangle}\big(|\widehat{H_N}(s)|\big) \approx 4\Delta s'_k$, as is obvious in Fig. 9.6, is narrower than band $[s'_{k-J}, s'_{k+J}]$ having width $20\Delta s'_k$ for $J = 10$, with $\Delta s'_k = 0.000122 = 1/(2N)$, $N = 4096$. In this example, moreover, $20\Delta s'_k \approx$

$0.0024 < 0.0026 \approx \mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)$, as is concluded from Table 9.4, and $V_m^{\vdash\tau\dashv}(s'_k) \approx 0$ for $|s_k| > J = 10$, as can be seen in Fig. 9.21.

The above assumptions imply that $\widehat{C}_X^{(d)}(s'_k)$, being discretely smoothed, is locally stationary with approximative moment functions $\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_k)\big) \approx C_X(s'_k)$ and $\mathrm{Cov}\big(\widehat{C}_X^{(d)}(s'_k), \widehat{C}_X^{(d)}(s'_{k+l})\big) \approx \big(C_X^2(s'_k)\big)\big(H{\circledast}H(l\Delta s_k)\big)$, as approximated in (9.12), (9.14) and (9.15).

In a first step, $\widehat{C}_X^{(d)}(s'_k) = \widehat{C}_X^{(d)}(k/(2N))$ is substituted with $Y(l/(2N))$, $k, l = -(N-1), \ldots, -1, 0, 1, \ldots, N$, and it is assumed that $Y(l/(2N))$ is a locally stationary stochastic process having $\mu_Y(l/(2N)) \approx C_X(l/(2N))$ and $c_Y(\ell/(2N)) \approx \big(C_X^2(\ell/(2N))\big)\big(H{\circledast}H(\ell/(2N))\big)$ as its moment functions. The parameter of this process (and thus also the argument of its moment functions) is $l/(2N)$, i.e., the discrete frequencies. The spectrum $C_Y(u)$ of $Y(l/(2N))$ is obtained as the Fourier +i-transform of $c_Y(\ell/(2N))$ in (9.131). A +i-transform is reconcilable with the conventions in Sect. 6.3, because $c_Y(\ell/(2N))$ is a function in the frequency domain.

$$\begin{aligned} \ell &= -(N-1), \ldots, -1, 0, 1, \ldots, N \\ u &= -(N-1), \ldots, -1, 0, 1, \ldots, N \end{aligned} \tag{9.130}$$

$$C_Y(u) \approx \frac{1}{2N} \sum_{\ell=-(N-1)}^{N} \hat{c}_Y(\ell/(2N)) \mathrm{e}^{\mathrm{i}2\pi(\ell/(2N))u} \tag{9.131}$$

$$\begin{aligned} &\approx \frac{1}{2N} \sum_{\ell=-(N-1)}^{N} \big(C_X^2(\ell/(2N))\big)\big(H{\circledast}H(\ell/(2N)\big)\mathrm{e}^{\mathrm{i}2\pi(\ell/(2N))u} \\ &\approx \big(C_X^2(\ell/(2N))\big)\big(h{\circledast}h(u)\big) \end{aligned} \tag{9.132}$$

(9.132) is obtained since $H{\circledast}H(\ell/(2N))$ and $h{\circledast}h(u)$, ℓ and u as in (9.130), are a discrete Fourier transform pair, as is concluded from $H{\circledast}H(r) = \sum_{u=-(N-1)}^{N-1} h{\circledast}h(u)\mathrm{e}^{-\mathrm{i}2\pi ru}$ as derived in (9.122). This property is shared with $\hat{c}_X^{(d)}(\tau)$ and $\widehat{C}_X^{(d)}(s)$ as derived in (9.6). As an aside, both pairs, $h{\circledast}h(u)$ and $H{\circledast}H(\ell/(2N))$ as well as $\hat{c}_X^{(d)}(\tau)$ and $\widehat{C}_X^{(d)}(s)$, are Fourier transform pairs as defined in (6.49) and (6.50), on condition that $h{\circledast}h(u)$ and $\hat{c}_X^{(d)}(\tau)$ are as required in (9.123) and (9.7).

In a second step, $\widehat{C}_X^{(d)}(s'_k) = \widehat{C}_X^{(d)}\big(k/(2N)\big)$ in (9.59) is substituted with $Y\big(l/(2N)\big)$, $k, l = -(N-1), \ldots, -1, 0, 1, \ldots, N$, i.e., with the locally stationary discrete-frequency stochastic process with spectrum $C_Y(u)$ as approximated in (9.132). This substitution results in the stochastic filter (9.133), $Y\big(l/(2N)\big)$ being its input, $V_m^{\vdash\tau\dashv}\big(\ell/(2N)\big)$ its impulse response function and $\widehat{C}_X^{\vdash\tau\dashv}\big(l/(2N)\big)$ its output.

$$\widehat{C}_X^{\vdash\tau\dashv}\big(l/(2N)\big) = \frac{1}{2N} \sum_{\ell=-(N-1)}^{N} \Big(V_m^{\vdash\tau\dashv}\big(\ell/(2N)\big)\Big)\Big(Y\big((l-\ell)/(2N)\big)\Big) \tag{9.133}$$

$$\text{spectrum of } \widehat{C}_X^{\vdash\tau\dashv}\big(l/(2N)\big) \approx \big(v_m^{\vdash\tau\dashv}(u)\big)^2\big(C_X^2(\ell/(2N))\big)\big(h^{\oplus}h(u)\big) \tag{9.134}$$

$$v_m^{\vdash\tau\dashv}(u) = \frac{1}{2N}\sum_{\ell=-(N-1)}^{N} V_m^{\vdash\tau\dashv}\big(\ell/(2N)\big)\mathrm{e}^{\mathrm{i}2\pi(\ell/(2N))u} \tag{9.135}$$

A stochastic LTI filter as defined in (7.84) for stationary continuous-time random functions $X(t)$ has properties (7.90), these properties, however, are shared by the filter in (9.133). Hence, the spectrum of the output $\widehat{C}_X^{\vdash\tau\dashv}\big(l/(2N)\big)$ in (9.134) is the product of the input spectrum $C_Y(u)$ derived above in (9.132) and the squared absolute value of the transfer function $v_m^{\vdash\tau\dashv}(u)$, i.e., the discrete Fourier transform of $V_m^{\vdash\tau\dashv}\big(\ell/(2N)\big)$ as obtained in (9.135). From a comparison of (9.135) with (9.57) it is concluded that $v_m^{\vdash\tau\dashv}(u)$ is identical with $v_m^{\vdash\tau\dashv}(\tau)$ and thus is a real-valued even function as required in (9.44).

In a third and final step, the variance of the lag window spectral estimator in (9.59) and (9.133), $\widehat{C}_X^{(d)}(s'_k) = \widehat{C}_X^{(d)}\big(k/(2N)\big) = Y\big(l/(2N)\big)$, $k, l = -(N-1), \ldots, -1, 0, 1, \ldots, N$, is obtained in (9.136) as the sum of the spectrum in (9.134). (9.137) follows from (9.123) when $h^{\oplus}h(u) = h^{\oplus}h(0)$ for $u = -(N-1), \ldots, -1, 0, 1, \ldots, N$ is assumed.

$$\mathrm{Var}\big(\widehat{C}_X^{\vdash\tau\dashv}(s'_k)\big) \approx \big(C_X^2(s'_k)\big)\sum_{u=-(N-1)}^{N}\big(v_m^{\vdash\tau\dashv}(u)\big)^2\big(h^{\oplus}h(u)\big) \tag{9.136}$$

$$\approx \big(C_X^2(s'_k)\big)\frac{\sum_{t=0}^{N}\big(h_N(t)\big)^4}{\Big(\sum_{t=0}^{N}\big(h_N(t)\big)^2\Big)^2}\sum_{u=-(N-1)}^{N}\big(v_m^{\vdash\tau\dashv}(u)\big)^2 \tag{9.137}$$

$$\approx \big(C_X^2(s'_k)\big)\frac{\sum_{t=0}^{N}\big(h_N(t)\big)^4}{\Big(\sum_{t=0}^{N}\big(h_N(t)\big)^2\Big)^2}\frac{1}{\mathcal{B}_\star\big(V_m^{\vdash\tau\dashv}(s)\big)} \tag{9.138}$$

The final approximation in (9.138) is arrived at by substituting the autocorrelation width of the smoothing kernel as defined in (9.129). The above derivation is borrowed from [108].

9.5.6 Spectral Estimation in R

When a continuous spectrum is estimated, tapering and the width of the smoothing kernel are the important issues. Tapering reduces the leakage due to the side lobes of the spectral window and thus allows for computing a direct spectral estimator that has a negligible bias for frequencies a larger distance away from the peaks in a spectrum having a large dynamic range, as is concluded from (9.12). The width of a smoothing kernel applied to a direct spectral estimator controls both (i) a possible bias due to smoothing as

implied by (9.72) and (ii) the variance of the estimator as implied by (9.77) and (9.78). The width of the smoothing kernel is obtained from diagnostics (6.115) and (6.116) on condition that the bandwidth of the spectrum to be estimated is known. If, however, the bandwidth is not known, then the width of the smoothing kernel can be assessed using the procedures introduced in Sect. 9.3.5.

Consequently, any spectral estimation method should allow for tapering and choosing the width of either the smoothing kernel or a corresponding lag window. The width of the smoothing kernel has to be chosen when a direct spectral estimator is discretely smoothed using definition (9.28). Alternatively, the width of the lag window is required for computing an estimator as proposed in the remarks to (9.44), i.e., by Fourier transforming the product of (i) the empirical covariance function calculated from tapered observations and (ii) the lag window.

In R, a continuous spectrum can readily be estimated using any of the following methods:

1. applying the techniques introduced in Sect. 5.4 to fit a linear process and thereafter substituting the empirical coefficients obtained for the theoretical ones in the spectra derived in Sect. 7.4.2 (parametric estimation, an example is given in Sect. 9.2.5)
2. pre-whitening as introduced in Sect. 9.2.5
3. computing the empirical correlation function $\hat{c}_X^{(d)}(\tau)$ (9.7) from tapered observations using `acf()` (Problem 9.5), multiplying $\hat{c}_X^{(d)}(\tau)$ with a lag window as required in (9.44), and finally applying `fft()` to Fourier transform the product thus obtained (as proposed in the remarks to (9.41), (9.42), (9.43), and (9.44))
4. applying `spec.pgram()` to compute a discretely smoothed direct spectral estimator.

`spec.pgram()` computes a direct spectral estimator (or a periodogram when no data taper is applied) and then smoothes the estimator thus obtained using either a modified discrete Daniell kernel (this is the default) or another kernel. `spec.pgram()` allows for computing spectral estimates for multivariate time series and also for plotting the estimates. Its code is available from the internet address given in [114]. However, `spec.pgram()` has the following disadvantages:

1. it computes the variance of a discretely smoothed direct spectral estimator without taking into account the inflation due to the data taper (i.e., assuming $\Xi(h) = 1$ in (9.77) and (9.78)): the degrees of freedom for the chi-square approximation are not reconcilable with (9.83) and (9.84)
2. the bandwidth as computed by `spec.pgram()` seems to be too small when compared with the forms of a discretely smoothed spectral estimate.

Hence, the author prefers `spec.univariate()` (its code is given below) to `spec.pgram()` provided that the time series under analysis is univariate. In

`spec.univariate()`, the variance (and thereafter the degrees of freedom in the chi-square approximation) of the smoothed estimator is computed using (9.77). There, the autocorrelation width of the discrete smoothing kernel is required, which can be computed using (9.129) as implemented in R function `awidthdsk()`.

```
awidthdsk <- function(k, N)
{
  ## H. Gilgen, Institute for Atmosphere and Climate Science ETH
  ## Version 1.0, May 2005
  ## computes the autocorrelation width of a discrete smoothing
  ## kernel assuming a unit sampling interval for approximating
  ## the variance of a lag window spectral estimator using (9.77)
  ## k: vector with an odd number of weights in the kernel
  ## N: number of observations
  M <- (length(k)-1)/2     #prepare for computing
  rpad <- N - (M+1)        #the lag window per-
  lpad <- N - M            #taining to this kernel
  NN <- 2*N
  v <- k
  vpad <- c(rep(0,lpad),v)
  vpad <- c(vpad,rep(0,rpad))
  vpad <- c(vpad[(N+1):NN], vpad[1:N])    #re-order for fft()
  m <- floor(NN/2)
  vlagw <- fft(vpad, inverse=T)           #lag window
  vlagw <- c(vlagw[(m+1):NN],vlagw[1:m])  #re-order
  #lags <- -m:(m-1)                       #for plotting
  #autocorrelation width of lag window (9.129)
  aw <- 1/sum( Re(vlagw[2:NN])*Re(vlagw[2:NN]) )
  #this is the lag window of a discretely smoothed direct spectral
  #estimator for frequencies with distance 1/(2*N) as in (9.59)
  aw <- 2*aw
  #return results
  awdsk.out <- list(kernel = k, N = N, aw = aw)
  class(awdsk.out) <- "awidthdsk"
  return(awdsk.out)
}
```

R function `awidthdsk()` as above is used in `spec.univariate()` for computing the variance of the smoothed estimator. Further, `spec.univariate()` applies `spec.taper()` (so does `spec.pgram()`) to compute a cosine taper (6.111,1). `spec.univariate()` provides for only one smoothing kernel (the modified discrete Daniell kernel (9.29)) because the width of the kernel is important, not its form, as argued in the remark to (9.44) and in the second last paragraph in Sect. 9.3.3.

```
spec.univariate <- function(x,taper=0.05,spans=1,pad=0,
```

```
                              confint=0.95,fast=FALSE)
{
  ## H. Gilgen, Institute for Atmospheric and Climate Science ETH
  ## Version 1.0, March 2003
  ## Version 1.1, May 2005
  ## discretely smoothed direct spectral estimator
  ## x:     a time series in R, demeaned and detrended
  ## taper: p-value of cosine taper
  ## spans: number of discrete frequencies in moving average
  ##        applied for smoothing the direct spectral estimator
  ## pad:   factor for zero-padding, i.e., 0 for no padding,
  ##        1 for padding with N zeros, N the length of x, etc.
  ## fast:  FALSE: calculate spectrum for N or N obtained
  ##               subsequent to padding
  ##        TRUE:  zero-pad to favour factorising for fft()
  xfreq <- frequency(x)
  x <- as.vector(x)
  N <- length(x)
  Nprime <- N                #no zero-padding yet
  ## construct variance normalised cosine taper (9.2,1)
  ## as discussed in the remarks to Fig. 9.4
  h <- rep(1,N)
  h <- as.vector(spec.taper(h, taper)) #taper = 0: periodogram
  sh2 <- sum(h*h)            #for normalising
  sh4 <- sum(h**4)           #for computing (9.74)
  h <- h/sqrt(sh2)           #variance normalised taper
  x <- x*h                   #now tapered
  ## zero-padding using definition (9.20) as discussed
  ## in the remarks to Fig. 9.25
  if (pad > 0) {
    NOpad <- N*pad
    Nprime <- N + NOpad
    x <- c(x,rep(0,NOpad))     #now zero-padded
  }
  ## additional zero-padding to favour factorising in fft()
  NN <- if(fast) nextn(Nprime) else Nprime
  x <- c(x,rep(0,(NN-Nprime)))
  ## frequencies, Fourier transform and direct spectral estimator
  ## interpolate estimator at frequency 0 because very small
  ## due to demeaning or detrending
  m <- floor(NN/2)
  sk <- seq(from=0, by=xfreq/NN,length = (m+1))     #for non-unit
  xt <- fft(x,inverse=T)                            #sampling
  spec <- (1/xfreq)*(Mod(xt))*(Mod(xt))             #interval
```

```
spec[1] <- 0.5*(spec[2] + spec[NN])
## smooth direct spectral estimator over spans frequencies
## using a modified discrete Daniell kernel having width 2*M
## in units of 1/NN, see the remarks to (9.29).
## variance reduction due to smoothing, calculated using
## autocorrelation width of kernel in units of (1/NN)
## chisqdf: degrees of freedom for chi-square approximation
## as required in (9.77), (9.78) and (9.84)
if (spans > 1) {
  M <- floor(spans/2)
  v <- rep( (1/(2*M)), 2*M+1)
  v[1] <- v[1]/2
  v[2*M+1] <- v[2*M+1]/2
  spec <- as.vector(filter(spec,v,method="convolution",
                      circular=T,sides=2))
  autocorrwidth <- awidthdsk(v,NN)
  awidth <- autocorrwidth$aw     #autocorrelation width (9.129)
  bwidth <- (2*M)*(1/NN)         #bandwidth (9.90)
  varred <- (1/N)*((N*sh4)/(sh2*sh2))*(1/awidth)
  chisqdf <- 2/varred
} else {
  bwidth <- (1/N)*((N*sh4)/(sh2*sh2))     #bandwidth (9.92)
  varred <- NA       #no smoothing: no reduction of variance
  chisqdf <- 2       #(8.24) and (9.19): not for first
  M <- 1             #and possibly last frequency in sk
  v <- NA
  awidth <- NA
}
awidth <- awidth*xfreq
bandwidth <- bwidth*xfreq
spec <- spec[1:(m+1)]     #for positive frequencies only
## confidence intervals (9.87) and (9.88)
pconf <- (1-confint)/2.0
plow <- rep((1-pconf),(m+1))
pupp <- rep(pconf,(m+1))
cilow <- (chisqdf/qchisq(plow,chisqdf))*spec
ciupp <- (chisqdf/qchisq(pupp,chisqdf))*spec
cidblow <- 10*log10(cilow[M+1]) - 10*log10(spec[M+1])
cidbupp <- 10*log10(ciupp[M+1]) - 10*log10(spec[M+1])
## above approximations apply for frequencies not close to
## first and (possibly) last frequency in sk, as required in
## (9.73,4) and (9.32,2)
for (i in 1:M) {
  cilow[i] <- NA
```

```
    cilow[(m+1)-i+1] <- NA
    ciupp[i] <- NA
    ciupp[(m+1)-i+1] <- NA
  }
  ## return results
  spg.uni.out <- list(freq = sk, spec = spec,
      confintvectorlow = cilow, confintvectorupp = ciupp,
      confintdblow = cidblow, confintdbupp = cidbupp,
      df = chisqdf, acwidth.df = awidth, bandwidth = bandwidth,
      kernel = v, N.timeseries = N, N.used = NN,
      numb.freq = m, p.costaper = taper, zeropad.factor= pad,
      numb.freq.smoothed = spans)
  class(spg.uni.out) <- "specuni"
  return(spg.uni.out)
}
```

Spectral estimates obtained from `spec.univariate()` and `spec.pgram()` are identical for the frequencies not excluded in (9.73,4) and (9.32,2), because identical tapers and smoothing kernels are applied; small differences are, however, found in the approximations of the degrees of freedom for the chi-square approximation and the bandwidth (subsequent to a multiplication of the bandwidth as obtained from `spec.pgram()` with $\sqrt{12}$).

`spec.univariate()` and `spec.pgram()` are distinct in their arguments and their results. Consequently, `plot.spec()` cannot be applied for plotting the estimates obtained from `spec.univariate()`. With a few modifications in its last lines however, `spec.univariate()` can be made reconcilable with `plot.spec()`.

9.6 Problems

9.1. Let $I_X(s)$ be a periodogram as defined in (8.1,4) using a segment in a discrete-time stationary stochastic process (X_t). Then (8.19) allows for calculating its expectation function $\mathrm{E}I_X(s)$ on condition that its covariance function $c_X(\tau)$ is known. Assume, additionally, that the (X_t) is reconcilable with model (9.1), having covariance function $c_X(\tau) = \sigma_W^2 \sum_{u=-\infty}^{\infty} b_u b_{u+\tau}$. $c_X(\tau)$ can be derived as in (2.26) even though the impulse response sequence (b_t) is two-sided in (9.1), since the sums do exist for all $-(N-1) \le h \le (N-1)$. Alternatively, by applying (7.90,4) and (6.48,6), $c_X(\tau)$ can also be calculated from spectrum (9.1,2).

Substituting $c_X(\tau) = \sigma_W^2 \sum_{u=-\infty}^{\infty} b_u b_{u+\tau}$ in (8.19),

$$\mathrm{E}\big(I_X(s)\big) = \sum_{\tau=-(N-1)}^{N-1} \Big(1 - \frac{|\tau|}{N}\Big) \left(\sigma_W^2 \sum_{u=-\infty}^{\infty} b_u b_{u+\tau}\right) \cos(2\pi s\tau) \qquad (9.139)$$

is arrived at as the Fourier transform of the product of $f(\tau) = (1 - |\tau|/N)$ and $g(\tau) = \sigma_W^2 \sum_{u=-\infty}^{\infty} b_u b_{u+\tau}$. Both functions are even (6.32): (i) $f(\tau)$ is the

even triangle sequence (6.56) and (ii) $g(\tau)$, being the autocorrelation of (b_t) in (9.1), is even as is concluded from the remarks to (6.103). Please derive, from (9.139), $\mathrm{E}\big(I_X(s)\big) = F_N(s) * C_X(s)$, for $-1/2 \le s \le 1/2$, $C_X(s)$ the spectrum of model (9.1).

9.2. Parseval's identity for Fourier transform pairs $(f_t) = \mathcal{F}_{-\mathrm{i}}\big(F(s)\big)$ as in (6.49) and (6.50) is used to derive the mean square approximation (6.127) of the low-pass filter. This version of Parseval's identity can also be used to derive $\int_{-1/2}^{1/2} \mathcal{H}_N^{(d)}(s)\mathrm{d}s = 1$, $\mathcal{H}_N^{(d)}(s)$ the spectral window as defined in (9.2,2).

9.3. Please find in file `/path/problem93.R` R expressions which can be used to generate cosine tapers, their normalised autocorrelations and their spectral windows as plotted in Fig. 9.2. Apply these R expressions to calculate the widths at half height of the main lobes of spectral windows for $N = 64$ and $p = 0.15, 0.20, 0.25, 0.30$ and 0.50. Then calculate and plot cosine tapers, their normalised autocorrelations and spectral windows for $p = 0.05$ and $N = 1024, 8196, 65536$.

9.4. Let $\widehat{h}_N(t)$ be a cosine taper as defined in (6.111,1). To $\widehat{h}_N(t)$ pertain its kernel $\widehat{H}_N(s) = \mathcal{F}_{-\mathrm{i}}\big(h_N(t)\big)$ calculated using definition (6.110) and its spectral window $\widehat{\mathcal{H}}_N^{(d)}(s) = \big(1/\Sigma(\widehat{h^2})\big)\widehat{H}_N(s)\overline{\widehat{H}_N(s)} = \big(1/\Sigma(\widehat{h^2})\big)\mathcal{F}_{-\mathrm{i}}\big(\widehat{h}_N \star \widehat{h}_N(t)\big)$, with $\Sigma(\widehat{h^2}) = \sum_{t=0}^{N-1}\big(\widehat{h}_N(t)\big)^2$, as defined in (9.2,2). Show that $\widehat{\mathcal{H}}_N^{(d)}(s)$, $N = 1, 2, 3, \ldots$ is a defining sequence for $\delta(s+n)$, $n = \ldots, -1, 0, 1, \ldots$, i.e., the periodic version of the delta function in the remarks to (6.62). Hints:

- $\widehat{H}_N(0) = \sum_{t=0}^{N-1} \widehat{h}_N(t)$ is implied by (6.48,1). Using $\cos(x) = -\cos(\pi - x)$, you will obtain $\sum_{t=0}^{q-1} \cos((\pi/(2q))(2t+1)) = 0$, q as defined in (6.111). Thus $\sum_{t=0}^{N-1} \widehat{h}_N(t) = N - q$ is arrived at.
- $\sum_{t=0}^{q-1}\big(\cos((\pi/(2q))(2t+1))\big)^2 = q/2$ can be obtained by applying the trigonometric identities and the sums in (6.10). You will arrive at $\Sigma(\widehat{h^2}) = \sum_{t=0}^{N-1}\big(\widehat{h}_N(t)\big)^2\big) = N - (5/4)q$.

9.5. Generate in R vector `x` a time slice of length $N = 4096$ in a simulated realisation of the AR[4] process (X_t) defined in Fig. 7.14. From `x`, calculate the empirical covariance function $\hat{c}_X^{(d)}(\tau)$ as defined in (9.7) by applying the following R expressions:

```
h <- rep(1, N)
h <- spec.taper(h, 0.10)          #cosine taper (6.111,1)
h <- h/sqrt(sum(h*h))             #normalised (9.2,1)
#autocorrelation (6.103) of x*h computed using acf(),
#albeit acf() normalises with (1/N) as required in (2.1,3),
#this is compensated below
ac <- acf(x*h, lag.max=N, plot=F, type="covariance")
```

```
lags <- -(N-1):N                #(9.25)
tmcov <- rev(ac$acf[2:N])       #revert order
tmcov <- c(tmcov, ac$acf)       #all lags
tmcov <- tmcov*N                #undo normalisation.
```

From $\hat{c}_X^{(d)}(\tau)$ thus obtained in `tmcov`, a direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ can be obtained using (9.23), which is identical to $\widehat{C}_X^{(d)}(s'_k)$, with $\Delta s'_k = 1/N' = 1/(2N)$, $N' = 2N$, calculated by multiplying `x` with a cosine taper and thereafter zero-padding as required in (9.20,1). Hint: R expressions for discrete Fourier transforms are given in Sect. 6.9.8.

9.6. R expressions that can be used to generate the plots of $H \circledast H(r)$ in Fig. 9.6 are available in `/path/problem96.R`. Use these expressions to generate plots of $H \circledast H(r)$ for cosine tapers having parameters $N = 128, 512, 2048$ and $p = 0.0, 0.05, 0.1, 0.3, 0.5$. Draw the frequencies $n\Delta s_k$, $n = 1/2, 1, 3/2, 2, 5/2, \ldots$, in your plots.

9.7. Simulate realisations of the AR[4] model in Fig. 7.14 with increasing lengths $N = 512$, $N = 1024$, $N = 2048$, etc. Using these realisations, obtain Burg's and ML estimates for the model parameters as demonstrated in Sects. 5.2.1 and 5.2.3. Substitute your estimates for the model parameters in the R expressions in the remarks to Fig. 7.14 and plot the empirical spectra thus obtained.

9.8. In Table 5.1, AR[2] model estimates are given for the de-meaned time series of the wind speeds in a turbulent atmospheric flow and of the amounts of the product of a chemical reaction. Substitute these estimates in (7.97) and plot the empirical spectra thus obtained.

9.9. Why are the convolutions in (9.28) not stochastic filters as defined in (7.84)?

9.10. Calculate a direct spectral estimator from a time slice of length $N = 16384$ in a realisation of the AR[4] model in Fig. 7.14. Apply a cosine taper with $p = 0.05$ and therafter zero-pad such that the estimator is obtained for frequencies s'_k with $\Delta s'_k = 1/N'$, $N' = 2N$. Thereafter smooth discretely over a band being approximately $(1/2)\mathcal{B}_\triangle\big(C_X(s)\big) = 0.00474/2$ wide, i.e., half of the bandwidth of the spectrum to be estimated.

9.11. Show that $\sum_{-\infty}^{\infty} v_m(\tau) = \sum_{-\infty}^{\infty} v_m^{\vdash\tau\dashv}(\tau) = 1$ with $v_m(\tau)$ and $v_m^{\vdash\tau\dashv}(\tau)$ as defined in (9.40), (9.41) and (9.44).

9.12. Approximation (9.38) for the variance of a discretely smoothed direct spectral estimator (repeated below) can be further reduced (second line below)

$$\begin{aligned}\mathrm{Var}\big(\widehat{C}_X^{(m)}(s'_k)\big) &\approx C_X^2(s'_k) \sum_{l=-M^{\oplus}}^{M^{\oplus}} \big(H \circledast H(|l|\Delta s'_k)\big) \sum_{\ell=-M}^{M-|l|} \big(V_m(s'_\ell)\big)\big(V_m(s'_{\ell+|l|})\big) \\ &\approx C_X^2(s'_k) \sum_{l=-M^{\oplus}}^{M} \big(H \circledast H(|l|\Delta s'_k)\big) \times \sum_{j=-M^{\oplus}}^{M} V_m(s'_j) \qquad (9.140)\end{aligned}$$

under suitable assumptions. Apply (9.140) to approximate the variance function of the smoothed direct spectral estimator as plotted in Figs. 9.15 and 9.16 and compare your result with the one obtained in the remarks to (9.38). Having calculated both approximations for this example, the assumptions under which (9.140) follows from (9.38) can be easily formulated.

9.13. Let $V_{dD,M}(s'_k)$ be the modified discrete Daniell kernel as plotted in Fig. 9.19 (b). Calculate a discrete Fourier transform of $V_{dD,M}(s'_k)$ using definition (6.22,4). Compare the resulting sequence with the one in Fig. 9.19 (a).

9.14. Show (9.128) using definition (6.101).

9.15. Calculate $\Xi(h)$ as in Table 9.3 for cosine tapers (6.111,1) implementing definition (9.74) in a few R expressions. Perform the calculations for $N = 64$, $N = 512$ and $N = 4096$. However, these $\Xi(h)$ could be readily calculated on condition that, in addition to $\Sigma(\widehat{h^2}) = \sum_{t=0}^{N-1}\big(\widehat{h_N}(t)\big)^2\big) = N - (5/4)q$ as obtained in Problem 9.3, the sum $\Sigma(\widehat{h^4}) = \sum_{t=0}^{N-1}\big(\widehat{h_N}(t)\big)^4$ could be expressed using the parameters of the cosine taper.

9.16. The seismogram of the earthquake as plotted above in Fig. 9.26 and available in file `/path/earthquake.dat` is not stationary. Assume that this seismogram is a deterministic function and calculate its spectrum.

9.17. On May 11, 1998, at 10:13:41.7 (hours:minutes:seconds UTC), India performed a nuclear test that produced an earthquake (27.11^0 N, 71.8^0 E and depth 0 km) having magnitude 5.1. The seismogram of this earthquake recorded at Nilore station is plotted in Fig. 9.37. This seismogram is, similar to the one in Fig. 9.26 recorded using the same seismograph at Nilore station, a plot of the speed of the earth in vertical direction (in ms^{-1} (meters/second)) against time (in s (seconds)). Its sampling interval is $\Delta t = 0.025$ s and it is available in file `/path/nucleartest.dat` downloaded from the IRIS data centre [72]. The distance between the nuclear explosion and the recording seismograph in Nilore is approximately 720 km.

In the seismogram of this nuclear test, no shear waves can be seen: the amplitudes of the shear waves are negligibly small when compared with the amplitudes of the compression waves. A seismogram with shear waves having small amplitudes is reconcilable with the simplest model of an explosion. In this model, a blasting composition mounted in a cavern is ignited. Subsequent

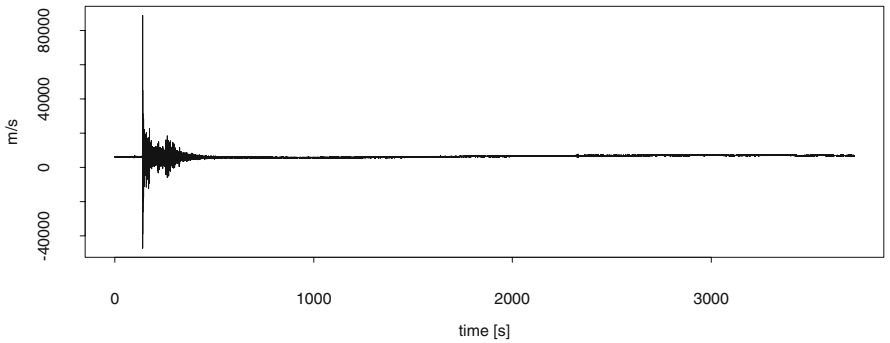


Fig. 9.37. Seismogram of a nuclear test performed May 11, 1998 in India (27.11^0 N, 71.8^0 E) recorded at Nilore, Pakistan (33.65^0 N, 73.26^0 E). The record starts 80 s before the compression waves emanated by the explosion arrive at the Nilore seismograph.

to the explosion, hot gases expand with high speed in the cavern and induce compression waves, but not shear waves, in the rock containing the cavern. There is a second difference between the seismograms in Fig. 9.26 and 9.37: in the seismogram of the nuclear explosion, the amplitude of the surface waves are much smaller than the amplitudes of the compression waves.

Assume that the seismogram in Fig. 9.37 is a deterministic function and calculate the spectrum of this function. Compare the spectrum obtained with the one calculated in Problem 9.16.

Seismograms of the Indian nuclear test performed on May 11, 1998, can be easily distinguished from a seismogram of an earthquake. Not all nuclear tests, however, leave "fingerprints" as clear as those of the Indian one in a seismogram [133].

9.18. Estimate the spectrum of marine microseisms using a time slice from the seismogram plotted in Fig. 9.37. Select a time slice which is longer that 65 s, i.e., the length of the time slice as plotted in Fig. 9.27 used to calculate the direct spectral estimators in Fig. 9.29.

If the mean of the seismogram is not constant in the time slice selected, then a spectrum can be estimated using the residuals of a linear model fitted as demonstrated in Chap. 3. It is also possible to detrend a time series by fitting local linear models as demonstrated in Fig. 6.32; differencing, however, changes the spectral properties of a time series, as is concluded from (6.122), (6.123) and Fig. 6.30 as well as the results obtained in Problem 9.24 and Sect. 10.3.2

9.19. Please fit an ARMA$[p, q]$ model to the vertical velocities in the turbulent atmospheric flow as plotted in Fig. 9.34 and then calculate the empirical spectrum of the estimate as proposed in Sect. 9.2.5. Thereafter compare the

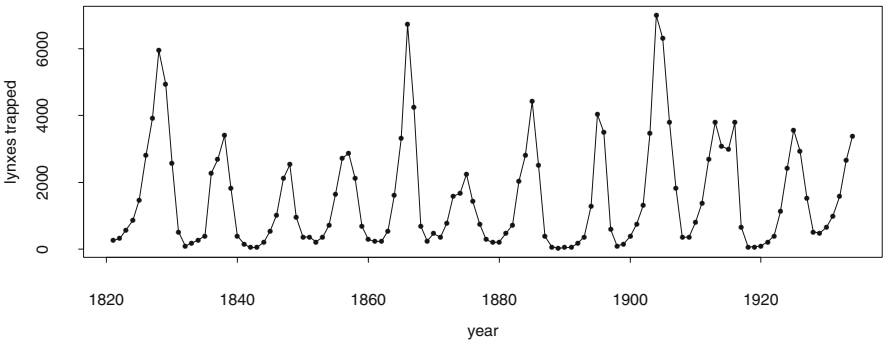


Fig. 9.38. Number of lynxes trapped in the Mackenzie River District in North West Canada in the years 1821 - 1934.

parametric estimate thus obtained with the smoothed temporal periodogram in Fig. 9.35.

9.20. The number of lynxes caught in the period from 1821 through to 1934 in the Mackenzie River District in North West Canada as plotted in Fig. 9.38 is obtained using R expressions

```
plot(lynx,type="n",xlab="year",ylab="lynx trapped")
lines(lynx,lty=1)
points(lynx)
```

since this time series is available in R. Time series `lynx` is comprehensively described in [28], and in [129] it is shown that the cyclic abundance of lynxes in Canada, which is the result of a predator-prey relationship, depends on the climatic region. Does this time series stem from a process reconcilable with model (9.1)? Please obtain an answer to this question by (i) fitting an ARMA$[p, q]$ model to the lynx time series and calculating a parametric estimate for the spectrum of this process and (ii) calculating a discretely smoothed direct spectral estimator.

9.21. Assume that the sun spot numbers as obtained in Problem 2.31 stem from a realisation of a stationary stochastic process. Estimate the spectrum of this process.

9.22. Parametric estimates for the spectrum of the SO index are plotted in Fig. 9.11 (b). These spectral estimates are calculated using the estimates in Table 5.3. Calculate a smoothed direct spectral estimator for this spectrum using the time series that is made available in R in the remarks to Fig. 5.14.

9.23. Estimate the spectrum of the amount of a product formed during a chemical reaction occurring in a batch reactor using the demeaned time series plotted in Fig. 5.3.

9.24. When a stochastic process is not stationary in its expectation function then it does not have a spectrum as defined in (7.51). Consequently, a smoothed direct spectral estimator calculated from observations stemming from a stochastic process having a trend and/or other systematic fluctuations is meaningless: it is simply not clear what it estimates. Nevertheless, non-stationarities can be removed from such observations using the procedures (5.58), i.e., detrending by calculating (usually the first) differences of the observations or fitting a linear model and then using its residuals for the further analysis.

If a time series (x_t), $t = 0, 1, \ldots, N-1$, becomes stationary in $\hat{\mu}_X(t)$ subsequent to detrending then it is assumed that also (X_t) becomes stationary in $\mu_X(t)$ subsequent to applying the operations successfully used to detrend (x_t). However, time series detrended using different methods are often not identical, as is obvious from Figs. 5.10 and 5.11 where the logarithm of the atmospheric CO_2 is detrended. These differences are inherited by empirical moment functions and also spectral estimators calculated from the detrended time series. For a demonstration, calculate smoothed direct spectral estimators using the detrended time series in Figs. 5.10 and 5.11. Compare your results with those obtained in Sect. 10.3.2 for estimators of a spectrum having a discrete part.

10 Estimators for a Spectrum Having a Discrete Part

This chapter introduces the reader to the estimation of mixed spectra as defined in (7.71,3). The models proposed in this chapter are the sum of (i) a harmonic process as defined in (7.30) (or alternatively in (7.38), (7.39) and (7.40)) and (ii) a process having a continuous spectrum. These models have a mixed spectrum. For example, the realisation in Fig. 7.9 stems from the most straightforward process having a mixed spectrum: it is the sum of a harmonic process with only one oscillation and a white noise process.

The discrete part of a mixed spectrum is readily estimated on condition that the frequencies of the harmonic oscillations are known to be identical with the Fourier frequencies as defined in (6.21,1). This property favouring an estimation is shared by processes having fluctuations with a known and constant period p: the fluctuations are approximated by a trigometric oscillation having frequency $1/p$ and the pertaining harmonics. These oscillations have frequencies which are in the set of the Fourier frequencies, on condition that $N = l \times p$, where l, p are integer numbers and N is the length of the time series observed. A model having oscillations with Fourier frequencies and a spectrum with a constant continuous part is introduced in Sect. 10.1. The estimation of such a model is an alternative to the estimation of a seasonal model as introduced in Sect. 3.4.1.

In the majority of cases however, the model to be estimated contains a harmonic process with trigonometric oscillations whose frequencies are not known. Such models are introduced in Sect. 10.2. The unknown frequencies of the oscillations in these models are approximated by trial and error: a direct spectral estimator is calculated, subsequent to tapering and zero-padding the observations, for frequencies s'_k having a small $\Delta s'_k$ as defined in (9.20), and in the estimator thus obtained, pilot frequencies for the unknown model frequencies are found by applying the diagnostics introduced in the remarks to (10.18). After substituting the pilot frequencies for the model frequencies, the trigonometric oscillations are estimated using linear least squares. Alternatively, a non-linear least squares estimation can help to find the unknown frequencies.

Sect. 10.3 contains the supplements and Sect. 10.4 the problems.

The estimation of a discrete or mixed spectrum is comprehensively dealt with in [108].

10.1 Estimating Oscillations with Fourier Frequencies

This section introduces a model having a mixed spectrum with a discrete part that can readily be estimated. This model is thereafter applied to the detrended Basel temperature series. Continuous, discrete and mixed spectra are defined in (7.71).

10.1.1 The Model

The harmonic process (X_t) is defined in (7.30), (X_t) being a finite sum of random trigonometric oscillations. If (X_t) is written as in (7.38) then coefficients A_j and B_j are random variables whereas frequencies s_j are not random. Since (X_t) is a discrete-time process, the frequencies s_j of its oscillations are all in interval $-1/2 \le s_p \le 1/2$ as required in (7.64). If A_j and B_j are normally distributed, then they are also mutually independent for all pairs (A_j, A_k) and (B_j, B_k), $j \neq k$, as well as (A_j, B_k), $j \neq k$ or $j = k$, as is concluded from (7.39), (7.40) and (7.41). (X_t) is stationary since each of its oscillations is stationary, a result obtained in Problem 7.6. The covariance function of the harmonic process, $c_X(\tau) = \sum_{j=1}^{n} \sigma_j^2 \cos(2\pi s_j \tau)$, is obtained from (7.31), using the result of Problem 7.6 and the moments of the A_j and B_j in (7.39), (7.40) and (7.41). Thereafter, $C_X(s) = \sum_{j=1}^{n} \big((\sigma_j^2/2)\delta(s - s_j) + (\sigma_j^2/2)\delta(s + s_j)\big)$ follows, using the result of Problem 6.19 as demonstrated in (7.73).

A harmonic process is often not in agreement with observations made in Geosciences. Hence, the sum of a harmonic process (7.38) and a white noise process (2.10) is defined in (10.1,1).

1. *Let* $X_t = \sum_{p=1}^{n} \big(A_p \cos(2\pi s_p t) + B_p \sin(2\pi s_p t)\big) + W_t$ *be a discrete-time stationary stochastic process with* $\mathrm{E}A_p = \mathrm{E}B_p = 0$, $\mathrm{Var}A_p = \mathrm{Var}B_p = \sigma_p^2$, $\mathrm{Cov}(A_p, B_p) = 0$, $\mathrm{Cov}(A_p, B_q) = \mathrm{Cov}(B_p, B_q) = \mathrm{Cov}(A_p, A_q) = 0$, *for* $p \neq q$, (W_t) *being a white noise process with* $\mathrm{E}W_t = 0$, $\mathrm{Var}W_t = \sigma_W^2$, $c_W(\tau) = \sigma_W^2$ *for* $\tau = 0$ *and* $c_W(\tau) = 0$ *for* $\tau \neq 0$, *and* $\mathrm{Cov}(W_t, A_p) = \mathrm{Cov}(W_t, B_p) = 0$. *Then*
2. $c_X(\tau) = \sum_{p=1}^{n} \sigma_p^2 \cos(2\pi s_p \tau) + c_W(\tau)$ *and* (10.1)
3. $C_X(s) = \sum_{p=1}^{n} \left(\frac{\sigma_p^2}{2}\delta(s - s_p) + \frac{\sigma_p^2}{2}\delta(s + s_p)\right) + \sigma_W^2$ *are the covariance function and the spectrum of* (X_t).
4. *It is assumed that frequencies* $s_p \in \{s_k\}$ *as in (6.21,1), i.e., are in the set of the Fourier frequencies pertaining to a time slice* (x_t), $t = 0, 1, \ldots, N-1$, *in a realisation of* (X_t).
5. *A process* (Y_t) *having a non-zero expectation function is reconcilable with model* (X_t) *provided that* $(X_t) = (Y_t) - \mu_Y(t)$.

The reason for defining model (10.1,1) is that a harmonic process does not provide for a measurement error and/or random fluctuations from one time

point to another, i.e., with period Δt, in the process under analysis. For example, the residuals of the component model fitted in Sect. 3.4.2 to the monthly values in the Basel temperature series are assumed to stem from a white noise process.

Since $\mathrm{Cov}(W_t, A_p) = \mathrm{Cov}(W_t, B_p) = 0$ is stipulated in (10.1,1), the A_p and B_p in the harmonic process, as well as the W_t in the white noise process, are required to be uncorrelated, borrowing from the definition of example process no. 4 in Sect. 7.3.4. Consequently, the covariance function and spectrum of the model can be derived as demonstrated in (7.80) and (7.81). The bandwidth $\mathcal{B}_{\triangle}\big(C_X(s)\big)$ of this mixed spectrum is the smallest distance between the frequencies s_p, as required in (7.83,2).

Assuming that observations (x_t), $t = 0, 1, \ldots, N-1$, stem from a realisation of model (10.1,1), can estimates for

1. the number n of oscillations in the model
2. their frequencies s_p and their power (the variance of their coefficients) σ_p^2 for $p = 1, \ldots, n$, and also
3. the variance of the white noise process σ_W^2

be calculated? An estimation is straightforward on condition that the s_p are in the set of the s_k, as required in (10.1,4). In this case, $n = m$ is arrived at, m as defined in (6.21,1), and all σ_k^2 as well as σ_W^2 are estimated, σ_k^2 pertaining to oscillations with Fourier frequencies s_k. Under this assumption, model (10.1,1) becomes, with s_k substituted for s_p and m for n, a linear model whose coefficients are estimated by minimising

$$\Sigma\big(\hat{r}_t^2\big) = \sum_{t=0}^{N-1} \left(x_t - \sum_{k=1}^{m} \big(A_k \cos(2\pi s_k t) + B_k \cos(2\pi s_k t) \big) \right)^2 \tag{10.2}$$

to arrive at the least squares estimates in (10.3) and (10.4):

$$\widehat{A}_X(s_k) = \begin{cases} (2/N) \sum_{t=0}^{N-1} X_t \cos(2\pi s_k t) & k = 1, \ldots, m-1 \ (m, N \text{ odd}) \\ (1/N) \sum_{t=0}^{N-1} X_t \cos(2\pi s_k t) & k = 0 \ (m, N \text{ even}) \end{cases} \tag{10.3}$$

$$\widehat{B}_X(s_k) = \begin{cases} (2/N) \sum_{t=0}^{N-1} X_t \sin(2\pi s_k t) & k = 1, \ldots, m-1 \ (m, N \text{ odd}) \\ 0 & k = 0 \ (m, N \text{ even}). \end{cases} \tag{10.4}$$

(10.3) and (10.4) are easily derived since, under assumption (10.1,4), the trigonometric functions in (10.1,1) become those in (6.9) and thus the orthogonality relations can be applied to the equations obtained from minimising (10.2). In practical applications, estimators $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ are computed using a discrete Fourier transform, as is concluded from (6.22) and (6.21).

The moments of estimators $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ in (10.3) and (10.4) are derived in (10.21) and (10.22) with the following results:

$$\mathrm{E}\widehat{A}_X(s_k) = A_k \qquad \text{and} \qquad \mathrm{E}\widehat{B}_X(s_k) = B_k \tag{10.5}$$

$$\mathrm{Var}\widehat{A}_X(s_k) = \begin{cases} (2/N)\sigma_W^2 & \text{for } -1/2 < s_k < 0 < s_k < 1/2 \\ (1/N)\sigma_W^2 & \text{for } s_k = -1/2, 0, 1/2 \end{cases} \tag{10.6}$$

$$\mathrm{Var}\widehat{B}_X(s_k) = \begin{cases} (2/N)\sigma_W^2 & \text{for } -1/2 < s_k < 0 < s_k < 1/2 \\ 0 & \text{for } s_k = -1/2, 0, 1/2. \end{cases} \tag{10.7}$$

From (10.5), (10.6) and (10.7) it is concluded that $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ are unbiased estimators for A_k and B_k and have variances that decrease with $1/N$ when N increases. In addition, the orthogonality relations can be used to show that $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ are not correlated, a desired property seeing that A_k and B_k are required to be non-correlated in (10.1,1).

The moments of $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ obtained above can be used to derive estimators $\hat{\sigma}_k^2$ for the variances σ_k^2 in model (10.1,1) under the above assumptions, i.e., with s_k substituted for s_p and m for n. Under these assumptions, $\sigma_k^2/2$ are the heights of the jump discontinuities located in frequencies $\pm s_k$ in the integrated spectrum $C_X^{(I)}(s)$ pertaining to model (X_t). Since model (10.1,1) requires that $\mathrm{Var}A_k = \mathrm{Var}B_k = \sigma_k^2$ and $\mathrm{E}A_k = \mathrm{E}B_k = 0$, and since exactly one value is obtained for A_k and B_k in each realisation, (10.8) is arrived at.

$$\hat{\sigma}_k^2 = \frac{1}{2}\big(\widehat{A}_X(s_k)\big)^2 + \frac{1}{2}\big(\widehat{B}_X(s_k)\big)^2 \qquad s_k \neq -1/2, 0, 1/2 \tag{10.8}$$

$$\mathrm{E}\hat{\sigma}_k^2 = \frac{1}{2}\left(\mathrm{Var}\widehat{A}_X(s_k) + \big(\mathrm{E}\widehat{A}_X(s_k)\big)^2 + \mathrm{Var}\widehat{B}_X(s_k) + \big(\mathrm{E}\widehat{B}_X(s_k)\big)^2\right) \tag{10.9}$$

$$= \frac{1}{2}\left(\frac{2}{N}\sigma_W^2 + A_k^2 + \frac{2}{N}\sigma_W^2 + B_k^2\right) = \frac{1}{2}\big(A_k^2 + B_k^2\big) + \frac{2}{N}\sigma_W^2 \tag{10.10}$$

The expectation in (10.9) holds since $\mathrm{E}X^2 = \mathrm{Var}X + (\mathrm{E}X)^2$ and (10.10) results when the moments obtained in (10.6) and (10.7) are substituted in (10.9).

The variance σ_W^2 of the white noise process in model (10.1,1) is the height of the continuous part in spectrum $C_X(s)$ and also the height of the integrated spectrum $C_X^{(I)}(s)$ for frequency $s = -1/2$. Under the above assumptions made to arrive at the estimators $\widehat{A}_X(s_k)$ in (10.3) and $\widehat{B}_X(s_k)$ in (10.4), the empirical variance $\hat{\sigma}_W^2$ of the residuals $\hat{r}_t$ in (10.2) becomes an estimator for σ_W^2. $\hat{\sigma}_W^2$ is calculated as required in (10.11).

$$\hat{\sigma}_W^2 = \frac{1}{N-2n}\Sigma\big(\hat{r}_t^2\big), \qquad \Sigma\big(\hat{r}_t^2\big) \text{ as in 10.2} \tag{10.11}$$

Model (10.1,1) contains $2n$ coefficients (except from σ_W^2). Hence, (10.11) accounts for the degrees of freedom associated with the empirical residuals $\hat{r}_t$ of this model, since the model coefficients are estimated using least squares. The degrees of freedom pertaining to the empirical residuals of a linear model are discussed in Sect. 3.3.1.

10.1.2 Diagnostics

Model (X_t) as defined in (10.1,1) contains, under assumption (10.1,4), $n \leq m$ trigonometric oscillations having frequencies $s_p \in \{s_k\}$, s_k being the Fourier frequencies as defined in (6.21) for an observed time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of (X_t). Often most oscillations are found to be negligibly small in their absolute values and thus $n \ll m$ ($\ll$ much smaller) is obtained. The absolute value of an oscillation with s_p is not negligibly small on condition that the periodogram as defined in (8.1,3) is much larger for frequency s_p than for the other frequencies: $I_X(s_p) \gg I_X(s_l) \approx I_X(s_\ell)$ with $s_p, s_l, s_\ell \in \{s_k\}$. This diagnostic is demonstrated for a deterministic example in Table 6.1 and Fig. 6.2.

In the case of model (X_t) as defined in (10.1,1), the application of this diagnostic becomes plausible when the expectation function $\mathrm{E}I_X(s)$ of a periodogram $I_X(s)$ is derived, $I_X(s)$ being a periodogram calculated as required in (8.1,4) using an observed time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of (X_t). Under general conditions, i.e., that (X_t) is a discrete-time stationary stochastic process, the expectation function $I_X(s)$ is derived in Sect. 8.5.1. Since model (10.1,1) is clearly reconcilable with these conditions, the results obtained in (8.35) and (8.21) also apply to a periodogram $I_X(s)$ calculated from a realisation of this model and thus (10.12) is arrived at by substituting the model spectrum (10.1,3) for $\mathrm{d}C_X^{(I)}(s)$ in (8.21). (10.13) follows because $\delta(x)$ is the identity element under convolution (2.28,1) for functions with a real argument as is concluded from (6.67) and (6.68).

$$\mathrm{E}I_X(s) = F_N(s) * \left(\sum_{p=1}^{n} \left(\frac{\sigma_p^2}{2}\delta(s - s_p) + \frac{\sigma_p^2}{2}\delta(s + s_p) \right) + \sigma_W^2 \right) \tag{10.12}$$

$$= \sum_{p=1}^{n} \frac{\sigma_p^2}{2}\big(F_N(s + s_p) + F_N(s - s_p)\big) + \sigma_W^2 \tag{10.13}$$

$$\mathrm{E}I_X(s_k) = \sum_{p=1}^{n} \frac{\sigma_p^2}{2}\Big(F_N(s_k + s_p) + F_N(s_k - s_p)\Big) + \sigma_W^2 \tag{10.14}$$

$\mathrm{E}I_X(s)$ in (10.13) is the sum of (i) the constant variance σ_W^2 of the white noise process and (ii) n pairs of Fejer kernels $F_N(s)$, which are displaced by s_p ($-s_p$, resp.) and multiplied with $\sigma_p^2/2$. σ_W^2, s_p and $\sigma_p^2/2$ are defined in (10.1,1), and $F_N(s)$ is defined in (6.57) with the properties enumerated in (6.58) and (6.59). Under assumption (10.1,4), the frequencies s_p of the trigonometric oscillations are in the set of the Fourier frequencies s_k. Consequently, a sum of Fejer kernels as in (10.14) becomes identically zero for $s_k \neq s_p$, since $F_N(s)$ is required to be identically zero for $s = \pm 1/N, \pm 2/N, \pm 3/N, \ldots$ in (6.58,4).

Using these results and (6.58,5), $\mathrm{E}I_X(s_k) = (\sigma_p^2/2)N + \sigma_W^2$ for $s_p = s_k$ as well as $\mathrm{E}I_X(s_k) = \sigma_W^2$ for all other s_k is arrived at in (10.14). Using (10.14), $\mathrm{E}I_X(s_k)$ can be plotted on condition that the spectrum (10.1,3) of the model

to be estimated is known. In such a plot, $\mathrm{E}I_X(s_k)$ is visibly larger than σ_W^2 for frequencies $s_k = s_p$, i.e., for Fourier frequencies which are identical with the frequencies of the oscillations in the model.

In applications, however, the spectrum of the model is not known and the frequencies of the oscillations in model (10.1,1) are thus identified in a plot of $I_X(s_k)$. $I_X(s_k)$ slightly deviates from $\mathrm{E}I_X(s_k)$, with deviations depending on the realisation observed. For example, the white noise process (W_t) which is part of model (X_t) in (10.1,1) can induce, in $I_X(s_k)$, values which are large enough to be erroneously attributed to a trigonometric oscillation in (X_t). This false identification is favoured by the distribution of the periodogram as proposed in (8.24,2), since model (10.1,1) with $n = 0$ and the W_t normally distributed is reconcilable with the assumptions made in (8.24) as well as in (8.8) and (8.7). Consequently, $I_X(s_k)$ can have large values for frequencies not being those of the oscillations in model (10.1,1), a possible pitfall when $I_X(s_k)$ is applied as diagnostic to identify trigonometric oscillations in (10.1,1). More reliable than this diagnostic is a test with Fisher's g statistic introduced in (8.15). Both Fisher's test as well as Schuster's test introduced in (8.14) assume the null hypothesis, that there are no trigonometric oscillations in model (10.1,1).

10.1.3 Example and Summary

It is assumed that the residuals of a first order linear model fitted to the monthly means in the Basel temperature series are a realisation of model (10.1,1). The monthly means in the Basel series are made available in the remarks to (3.19) in R vector `temp`, a time slice is plotted in Fig. 3.6 (time in years with origin in 1755), time series of the January, April and July values in Fig. 3.7 (time in months with origin in June, 1856). This time series is not stationary in the first moment function since temperatures measured in Basel have increased since 1755, as is seen in Fig. 3.3. The Basel temperature series is therefore detrended by means of computing the residuals of a first order linear model for the secular trend, as proposed in (5.58,2). These residuals are plotted above in Fig. 10.1. Thereafter, the periodogram of the residuals is computed as demonstrated in the remarks to Fig. 8.1 and plotted below in Fig. 10.1.

In the periodogram plotted in Fig. 10.1 (b), two values being much larger than the others can be seen for frequencies $s_k = 0.0833 = 203/2436 = 1/12$ month^{-1} and $s_k = 0.1667 = 406/2436 = 1/6$ month^{-1}. If this periodogram is calculated using the R expressions in the remarks to (8.1), then $I_X(s_k = 1/12) = 47.28$ dB and $I_X(s_k = 1/6) = 20.63$ dB as well as $\widehat{A}_X(s_k = 1/12) = -9.3008$, $\widehat{B}_X(s_k = 1/12) = -0.1054$, $\widehat{A}_X(s_k = 1/6) = 0.0584$ and $\widehat{B}_X(s_k = 1/6) = 0.4317$ are obtained as estimators (10.3) and (10.4) for the coefficients of these oscillations. Except for $I_X(s_k)$, $s_k = 1/12$ and $s_k = 1/6$ month^{-1}, no other large values are visible in Fig. 10.1 (b),

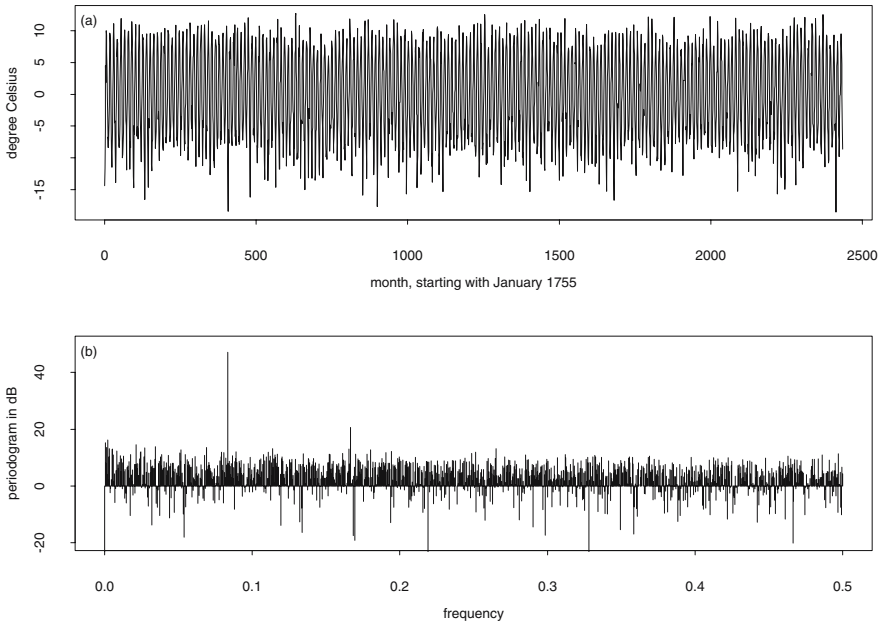


Fig. 10.1. Above (a) residuals of a first order linear model for the secular trend in the monthly means in the Basel temperature series, below (b) periodogram calculated from the residuals in plot (a).

and thus the model contains only $n = 2$ oscillations, as is concluded from the remarks to (10.14). Summarising, in a model (10.1,1) for the Basel temperature series, there are $n = 2$ trigonometric oscillations with frequencies $\hat{s}_1 = 1/12$ month^{-1}, $\hat{s}_2 = 1/6$ month^{-1} and coefficients $\widehat{A}_1$, $\widehat{B}_1$, $\widehat{A}_2$ and $\widehat{B}_2$ as obtained above.

$1/12$ month^{-1} is the known frequency of the annual cycle in the Basel temperature series for which the seasonal model in Table 3.3 is estimated, and $s_k = 1/6$ is the first harmonic to $s_k = 1/12$. Hence, it is concluded from the remarks to Table 6.1 as well as to Figs. 6.2 and 6.3 that the annual cycle in the Basel series can be represented by the sum of the oscillations estimated above. This sum is an estimate for the harmonic process in (10.1,1), and therefore, the white noise process in this model remains to be estimated.

An estimate for s_W^2 in (10.1,1) is the variance of the empirical residuals $\hat{r}_t$ in (10.2), which is readily obtained as the variance of the residuals in a linear model. The following R expressions estimate the coefficients of the trigonometric oscillations with frequencies $\hat{s}_1 = 1/12$ month^{-1} and $\hat{s}_2 = 1/6$ month^{-1}.

```
#R vector baselmon1: months 0,1, ..., 2435
#R vector baseltempres: residuals as plotted in Fig. 10.1 (a) of a
```

```
#first order linear model for the trend in the Basel series
s1 <- 0.08333333 #frequency of first oscillation
s2 <- 0.16666667 #frequency of second oscillation
#R vectors for predictor variables
A1 <- cos(2*pi*s1*baselmon1)
B1 <- sin(2*pi*s1*baselmon1)
A2 <- cos(2*pi*s2*baselmon1)
B2 <- sin(2*pi*s2*baselmon1)
#-1 requires a linear model with zero intercept, as introduced in
#the comments on the formulation of linear models in Sect. 3.2.1
baselfitharmproc <- lm(baseltempres ~ A1 + B1 + A2 + B2 -1)
```

The estimates for the coefficients computed using R expressions as above

```
     Value Std.-Error   t-value Pr(>|t|)
A1 -9.3008     0.0550 -169.0752   0.0000
B1 -0.1052     0.0550   -1.9119   0.0560
A2  0.0584     0.0550    1.0611   0.2887
B2  0.4317     0.0550    7.8885   0.0000
```

have only small differences when compared with those obtained in the remarks to Fig. 10.1 (b). There, these coefficients result as a by-product when the periodogram is computed using the R expressions in the remarks to (8.1). The variance s_W^2 of the white noise process in model (10.1,1) is estimated from the residuals, i.e., for s_W

```
Residual standard error: 1.92 on 2432 degrees of freedom
```

is obtained, accounting for $N - 2\hat{n} = 2436 - 2 \times 2$ degrees of freedom. Further results are an R^2 statistic 0.9218 and correlations between the estimated coefficients which are identically zero, since the predictor variables are orthogonal.

The harmonic model fitted above to the residuals of a first order linear model for the trend in the Basel temperature series (plotted for the period January 1851 through to December 1890 in Fig. 10.2) has a standard error of the residuals $\hat{\sigma}_r = 1.92^0$ Celsius on 2432 degrees of freedom. A similar residual error, i.e., $\tilde{\sigma}_r = 1.921^0$ Celsius on 2423 degrees of freedom, is arrived at when an additive component model for the trend and annual cycle in the Basel temperature series is fitted in Sect. 3.4.2. The first estimate, however, has a smaller number of coefficients: two in the model for the secular trend and four in (10.1,1), since the annual cycle is captured with an oscillation having frequency 1/12 month^{-1} together with its first harmonic. The component model, in contrast, requires 13 coefficients: the regression coefficient for the trend, and 12 for the intercepts, one for each month.

Model (10.1,1) is stationary. Consequently, the Basel temperature series is detrended by estimating a first order linear model for the trend and thereafter model (10.1,1) is fitted to the residuals, as demonstrated above and as recommended in (5.58,2). The alternative detrending method, i.e., computing differences of the time series as proposed in (5.58,1), changes the spectrum

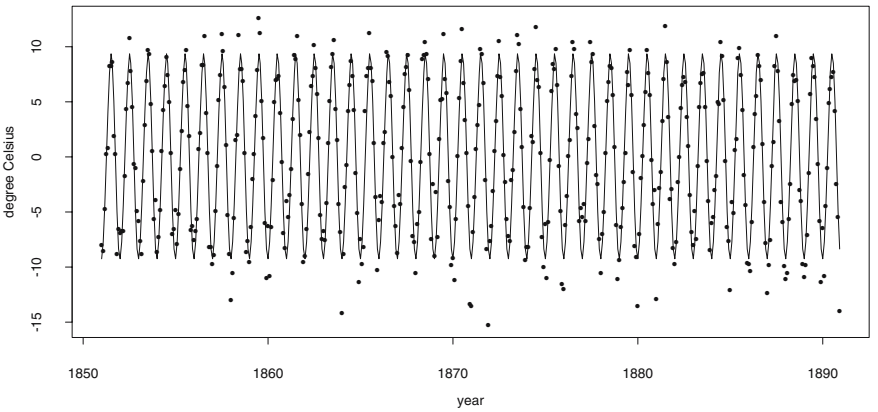


Fig. 10.2. Residuals (•) of a first order linear model for the secular trend in the monthly means in the Basel temperature series, together with a fitted model (10.1,1) having two oscillations (solid line) in the time slice from 1851 through to 1891.

of the original series, as is concluded from the gain and phase shift functions in Fig. 6.30, and therefore it is not recommended to compute a spectral estimator subsequent to differencing a time series. This applies when estimating a continuous spectrum assuming model (9.1) as well as a spectrum having a discrete part assuming model (10.1,1). Examples are given in Problem 9.24 and in Sect. 10.3.2.

In Fig. 10.1 (b), two periodogram values being larger than the others are clearly visible and thus the number as well as the frequencies of the trigonometric oscillations in model (10.1,1) are readily obtained using the diagnostics in the remarks to (10.14). If it is doubtful whether a periodogram value is larger than the remaining ones, then Fisher's g test as proposed in (8.17) — or one of the tests in the remarks to (8.17) — can be performed.

This section is summarised as follows. The observations are assumed to be a time slice in a realisation of a discrete-time stochastic process being reconcilable with model (10.1). A stochastic process for example, which has seasonal fluctuations with a constant and known period p that can be captured using a seasonal model (3.18), is reconcilable with model (10.1) on condition that the number of observations N is a multiple of p.

Assuming model (10.1), $\widehat{A}(s_k = s_p)$ and $\widehat{B}(s_k = s_p)$ as proposed in (10.3) and (10.4), s_k being the Fourier frequencies in (6.21,1), are least squares estimators for the coefficients A_p and B_p of the trigonometric oscillations in model (10.1). These estimators are unbiased on condition that the realised values of A_p and B_p depend only on the realisation and not on the time, i.e., remain constant for all time points in a realisation. $\widehat{A}(s_k = s_p)$ and $\widehat{B}(s_k = s_p)$ are computed using a discrete Fourier transform, as is concluded from (6.21) and (6.22). From a plot of a periodogram calculated as required

in (8.1,3), diagnostics for the oscillations in model (10.1) can be obtained, and periodogram tests as introduced in Sect. 8.2.2 can be applied to clarify any remaining uncertainties.

10.2 Estimating Oscillations with Unknown Frequencies

Often, a stochastic process (X_t) is assumed to be the sum of $(X_t^{(h)})$ and $(X_t^{(r)})$, where $(X_t^{(h)})$ is a harmonic process as defined in (7.38) contributing to the discrete part of the spectrum of (X_t), and $(X_t^{(r)})$ a stationary process contributing to the continuous part of the spectrum of (X_t). If the frequencies of the oscillations in $(X_t^{(h)})$ are in the set of Fourier frequencies (6.21,2) and $(X_t^{(r)})$ is a white noise process (2.10) then (X_t) is reconcilable with model (10.1). Two less restrictive models are introduced in Sects. 10.2.1 and 10.2.2. In both cases, the diagnostics introduced in the remarks to (10.18) are applied to an illustrative example.

10.2.1 Sum of a Harmonic Process and White Noise

In this section, it is assumed that the stochastic process (X_t) under analysis has a mixed spectrum with (i) a continuous part that is constant and (ii) a discrete part that shows spikes due to trigonometric oscillations which have frequencies $-1/2 \leq s_p \leq 1/2$. Such a process is reconcilable with the model defined in (10.15).

Let (X_t) be a model as defined in (10.1) but with $-1/2 \leq s_p \leq 1/2$ substituted for (10.1,4). (10.15)

(X_t) as defined in (10.15) is not a linear model since frequencies s_p in (10.1,1) are not coefficients in a linear combination of predictor variables. To circumvent this difficulty, it is assumed that a procedure exists which allows for estimating the s_p and therefore also n. Substituting the $\hat{s}_p$ thus obtained for s_p, (10.15) becomes a linear model and least-squares estimates $\widehat{A}_X(\hat{s}_p)$ and $\widehat{B}_X(\hat{s}_p)$ for $A_X(s_p)$ and $B_X(s_p)$ can be computed using the techniques introduced in Chap. 3, as is demonstrated in Sect. 10.1.3. Thereafter, the residual variance of the fitted model becomes an estimate for the variance σ_W^2 of the white noise process in (10.15).

Consequently, the problem of estimating model (10.15) becomes the problem of estimating (i) the frequencies s_p of the oscillations and (ii) the number n of oscillations in the model. A possible solution to this problem is to estimate a non-linear regression model using R function `nls()`. `nls()` requires initial values s_p^* for the unknown s_p. With s_p^* being close to s_p, $p = 1, 2, \ldots, n$, the number of iterations required to arrive at non-linear least squares estimates is small and an estimation becomes feasible [31]. Can initial values s_p^* be obtained from a diagnostic for the unknown frequencies s_p?

If a direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$, s'_k as defined in (9.20), is computed from a time slice (x_t), $t = 0, 1, \ldots, N-1$, in a realisation of model (10.15), then frequencies s'_k are assumed to be pilot frequencies for the unknown model frequencies, for the reason that $\widehat{C}_X^{(d)}(s'_k)$ is proportional to the squared absolute values of the trigonometric oscillations in the Fourier representation of the tapered observations $(x_t) \times \big(h_N^{(d)}(t)\big)$. $h_N^{(d)}(t)$ is a variance-normalised data taper as defined in (9.2,1). If a pilot frequency s'_k is close to a model frequency s_p then the direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ is large, whereas $\widehat{C}_X^{(d)}(s'_k)$ is small for s'_k a larger distance away from any of the model frequencies, as is demonstrated in Sect. 6.1.1.

When is a direct spectral estimator large and when is it small? When is a pilot frequency close to and when is it a larger distance away from a model frequency? An answer to these questions can be obtained from an analysis of $\mathrm{E}\widehat{C}_X^{(d)}(s'_k)$, i.e., the expectation of a direct spectral estimator calculated from observations in a realisation of model (10.15). This expectation is derived below from $\mathrm{E}\widehat{C}_X^{(d)}(s)$ in (9.11). A derivation using $\mathrm{E}\widehat{C}_X^{(d)}(s)$ in (9.11) is feasible since this expectaton is obtained in Sect. 9.5.2 under the assumptions that the process under analysis (i) is stationary and (ii) has an expectation function being identically zero, both reconcilable with model (10.15). Consequently, the spectrum of model (10.15), which is to be found in (10.1,3), is substituted in (9.11) and (10.16) is arrived at.

$$\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big) = \mathcal{H}_N^{(d)}(s) * \left(\sum_{p=1}^{n} \left(\frac{\sigma_p^2}{2}\delta(s - s_p) + \frac{\sigma_p^2}{2}\delta(s + s_p) \right) + \sigma_W^2 \right) \tag{10.16}$$

$$= \sum_{p=1}^{n} \frac{\sigma_p^2}{2} \Big(\mathcal{H}_N^{(d)}(s + s_p) + \mathcal{H}_N^{(d)}(s - s_p) \Big) + \sigma_W^2 \tag{10.17}$$

$$\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_k)\big) = \sum_{p=1}^{n} \frac{\sigma_p^2}{2} \Big(\mathcal{H}_N^{(d)}(s'_k + s_p) + \mathcal{H}_N^{(d)}(s'_k - s_p) \Big) + \sigma_W^2 \tag{10.18}$$

(10.17) follows since $\delta(x)$ is the identity element under convolution as is concluded from (6.67) and (6.68) and, by substituting s'_k for s, the expectation of a direct spectral estimator computed from observations in a realisation of model (10.15) is arrived at in (10.18).

$\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big)$ in (10.17) is the sum of σ_W^2 and of n spectral windows $\mathcal{H}_N^{(d)}(s)$ which are displaced by s_p ($-s_p$ resp.) and multiplied with $\sigma_p^2/2$. Even in the case of estimating model (10.15) without tapering the observations, i.e., when $\mathcal{H}_N^{(d)}(s)$ in (10.17) becomes $F_N(s)$, the Fejer kernels have main lobes which are not centred in frequencies s_k and thus do not become identically zero for the $s_k \neq s_p$, and therefore the sum of the displaced and scaled spectral windows in (10.17) is a function whose intricacy increases with n.

The most straightforward model has $n = 1$, i.e., only one oscillation with frequency s_1 and a squared absolute value σ_1^2. A direct spectral estimator calculated from observations in a realisation of this model has expectation

1. $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big)(s'_k \approx s_1) \approx (\sigma_1^2/2)\mathcal{H}_N^{(d)}(0) + \sigma_W^2$ for $s'_k \approx s_1$, i.e., pilot frequencies s'_k being close to model frequency s_1, or
2. $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big)(s'_k \bowtie s_1) \approx \sigma_W^2$ for $s'_k \bowtie s_1$, i.e., pilot frequencies s'_k a larger distance away from s_1,

on condition that $\sigma_1^2/2 \gg \sigma_W^2$, i.e., that the oscillation has a squared absolute value which is much larger than the variance of the white noise in the model. $\sigma_1^2/2 \gg \sigma_W^2$ is required since an oscillation with $\sigma_1^2/2 \approx \sigma_W^2$ remains hidden in the model noise as is demonstrated in Problem 10.1.

If the model contains a variety of trigonometric oscillations with unknown frequencies s_p then $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big)$ in (10.17) becomes an intricate function. The unknown frequencies s_p can be traced in a plot of $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big)$ provided that

1. locally in the neighbourhoods of the s_p, the main lobe of the spectral window $\mathcal{H}_N^{(d)}(s)$ dominates $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big)$ (this condition is fulfilled when superimposed side lobes of the spectral window displaced by s_p or $-s_p$, $p = 1, \ldots, n$, do not take the form of the main lobe), and
2. $\mathcal{B}_\triangle^\cap\big(\mathcal{H}_N^{(d)}(s)\big) < (1/2)\mathcal{B}_\triangle\big(C_X^{(I)}(s)\big)$ adopted from diagnostic (6.115), with $\mathcal{B}_\triangle^\cap\big(\mathcal{H}_N^{(d)}(s)\big)$ being the bandwidth of a direct spectral estimator as defined in (9.89) under the assumption that $B = 1$ and that $\mathcal{B}_\triangle\big(C_X^{(I)}(s)\big)$ is the bandwidth (7.83,2) of the spectrum to be estimated.

The widths which are compared in the second item above depend on N since a data taper and therefore also the pertaining spectral window, both as defined in (9.2,1), depend on N, and not on N', as required in (9.20). Consequently, the probability of resolving two model frequencies within a short distance of each other increases with increasing N and decreasing $\mathcal{B}_\triangle^\cap\big(\mathcal{H}_N^{(d)}(s)\big)$.

However, $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big)$ is not available when model (10.15) is estimated and thus the unknown model frequencies have to be detected in a plot of a direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$. $\widehat{C}_X^{(d)}(s'_k)$ computed from observed time slices in realisations of model (10.15) fluctuates about $\mathrm{E}\big(\widehat{C}_X^{(d)}(s)\big)$ and the white noise process in this model can induce values in $\widehat{C}_X^{(d)}(s'_k)$ which are large enough to be identified erroneously as being due to the oscillations in the model. This false identification is favoured by the $\chi^2_{(n)}$ distributions (9.19) and the rapidly decaying correlation (9.17). Similar properties are shared by a periodogram computed from observations in a realisation model (10.1). Hence, the comment on pseudo peaks in a periodogram as in the remarks to (10.14) also applies to a direct spectral estimator.

In the remarks to Figs. 9.3 and 9.4, it is recommended to multiply the observations with a variance-normalised data taper (9.2) prior to estimating

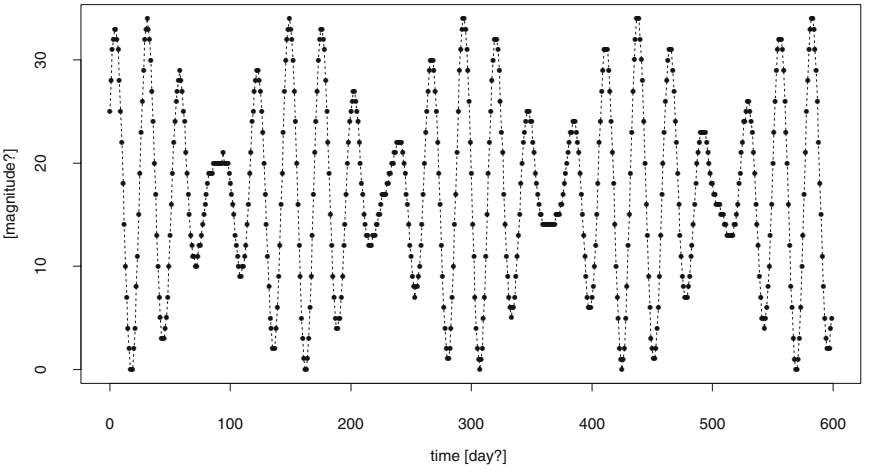


Fig. 10.3. A time slice in a realisation of model (10.15) (subsequent to adding a mean) or observations of the magnitude (apparent brightness) of a variable star (?)

a continuous spectrum (9.1,2) since a data taper efficiently reduces a possible bias due to leakage. Tapering is also compulsory when estimating the spectrum of model (10.15). In the case of estimating a spectrum having a discrete part, trying a sequence of variance-normalised data tapers not only reduces a possible leakage but also often promotes the detection of the model oscillations for the following reason: there are models with oscillations that can be successfully traced using a first data taper but not using a second one, depending on the distances between the unknown frequencies s_p and the width of the spectral windows pertaining to the tapers applied.

Tapering and also zero-padding, i.e., computing a direct spectral estimator for frequencies s'_k as defined in (9.20), are often beneficial when a continuous spectrum is estimated. Both methods can also help to estimate model (10.15) having a spectrum with a discrete part, as demonstrated by the following example.

Whittaker and Robinson in [146] give the magnitudes of a variable star for 600 consecutive days, as plotted in Fig. 10.3. The magnitude of an object in the sky (star, planet, etc.) is a logarithmic measure of the amount of light received from that object, i.e., its apparent brightness. Thus the magnitude depends on the object observed, the atmospheric opacity and the observational methods. The author believes that the Whittaker and Robinson magnitudes are not actual observations for the following reasons:

1. it is difficult to observe a circumpolar star for 600 consecutive days (no values are missing in this series) with a constant sampling interval (one day) even when many observers in various locations around the world

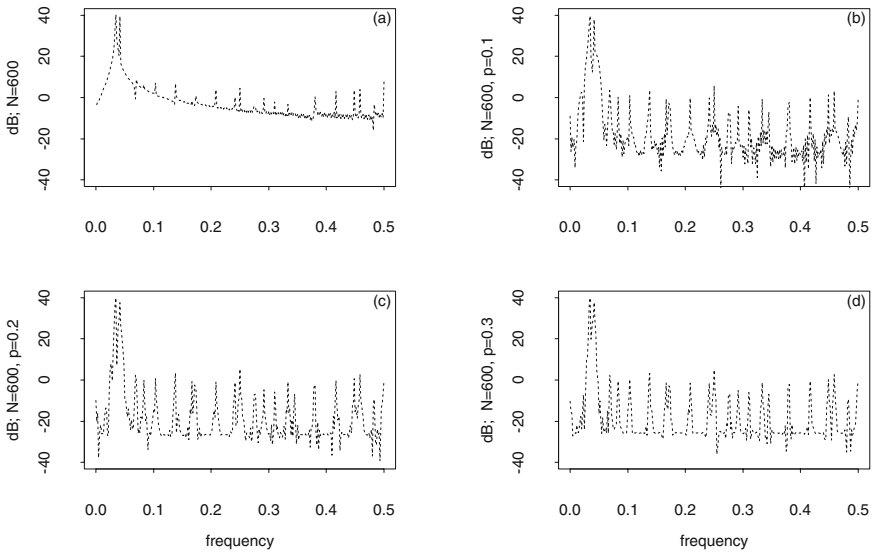


Fig. 10.4. Periodogram (a) and direct spectral estimators computed for Fourier frequencies s_k subsequent to applying cosine tapers for $p = 0.1$, $p = 0.2$ and $p = 0.3$ to the demeaned simulation in Fig. 10.3.

are involved in the experiment, since the probability of having good sky transparency for 600 consecutive days is small [138], and

2. a time slice in a realisation of model (10.15) is readily simulated even when no computer is available.

Nevertheless, this (with a high probability simulated) time series appears as observations in the time series literature (e.g., [14]) and can be downloaded from several websites (e.g., [91]). Here it is assumed that the "data" as plotted in Fig. 10.3 and as available in file `/path/allegedstar.dat` do not contain any information about a variable star but are, subsequent to de-meaning, merely a time slice in a simulated realisation of model (10.15). Under this assumption, the parameters in this model are estimated in the following paragraphs.

In Fig. 10.3, there are more than three and less than five oscillations within a period of 100 u (u is used here for the time unit) and consequently, the underlying model has trigonometric oscillations with periods between 20 and 33 u or frequencies between 0.033 and 0.05 u^{-1}. A peak in this band clearly dominates the periodogram in Fig. 10.4 (a) calculated from the demeaned simulation. This periodogram has a dynamic range of more than 40 dB and thus is presumably biased due to leakage.

This suspicion induces an investigation to be made following the recommendations in the remarks to Figs. 9.3 and 9.4 which also apply when estimating a spectrum having a discrete part: the simulation as plotted in

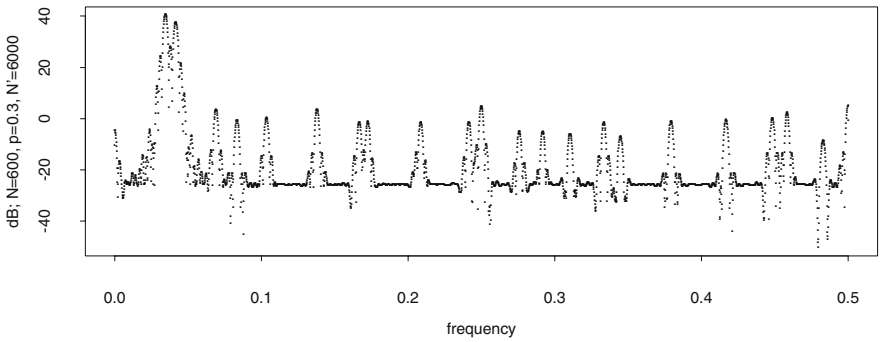


Fig. 10.5. Direct spectral estimator calculated from the demeaned simulation in Fig. 10.3 for frequencies s'_k with $N' = 6000$ as defined in (9.20) subsequent to multiplying the observations using a cosine taper with $p = 0.3$.

Fig. 10.3 is multiplied with cosine tapers for $p = 0.1$, $p = 0.2$ and $p = 0.3$, and the direct spectral estimators computed from the tapered observations are plotted in Fig. 10.4 (b), (c) and (d). Obviously, the periodogram and the direct spectral estimators obtained using cosine tapers with $p = 0.1$ and $p = 0.2$ are biased due to leakage, whereas the direct spectral estimator as plotted in Fig. 10.4 (d) is considered to be unbiased, in accordance with the remarks to diagnostic (6.117). This estimator is constant for approximately -25 dB and has a primary twin peaks in band $0.033 \leq s \leq 0.05$ and, albeit smaller by orders of magnitude, also many secondary peaks for higher frequencies.

This direct spectral estimator is reconcilable with model (10.15) on condition that (i) it is constant in segments between the peaks due to a constant (approximately -25 dB) continuous white noise spectrum and (ii) its primary twin peaks as well as its secondary peaks are due to the discrete part of the model spectrum. Under these and the additional condition that the main lobe of the spectral window is not too wide (relative to the bandwidth (7.83,2) of the model spectrum), the peaks in the direct spectral estimators become, in the mean over a large number of realisations, identical with the main lobe of the spectral window pertaining to the cosine taper applied. This conclusion drawn from (10.17) implies that the peaks in a direct spectral estimator resemble the main lobe of the spectral window. If such a similarity has been found in a diagnostic plot then the pilot frequencies guessed are assumed to be an estimate for the unknown frequencies of the trigonometric oscillations in the model. Such diagnostic plots are obtained in Figs. 10.6 and 10.7 subsequent to the preparations made in Fig. 10.5 and Table 10.1.

Hence, the direct spectral estimator as plotted in Fig. 10.4 (d) is recomputed: (i) the $N = 600$ simulated values are multiplied with a variance-normalised cosine taper $\widehat{h_N^{(d)}}(t)$ for $p = 0.3$ with pertaining spectral window

Table 10.1. Frequencies of the primary twin peaks (left $\hat{s}_1 = .0345$ and right $\hat{s}_2 = .041666$ u^{-1}) and the secondary peaks in the direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ as plotted in Fig. 10.5.

primary peak	u^{-1}	.03400	.03416	.03433	.03450	.03466	.03483	.0350
left	dB	40.05	40.43	40.65	40.72	40.65	40.43	40.05
primary peak	u^{-1}	.0410	.04116	.04133	.0415	.04167	.04183	.0420
right	dB	36.67	37.15	37.48	37.66	37.71	37.36	36.96

		primary peaks	secondary peaks						
l		1	2	3	4	5	6	7	8
$l \times \hat{s}_1$	u^{-1}	.034500	.069	.1035	.138	$\approx .1725$	.207	.2415	.276
$\widehat{C}_X^{(d)}(s'_k)$	dB	40.72	3.61	0.51	3.62	−1.12	−10.41	−1.25	−4.84
$l \times \hat{s}_2$	u^{-1}	.041667	.08333	.12500	.16666	$\approx .2085$	.25000	.29166	.33333
$\widehat{C}_X^{(d)}(s'_k)$	dB	37.71	−0.40	−25.57	−1.20	−1.36	4.89	−5.05	−1.38

$\widehat{\mathcal{H}}_N^{(d)}(s)$, (ii) the values thus obtained are zero-padded as defined in (9.20) to arrive at $N' = 6000$, and (iii) the direct spectral estimator $\widehat{C}_X^{(d)}(s'_k)$ is arrived at. $\widehat{C}_X^{(d)}(s'_k)$ is plotted using fine dots in Fig. 10.5. This plot confirms that $\widehat{C}_X^{(d)}(s'_k)$ amounts to approximately −25 dB for frequencies at a larger distance from its peaks, a property being in agreement with the continuous part of the model spectrum as defined in (10.15).

Are also the peaks in $\widehat{C}_X^{(d)}(s'_k)$ as plotted in Fig. 10.5 reconcilable with model (10.15)? The unknown frequencies s_p of the oscillations in this model are approximated using the frequencies of the peaks in the direct spectral estimator given in Table 10.1. This table is to be completed for frequencies larger that 0.3 u^{-1} in Problem 10.2. From Fig. 10.5, Table 10.1 and the results of Problem 10.2, it is concluded that

1. in the primary twin peaks, the left peak is symmetric with respect to frequency 0.0345 u^{-1} whereas the right one is not symmetric to frequency 0.0416667 u^{-1}
2. the primary twin peaks are by order of magnitudes larger than the secondary peaks, and
3. the frequencies associated with the maxima of the secondary peaks are multiples of the frequencies associated with both maxima in the primary twin peaks.

These conclusions suggest that the model used to simulate the time slice in Fig. 10.3 is a sum of (i) a white noise process, (ii) two trigonometric oscillations having frequencies 0.0345 as well as 0.0416667 u^{-1}, and (iii) an unknown number of harmonics pertaining to these frequencies.

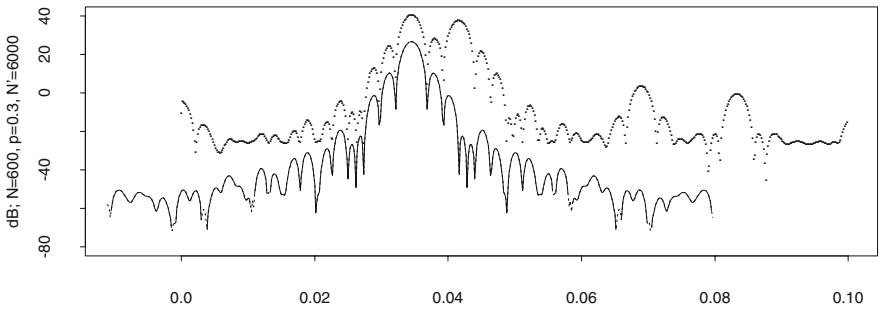


Fig. 10.6. Direct spectral estimator as in Fig. 10.5 together with the spectral window pertaining to the cosine taper having parameters $p = 0.3$ and $N = 600$ plotted beneath the left peak.

These suggestions can be corroborated using diagnostic plots as demonstrated in Figs. 10.6 and 10.7. In Fig. 10.6, $\widehat{\mathcal{H}}_N^{(d)}(s - \hat{s}_1)$ is plotted using a solid line below the maximum of $\widehat{C}_X^{(d)}(s'_k)$, with displacement $\hat{s}_1 = 0.0345$ u^{-1} being the frequency for which $\widehat{C}_X^{(d)}(s'_k)$ attains is maximum as can be seen in Table 10.1. In the neighbourhood of this frequency, $\widehat{C}_X^{(d)}(s'_k)$ and $\widehat{\mathcal{H}}_N^{(d)}(s - \hat{s}_1)$ are very close to each other (except for a small vertical displacement). This perfect match is sufficient reason for assuming that there is a trigonometric oscillation with frequency $\hat{s}_1 = 0.0345$ u^{-1} in the model. A less perfect but still close match is obtained in the case of the trigonometric oscillation with frequency $\hat{s}_2 = 0.0416667$ u^{-1} (not shown). For these frequencies, the direct spectral estimator and the spectral windows match for the following reasons: (i) approximations $\hat{s}_1$ and $\hat{s}_2$ are close to model frequencies s_1 and s_2 and (ii) the continuous constant part of the model spectrum $\sigma_W^2 \approx -25$ dB is small as compared with the squared absolute values σ_1^2 and σ_1^2 of the oscillations in model (10.15). Clearly, the identification of the unknown frequencies in such a model becomes much more difficult when the variance of the white noise process increases and the absolute values of the oscillations decrease.

The heights of the jump discontinuities $\sigma_1^2/2$ and $\sigma_2^2/2$ in the integrated model spectrum are approximated using the expectation derived in (10.18): $\hat{\sigma}_p^2/2 = \mathrm{E}\big(\widehat{C}_X^{(d)}(s'_k \approx s_p)\big)/\widehat{\mathcal{H}}_N^{(d)}(s = 0)$ is arrived at on condition that (i) $\hat{\sigma}_W^2$ is neglected and (ii) $\mathrm{E}\big(\widehat{C}_X^{(d)}(s'_k \approx s_p)\big)$ is substituted with $\widehat{C}_X^{(d)}(s'_k \approx s_p)$. Clearly, the second condition requires that $\hat{s}_p$ be close to s_p and that the bias of the direct spectral estimator is small. In the case of the simulation example in Fig. 10.3, $\hat{\sigma}_1^2/2 = 11824.01/470.40 = 25.13609$ for frequency $\hat{s}_1 = 0.0345$ u^{-1} as well as $\hat{\sigma}_2^2/2 = 5896.365/470.40 = 12.53478$ for frequency $\hat{s}_2 = 0.041667$ u^{-1} are arrived at. 470.40 is the height of the main lobe in the

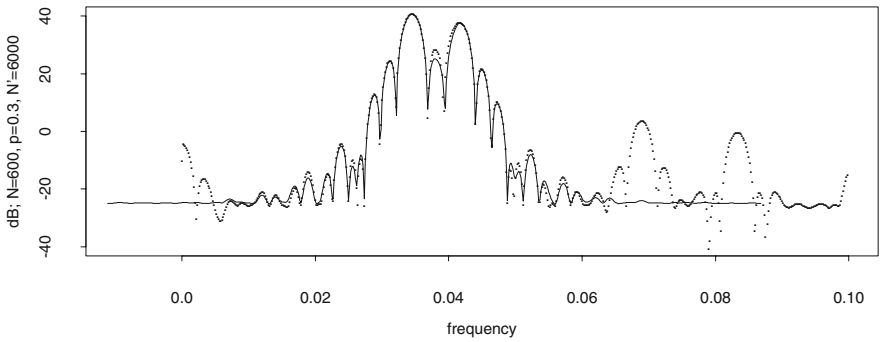


Fig. 10.7. Direct spectral estimator as in Fig. 10.5 together with a linear combination of two spectral windows.

spectral window pertaining to a cosine taper for $N = 600$ and $p = 0.3$, which can be obtained using the hints in Problem 9.3.

Using these values in linear combination $(\hat{\sigma}_1^2/2)\widehat{\mathcal{H}}_N^{(d)}(s-\hat{s}_1)+(\hat{\sigma}_2^2/2)\widehat{\mathcal{H}}_N^{(d)}(s-\hat{s}_2)+\hat{\sigma}_W^2$, the solid line plotted in Fig. 10.7 is arrived at. Obviously, the linear combination and the direct spectral estimator match perfectly in the neighbourhood of $\hat{s}_1 = 0.0345$ u^{-1}, whereas the functions show a slight deviation in the neighbourhood of $\hat{s}_2 = 0.041667$ u^{-1}. This slight deviation is in accordance with the values given in Table 10.1, where the spectral estimator is not symmetric on both sides of $\hat{s}_2 = 0.041667$ u^{-1}.

From (i) the direct spectral estimator as plotted in Fig. 10.5, (ii) its peak frequencies in Table 10.1 and Problem 10.2, as well as (iii) the diagnostic plots in Figs. 10.6 and 10.7, it is concluded that the simulation in Fig. 10.3 is in agreement with model (10.15). This model contains (i) two dominating trigonometric oscillations having frequencies $s_1 \approx \hat{s}_1 = 0.0345$ u^{-1} and $s_2 \approx \hat{s}_2 = 0.0416667$ u^{-1}, (ii) an unknown number of minor oscillations with frequencies being multiples of s_1 and s_2 (the harmonics to the dominating frequencies as introduced in the remarks to Tab. 6.1), and (iii) a white noise process having a variance of approximately -25 dB. How many harmonics pertain to each dominating oscillation?

The most straightforward model has no harmonics, i.e., it is the sum of two trigonometric oscillations with frequencies s_1 and s_2 as approximated above and a white noise process: $n = 2$, $s_1 = \hat{s}_1 = 0.0345$ u^{-1} and $s_2 = \hat{s}_2 = 0.041666$ u^{-1} are assumed in model (10.15). The following least squares estimates for the coefficients of the oscillations in this model

```
             Value Std.Error   t value Pr(>|t|)
(Intercept) 17.0864    0.0135 1261.7300   0.0000
         A1  7.4262    0.0192  387.7588   0.0000
         B1  6.7376    0.0192  350.5824   0.0000
         A2 -0.0019    0.0192   -0.0998   0.9206
```

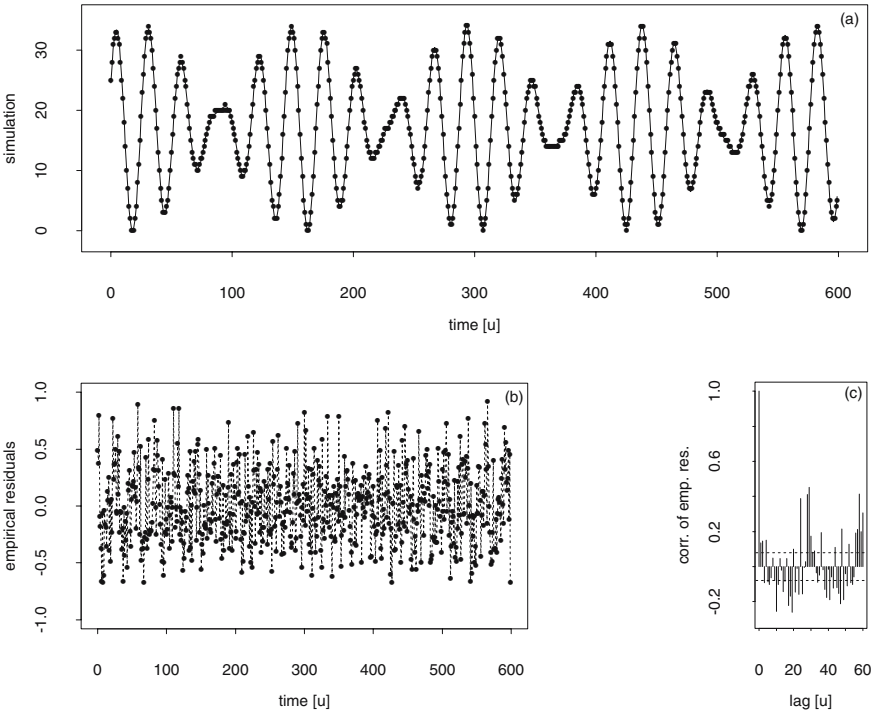


Fig. 10.8. Linear model with a mean and two trigonometric oscillations having frequencies $\hat{s}_1 = 0.0345$ u^{-1} and $\hat{s}_2 = 0.0416667$ u^{-1} fitted to the simulation in Fig. 10.8 (above), together with its empirical residuals and their correlations (below).

```
      B2  7.0987    0.0192  369.9303    0.0000
```

and for the standard error of the white noise process

```
Residual standard error: 0.3317 on 595 degrees of freedom
```

are obtained by applying a linear model as demonstrated in Sect. 10.1.3. The simulated values and the linear model thus obtained are plotted in Fig. 10.8 (a), the empirical residuals and their correlations in Fig. 10.8 (b) and (c).

Using these estimates $\tilde{\sigma}_1^2/2 + \tilde{\sigma}_W^2 = (1/4)(7.4262^2 + 6.7376^2) = 25.13593$, $\tilde{\sigma}_2^2/2 + \tilde{\sigma}_W^2 = (1/4)(-0.0019^2 + 7.0987^2) = 12.59789$ and $\tilde{\sigma}_W^2 = 0.3317^2 = 0.11 = -9.59$ dB are obtained as estimates for the model spectrum. $\tilde{\sigma}_1^2/2$ and $\tilde{\sigma}_2^2/2$ are close to the estimates $\hat{\sigma}_1^2/2$ and $\hat{\sigma}_2^2/2$, established in the remarks to Fig. 10.6, whereas $\tilde{\sigma}_W^2 = -9.59$ dB is much too large, since the direct spectral estimator is constant at the level of approximately $\tilde{\sigma}_W^2 = -25$ dB in Figs. 10.5, 10.6 and 10.7. This overestimation is due to the missing harmonics in the model with $n = 2$: the variance due to the harmonics pertaining to both dominant oscillations is attributed to the model noise. This finding is

in line with the correlations of the empirical residuals as in Fig. 10.8 (c). An estimate having a similar residual variance (−10.37 dB) is obtained in Problem 10.5. There, $n = 2$ is assumed as above and the frequencies of the two dominating oscillations are approximated using R function `nls()` (non-linear least squares).

Estimates having a clearly smaller residual variance than $\tilde{\sigma}_W^2 = -9.59$ dB can be obtained in Problem 10.3 using the techniques introduced in Sects. 3.2 and 3.3 together with R function `lm()`, on condition that the model is assumed to have more than $n = 2$ trigonometric oscillations.

The frequencies estimated above in the remarks to Table 10.1 and Figs. 10.6 and 10.7, as well as those estimated in Problem 10.5, are very close to $1/29 = 0.03448276$ u^{-1} and to $1/24 = 0.04166667$ u^{-1}. The author therefore believes that the time series in Fig. 10.3 was simulated (without using electronic equipment) as the sum of

1. a realisation of a white noise process with variance $\sigma_W^2 \approx -25$ dB
2. two trigonometric oscillations having frequencies $s_1 = 1/29$ u^{-1} and $s_2 = 1/24$ u^{-1}, with coefficients $A_1 = 7.65$, $B_1 = 6.50$, $A_2 = 0.00$ and $B_2 = 7.10$, and
3. an unknown (to be found in Problem 10.3) number of harmonics pertaining to the above oscillations.

Another hypothesis is that the time series in Fig. 10.3 was obtained by interpolating the gaps in actual observations, considering that astronomical observations are often subject to uneven sampling. For unevenly sampled time series, the usual methods (as introduced in this book, an exception are the Geostatistical ones in Chap. 4) do not apply. Consequently, astronomers have developed methods for the analysis of unevenly sampled time series. Software based on such methods is available at no charge from the American Association of Variable Star Observers (AAVSO) [138].

10.2.2 Sum of a Harmonic Process and Coloured Noise

In this section, it is assumed that the discrete-time stochastic process under analysis has a mixed spectrum with a discrete part due to a harmonic process and a non-constant continuous part due to a linear process. If both stationary processes in this model, the harmonic and the linear one, are not correlated as required in (10.19,1), then (10.19,2,3) are readily obtained in Problem 10.6.

Model (10.19) is not a linear model, a property shared with model (10.15). In both models, frequencies s_p are unknown and, unfortunately, are not coefficients of the predictor variables. Approximations $\hat{s}_p$ for s_p are obtained using the diagnostics in the remarks to (10.18), as demonstrated in Table 10.1 as well as in Figs. 10.6 and 10.7. Substituting $\hat{s}_p$ for s_p, a linear model for the coefficients of the trigonometric oscillations contributing to the discrete part of the model is arrived at. This linear model is estimated as demonstrated

in Sect. 10.1.3. Thereafter, the empirical residuals of the estimate thus obtained are assumed to be a realisation of the linear process (X_t) in (10.19) and $C_X(s)$ is estimated using the methods introduced in Chap. 9.

1. *Let* $Y_t = \sum_{p=1}^{n} \big(A_p \cos(2\pi s_p t) + B_p \sin(2\pi s_p t)\big) + X_t$, *with* $\mathrm{E}A_p = \mathrm{E}B_p = 0$, $\mathrm{Var}A_p = \mathrm{Var}B_p = \sigma_p^2$, $\mathrm{Cov}(A_p, B_p) = 0$, $\mathrm{Cov}(A_p, B_q) = \mathrm{Cov}(B_p, B_q) = \mathrm{Cov}(A_p, A_q) = 0$, *for* $p \neq q$, *and* (X_t) *being a model having a non-constant continuous spectrum as defined in (9.1), such that* $\mathrm{Cov}(X_t, A_p) = \mathrm{Cov}(X_t, B_p) = 0$. *Then model* (Y_t) *will have* (10.19)
2. *covariance function* $c_Y(\tau) = \sum_{p=1}^{n} \sigma_p^2 \cos(2\pi s_p \tau) + c_X(\tau)$ *and*
3. *spectrum* $C_Y(s) = \sum_{p=1}^{n} \left(\frac{\sigma_p^2}{2} \delta(s - s_p) + \frac{\sigma_p^2}{2} \delta(s + s_p) \right) + C_X(s)$.

Usually, a direct spectral estimator for a mixed spectrum is calculated subsequent to (i) tapering and (ii) zero-padding the observations, as demonstrated in Sect. 10.2.1. In this section, it is demonstrated that a data taper can also help to approximate the unknown frequencies s_p in the discrete part of the spectrum to be estimated. The process under analysis is the discharge of an alpine river.

The water of River Inn at St. Moritz stems from a catchment area which is not strongly glaciated and has a reservoir in the form of lakes in the upper Engadine valley. Downriver of St. Moritz through to Cinous-chel where the next gauge is located, the Inn receives water from tributaries which drain subbasins with large glaciers and there is no reservoir. The discharge of the Inn catchment measured at Cinous-chel for the period 1974 through to 1990 (to exclude the years with accelerated climate change following 1990) is plotted in Figs. 10.9 and 10.10. The Inn discharge data are a courtesy of [132].

The discharge of River Inn is measured at five gauges located in the Engadine (in the most south-easterly part of Switzerland (Südbünden), a map is available from [132]). The hourly discharge averages for the period 1974 through to 1990 are made available in file `/path/inn7490.dat`. Columns nos. 1, 2, 3 and 4 in this file give the date (day, month, year, hour), and columns nos. 5, 6, 7, 8 and 9 contain discharge values for River Inn measured at gauges located in: (i) Martina (at the Austrian/Swiss border), (ii) St. Moritz, (iii) the channel feeding the turbines of the small St. Moritz hydro-electrical power station (only in summer), as well as both (iv) and (v) Cinous-chel. (iv) is the amount of water not caught by the Engadine Power Company and (v) is the amount of water feeding the turbines to generate hydro-electricity. Clearly, the discharge of the Inn catchment at Cinous-chel is the sum of the water caught and the water remaining in the river. These hourly values are plotted in Fig. 10.9 for the year 1974.

From Fig. 10.9, it becomes obvious that the observations are missing for day no. 213 in year 1974. The missing values are interpolated by calculating, for each hour, the mean of the values measured on days no. 212 and 214. This

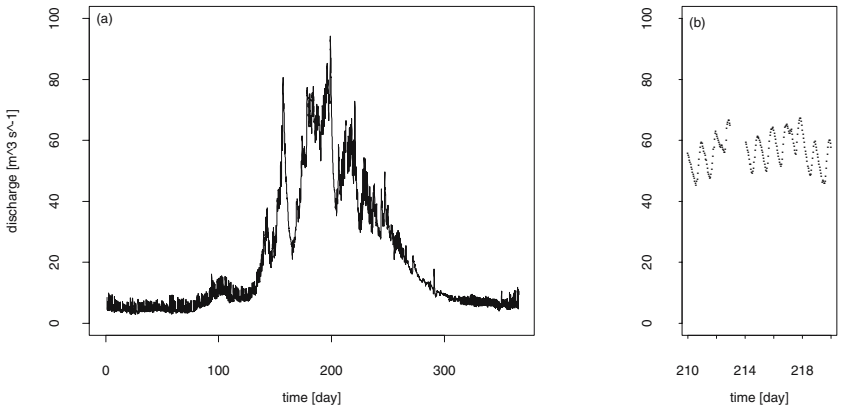


Fig. 10.9. Discharge of the Inn catchment at Cinous-chel: hourly averages in m^3s^{-1} for the year 1974.

is the only large gap in the time series, smaller ones (one to three consecutive hourly values missing) occur at the turn of the year due to maintenance of the gauges and the pertaining data acquisition systems. These gaps are filled by substituting the mean of the two adjacent values. The result of these interpolations is a complete discharge time series that is plotted on the logarithmic scale in Fig. 10.10 and made available in file `/path/inn7490nogaps.dat`. This discharge time series contains $17 \times 365 \times 24 + 4 \times 24 = 149016$ hourly values, since the observational period includes four leap years. It is called the Inn discharge in the following paragraphs. The Inn discharge is assumed to stem from a stationary stochastic process.

River Inn at Cinous-chel drains an alpine catchment area that has undergone only minor changes due to the small St. Moritz power station. This catchment area is therefore near-natural with respect to river discharge (clearly not with respect to touristic use) and its discharge shows the typical annual cycle of an alpine river, as is obvious from Figs. 10.9 and 10.10: in midsummer, the discharge is between 40 and 100 m^3s^{-1} (cubic meters per second) whereas in winter, it is less than 5 m^3s^{-1}. The annual cycle in the Inn discharge is induced by the annual cycle in temperature. The Engadine winters are cold enough such that precipitation usually falls in the form of snow and is stored in a growing seasonal snow cover (and in a growing perennial snow cover in the high alpine areas), which then melts in the warm seasons: the depletion of the snow cover results in a steep increase in the river's discharge in spring and early summer with a more exponential-like decrease in late summer and autumn.

On midsummer days with fair weather, the discharge has a mininum in the late morning and a maximum after midnight due to melting of (i) snow and ice in the high alpine areas and (ii) the large glaciers in the catchment

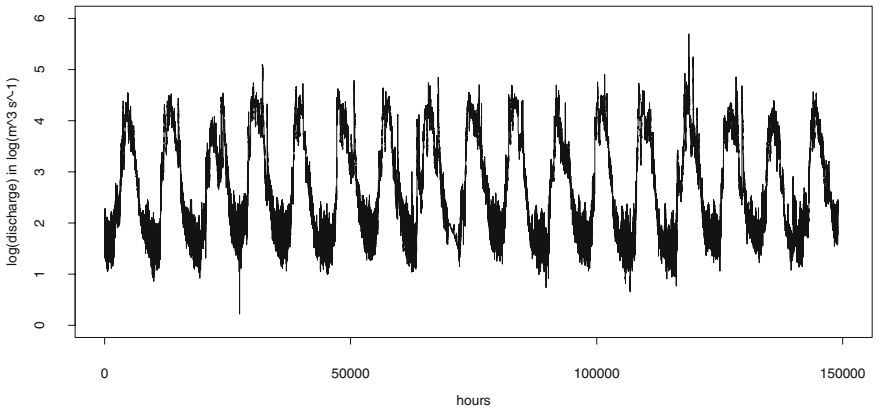


Fig. 10.10. Discharge of the Inn catchment at Cinous-chel: logarithms of the hourly averages in m^3s^{-1} in the period 1974 through to 1990 against time in hours since 1.1.1974 00.00 [day.month.year hours:minutes].

area. On days with fair weather in late spring and early summer, the amplitude of the diurnal cycle is smaller, with extremes in the morning and late evening. In winter, the diurnal cycle is due to the production of power to supply peak demand at noon: the small St. Moritz power station uses up to 5 m^3s^{-1} to drive its turbines, thus increasing the Cinous-chel discharge by approximately 3 m^3s^{-1} in the evening. The above lags in the peak discharge are due to different flow times from the sources having a diurnal cycle (high alpine areas with perennial snow cover and glaciers, areas at medium altitudes with seasonal snow cover, the St. Moritz power station) to the gauges at Cinous-chel.

These periodicities in the Inn discharge suggest that this time series stems from a process having a spectrum with a discrete and a non-constant continuous part, reconcilable with model (10.19). For approximate frequencies 1 y^{-1} (1/year) and 1 d^{-1} (1/day), as well as multiples thereof, the discrete part of the spectrum is supposed to have jump discontinuities, and the continuous part to have peaks. Can this spectrum be estimated using direct spectral estimators calculated from this time series? The answer is given below.

The Inn discharge is multiplied with cosine tapers having parameters $p = 0.0$ (the "default" taper used when a periodogram is calculated), $p = 0.05$, $p = 0.10$, and $p = 0.50$. Thereafter, direct spectral estimators are calculated for Fourier frequencies s_k with $\Delta s_k = 1/N = 1/149016$ h^{-1}. Two of the estimators calculated are shown in Fig. 10.11. All estimators calculated (including the periodogram) are not biased due to leakage generated by the side lobes in the spectral windows, a result obtained when plots of the estimators are compared. Estimators calculated for the Fourier frequencies are assumed to be favourable for detecting the unknown frequencies s_p in

Table 10.2. Frequencies of the primary and secondary peaks in direct spectral estimators calculated from the Inn discharge.

cosine	primary peak		secondary peaks						
taper	h^{-1}	dB	multiples of $1/24 = 0.041667$ h^{-1}						
$p = 0.00$	1/8765.647	47.77	1.0000	1.9973	2.9784	3.9908	4.9665	5.9647	6.9957
$p = 0.05$	1/8765.647	47.66	1.0000	1.9973	2.9918	3.9908	4.9665	5.9647	6.9821
$p = 0.10$	1/8765.647	47.52	1.0000	1.9973	2.9918	3.9908	4.9665	5.9647	6.9821
$p = 0.50$	1/8765.647	46.19	1.0000	1.9973	2.9720	4.0000	4.9665	5.9647	6.9576

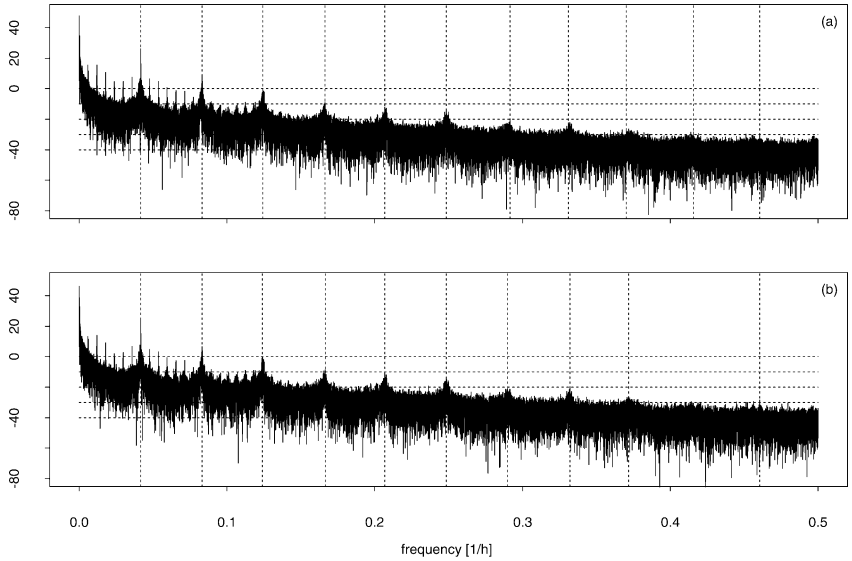


Fig. 10.11. Direct spectral estimators calculated subsequent to applying cosine tapers with $p = 0.0$ (the periodogram, in plot (a)) and $p = 0.5$ (in plot (b)) to logarithms of the hourly averages of the Inn discharge as plotted in Fig. 10.10. The frequencies of the maxima for approximate multiples of $1/24 = 0.041667$ h^{-1} as in Table 10.2 are plotted using vertical broken lines.

model (10.19), i.e., it is assumed that the s_p are in the set of Fourier frequencies, seeing that the time units (year, day, hour) correspond to natural periodicities. In the estimators thus obtained, the frequencies of the peaks as given in Table 10.2 (and plotted in Fig. 10.11 with vertical broken lines) are obtained using the methods introduced in Sect. 7.1.3 and Problem 7.2.

In all estimators calculated, the primary peak attains a height of approximately 47 dB for frequency $1/8765.647$ h^{-1}, as can be seen in Table 10.2. This frequency is the inverse of the average duration of a year in the observational period, since $N/17 = 149016/17 = (17 \times (365 \times 24) + 4 \times 24)/17 = 8765.647$ h,

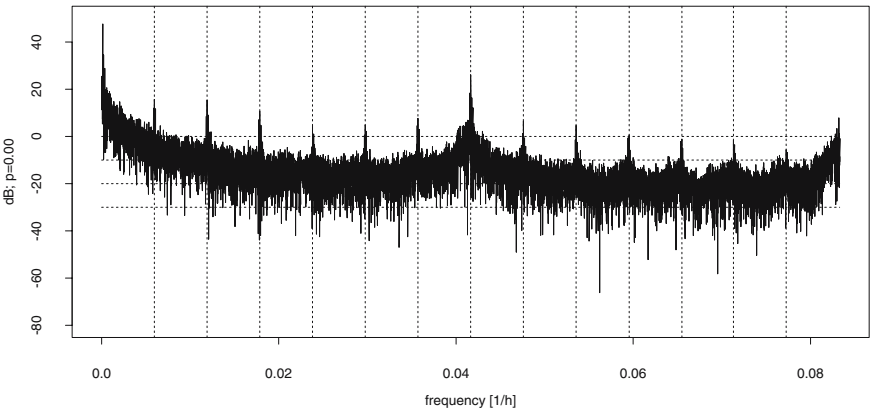


Fig. 10.12. Periodogram as in Fig. 10.11 (a) for the lower frequencies. The frequencies plotted with broken vertical lines can be approximated using multiples of frequency $0.00551422 \approx 1/168 = 1/(7 \times 24)$ h^{-1}.

as obtained in the remarks to Fig. 10.10. Consequently, the primary peak is attributable to the annual cycle in the discharge. Although the heights of the secondary peaks are not given in Table 10.2, it is obvious from Fig. 10.11 that they are smaller than the primary peak. The secondary peaks attain heights of approximately 30 dB and 10 dB for frequencies $1.0000 \times (1/24)$ h^{-1} and $1.9973 \times (1/24)$ h^{-1} (in all estimators calculated), and heights of less than 0 dB for higher frequencies, as can be seen in Fig. 10.11. Thus, the Inn discharge has both, an annual and a diurnal cycle, a finding reconcilable with the discussion in the remarks to Figs. 10.9 and 10.10.

Besides the peaks resulting from the annual and diurnal cycles, a variety of sharp peaks for frequencies being approximate multiples of $0.005952 \approx 1/168 = 1/(7 \times 24)$ h^{-1} can be seen in Fig. 10.11 and, more clearly, in Fig. 10.12, owing to its higher resolution. From these peaks it is concluded that the Inn discharge contains fluctuations which last approximately one week, in addition to the fluctuations having periods of 1 y (year) and 1 d (day) found above. This finding is reconcilable with the result obtained in Problem 10.7 from logarithmic plots of the discharge time series in the winter months. In the Inn's catchment area upstream of Cinous-chel, however, no natural phenomena exist (neither in the atmosphere, nor in the snow cover, the glaciers, the soil, the lakes or rivers) which are subject to fluctuations with a period of one week and which could thus induce a weekly cycle in the discharge. Hence, the weekly fluctuations are assumed to originate from the production of power to supply peak demand by the St. Moritz power station.

The weekly fluctuations in the discharge time series presumably do not have the form of trigonometric functions, neither do the diurnal and annual cycles, as shown in Figs. 10.9 and 10.10. However, these fluctuations can be

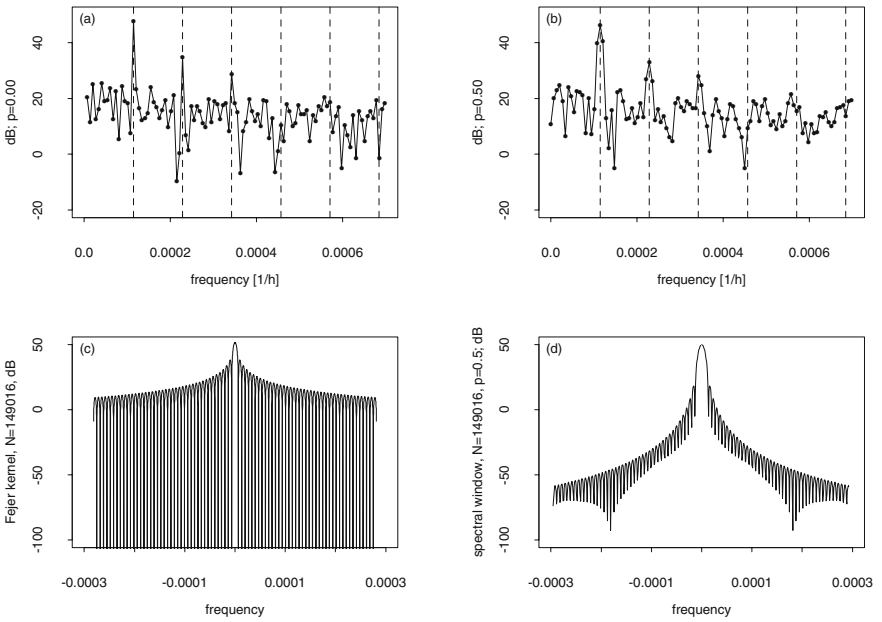


Fig. 10.13. Above, direct spectral estimators as in Fig. 10.11 for $p = 0.00$ (plot a) and $p = 0.50$ (plot b). Multiples of frequency $0.000114 = 1/8765.647$ h^{-1} (the frequency of the primary peak in Table 10.2) are given with broken vertical lines. Below, Fejer kernel for $N = 149016$ (plot c) and spectral window of the cosine taper having parameters $N = 149016$ and $p = 0.50$ (plot d).

represented using sums of trigonometric functions (having frequencies being the inverses of their periods) and the pertaining harmonics, as demonstrated in Sect. 6.1.1. The weekly and the diurnal fluctuations generate peaks in a direct spectral estimator calculated from the river Inn discharge measurements, reconcilable with those that can be seen in Figs. 10.11 and 10.12, whereas the resolution in Fig. 10.12 is not high enough to show distinct peaks for multiples of $1/8765.647$ h^{-1}, i.e., the frequency of the main peak.

From Figs. 10.11 and 10.12 it is obvious that the Inn discharge is not in agreement with model (10.15), seeing that the continuous part of the spectrum is not constant. Is the Inn discharge reconcilable with (10.19), i.e., does it have a discrete part in its spectrum? If the discharge spectrum is reconcilable with model (10.19) then it is very likely that the unknown s_p in this model are multiples of approximately $1/8765.647$, $1/168$ and $1/24$ h^{-1}, the harmonics pertaining to the trigonometric oscillations with periods of (approximately) one year, one week and one day, since fluctuations with these periods are found above, in Figs. 10.9, 10.10, 10.11 and 10.12.

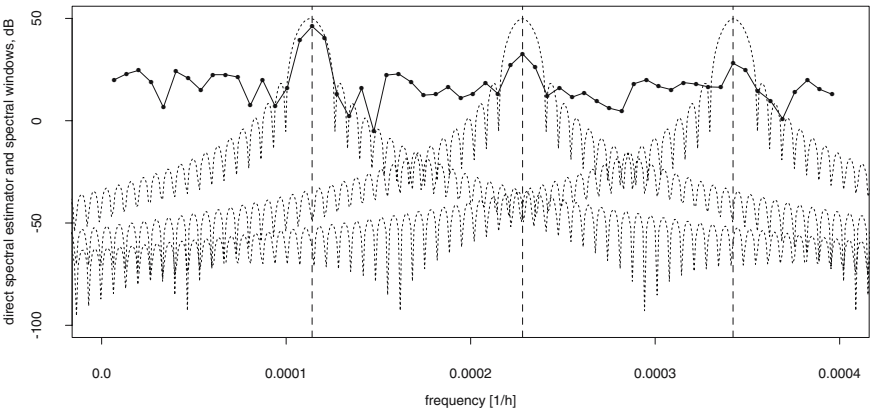


Fig. 10.14. Direct spectral estimator as in Fig. 10.13 (b) together with the spectral window pertaining to a cosine taper with parameters $N = 149016$ and $p = 0.5$ (fine points), displaced such that it attains its maximum for frequencies $\hat{s}_1 = 1/8765.647$, $\hat{s}_2 = 2/8765.647$ and $\hat{s}_3 = 3/8765.647$ h^{-1} (vertical broken lines).

Hence, oscillations in the discrete part of the discharge spectrum are sought in the spectral estimators close to frequencies being multiples of $1/8765.647$ h^{-1} (the frequency of the annual cycle). These frequencies are highlighted with vertical broken lines in Fig. 10.13, plots (a) and (b), which show the spectral estimators as in Fig. 10.11 for the very low frequencies, together with the pertaining spectral windows in plots (c) and (d). The frequencies pointed out in Fig. 10.13, plots (a) and (b), are assumed to be pilot frequencies for the unknown frequencies s_p in model (10.19). In Fig. 10.13 and also in Fig. 10.14, the "footprints" of the spectral windows can be seen in the estimators for frequencies $\hat{s}_1 = 1 \times (1/8765.647)$, $\hat{s}_2 = 2 \times (1/8765.647)$ and, albeit less clearly, $\hat{s}_3 = 3 \times (1/8765.647)$ h^{-1}, using the diagnostics introduced in the remarks to (10.18). These "footprints" become, in the mean over a large number of realisations, the spectral window of the estimator, as is concluded from (10.18) (with $C_X(s)$ substituted for σ_W^2 in the case of model (10.19)).

When the diagnostics in Figs. 10.13 and 10.14 are compared with those in Figs. 10.6 and 10.7, it becomes obvious that tracing the unknown frequencies s_p in the discrete part of the spectrum under analysis is more difficult when $(\sigma_p^2/2)$ is not manifestly larger than the continuous part of the spectrum. In this situation, it is recommended to search for the unknown model frequencies in at least two direct spectral estimators, calculated for data tapers having distinct spectral windows which produce distinct "footprints" in the estimators.

The periodogram and the direct spectral estimator in Fig. 10.13 (a) and (b) attain almost 30 dB for frequency $\hat{s}_3 = 3 \times (1/8765.647)$ h^{-1}, distinctly

(but not by orders of magnitudes) larger than the values of the estimators for the neighbour frequencies. Does an oscillation with frequency $s_p = \hat{s}_3$ exist in the discrete part of the discharge spectrum? Unfortunately, Fisher's g statistic in (8.15) cannot be applied in the case of model (10.19) to test whether the periodogram value for this frequency is significantly larger than those values for the other frequencies, since the continuous part of the spectrum is not constant, i.e., the Inn discharge is not reconcilable with model (10.1). A test for oscillations in the discrete part of the spectrum in (10.19) is introduced in [108]. This test requires that the spectrum be estimated with the multi-taper method, using the DPSSs in the remarks to (6.127) as data tapers.

However, DPSSs are not (yet) available in R. Hence, the results obtained in the above case cannot be corroborated using a formal test. Nevertheless, it is assumed that the annual cycle in the Inn discharge generates three oscillations in the discrete part of its spectrum. The frequencies of these oscillations are $\hat{s}_1 = 1/8765.647$ h^{-1}, $\hat{s}_2 = 2/8765.647$ h^{-1}, and $\hat{s}_3 = 3/8765.647$ h^{-1}, as found in Figs. 10.13 and 10.14. Summarising, it is assumed that (i) the discharge time series is reconcilable with model (10.19) and (ii) the main peak of the estimators in Fig. 10.11 and Table 10.2 is due to both, the discrete and the continuous part, in model spectrum (10.19,3).

The further peaks in Fig. 10.11 are associated, in the remarks to Fig. 10.12, with the weekly and diurnal cycles in the Inn discharge. Do both parts of the model spectrum, discrete and continuous, contribute to these peaks? Answers to this question can be found when diagnostics as proposed in the remarks to (10.18) and demonstrated above in Figs. 10.13 and 10.14 are applied when analysing these peaks. In the case of the peaks pertaining to the weekly cycle, three spikes having estimated frequencies $\hat{s}_4 = 0.005838299 = 1/171.2828$, $\hat{s}_5 = 0.005952381 = 1/168 = 1/(7 \times 24)$ and $\hat{s}_6 = 0.006066463 = 1/164.8407$ h^{-1} can be seen in Fig. 10.15 (a). The distance between these frequencies is $\hat{s}_5 - \hat{s}_4 = \hat{s}_6 - \hat{s}_5 = 17/149016 = 17\Delta s_k$, Δs_k as defined in (6.21,1). The periods of these oscillations are $149016/870 \approx 171.2828$, $149016/887 \approx 168$ and $149016/904 \approx 164.8407$ h.

Triads of spikes can also be seen in the discharge periodogram in bands with centres in frequencies $\ell \times \hat{s}_5 = \ell \times (1/168) = \ell \times 1/(7 \times 24)$ h^{-1}, $\ell = 2, \ldots, 12$, i.e., the harmonics to frequency $1/(7 \times 24)$ h. In the band of the first harmonic, for example, the spikes have frequencies $\hat{s}_7 = 2/168 - 17\Delta s_k$, $\hat{s}_8 = 2/168$ and $\hat{s}_9 = 2/168 + 17\Delta s_k$ h^{-1}. In all these bands, the distance between the spikes remains constant $17\Delta s_k$. Consequently, $\hat{s}_8 = 2/168$, $\hat{s}_9 = 3/168$, ..., are the harmonics to $\hat{s}_5 = 1/(7 \times 24)$ h, whereas frequencies $\hat{s}_4 = 1/168 - 17\Delta s_k$ and $\hat{s}_6 = 1/168 + 17\Delta s_k$ have no harmonics.

For $\ell = 7$, the band of the diurnal cycle is arrived at. In this band, the periodogram has four spikes, as can be seen in Fig. 10.15 (b). Of these, the spikes with frequencies $7/168$, $7/168 - 17\Delta s_k$ and $(7/168) + 17\Delta s_k$ are assumed to be owed to the weekly cycle. The fourth peak will be discussed in the remarks to Fig. 10.19. Hence, from the discharge periodogram as plotted

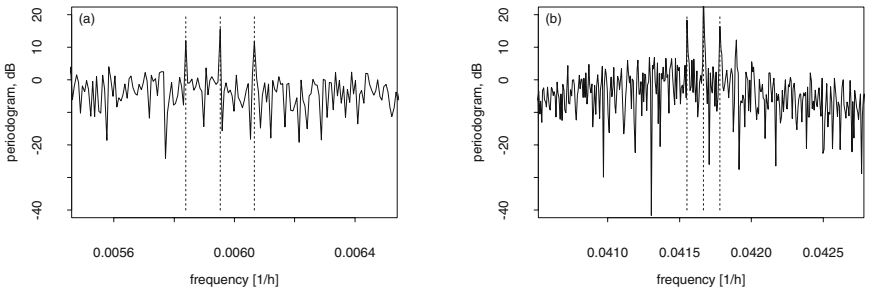


Fig. 10.15. Periodogram as in Fig. 10.11 (a) and Fig. 10.12 for bands with centres in $1/(7 \times 24) = 1/168 = 0.005952381$ h^{-1} (plot (a)) and $7/168 = 1/24 = 0.04166667$ h^{-1} (plot (b)), i.e., in the bands associated with the weekly and the daily cycles in the Inn discharge.

in Fig. 10.15 it becomes evident that (i) the weekly cycle in the river Inn discharge time series contributes to the discrete part of the discharge spectrum for frequencies smaller than $0.072 \approx 12 \times (1/(7 \times 24)$ h^{-1}, and (ii) the diurnal cycle in this time series is likely to contribute only to the continuous part of the discharge spectrum.

The results obtained above using the diagnostics in Figs. 10.13, 10.14 and 10.15 are summarised as follows:

1. model (10.19) is reconcilable with the Inn discharge
2. the discrete part of the model spectrum contains $\hat{n} = 39$ trigonometric oscillations having frequencies $\hat{s}_p$, $p = 1, \ldots, \hat{n}$
3. three oscillations, having estimated frequencies $\hat{s}_1 = 1/8765.647$, $\hat{s}_2 = 2/8765.647$ and $\hat{s}_3 = 3/8765.647$ h^{-1}, are due to the annual cycle
4. 36 oscillations, having estimated frequencies $\hat{s}_5, \ldots, \hat{s}_{38} = \ell \times (1/(7 \times 24))$ h^{-1}, $\hat{s}_4, \ldots, \hat{s}_{37} = \ell \times (1/(7 \times 24)) - 17\Delta s_k$ h^{-1} and $\hat{s}_6, \ldots, \hat{s}_{39} = \ell \times (1/(7 \times 24)) + 17\Delta s_k$ h^{-1}, with $\ell = 1, \ldots, 12$, are due to the weekly cycle and
5. it is very likely that the diurnal cycle does not contribute to the discrete part of the model spectrum.

These results are used in the following paragraphs to fit model (10.19) to the Inn discharge. In a first step, the coefficients of trigonometric oscillations with frequencies $\hat{s}_p$, $p = 1, 2, \ldots, \hat{n}$ and $\hat{n} = 39$ as obtained above, are estimated. The periodogram values for these frequencies, as plotted in Fig. 10.16, are calculated in Problem 10.9 and estimates $\widehat{A}(s_k = \hat{s}_p)$ and $\widehat{B}(s_k = \hat{s}_p)$ are obtained in Problem 10.10. Hereby, the discrete part of this model spectrum is estimated. In a second step, using the coefficients thus obtained in (10.20),

$$\hat{x}_t = y_t - \sum_{p=1}^{\hat{n}} \Big(\big(\widehat{A}(\hat{s}_p) \cos(2\pi \hat{s}_p t) + \widehat{B}(\hat{s}_p) \sin(2\pi \hat{s}_p t) \big) \Big), \quad \begin{array}{l} \hat{s}_p \text{ and } \hat{n} \\ \text{as above} \end{array} \qquad (10.20)$$

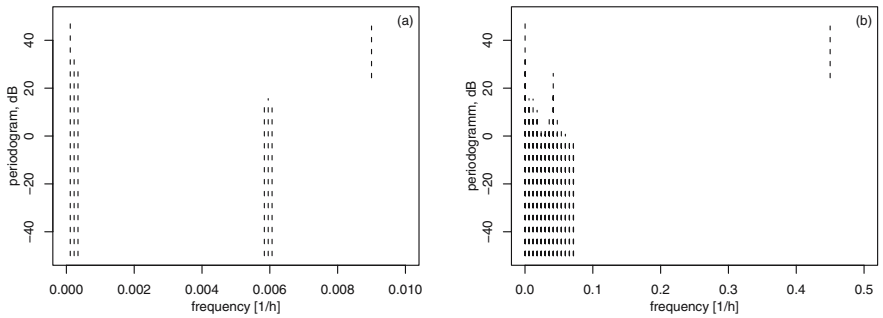


Fig. 10.16. Periodogram as in Fig. 10.11 (a) (together with its .95 confidence interval) for three frequencies $l \times (1/8765.647)$ h^{-1}, $l = 1, 2, 3$, and 36 frequencies $(\ell \times (1/168)) - 17\Delta s_k$ h^{-1}, $\ell \times (1/168)$ h^{-1} and $(\ell \times (1/168)) + 17\Delta s_k$ h^{-1}, $\ell = 1, \ldots, 12$. The oscillations with these frequencies are in the discrete part of (10.19,3), i.e., the spectrum of the model found to be reconcilable with the Inn discharge as plotted in Fig. 10.10.

empirical residuals $(\hat{x}_t)$ will be computed in Problem 10.10. The $\hat{x}_t$, $t = 0, 1, \ldots, N-1$, $N = 149016$ thus obtained are a time slice in a realisation of (X_t) in model (10.19,1), on condition that this model is reconcilable with the discharge time series, as assumed above.

The continuous spectrum $C_X(s)$ of (X_t) is estimated using methods as introduced in Chap. 9, subsequent to approximating $\mathcal{B}_{\triangle}\big(C_X(s)\big)$, i.e., the bandwidth of the continuous part of the discharge spectrum as defined in (7.83,1), using diagnostic (9.93,6) as follows. From Figs. 10.12, 10.13, 10.14, and 10.15 it is concluded that the continuous part of the discharge spectrum has its narrowest peak in the band of the daily cycle. Hence, the periodogram and the direct spectral estimator shown in Fig. 10.11 are plotted on the non-logarithmic scale for the band of the daily cycle (not shown). From these plots it is concluded that the bandwidth of the continuous part is approximately the distance of the $\hat{s}_p$ in the frequency triads in the discrete part of the discharge spectrum, i.e., $\mathcal{B}_{\triangle}\big(C_X(s)\big) \approx 17\Delta s_k$, $\Delta s_k = 1/N$. Hence, a periodogram $I_X(s_k)$ is computed from the residual time series (x_t) as defined in (10.20) and thereafter smoothed using a modified Daniell kernel (9.29) with parameter $M = 5$. The smoothed periodogram $I_X^{(m)}(s_k)$ thus arrived at is an estimator for the continuous part of the discharge spectrum. Its bandwidth is approximately $2M\Delta s_k = 10\Delta s_k$, calculated as required in (9.90).

$I_X^{(m)}(s_k)$ for $M = 5$ is plotted in Figs. 10.17 and 10.18. Both figures also show differences $I_Y(s_k) - I_X^{(m)}(s_k)$, $I_Y(s_k)$ as in Fig. 10.16, for $\hat{n} = 39$ frequencies $s_k = \hat{s}_p$, $p = 1, \ldots, \hat{n}$, as estimates of the discrete part of the discharge spectrum (triads of broken vertical lines). The bandwidth of the estimator for the continuous part of the spectrum is plotted as the horizonal bar in the solid cross (not resolved in Fig. 10.18), and .95 confidence intervals

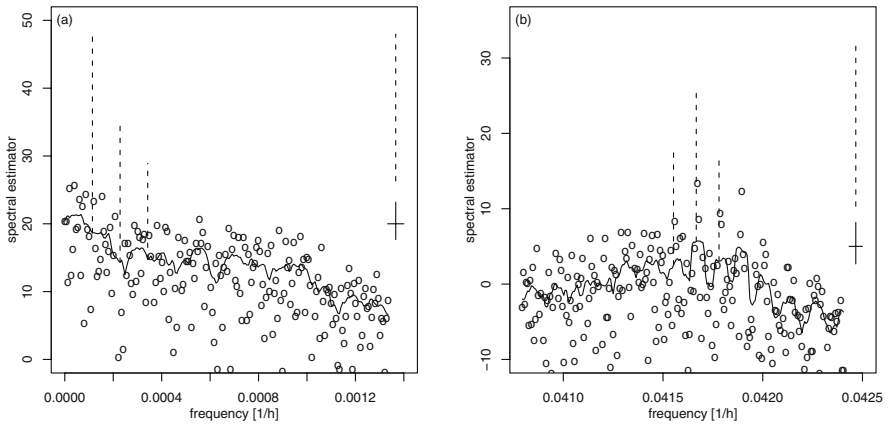


Fig. 10.17. Spectrum of the Inn discharge estimated using a periodogram calculated for the Fourier frequencies in the bands of the annual (plot a) and the diurnal (plot b) cycles. The periodogram as in Fig. 10.16 is the estimator for the discrete part (triads of spikes plotted with broken lines). The periodogram calculated subsequent to subtracting the oscillations in the discrete part from the time series (plotted with symbol ∘) becomes an estimator for the continuous part when smoothed using a modified Daniell kernel with $M = 5$ (solid line).

are plotted using vertical bars: a solid one in the cross for $I_X^{(m)}(s_k)$ and, above the cross, a broken one for $I_Y(s_k)$. Consequently, Figs. 10.17 and 10.18 show an estimate for the spectrum of model (10.19), which is assumed to be in agreement with the Inn discharge as plotted in Fig. 10.10.

In Fig. 10.17 (a), the continuous discharge spectrum attains approximately 20 dB for the very low frequencies. There, the river Inn discharge has also three trigonometric oscillations in its discrete spectrum: the one having frequency $\hat{s}_1 = 1/8765.647$ h^{-1} (the inverse of the average duration of a year in the observational period, as in the remarks to Table 10.2 and Fig. 10.11) as well as those for two harmonics of $\hat{s}_1$, i.e., $\hat{s}_2 = 2/8765.647$ and $\hat{s}_3 = 3/8765.647$ h^{-1}. Hence, the annual cycle in the discharge contributes to both parts in the model spectrum, the continuous and the discrete ones.

The weekly cycle in the river Inn discharge also contributes to both parts of the discharge spectrum in the following manner:

1. the band with centre in frequency $\hat{s}_5 = 1/168 = 1/(7 \times 24)$ h^{-1} contains, in addition, $\hat{s}_4 = 1/168 - 17\Delta s_k$ and $\hat{s}_6 = 1/168 + 17\Delta s_k$, i.e., the frequencies of three oscillations in the discrete spectrum, yet no peak can be seen in the continuous spectrum in this band
2. bands with centres in frequencies $\ell \times \hat{s}_5$, $\ell = 2, \ldots, 12$ contain triads of frequencies in the discrete spectrum and peaks in the continuous spectrum, and

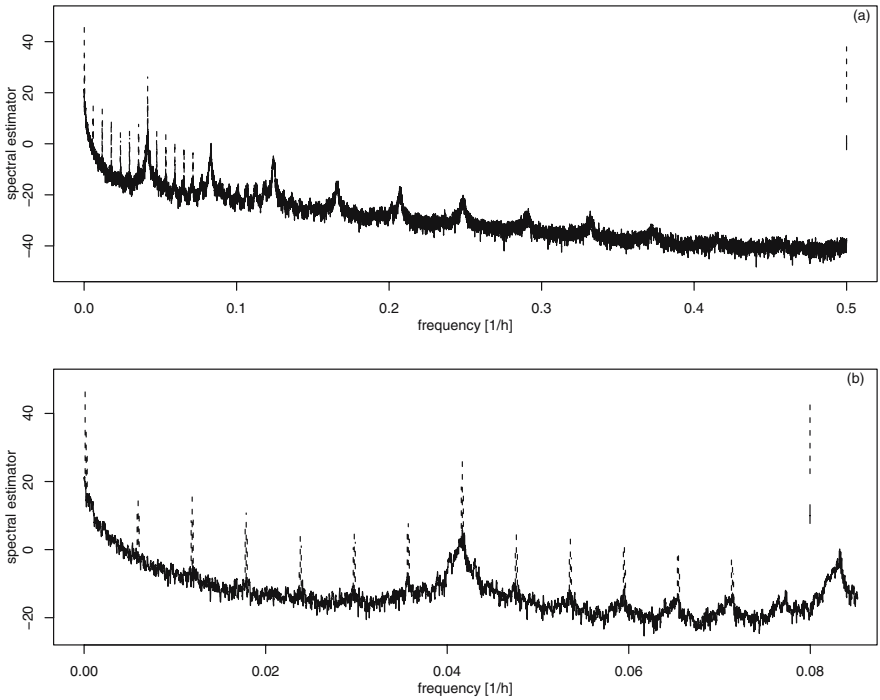


Fig. 10.18. Spectrum of the Inn discharge as in Fig. 10.17 (the values plotted with ∘ in Fig. 10.17 are not plotted here).

3. bands with centres in frequencies $\ell \times \hat{s}_5$, $\ell = 13, \ldots, 22$ contain peaks in the continuous spectrum

as can be seen in Figs. 10.18, 10.11, 10.12 and 10.15. From this estimate for the spectrum of the weekly cycle it is concluded that St. Moritz power station produces fluctuations in the Inn discharge which have a period of approximately one week.

The discrete spectrum due to the weekly cycle contains, on both sides of the multiples of $\hat{s}_5 = 1/168 = 1/(7 \times 24)$ h^{-1}, additional frequencies in distances of $17\Delta s_k$, as plotted in Fig. 10.18 (b), and consequently, the discrete spectrum due to the weekly cycle in the discharge of the Inn contains triads of frequencies as traced in Figs. 10.12 and 10.15 and summarised in the remarks to Fig. 10.15. These triads are presumably owed to changes from normal to daylight saving time in spring and back to normal time in autumn, since these changes displace the time of the daily and weekly cycles in electricity consumption against the time of the discharge measurements (always in normal time).

The estimated discharge spectrum of river Inn for the band pertaining to the daily cycle is plotted in Fig. 10.17 (b). There, $I_X(s_k)$ (plotted using ∘) con-

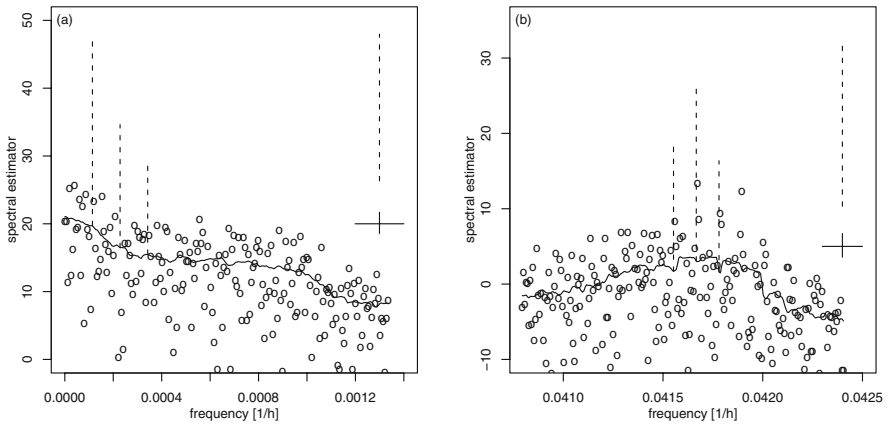


Fig. 10.19. As Fig. 10.17, except for smoothing with a modified Daniell kernel having parameter $M = 15$.

ceals a variety of relatively small peaks which become visible in $I_X^{(m)}(s_k)$ (solid line) obtained by smoothing $I_X(s_k)$ using a modified Daniell kernel having parameter $M = 5$. The distance between one peak and the other is approximately $17\Delta s_k$, i.e., the bandwidth of the continuous part of the discharge spectrum as approximated in the remarks to (10.20). Three of these peaks are approximately centred in frequencies of oscillations in the discrete part of the spectrum due to the weekly cycle, namely $(7 \times (1/168)) - 17\Delta s_k$ h^{-1}, $7 \times (1/168)$ h^{-1} and $(7 \times (1/168)) + 17\Delta s_k$ h^{-1}. The other small peaks in the continuous spectrum do not have oscillations with approximately identical frequencies in the discrete spectrum. It is therefore assumed that the small peaks in the continuous part of the spectrum as estimated in Fig. 10.17 (b) are due to the changes from normal to daylight saving time in spring and back to normal time in autumn for the reason given above in the discussion of the triads in the discrete part of the discharge spectrum (Problem 10.12). These peaks disappear when the bandwidth of $I_X^{(m)}(s_k)$ increases, as is demonstrated in Fig. 10.19 (b). There, $I_X(s_k)$ is smoothed using a modified discrete Daniell kernel having parameter $M = 15$.

The Inn discharge has a spectrum that is continuous in the band of the daily cycle (except for the oscillations with frequencies $(7 \times (1/168)) - 17\Delta s_k$ h^{-1}, $7 \times (1/168)$ h^{-1} and $(7 \times (1/168)) + 17\Delta s_k$ h^{-1} due to the weekly cycle) on condition that the catchment area upstream of Cinous-chel acts as a stochastic LTI filter and the daily disturbances have a continuous spectrum. The first assumption is quite often reconcilable with catchment areas in general and the second one is not unrealistic for the following reasons:

1. the melting of the snow cover and of glacier ice feeding the Inn depends on temperature as well as on other variables (solar radiation, cloud cover,

precipitation, wind speed, etc.) which all have interdependencies and thus can have (more or less) pronounced diurnal cycles, and
2. the St. Moritz power station starts and stops its turbines producing power to supply peak demand at not exactly identical time points each day.

Hence, the diurnal cycle in the Inn discharge is assumed to have a continuous spectrum.

10.2.3 Summary

Models (10.15) and (10.19) for discrete-time stationary stochastic processes having a mixed spectrum contain trigonometric oscillations with unknown frequencies s_p, $p = 1, 2, \ldots, n$. Subsequent to calculating direct spectral estimators, approximations $\hat{s}_p$ for s_p can be obtained using the diagnostics introduced in the remarks to (10.18), and thereafter, these models can be estimated. The estimation of these models is summarised as follows:

1. direct spectral estimators $\widehat{C}_X^{(d)}(s_k)$ or $\widehat{C}_X^{(d)}(s'_k)$ (in the case of model (10.15)) and $\widehat{C}_Y^{(d)}(s_k)$ or $\widehat{C}_Y^{(d)}(s'_k)$ (in the case of model (10.19)) are computed using a variety of data tapers from time slices in realisations of the process under analysis for Fourier frequencies s_k, and/or subsequent to zero-padding, for frequencies s'_k with $\Delta s'_k < \Delta s_k$
2. the frequencies of the maxima in these estimators become pilot frequencies $\hat{s}_p$, i.e., preliminary approximations for the model frequencies
3. displaced spectral windows $H_N^{(d)}(s - \hat{s}_p)$ are plotted together with $\widehat{C}_X^{(d)}(s'_k)$ or $\widehat{C}_Y^{(d)}(s'_k)$. If $\hat{s}_p$ is close to s_p, then the spectral estimator plotted is similar to the spectral window, seeing that, in the mean over many realisations, the spectral estimators become identical with the spectral window
4. alternatively, the unknown s_p can be approximated using non-linear least squares
5. substituting the $\hat{s}_p$ thus found for the s_p in the models, the coefficients of the trigonometric oscillations can be estimated using linear least squares
6. from the residuals of the sum of the trigonometric oscillations thus obtained, the continuous part of the model spectrum can be estimated (using the methods introduced in Chap. 9 in the case of model (10.19)),

and is demonstrated by estimating two example spectra in Sects. 10.2.1 and 10.2.2.

Do trigonometric oscillations, found and estimated using the above procedure, exist in the stochastic process under analysis? Does the process under analysis have an integrated spectrum that jumps for frequencies $\hat{s}_p$ obtained using the above diagnostics? A test for oscillations in the discrete part of the spectrum in (10.15) and (10.19) is introduced in [108]. This test requires that the spectrum be estimated with the multi-taper method, using the DPSSs in the remarks to (6.127) as data tapers. This is clearly an asset attached to

the multi-taper method as compared with the classical methods introduced in this chapter. DPSSs are however not (yet) available in R.

10.3 Supplements

The moments of coefficients in the trigonometric oscillations in model (10.1) are derived in Sect. 10.3.1, and in Sect. 10.3.2 the effects of detrending on spectral estimators is demonstrated for the case of model (10.1).

10.3.1 Estimating the Model defined in Sect. 10.1

The expectations of $\widehat{A}_X(s_k)$ and $\widehat{B}_X(s_k)$ in (10.3) and (10.4) are derived by substituting model (10.1,1) for X_t in the right sides of (10.3) and (10.4). Thereafter, orthogonality relations (6.11), (6.12) and (6.13) allow for reducing the intricacy of the expressions, as is demonstrated for the cases $k \neq 0$ and $k \neq N/2$ below. Thereafter, the rules for calculating expectations as enumerated in (1.15) and (1.16) are applied to obtain the expectations in (10.21) since, in the segment of the stochastic process observed, the W_t are random variables depending on time and the A_k and B_k are constants for each realisation.

$$
\begin{aligned}
&\mathrm{E}\widehat{A}_X(s_k) = \mathrm{E}\left(\frac{2}{N}\sum_{t=0}^{N-1} X_t \cos(2\pi s_k t)\right) \qquad -1/2 < s_k < 0 < s_k < 1/2 \\
&= \mathrm{E}\left(\frac{2}{N}\sum_{t=0}^{N-1}\left(\sum_{k=1}^{m}\big(A_k\cos(2\pi s_k t) + B_k \sin(2\pi s_k t)\big) + W_t\right)\cos(2\pi s_k t)\right) \\
&= \mathrm{E}\left(\frac{2}{N}\sum_{t=0}^{N-1}\left(\sum_{k=1}^{m}\big(A_k\cos(2\pi s_k t) + B_k \sin(2\pi s_k t)\big)\right)\cos(2\pi s_k t)\right) + \\
&\qquad\qquad \mathrm{E}\left(\frac{2}{N}\sum_{t=0}^{N-1} W_t \cos(2\pi s_k t)\right) \\
&= \mathrm{E}\left(\frac{2}{N}\frac{N}{2}A_k\right) + \frac{2}{N}\sum_{k=1}^{m}\big(\mathrm{E}W_t\cos(2\pi s_k t)\big) \\
&= A_k \qquad \text{and similarly} \qquad \mathrm{E}\widehat{B}_X(s_k) = B_k
\end{aligned}
\tag{10.21}
$$

For the very same reason, the variance of the A_k and B_k becomes identically zero for each realisation and therefore, subsequent to applying the orthogonality relations, the variances of the estimators are arrived at in (10.22).

$$
\mathrm{Var}\widehat{A}_X(s_k) = \mathrm{Var}\left(\frac{2}{N}\sum_{t=0}^{N-1} X_t \cos(2\pi s_k t)\right) \qquad -1/2 < s_k < 0 < s_k < 1/2
$$

$$= \mathrm{Var}\left(\frac{2}{N}\sum_{t=0}^{N-1}\left(\sum_{k=1}^{m}\big(A_k\cos(2\pi s_k t) + B_k\sin(2\pi s_k t)\big)\right)\cos(2\pi s_k t)\right) + \mathrm{Var}\left(\frac{2}{N}\sum_{t=0}^{N-1} W_t\cos(2\pi s_k t)\right)$$

$$= 0 + \left(\frac{2}{N}\right)^2 \sigma_W^2 \sum_{k=1}^{m}\big(\cos(2\pi s_k t)\big)^2$$

$$= \frac{2}{N}\sigma_W^2 \qquad \text{and similarly} \qquad \mathrm{Var}\widehat{B}_X(s_k) = \frac{2}{N}\sigma_W^2 \tag{10.22}$$

Clearly, expressions for $\mathrm{E}\widehat{A}_X(s_k)$, $\mathrm{E}\widehat{B}_X(s_k)$, $\mathrm{Var}\widehat{A}_X(s_k)$ and $\mathrm{Var}\widehat{B}_X(s_k)$ become much more intricate when these estimators are calculated for frequencies $s \neq s_k$, s_k the Fourier frequencies as defined in (6.21,1). The reason is that the orthogonality relations can only be applied as shown above on condition that the estimators are calculated for the s_k in (6.21,1).

10.3.2 Detrending and Spectral Estimation: an Example

Temperature measured in Basel increases since 1755 in the secular mean, as can be seen in Fig. 3.3. Hence, the monthly means in the Basel temperature series are not in agreement with model (10.1,1), because this model is stationary. Is this model reconcilable with the residuals of a first order linear model fitted to the monthly means in the Basel series, as assumed in Sect. 10.1.3? In Fig. 5.7 it can be seen that decadal changes remain in the residuals of a first order linear model for the secular trend fitted to the yearly values in the Basel temperature series. Hence, the residuals of a first order linear model fitted to the monthly values in the Basel series as in Fig. 10.1 (a) also have decadal changes in their mean, and thus are not in agreement with model (10.1). This disagreement is supected of being the source of extra low-frequency oscillations in the periodogram in Fig. 10.1 (b).

As an alternative to de-trending a time series by fitting a linear model and thereafter calculating the residuals, (5.58,1) recommends to calculate its differences. For example, first order differences of the yearly values in the Basel temperature series are plotted in Fig. 5.8. In these differences, no decadal fluctuations can be seen, unlike in the residuals of the linear model in Fig. 5.7. Differencing is often superior in performance as compared to calculating the residuals of a linear model, as demonstrated by the second example given in Sect. 5.4.1: first order differences calculated with lag 12 (above) of the monthly values in the Hawaii atmospheric CO_2 as plotted in Fig. 5.11 are less subject to decadal fluctuations than the residuals of a component model as defined in (3.19) for the decadal trend and annual cycle as plotted in Fig. 5.10.

If a time series stems from a process with a constant and known period M then fluctuations having this period are removed together with possible trends

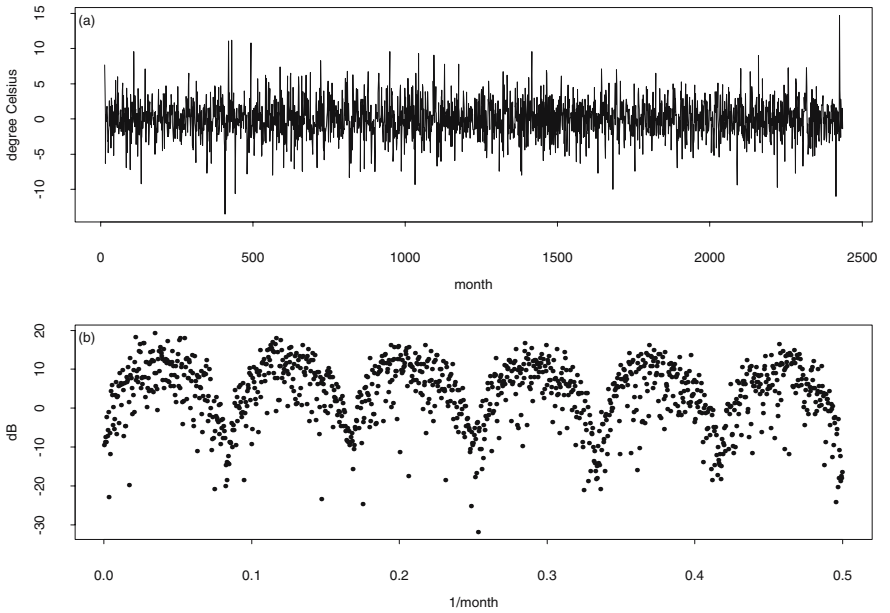


Fig. 10.20. Monthly values in the Basel temperature series introduced in the remarks to Fig. 2.13: differences in lag 12 (above, plot (a)) together with their periodogram (below, plot (b)).

when differences in lag M are calculated. In the case of the monthly means in the Basel temperature series, differences with lag 12 are calculated and plotted in Fig. 10.20 (a). Thereafter, the periodogram of these differences is calculated and plotted in Fig. 10.20 (b). When this periodogram is compared with the one in Fig. 10.1, it becomes obvious that differencing

1. removes the annual cycle in the temperature time series since no large periodogram values for $s_k = 1/12$ and $s_k = 1/6$ month^{-1} can be seen in Fig. 10.20 (b), and
2. generates a periodic periodogram in accordance with the properties of the difference filter as derived in (6.122) and (6.123) and demonstrated in Fig. 6.30.

The properties of a periodogram calculated from a difference time series are further demonstrated in Problem 10.13.

The undesired periodicity of the periodogram in Fig. 10.20 (b), owed to differencing (with lag 12) the time series under analysis, can be avoided on condition that the monthly values in the Basel temperature series are detrended by means of an alternative method. For example, decadal fluctuations in the Basel temperature series are captured using local (2.55) linear models estimated in the time slices found in Fig. 6.32. The residuals of these

Table 10.3. Periodograms in Figs. 10.1 (b) and 10.21 (b) for frequencies 0.0833 and 0.1666 month^{-1} (the frequency of the annual cycle and its first harmonic) and for frequencies ≤ 0.0030 month^{-1}.

s_k	.0833	.1666	.0005	.0008	.0012	.0016	.0020	.0026	.0029
Fig. 10.1 (b), dB	47.21	20.62	15.13	7.57	13.84	3.57	16.23	2.33	10.40
Fig. 10.21 (b), dB	47.25	20.64	−20.75	−5.54	5.25	0.94	4.71	1.21	2.00

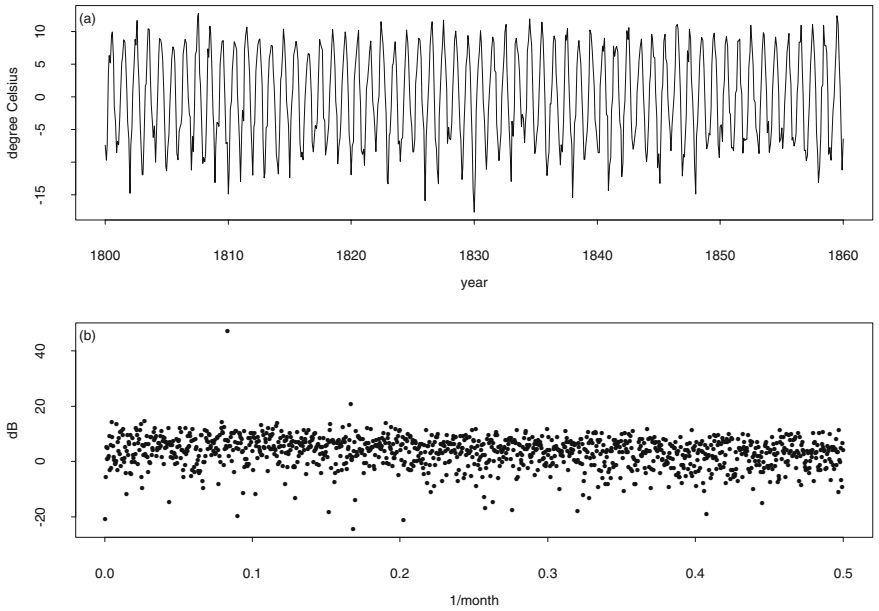


Fig. 10.21. Residuals of the local models for the decal trends as estimated in Fig. 6.32 in the Basel temperature series (plotted for the time slice from 1800 through to 1860) together with the periodogram computed using all residuals available, i.e., those within the observational period from 1755 through to 1957.

local linear models fitted to the Basel temperature monthly means are plotted in Fig. 10.21 (a). From these residuals, a periodogram is calculated for the Fourier frequencies and plotted in Fig. 10.21 (b), and in Table 10.3 the values in this periodogram are compared with those in the periodogram plotted in Fig. 10.1 calculated from the residuals of a global first order linear model for the secular trend in the Basel temperature series.

In Table 10.3 it can be seen that oscillations with frequencies between 0.0 and 0.003 month^{-1} (corresponding to periods of more than 30 years) have larger absolute amplitudes in the residuals of the global first order linear

model for the secular trend (in Fig. 10.1 (a)) than in the residuals of local linear models (in Fig. 10.21 (a)). It thus becomes evident that fluctuations having periods of more than 30 years in the Basel temperature series are captured by the local linear models as plotted in Fig. 6.32 and, consequently, are removed when the residuals of these local models are calculated.

If a periodogram of the monthly means in the Basel temperature series is calculated as demonstrated in Fig. 10.1 then it is believed that the decadal fluctuations in this series should be represented using the trigonometric oscillations in model (10.1,1). If, in contrast, a periodogram of this time series is calculated as shown in Fig. 10.21 then it is believed that the decadal fluctuations in temperature are not a stationary phenomenon and therefore should be removed prior to estimating (10.1,1). In both cases nevertheless, identical results are obtained, namely that there is an annual cycle in the Basel temperature series, obviously a finding not new. New, however, is that the annual cycle can be represented using two trigonometric oscillations once the Basel temperature series is assumed to be in agreement with model (10.1,1).

10.4 Problems

10.1. Apply R expressions similar to those in `/path/harmproc.R` introduced in Problem 7.7 to generate simulations of model (10.1,1). Try a sequence of increasing values for the variance σ_W^2 of the white noise process together with a sequence of decreasing squared absolute values σ_p^2 for the trigonometric oscillations. For which ratio σ_p^2/σ_W^2 are the oscillations no longer visible in the simulated realisations?

10.2. Estimate primary and secondary peaks in the direct spectral estimator in Fig. 10.5 calculated by applying a cosine taper with $p = 0.30$ to the simulation in Fig. 10.3. Compile a table showing the frequencies and the heights of the peaks in this estimator for frequencies larger than 0.3 u^{-1}. Hint: the heights of the peaks for frequencies smaller than 0.3 u^{-1} are given in Table 10.1.

10.3. It is assumed that the simulation in Fig. 10.3 is reconcilable with model (10.15) having two dominant trigonometric oscillations as estimated in Fig. 10.8 and harmonics pertaining to these oscillations. Under this assumption, approximate the unknown frequencies using the results in Table 10.1 as well as those obtained in Problem 10.2 and thereafter apply linear models as demonstrated in Sect. 10.1.3 to obtain least squares estimates for the model parameters. First try a model having $n = 6$ oscillations and thereafter repeat the estimates for models having an increasing number of oscillations. Apply the diagnostics introduced in Sect. 3.3 to your estimates. Do you arrive at empirical residuals reconcilable with both, non-correlation for lags $\tau \neq 0$ as required in (3.11) and a variance of -25 dB as required by the constant continuous part of the direct spectral estimator in Figs. 10.5, 10.6 and 10.7?

10.4. Estimate the jump discontinuities $\sigma_p^2/2$, $p = 1, \ldots, n$, in the integrated spectra of the models you have estimated in Problem 10.3.

10.5. The following R expressions approximate, by applying non-linear least squares, the unknown frequencies of two trigonometric oscillations in a model (10.15) for the simulation in Fig. 10.3.

```
fs <- ts(scan("/path/allegedstar.dat"),start=0,frequency=1,)
y <- as.array(fs)
t <- as.array(time(fs))
#s1 <- 0.03450   as in Table 10.1: error in nls(s),
#s2 <- 0.04167   too close to frequencies of local maximum?
s1 <- 0.035    #0.034442 in [14]
s2 <- 0.041    #0.041724 in [14]
a <- 2*pi*s1*t
x <- cbind(cos(a),sin(a))
a <- 2*pi*s2*t
x <- cbind(x, cos(a), sin(a))
l <- lsfit(x, y)
n <- nls(y ~ mu + a1*cos(2*pi*s1*t) + b1*sin(2*pi*s1*t) +
                  a2*cos(2*pi*s2*t) + b2*sin(2*pi*s2*t),
             data = data.frame(y,t),
             start = list(mu = l$coef[1],
                          a1 = l$coef[2], b1 = l$coef[3],
                          a2 = l$coef[4], b2 = l$coef[5],
                          s1 = s1, s2 = s2))
```

These R expressions being identical with the Splus expressions given in [14], with `nls()` being described in [31], produce the following results:

```
           mu           a1           b1           a2           b2
17.085781782  7.647775769  6.490568484  0.001119107  7.084562755
           s1           s2
 0.034482436  0.041666485
```

which are quite similar to those obtained using `lm()` in the remarks to Fig. 10.7, and

```
residual sum-of-squares: 54.67214
```

from which the variance of the white noise process is obtained: $\breve{\sigma}_W^2 = 54.67214/595 = 0.09988595$ and thus $\breve{\sigma}_W^2 = -10.36751$ dB as well as $\breve{s}_W = 0.3031270$, i.e., also not too far from $\tilde{\sigma}_W = 0.3317$ obtained in the remarks to Fig. 10.7.

Try other initial values for the frequencies of the trigonometric oscillations. Thereafter discuss the following statement: a non-linear least-squares estimation of the unknown frequencies in model (10.15) is inferior in performance when compared with the diagnostics introduced in the remarks to (10.18) and demonstrated in Figs. 10.6 and 10.7.

10.6. Example process no. 4 in Sect. 7.3.4 is the sum of two stationary processes which are not correlated, a property that favours the derivation of its correlation function and its spectrum in (7.80) and (7.81). Borrowing from this derivation, obtain the covariance function and the spectrum of model (10.19).

10.7. Hourly values of the Inn discharge as plotted in Fig. 10.10 are available in file `/path/inn7490nogaps.dat`. This file is read with

```
engadin <- read.table("/path/inn7490nogaps.dat", header=T)
#logarithms of Inn discharge at Cinous-chel
inn <- log(engadin$cich + engadin$ewci)
```

Plot time slices from December through to April for a few winter seasons on the logarithmic scale.

10.8. Compute the periodogram of the Inn discharge using

```
x <- inn - mean(inn)
innperio <- spec.univariate(x,taper=0.0,spans=1,
                                  pad=0,confint=0.95,fast=F)
par(mfrow=c(1,2))
plot(innperio$freq,10*log10(innperio$spec),pch=".")
```

and, alternatively, using

```
N <- length(x)
m <- N/2
xt <- fft(x,inverse=T)
xtsc <- ((xt/N)*2)[1:(m+1)]
xtsc[1] <- xtsc[1]/2
xtsc[m+1] <- xtsc[m+1]/2
ampl2 <- (Mod(xtsc))*(Mod(xtsc))
sk <- (0:m)/N
pgr <- (N/4)*ampl2
pgr[1] <- N*ampl2[1]
pgr[m+1] <- N*ampl2[m+1]
pgrdb <- 10*log10(pgr)
pgrdb[1] <- NA
plot(sk,pgrdb,pch=".").
```

Are the plots thus obtained identical? Compare the above R expressions with those in `spec.univariate()` as introduced in Sect. 9.5.6 and with definition (8.1). Enhance the above R expressions with comments.

10.9. Use the scaled Fourier transform `xtsc` as calculated above in Problem 10.8 to estimate the discrete part in the discharge spectrum, i.e., the oscillations having frequencies as approximated in the remarks to Figs. 10.13, 10.14 and 10.15. Hint: In R vector `sk` as calculated above in Problem 10.8, the frequencies of the annual, the weekly and the diurnal cycles have indices $17 + 1$, $887 + 1$ and $6209 + 1$. The following R expressions therefore,

```
skky <- c(1*17+1,2*17+1,3*17+1)
```

```
skkw1 <- c(1*887+1,2*887+1,...,11*887+1,12*887+1)
skkw2 <- c(1*887+1-17,2*887+1-17,...,11*887+1-17,12*887+1-17)
skkw3 <- c(1*887+1+17,2*887+1+17,...,11*887+1+17,12*887+1+17)
skk <- concat(skky,skkw1)
skk <- concat(skk,skkw2)
skk <- concat(skk,skkw3)
```

make these indices available in R vector `skk`. By using `skk` in the following loop,

```
i <- 1
for(i in 1:length(skk)) {
lines(c(sk[skk[i]],sk[skk[i]]),c(-50,10*log10(
  (N/4)*(Re(xtsc[skk[i]])**2) +
  (N/4)*(Im(xtsc[skk[i]])**2))),lty=2)          }
```

the plots in Fig. 10.16 are readily obtained.

10.10. Compute all oscillations pertaining to the discrete part of the Inn discharge spectrum for $N = 149016$ time points in the observational period. Add these oscillations as demonstrated in Sect. 6.1.1. Thereafter compute the difference of the sum thus obtained and the observations.

10.11. Estimate the spectrum of the differences obtained in Problem 10.10.

10.12. Changes from normal to daylight saving time in spring and back to normal time in autumn displace the time of the daily and weekly cycles in electricity consumption against the time of the discharge measurements (always in normal time). These displacements are supposed, in the remarks to Fig. 10.17 and 10.18, to generate the triads of frequencies in the discrete part of the estimator for the discharge spectrum.

Construct two time series from the hourly means of the Inn discharge as plotted in Fig. 10.10: the first one by appending time slices for the winter months (November 1 through to March 20, in the following year) and the summer months (April 1 through to September 20). Are the time series thus obtained stationary? Calculate estimators for the discrete and also the continuous part of the discharge spectra in summer and winter. Compare these estimators with those in Figs. 10.17, 10.18 and in Fig. 10.19. Which of the estimators for the continuous spectrum of the discharge do you prefer: the one in Fig. 10.17 (b) or the one Fig. 10.19 (b)?

10.13. Calculate first order differences as defined in (2.39,2) (i.e., in lag 1) of the monthly values in the Basel temperature series, which are made available in the remarks to (3.19) in R vector `temp`. Thereafter, calculate the periodogram and compare your result with the one in Fig. 10.20 (b).

A Answers to Problems

1.2 Constant a is taken as random variable X with $X = a$ such that $\Pr(X = a) = 1$. Then

$$\mathrm{E}a = \mathrm{E}X = \int_{x=-\infty}^{\infty} x\mathrm{d}F(x) = a \times 1 = a$$

with $\int_{x=-\infty}^{\infty} x\mathrm{d}F(x)$ being the Riemann-Stieltjes integral of x with respect to $F(x)$. The Riemann-Stieltjes integral (7.19) is often used in an introduction to Statistics to define the expectation $\mathrm{E}X$ of a random variable X having probability distribution $F(x)$.

If Y is a random variable having $\mathrm{E}Y$, and both a as well as b are arbitrary constants, then

$$\begin{aligned}\mathrm{E}(aY + b) &= \int_{y=-\infty}^{\infty} (ay + b)\mathrm{d}F(y) = \int_{y=-\infty}^{\infty} ay\mathrm{d}F(y) + \int_{x=-\infty}^{\infty} x\mathrm{d}F(x) \\ &= a\int_{y=-\infty}^{\infty} y\mathrm{d}F(y) + \mathrm{E}b = a\mathrm{E}Y + b\end{aligned}$$

assuming that $X = b$ such that $\Pr(X = b) = 1$, as above.

1.3 (1.15,4) is obtained from (1.15,1,2,3) by way of induction, and (1.15,7) follows from definition (1.11).

1.7 Let random variables X_1' and X_2', both with expectation 0 and variance 1, be independent. Then construct $X_1 = \mu_1 + \sigma_1 X_1'$ and $X_2 = \mu_2 + \rho\sigma_2 X_1' + \sqrt{1-\rho^2}\sigma_2 X_2'$ using (1.13), and calculate variance and covariance of X_1 and X_2 as follows:

$$\begin{aligned}\mathrm{Var}X_1 &= \sigma_1^2 \mathrm{Var}X_1' = \sigma_1^2 \\ \mathrm{Var}X_2 &= \rho^2\sigma_2^2 \mathrm{Var}X_1' + (1-\rho^2)\sigma_2^2 \mathrm{Var}X_2' \\ &= \rho^2\sigma_2^2 + (1-\rho^2)\sigma_2^2 = \sigma_2^2\end{aligned}$$

$$\begin{aligned}\mathrm{Cov}(X_1, X_2) &= \mathrm{E}\big((\mu_1 + \sigma_1 X_1' - \mu_1)(\mu_2 + \rho\sigma_2 X_1' + \sqrt{1-\rho^2}\sigma_2 X_2' - \mu_2)\big) \\ &= \mathrm{E}\big(\sigma_1 X_1'(\rho\sigma_2 X_1' + \sqrt{1-\rho^2}\sigma_2 X_2')\big) \\ &= \sigma_1\rho\sigma_2 \mathrm{E}(X_1' X_1') + \sigma_1\sigma_2\sqrt{1-\rho^2}\mathrm{E}(X_1' X_2')\end{aligned}$$

$$
\begin{aligned}
&= \sigma_1\rho\sigma_2 \mathrm{Var}X_1' + \sigma_1\sigma_2\sqrt{1-\rho^2}(\mathrm{E}X_1')(\mathrm{E}X_2') \quad \text{using (1.16,1)} \\
&= \sigma_1\rho\sigma_2\big(\mathrm{E}X_1'^2 - (\mathrm{E}X_1')^2\big) + (\sigma_1\sigma_2\sqrt{1-\rho^2}) \times 0 \times 0 \qquad \text{(A.1)} \\
&= \sigma_1\sigma_2\rho \times (1-0) \quad \text{using the assumptions} \\
\rho_{X_1,X_2} &= (\sigma_1\sigma_2\rho)/(\sigma_1\sigma_2) = \rho
\end{aligned}
$$

(A.1) follows using the assumptions since $\mathrm{Var}Y = \mathrm{E}Y^2 - \mathrm{E}^2Y$.

1.12 The error integral is obtained as follows:

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-b(x^2+y^2)}\mathrm{d}y\mathrm{d}x \\
&= \int_0^{R\to\infty}\int_0^{2\pi} \mathrm{e}^{-b\rho^2}\rho\mathrm{d}\rho\mathrm{d}\phi = \int_0^{R\to\infty} \mathrm{e}^{-b\rho^2}\rho\mathrm{d}\rho \int_0^{2\pi}\mathrm{d}\phi \\
&= 2\pi\int_0^{R\to\infty} \mathrm{e}^{-b\rho^2}\rho\mathrm{d}\rho = 2\pi\left[-\frac{1}{2b}\mathrm{e}^{-b\rho^2}\right]_0^{R\to\infty} \\
&= 2\pi\lim_{R\to\infty}\left[\frac{1}{2b} - \frac{1}{2b}\mathrm{e}^{-bR^2}\right] = \frac{\pi}{b} \\
&= \int_{-\infty}^{\infty} \mathrm{e}^{-bx^2}\mathrm{d}x \times \int_{-\infty}^{\infty} \mathrm{e}^{-by^2}\mathrm{d}y = \frac{\pi}{b} \qquad \begin{matrix}\text{since } \mathrm{e}^{-b(x^2+y^2)} \\ = \mathrm{e}^{-bx^2} \times \mathrm{e}^{-by^2}\end{matrix} \\
\int_{-\infty}^{\infty} \mathrm{e}^{-bx^2}\mathrm{d}x &= \frac{\sqrt{\pi}}{\sqrt{b}}
\end{aligned}
$$

2.4 The sums in (2.18) are written for finite $t = 1, 2, 3, \ldots, M$ and $N \le t$, resulting in (i) $X_t = \sum_{u=-N}^{t} a^{t-u}W_u$ and (ii) $X_t = \sum_{v=0}^{t+N} a^v W_{t-v}$. The terms in both sums are identical as can be shown by substituting arbitrary t and $N \le t$. For example, substituting $t = 5$ and $N = 2$, sum (i) becomes $a^7W_{-2} + \ldots + a^0W_5$ and sum (ii) $a^0W_5 + \ldots + a^7W_{-2}$. When M increases, and also in the limit for $M \to \infty$, both sums still contain identical terms.

2.10 The moment functions of the MA[1] model $Y_t = W_t - bW_{t-1} = b_0W_t - b_1W_{t-1}$ are readily calculated since the model contains only two terms:

$$
\begin{aligned}
\mathrm{E}(Y_t) &= \mu_W(1-b) = 0 \qquad \text{for } \mu_W = 0 \\
c_Y(0) &= \sigma_W^2(b_0b_0 + b_1b_1) = \sigma_W^2(1+b^2) \\
c_Y(1) &= \sigma_W^2(-b_1b_0 + 0b_1) = -\sigma_W^2 b \\
c_Y(-1) &= -\sigma_W^2 b \\
c_Y(\tau) &= 0 \qquad \text{for } |\tau| > 1
\end{aligned}
$$

(2.18) and (2.19) imply that an AR[1] model is stationary on condition that (i) $|a| < 1$ and (ii) an infinitely long time has elapsed since

its initialisation. Therefore, its moment functions can be calculated using $X_t = \sum_{u=0}^{\infty} a^u W_{t-u}$:

$$\begin{aligned}
\mathrm{E}(X_t) &= \mu_W \sum_{u=0}^{\infty} a^u = \frac{1}{1-a}\mu_W \\
c_Y(\tau) &= \sigma_W^2 \sum_{u=\tau}^{\infty} a^u a^{u-\tau} = \sigma_W^2 \big(a^\tau a^0 + a^{\tau+1}a^1 + a^{\tau+2}a^2 + \ldots\big) \\
&= \sigma_W^2 a^\tau \big(1 + a^2 + a^4 + \ldots\big) = \sigma_W^2 a^\tau \frac{1}{1-a^2}
\end{aligned}$$

2.29 The moment functions of (Y_t) are easily derived: $\mathrm{E}Y_t = \mathrm{E}X_t + \mathrm{E}V = \mu_X + \mu_V = 0$ as well as $c_Y(\tau) = c_X(\tau) + \sigma_V^2$ are arrived at. Hence, (Y_t) is stationary.

Let $(x_1^{(1)} + v^{(1)}, x_2^{(1)} + v^{(1)}, \ldots) = (x_t^{(1)}) + v^{(1)} = (y_t^{(1)})$ with $v^{(1)} = 1.25$ be a first realisation of (Y_t) and $(x_t^{(2)}) + v^{(2)} = (y_t^{(2)})$ with $v^{(2)} = 0.1$ be a second one, and so on. Then, applying (2.1,1) and (2.56), $\mathrm{E}Y_t$ is estimated from the first realisation and $(1/N)\sum_{t=1}^{N} y_t^{(1)} = \hat{\mu}_X + v^{(1)}$, with $\hat{\mu}_X = (1/N)\sum_{t=1}^{N} x_t^{(1)} \approx 0$, and thus $(1/N)\sum_{t=1}^{N} y_t^{(1)} \approx 0 + v^{(1)} \approx 1.25$ is arrived at. From the second realisation, $(1/N)\sum_{t=1}^{N} y_t^{(2)} \approx 0.1$ is obtained.

The variance of this estimator is obtained using (1.16,3) and the variance of the empirical mean of a time series as proposed in (2.56): $\mathrm{Var}\hat{\mu}_X + \mathrm{Var}V \approx (1/N)\sum_{\tau=-(N-1)}^{N-1} \big((N-\tau)/N\big)|c_X(\tau)| + \sigma_V^2$. This sum comes closer to σ_V^2 with increasing N and thus this estimator is not consistent as is required in (1.4) and therefore $Y(t)$ is not ergodic, as is concluded from Sect. 2.6.2.

3.1 The sum of the squared residuals (3.1,2) is differentiated with respect to coefficients b_0 and b_1. Substituting b_0 and b_1 with $\hat{b}_0$ and $\hat{b}_1$, and equating the partial derivatives thus obtained with zero, the normal equations

$$\sum_{i=1}^{N} \big(Y_i - \hat{b}_0 - \hat{b}_1 x_i\big) = 0 \quad \text{and} \quad \sum_{i=1}^{N} x_i \big(Y_i - \hat{b}_0 - \hat{b}_1 x_i\big) = 0$$

are arrived at, which thereafter are solved in order to obtain the estimators in (3.2).

3.2 (3.10) also applies in case of simple linear regression. An alternative is to use (1.15), subsequent to writing $\hat{b}_1$ as a linear combination of Y_i as demonstrated in Problem 3.3, since $\mathrm{E}Y_i = \mu_Y(x_i) = b_0 + b_1 x_i$.

3.3 Write $\hat{b}_1$ in (3.2) as a linear combination of Y_i

$$\hat{b}_1 = \frac{\sum_{i=1}^N (Y_i - \overline{Y})(x_i - \overline{x})}{\sum_{i=1}^N (x_i - \overline{x})^2} = \frac{\sum_{i=1}^N Y_i(x_i - \overline{x}) - \sum_{i=1}^N \overline{Y}(x_i - \overline{x})}{\sum_{i=1}^N (x_i - \overline{x})^2}$$
$$= \frac{\sum_{i=1}^N Y_i(x_i - \overline{x})}{\sum_{i=1}^N (x_i - \overline{x})^2} = \frac{1}{\sum_{i=1}^N (x_i - \overline{x})^2} \sum_{i=1}^N (x_i - \overline{x}) Y_i$$

and thereafter calculate the variance of $\hat{b}_1$ as the variance of the linear combination of random variables obtained above using (1.16,4), since (3.1,3) requires that the observations be independent and have constant variance $\mathrm{Var}Y_i = \sigma_e^2$.

$$\mathrm{Var}\hat{b}_1 = \left(\frac{1}{\sum_{i=1}^N (x_i - \overline{x})^2} \sum_{i=1}^N (x_i - \overline{x}) \right)^2 \mathrm{Var}Y_i = \ldots = \frac{\sigma_e^2}{\sum_{i=1}^N (x_i - \overline{x})^2}$$

3.4 A linear combination of normally distributed random variables is always normally distributed as is concluded from the remarks to (1.34) and has moments that can be calculated using (1.16) under the additional assumption that the variables are independent. Y_i is required to be independent and normally distributed in (3.1,3,4) with $\mathrm{E}Y_i = b_0 + b_1 x_i$ and $\mathrm{Var}Y_i = \sigma_e^2$. Hence, $V_i = a_i Y_i$ with $a_i = (x_i - \overline{x}) / \sum_{i=1}^N (x_i - \overline{x})^2$ as obtained above in the solution to Problem 3.3, is normally distributed with expectation $a_i(b_0 + b_1 x_i)$ and variance $a_i^2 \sigma_e^2$. Consequently,

$$\begin{aligned} \hat{b}_1 \quad & \text{is normally distributed} \\ \mathrm{E}\hat{b}_1 &= \sum_{i=1}^N a_i(b_0 + b_1 x_i) = b_1 \sum_{i=1}^N a_i x_i + b_0 \sum_{i=1}^N a_i \\ \mathrm{Var}\hat{b}_1 &= \sum_{i=1}^N a_i^2 \sigma_e^2 \end{aligned}$$

are arrived at.

3.7 From (3.4) and (3.2) result:

$$\begin{aligned} SS_Y &= SS_M + SS_E \\ \sum_{i=1}^N e_i^2 &= \sum_{i=1}^N \big(Y_i - \hat{b}_0 - \hat{b}_1 x_i\big)^2 = \sum_{i=1}^N \big(Y_i - \overline{Y} + \hat{b}_1 \overline{x} - \hat{b}_1 x_i\big)^2 \\ &= \sum_{i=1}^N \big(Y_i - \overline{Y} - \hat{b}_1 (x_i - \overline{x})\big)^2 \\ &= \sum_{i=1}^N \big(Y_i - \overline{Y}\big)^2 - 2\hat{b}_1 \sum_{i=1}^N \big(Y_i - \overline{Y}\big)\big(x_i - \overline{x}\big) + b_1^2 \sum_{i=1}^N \big(x_i - \overline{x}\big)^2 \end{aligned}$$

$$= \sum_{i=1}^{N} \big(Y_i - \overline{Y}\big)^2 - \hat{b}_1^2 \sum_{i=1}^{N} \big(x_i - \overline{x}\big)^2$$

$$SS_E = \quad SS_Y \quad - \quad SS_M$$

6.7 Applying (6.28,2), the Fourier coefficients pertaining to $f(x) = |x|$, $-2 \le x \le 2$, are computed as follows:

$$\begin{aligned}
A_k &= \frac{2}{T} \int_{-T/2}^{T/2} f(x) \cos(2\pi s_k x) \mathrm{d}x = 2\frac{2}{4} \int_0^2 x \cos(2\pi \frac{k}{4} x) \mathrm{d}x \\
&= \left[\frac{2}{\pi k} x \sin\left(\frac{\pi k}{2} x\right) - \frac{-4}{\pi^2 k^2} x \cos\left(\frac{\pi k}{2} x\right) \right]_0^2 \\
&= \frac{4}{\pi k} \sin(\pi k) + \frac{4}{\pi^2 k^2} \cos(\pi k) + 0 - \frac{4}{\pi^2 k^2} \\
&= \frac{4}{\pi^2 k^2} \left(\cos(\pi k) - 1\right), \ k = 1, 2, \ldots \\
A_0 &= 2\frac{2}{4} \int_0^2 x \mathrm{d}x = 2, \quad \text{since } \cos(2\pi \frac{0}{4} x) = 1 \\
B_k &= 0, \quad \text{since } f(x) \sin(2\pi \frac{0}{4} x) = 1
\end{aligned}$$

and, using these coefficients, the Fourier series

$$\begin{aligned}
f_M(x) &= A_0/2 + \sum_{k=1}^{M} \left(A_k \cos(2\pi s_k x) + B_k \sin(2\pi s_k x)\right) \\
&= 1 + \sum_{k=1}^{M} \frac{4}{\pi^2 k^2} \left(\cos(\pi k) - 1\right) \cos\left(\frac{\pi k}{2} x\right) \\
&= 1 + \frac{4}{\pi^2} \left(\frac{-1-1}{1^2} \cos(\frac{\pi}{2} x) + \frac{1-1}{2^2} \cos(\frac{2\pi}{2} x) + \frac{-1-1}{3^2} \cos(\frac{3\pi}{2} x) + \ldots \right) \\
&= 1 + \frac{4}{\pi^2} \left(\frac{-2}{1^2} \cos(\frac{\pi}{2} x) + 0 + \frac{-2}{3^2} \cos(\frac{3\pi}{2} x) + \ldots \right) \\
&= 1 - \frac{8}{\pi^2} \left(\cos(\frac{\pi}{2} x) + \frac{1}{3^2} \cos(\frac{3\pi}{2} x) + \frac{1}{5^2} \cos(\frac{5\pi}{2} x) + \ldots \right)
\end{aligned}$$

is arrived at.

6.10 Differentiating,

$$\begin{aligned}
g(u) &= \frac{1}{p^2 + q^2} \mathrm{e}^{pu} \big(p \sin(qu) - q \cos(qu)\big) \\
g'(u) &= \frac{1}{p^2 + q^2} \mathrm{e}^{pu} \big(pq \cos(qu) + q^2 \sin(qu)\big) + p\mathrm{e}^{pu} \big(p \sin(qu) + q \cos(qu)\big)
\end{aligned}$$

$$= \frac{1}{p^2+q^2} \mathrm{e}^{pu} \sin(qu)\big(p^2+q^2\big) = \mathrm{e}^{pu}\sin(qu)$$

is obtained. Thereafter, the following two-fold integral

$$\begin{aligned}
I_1 &= \int_{y=0}^{\infty} \left(\int_{x=0}^{\infty} \mathrm{e}^{-xby} \sin(by) \mathrm{d}x \right) \mathrm{d}y \\
&= \int_{x=0}^{\infty} \left(\int_{y=0}^{\infty} \mathrm{e}^{-xby} \sin(by) \mathrm{d}y \right) \mathrm{d}x = I_2 \\
I_1 &= \int_{y=0}^{\infty} \left[-\frac{1}{by} \mathrm{e}^{-xby} \sin(by) \right]_{x=0}^{x\to\infty} \mathrm{d}y = \int_{y=0}^{\infty} \frac{1}{by} \sin(by) \mathrm{d}y \\
I_2 &= \int_{x=0}^{\infty} \frac{1}{|b|} \left(\int_{u=0}^{\infty} \mathrm{e}^{-xu} \sin(u) \mathrm{d}u \right) \mathrm{d}x \qquad \text{with } u = by, \frac{\mathrm{d}u}{\mathrm{d}y} = b, \mathrm{d}y = \frac{1}{b}\mathrm{d}u \\
&= \int_{x=0}^{\infty} \frac{1}{|b|} \left[\frac{1}{x^2+1} \mathrm{e}^{-xu} \big(-x\sin(u) - \cos(u) \big) \right]_{u=0}^{\infty} \mathrm{d}x \qquad \begin{array}{l}\text{with } p = -x \\ \text{and } q = 1\end{array} \\
&= \frac{1}{|b|} \int_{x=0}^{\infty} \left[0 - \frac{1}{x^2+1} \times (1) \cdot (-x \times 0 - 1) \right] \mathrm{d}x \\
&= \frac{1}{|b|} \int_{x=0}^{\infty} \frac{1}{x^2+1} \mathrm{d}x = \frac{1}{|b|} \lim_{r\to\infty} \big[\arctan(x) \big]_0^r = \frac{\pi}{|2b|}
\end{aligned}$$

is evaluated in two ways.

6.12 The Fourier transform $F(s) = \mathcal{F}_{-\mathrm{i}}\big(f(t)\big)$ of the triangle function $f(t)$ is obtained as $F(s) = G(s) - H(s)$. $G(s)$ is the Fourier transform of a rectangle function $g(t)$ having height a and width $2b$

$$g(t) = \begin{cases} a & \text{for } -b \le t \le b \\ 0 & \text{for } |t| > b \end{cases}$$

$$\mathcal{F}_{+i}\big(g(t)\big) = G(s) = 2a \frac{1}{2\pi s} \sin(2\pi s t) \Big|_{t=0}^{t=b} = 2ab \frac{\sin(2\pi b s)}{2\pi b s}$$

and $H(s)$ is obtained using integration by parts

$$\begin{aligned}
H(s) &= \frac{2}{b} \frac{t}{2\pi s} \sin(2\pi s t) \Big|_{t=0}^{t=b} - \frac{2}{b} \int_0^b \frac{1}{2\pi s} \sin(2\pi s t) \mathrm{d}t \\
&= \frac{2}{b} \frac{t}{2\pi s} \sin(2\pi s t) \Big|_{t=0}^{t=b} + \frac{2}{b} \frac{1}{(2\pi s)^2} \cos(2\pi s t) \Big|_{t=0}^{t=b} \\
&= \frac{2}{b} \frac{b}{2\pi s} \sin(2\pi b s) + \frac{2}{b} \frac{1}{(2\pi s)^2} \big(\cos(2\pi b s) - 1 \big) \\
&= \frac{2\sin(2\pi b s)}{2\pi s} - \frac{2}{b} \frac{1}{(2\pi s)^2} \big(1 - \cos(2\pi b s) \big)
\end{aligned}$$

Substituting,

$$
\begin{aligned}
F(s) &= 2ab\frac{\sin(2\pi bs)}{2\pi bs} - 2\frac{\sin(2\pi bs)}{2\pi s} + 2\frac{1}{b}\frac{1}{(2\pi s)^2}\big(1-\cos(2\pi sb)\big)\\
&= 2\frac{\sin(2\pi bs)}{2\pi s}(a-1) + \frac{2}{b}\frac{1}{(2\pi s)^2}2\big(\sin(\pi bs)\big)^2\\
&= 2\frac{\sin(2\pi bs)}{2\pi s}(a-1) + \frac{1}{b}\left(\frac{\sin(\pi bs)}{\pi s}\right)^2
\end{aligned}
$$

is arrived at.

6.17 For $s = \ldots, -1, 0, 1, \ldots$

$$
\begin{aligned}
F_N(s) &= \frac{1}{m}\sum_{t=-(m-1)}^{m-1}(m-|t|)\big(\cos(2\pi st) + \mathrm{i}\times\sin(2\pi st)\big)\\
&= \sum_{t=-(m-1)}^{m-1}\big(1-\frac{|t|}{m}\big) = \sum_{t=-(m-1)}^{m-1} 1 \;-\; \frac{1}{m}\sum_{t=-(m-1)}^{m-1}|t|\\
&= 2(m-1)+1 \;-\; \frac{1}{m}2\big(\frac{1}{2}(m-1)m\big)\\
&= 2m-1-(m-1) = m
\end{aligned}
$$

6.19 The Fourier transform of $f(t) = a\cos(2\pi ut)$

$$
\begin{aligned}
F(s) &= \lim_{r\to\infty}\int_{-r}^{r} a\cos(2\pi ut)\mathrm{e}^{\mathrm{i}2\pi st}\mathrm{d}t\\
&= \lim_{r\to\infty}\int_{0}^{r} 2a\cos(2\pi st)\cos(2\pi ut)\mathrm{d}t\\
&= \lim_{r\to\infty}\int_{0}^{r} 2\frac{a}{2}\Big(\cos\big(2\pi(s-u)t\big) + \cos\big(2\pi(s+u)t\big)\Big)\mathrm{d}t\\
&= \lim_{r\to\infty} 2\frac{a}{2}\left(\frac{\sin\big(2\pi(s-u)t\big)}{2\pi(s-u)}\bigg|_{t=0}^{r} + \frac{\sin\big(2\pi(s+u)t\big)}{2\pi(s+u)}\bigg|_{t=0}^{r}\right)\\
&= \lim_{r\to\infty}\frac{a}{2}\left(2\frac{\sin\big(2\pi(s-u)r\big)}{2\pi(s-u)} + 2\frac{\sin\big(2\pi(s+u)r\big)}{2\pi(s+u)}\right)\\
&= \lim_{r\to\infty}\frac{a}{2}\left(2r\frac{\sin\big(2\pi(s-u)r\big)}{2\pi(s-u)r} + 2r\frac{\sin\big(2\pi(s+u)r\big)}{2\pi(s+u)r}\right)
\end{aligned}
$$

can be obtained as the limiting case of two $(\sin x)/x$ functions as defined in (6.39).

6.21 The first order linear differential equation $x(t) + ax'(t) = w(t)$ with $\delta(t)$ substituted for $w(t)$:

$$
\begin{aligned}
y(t) &= \int_0^t x(u)\frac{1}{\tau}\mathrm{e}^{-(t-u)/\tau}\mathrm{d}u = \int_0^t \delta(u)\frac{1}{\tau}\mathrm{e}^{-(t-u)/\tau}\mathrm{d}u \\
&= \int_0^t \frac{1}{\tau}\mathrm{e}^{-t/\tau}\delta(u)\mathrm{e}^{u/\tau}\mathrm{d}u = \frac{1}{\tau}\mathrm{e}^{-t/\tau}\int_0^t \delta(u)\mathrm{e}^{u/\tau}\mathrm{d}u \\
&= \frac{1}{\tau}\mathrm{e}^{-t/\tau}1 = \frac{1}{\tau}\mathrm{e}^{-t/\tau}
\end{aligned}
$$

6.23 Fourier transform of damped oscillation $f(t) = 2\mathrm{e}^{-|at|}\cos(2\pi rt)$:

$$
\begin{aligned}
F(s) &= \mathcal{F}_{-i}\left(2\mathrm{e}^{-|at|}\cos(2\pi rt)\right) \\
&= \int_{-\infty}^{\infty} \mathrm{e}^{-|at|}2\cos(2\pi rt)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \mathrm{e}^{-|at|}\left(\mathrm{e}^{\mathrm{i}2\pi rt} + \mathrm{e}^{-\mathrm{i}2\pi rt}\right)\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\
&= \int_{-\infty}^{\infty} \mathrm{e}^{-|at|}\mathrm{e}^{\mathrm{i}2\pi rt}\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t + \int_{-\infty}^{\infty} \mathrm{e}^{-|at|}\mathrm{e}^{-\mathrm{i}2\pi rt}\mathrm{e}^{-\mathrm{i}2\pi st}\mathrm{d}t \\
&= 2\mathrm{re}\int_0^{\infty} \mathrm{e}^{-t(a-\mathrm{i}2\pi r+\mathrm{i}2\pi s)}\mathrm{d}t + 2\mathrm{re}\int_0^{\infty} \mathrm{e}^{-t(a+\mathrm{i}2\pi r+\mathrm{i}2\pi s)}\mathrm{d}t \\
&= 2\mathrm{re}\int_0^{\infty} \mathrm{e}^{-t(a+\mathrm{i}2\pi(s-r))}\mathrm{d}t + 2\mathrm{re}\int_0^{\infty} \mathrm{e}^{-t(a+\mathrm{i}2\pi(s+r))}\mathrm{d}t \\
&\cdots \\
&= \frac{2a}{a^2 + \big(2\pi(s-r)\big)} + \frac{2a}{a^2 + \big(2\pi(s+r)\big)}
\end{aligned}
$$

7.5 $A_j = R_j\cos(P_j)$ and $B_j = R_j\sin(P_j)$, with R_j being real constants and P_j independent random variables having a uniform (or rectangular) distribution $1/(2\pi)$ in $[-\pi, \pi]$. The moments of the A_j and B_j are obtained by integration:

$$
\begin{aligned}
\mathrm{E}A_j &= R_j\mathrm{E}\big(\cos(P_j)\big) = R_j\int_{-\pi}^{\pi}\frac{1}{2\pi}\cos(P_j)\mathrm{d}P_j = \frac{R_j}{2\pi}\big[\sin(P_j)\big]_{-\pi}^{\pi} = 0 \\
\mathrm{E}B_j &= -\frac{R_j}{2\pi}\big[\cos(P_j)\big]_{-\pi}^{\pi} = -\frac{R_j}{2\pi}\big[-1-(-1)\big]_{-\pi}^{\pi} = 0 \\
\mathrm{Var}A_j &= \mathrm{Var}\big(R_j\cos(P_j)\big) = R_j^2\mathrm{E}\big(\cos(P_j)\big)^2 = R_j^2\int_{-\pi}^{\pi}\frac{1}{2\pi}\big(\cos(P_j)\big)^2\mathrm{d}P_j \\
&= \frac{R_j^2}{2\pi}\int_{-\pi}^{\pi}\frac{1}{2}\big(1+\cos(2P_j)\big)\mathrm{d}P_j \\
&= \frac{R_j^2}{2\pi}\left(\int_0^{\pi}\mathrm{d}P_j + \int_0^{\pi}\cos(2P_j)\mathrm{d}P_j\right)
\end{aligned}
$$

$$= \frac{R_j^2}{2\pi}\Big(\big[P_j\big]_0^\pi + (1/2)\big[\sin(2P_j)\big]_0^\pi \Big) = \frac{R_j^2}{2\pi}(\pi - 0) = R_j^2/2$$

$$\mathrm{Var}B_j = \mathrm{Var}\big(R_j \sin(P_j)\big) = R_j^2 \mathrm{E}\big(\sin(P_j)\big)^2 = R_j^2 \int_{-\pi}^{\pi} \frac{1}{2\pi}\big(\sin(P_j)\big)^2 \mathrm{d}P_j$$

$$= \frac{R_j^2}{2\pi}\int_{-\pi}^{\pi} \frac{1}{2}\big(1 - \cos(2P_j)\big)\mathrm{d}P_j$$

$$= \frac{R_j^2}{2\pi}\left(\int_0^\pi \mathrm{d}P_j - \int_0^\pi \cos(2P_j)\mathrm{d}P_j \right)$$

$$= \frac{R_j^2}{2\pi}\Big(\big[P_j\big]_0^\pi - (1/2)\big[\sin(2P_j)\big]_0^\pi \Big) = \frac{R_j^2}{2\pi}(\pi - 0) = R_j^2/2$$

$$\mathrm{Cov}(A_j, B_j) = \mathrm{Cov}\big(R_j \cos(P_j), R_j \sin(P_j)\big) = R_j^2 \int_{-\pi}^{\pi} \cos(P_j)\sin(P_j)\mathrm{d}P_j$$

$$= \frac{R_j^2}{4\pi}\int_{-\pi}^{\pi} \sin(2P_j)\mathrm{d}P_j = 0 \quad \text{since odd}$$

$$\mathrm{Cov}(A_j, B_k) = \mathrm{Cov}\big(R_j \cos(P_j), R_k \sin(P_k)\big) \quad j \neq k$$

$$= R_j^2 \int_{-\pi}^{\pi} \cos(P_j)\sin(P_j)\mathrm{d}P_j = 0 \qquad \begin{array}{l}\text{since } P_j \text{ and} \\ P_k \text{ independent}\end{array}$$

7.6 A random oscillation $X(t) = A\cos(2\pi st) + B\sin(2\pi st)$, A and B being random variables, has expectation function

$$\mathrm{E}X(t) = \mathrm{E}\big(A\cos(2\pi st) + B\sin(2\pi st)\big) = \cos(2\pi st)\mathrm{E}A + \sin(2\pi st)\mathrm{E}B$$

which is constant on condition that $\mathrm{E}A = 0$ and $\mathrm{E}B = 0$.

The covariance function of $X(t) = A\cos(2\pi st) + B\sin(2\pi st)$, A and B being random variables with expectations $\mathrm{E}A = \mu_A$ and $\mathrm{E}B = \mu_B$, is derived as follows:

$$c_X(t,u) = \mathrm{Cov}\big(X(t), X(u)\big) = \mathrm{E}\Big(\big(X(t) - \mu_X(t)\big)\big(X(u) - \mu_X(u)\big) \Big)$$

$$= \mathrm{E}\Big(\big(A\cos(2\pi st) + B\sin(2\pi st) - \mu_A\cos(2\pi st) - \mu_B\sin(2\pi st)\big)$$

$$\times \big(A\cos(2\pi su) + B\sin(2\pi su) - \mu_B\cos(2\pi su) - \mu_B\sin(2\pi su)\big)\Big)$$

$$= \mathrm{E}\Big(A^2\cos(2\pi st)\cos(2\pi su) + AB\cos(2\pi st)\sin(2\pi su)$$

$$+B^2\sin(2\pi st)\cos(2\pi su) + AB\sin(2\pi st)\cos(2\pi su) \qquad \text{(A.2)}$$

$$-A\cos(2\pi st)\mu_A\cos(2\pi su) - A\cos(2\pi st)\mu_B\sin(2\pi su)$$

$$-A\cos(2\pi st)\mu_A\cos(2\pi su) - A\sin(2\pi st)\mu_B\cos(2\pi su) \qquad \text{(A.3)}$$

$$-B\sin(2\pi st)\mu_A\cos(2\pi su) - B\sin(2\pi st)\mu_B\sin(2\pi su)$$

$$-B\cos(2\pi st)\mu_A\sin(2\pi su) - B\sin(2\pi st)\mu_B\sin(2\pi su) \qquad \text{(A.4)}$$

$$+(\mu_A)^2\cos(2\pi st)\cos(2\pi su) + \mu_A\mu_B\cos(2\pi st)\sin(2\pi su)$$

$$+(\mu_B)^2 \sin(2\pi st)\sin(2\pi su) + \mu_A\mu_B \sin(2\pi st)\cos(2\pi su)\Big) \tag{A.5}$$

$$= \cos(2\pi st)\cos(2\pi su)\mathrm{E}A^2 \tag{A.6}$$

$$+\big(\cos(2\pi st)\sin(2\pi su) + \sin(2\pi st)\cos(2\pi su)\big)\mathrm{E}AB \tag{A.7}$$

$$+\sin(2\pi st)\sin(2\pi su)\mathrm{E}B^2 \tag{A.8}$$

$$-2\mu_A \cos(2\pi st)\cos(2\pi su)\mathrm{E}A \tag{A.9}$$

$$-\mu_B\big(\cos(2\pi st)\sin(2\pi su) + \sin(2\pi st)\cos(2\pi su)\big)\mathrm{E}A \tag{A.10}$$

$$-\mu_A\big(\sin(2\pi st)\cos(2\pi su) + \cos(2\pi st)\sin(2\pi su)\big)\mathrm{E}B \tag{A.11}$$

$$-2\mu_B \sin(2\pi st)\sin(2\pi su)\mathrm{E}B \tag{A.12}$$

$$+(\mu_A)^2 \cos(2\pi st)\cos(2\pi su) \tag{A.13}$$

$$+\mu_A\mu_B\big(\cos(2\pi st)\sin(2\pi su) + \sin(2\pi st)\cos(2\pi su)\big) \tag{A.14}$$

$$+(\mu_B)^2 \sin(2\pi st)\sin(2\pi su) \tag{A.15}$$

Since $\mathrm{E}A = \mu_A$ and $\mathrm{E}B = \mu_B$, (i) the sum of (A.9) and (A.13) becomes $-(\mu_A)^2\cos(2\pi st)\cos(2\pi su)$ and (ii) the sum of (A.10) and (A.14) becomes identically zero. Further, the sum of (A.12) and (A.15) becomes $-(\mu_B)^2\sin(2\pi st)\sin(2\pi su)$ and (A.11) becomes $\mu_A\mu_B\big(\sin(2\pi st)\cos(2\pi su) + \cos(2\pi st)\sin(2\pi su)\big)$:

$$c_X(t,u) = \cos(2\pi st)\cos(2\pi su)\mathrm{E}A^2 \tag{A.16}$$

$$+\big(\cos(2\pi st)\sin(2\pi su) + \sin(2\pi st)\cos(2\pi su)\big)\mathrm{E}AB \tag{A.17}$$

$$+\sin(2\pi st)\sin(2\pi su)\mathrm{E}B^2 \tag{A.18}$$

$$-(\mu_A)^2\cos(2\pi st)\cos(2\pi su) \tag{A.19}$$

$$-\mu_A\mu_B\big(\sin(2\pi st)\cos(2\pi su) + \cos(2\pi st)\sin(2\pi su)\big) \tag{A.20}$$

$$-(\mu_B)^2\sin(2\pi st)\sin(2\pi su) \tag{A.21}$$

If $\mathrm{Cov}(A,B) = 0$ then $\mathrm{E}AB = \mathrm{E}A \times \mathrm{E}B = \mu_A\mu_B$ and the sum of (A.16) and (A.17) becomes identically zero, implying

$$\begin{aligned} c_X(t,u) = \mathrm{E}A^2\frac{1}{2}\Big(\cos\big(2\pi s(t-u)\big) + \cos\big(2\pi s(t+u)\big)\Big) \\ +\mathrm{E}B^2\frac{1}{2}\Big(\cos\big(2\pi s(t-u)\big) - \cos\big(2\pi s(t+u)\big)\Big) \\ -(\mu_A)^2\frac{1}{2}\Big(\cos\big(2\pi s(t-u)\big) + \cos\big(2\pi s(t+u)\big)\Big) \\ -(\mu_B)^2\frac{1}{2}\Big(\cos\big(2\pi s(t-u)\big) - \cos\big(2\pi s(t+u)\big)\Big) \end{aligned}$$

If additionally $\mathrm{E}A^2 = \mathrm{E}B^2$ and $(\mu_A)^2 = (\mu_B)^2$, then

$$c_X(t,u) = \mathrm{E}A^2\cos\big(2\pi s(t-u)\big) - (\mu_A)^2\cos\big(2\pi s(t-u)\big)$$

and the covariance of the random trigonometric oscillation with constant frequency s becomes a function of $t-u$, and not of t and/or u. Consequently,

under the above assumptions, random oscillation $X(t) = A\cos(2\pi st) + B\sin(2\pi st)$ is stationary in its second moment function.

8.2 Let X be an exponentially distributed random variable with $F(x) = 1 - \mathrm{e}^{-\lambda x}$ for $x \geq 0$ and $F(x) = 0$ for $x < 0$ as defined in (7.4), and let $Y = \log_{\mathrm{e}}(X) = \ln(X)$. Then the distribution $F(y)$ of Y is obtained by evaluating the following Stieltjes integral:

$$\begin{aligned} F(y) &= \int_{ln(x)\leq y} \mathrm{d}F(x) = \int_0^{\mathrm{e}^y} \big(\mathrm{d}(1-\mathrm{e}^{-\lambda x})/\mathrm{d}x\big)\mathrm{d}x \\ &= \Big[1-\mathrm{e}^{-\lambda x}\Big]_0^{\mathrm{e}^y} = 1-\mathrm{e}^{-\lambda \mathrm{e}^y} \end{aligned}$$

If $Y = 10\log_{10}(X)$ then

$$\begin{aligned} F(y) &= \int_{10\log_{10}(x)\leq y} \mathrm{d}F(x) = \Big[1-\mathrm{e}^{-\lambda x}\Big]_0^{10^{x/10}} = 1-\mathrm{e}^{-\lambda 10^{x/10}} \\ f(y) &= -\mathrm{e}^{-\lambda 10^{x/10}}\big(-\lambda 10^{x/10}\big)\big(\ln(10)\big)\frac{1}{10} = \frac{\lambda\ln(10)}{10}\big(10^{x/10}\big)\mathrm{e}^{-\lambda(10^{x/10})} \end{aligned}$$

and, substituting $\lambda = 1/2$, the probability density pertaining to a periodogram of a white noise process subsequent to a logarithmic transformation is arrived at.

9.1 Using the convolution theorem derived in (6.144) and (6.145),

$$\begin{aligned} \mathrm{E}\big(I_X(s)\big) &= \sum_{\tau=-(N-1)}^{N-1}\Big(1-\frac{|\tau|}{N}\Big)\left(\sigma_W^2\sum_{u=-\infty}^{\infty} b_u b_{u+\tau}\right)\cos(2\pi s\tau) \\ &= \left(\sum_{\tau=-(N-1)}^{N-1}\Big(1-\frac{|\tau|}{N}\Big)\cos(2\pi s\tau)\right) * \\ &\qquad \left(\sigma_W^2\sum_{\tau=-(N-1)}^{N-1}\left(\sum_{u=-\infty}^{\infty} b_u b_{u+\tau}\right)\cos(2\pi s\tau)\right) \\ &= F_N(s) * \left(\sigma_W^2\Big(\sum_{u=-\infty}^{\infty} b_u e^{-i2\pi su}\Big)\overline{\Big(\sum_{u=-\infty}^{\infty} b_u e^{-i2\pi su}\Big)}\right) \\ &= F_N(s) * \Big(B(s)\overline{B(s)}\sigma_W^2\Big) = F_N(s) * C_X(s) \end{aligned}$$

is arrived at.

9.4 To a cosine taper $h_N(t)$ as defined in (6.111,1) pertains its spectral window $\widehat{\mathcal{H}}_N^{(d)}(s) = \big(1/\Sigma(\widehat{h^2})\big)\widehat{H}_N(s)\overline{\widehat{H}_N(s)} = \big(1/\Sigma(\widehat{h^2})\big)\mathcal{F}_{-\mathrm{i}}\big(\widehat{h}_N \star \widehat{h}_N(t)\big)$.

$\widehat{\mathcal{H}}_N^{(d)}(s) \to \delta(s)$ for $N \to \infty$ on condition that $\widehat{\mathcal{H}}_N^{(d)}(s)$ has the properties required in (6.60). $\int_{-\infty}^{\infty} \widehat{\mathcal{H}}_N^{(d)}(s)\mathrm{d}s = 1$ for $N = 1, 2, 3, \ldots$ follows directly from definition (9.2). It remains to show that $\widehat{\mathcal{H}}_N^{(d)}(s) \to 0$, $s \neq 0$, and that $\widehat{\mathcal{H}}_N^{(d)}(s) \to \infty$, $s = 0$, for $N \to \infty$.

In a first step, the sum of the weights in a cosine taper $\widehat{h_N}(t)$, with N, p and q as defined in (6.111,1), is calculated. For this calculation, q is substituted with p_1.

$$\begin{aligned}\sum_{t=0}^{N-1} \widehat{h_N}(t) &= \sum_{t=0}^{p_1-1} \frac{1}{2}\left(1 - \cos\Big(\frac{\pi}{2p_1}(2t+1)\Big)\right) + \sum_{t=p_1}^{N-(p_1+1)} 1 \\ &\quad + \sum_{t=N-p_1}^{N-1} \frac{1}{2}\left(1 - \cos\Big(\frac{\pi}{2p_1}\big(2(N-1-t)+1\big)\Big)\right) \\ &= S_1^{(1)} + S_2^{(1)} + S_3^{(1)} \qquad \text{with } S_2^{(1)} = N - 2p_1 \text{ and } S_1^{(1)} = S_3^{(1)} \end{aligned} \tag{A.22}$$

$S_1^{(1)}$ is obtained by assuming that below, in (A.25), (i) p_2 is the largest integer number which is less than or equal to $p_1/2$ and (ii) the last term $\cos\big(\big(\pi(2p_2+1)\big)/(2p_1)\big)$ is identically zero for even p_1. (A.25) implies (A.26) seeing that $\cos(x) = -\cos(\pi - x)$.

$$S_1^{(1)} = \sum_{t=0}^{p_1-1} \widehat{h_N}(t) = \sum_{t=0}^{p_1-1} \frac{1}{2}\left(1 - \cos\Big(\frac{\pi}{2p_1}(2t+1)\Big)\right) \tag{A.23}$$

$$= \frac{1}{2}\sum_{t=0}^{p_1-1} 1 - \frac{1}{2}\sum_{t=0}^{p_1-1} \cos\Big(\frac{\pi}{2p_1}(2t+1)\Big) \tag{A.24}$$

$$= \frac{p_1}{2} - \frac{1}{2}\sum_{t=0}^{p_2-1}\left(\cos\Big(\frac{\pi(2t+1)}{2p_1}\Big) + \cos\Big(\frac{\pi(2p_1-(2t+1))}{2p_1}\Big)\right) + \cos\Big(\frac{\pi(2p_2+1)}{2p_1}\Big) \tag{A.25}$$

$$= \frac{p_1}{2} - \frac{1}{2}\sum_{t=0}^{p_2-1}\left(\cos\Big(\frac{\pi(2t+1)}{2p_1}\Big) + \cos\Big(\pi - \frac{\pi(2t+1)}{2p_1}\Big)\right) + \cos\Big(\frac{\pi p_1}{2p_1}\Big) \tag{A.26}$$

$$= \frac{p_1}{2} - \frac{1}{2}\sum_{t=0}^{p_2-1}(0) + 0 = \frac{p_1}{2} \tag{A.27}$$

This sum is substituted in (A.22) and thus

$$\sum_{t=0}^{N-1} \widehat{h_N}(t) = p_1/2 + (N - 2p_1) + p_1/2 = N - p_1 \tag{A.28}$$

is obtained. From this result and (6.48,1), it is concluded that $\widehat{H}_N(0) = \sum_{t=0}^{N-1} \widehat{h}_N(t) = N - p_1$ since $\widehat{H}_N(s)$ is the kernel pertaining to the cosine taper as defined in (6.110,2): $\widehat{H}_N(s) = \mathcal{F}_{-\mathrm{i}}\big(\widehat{h}_N(t)\big)$.

In a second step, the sum of the squared weights in a cosine taper $\widehat{h}_N(t)$, with N, p and q as defined in (6.111,1), is calculated. For this calculation, q is substituted with p_1.

$$\begin{aligned}\sum_{t=0}^{N-1}\big(\widehat{h}_N(t)\big)^2 &= \sum_{t=0}^{p_1-1}\left(\frac{1}{2}\left(1-\cos\Big(\frac{\pi}{2p_1}(2t+1)\Big)\right)\right)^2 + \sum_{t=p_1}^{N-(p_1+1)} 1^2 \\ &\quad + \sum_{t=N-p_1}^{N-1}\left(\frac{1}{2}\left(1-\cos\Big(\frac{\pi}{2p_1}\big(2(N-1-t)+1\big)\Big)\right)\right)^2 \\ &= S_1^{(2)} + S_2^{(2)} + S_3^{(2)} \qquad \text{with } S_2^{(2)} = N - 2p_1 \text{ and } S_1^{(2)} = S_3^{(2)} \end{aligned} \tag{A.29}$$

Below in (A.31), the second term becomes identically zero as is concluded from (A.24) and (A.27) and thus (A.32) is arrived at.

$$S_1^{(2)} = \frac{1}{4}\sum_{t=0}^{p_1-1}\left(1-\cos\Big(\frac{\pi}{2p_1}(2t+1)\Big)\right)^2 \tag{A.30}$$

$$= \frac{1}{4}\sum_{t=0}^{p_1-1} 1^2 - \frac{1}{2}\sum_{t=0}^{p_1-1}\cos\Big(\frac{\pi}{2p_1}(2t+1)\Big) + \frac{1}{4}\sum_{t=0}^{p_1-1}\left(\cos\Big(\frac{\pi}{2p_1}(2t+1)\Big)\right)^2 \tag{A.31}$$

$$= \frac{p_1}{4} - 0 + \frac{1}{4}S_{12}^{(2)} \tag{A.32}$$

$S_{12}^{(2)}$ is derived below: (A.34) and (A.37) follow from trigonometric identities, and (A.39) using the sums in (6.10).

$$S_{12}^{(2)} = \sum_{t=0}^{p_1-1}\left(\cos\Big(\frac{\pi}{2p_1}(2t+1)\Big)\right)^2 \tag{A.33}$$

$$= \frac{1}{2}\sum_{t=0}^{p_1-1}\left(\cos\Big(\big(\frac{\pi}{2p_1}+\frac{\pi}{2p_1}\big)(2t+1)\Big) + \cos\Big(\big(\frac{\pi}{2p_1}-\frac{\pi}{2p_1}\big)(2t+1)\Big)\right) \tag{A.34}$$

$$= \frac{1}{2}\sum_{t=0}^{p_1-1}\cos\Big(\frac{\pi}{p_1}(2t+1)\Big) + \frac{1}{2}\sum_{t=0}^{p_1-1}\cos(0) \tag{A.35}$$

$$= \frac{p_1}{2} + \frac{1}{2}\sum_{t=0}^{p_1-1}\cos\Big(2\pi\frac{1}{p_1}t + \frac{\pi}{p_1}\Big) \tag{A.36}$$

$$= \frac{p_1}{2} + \frac{1}{4}\sum_{t=0}^{p_1-1}\left(\cos\Big(2\pi\frac{1}{p_1}t\Big)\cos\Big(\frac{\pi}{p_1}\Big) + \sin\Big(2\pi\frac{1}{p_1}t\Big)\sin\Big(\frac{\pi}{p_1}\Big)\right) \tag{A.37}$$

$$= \frac{p_1}{2} + \frac{1}{4}\cos\Big(\frac{\pi}{p_1}\Big)\sum_{t=0}^{p_1-1}\cos\Big(2\pi\frac{1}{p_1}t\Big) + \frac{1}{4}\sin\Big(\frac{\pi}{p_1}\Big)\sum_{t=0}^{p_1-1}\sin\Big(2\pi\frac{1}{p_1}t\Big) \tag{A.38}$$

$$= \frac{p_1}{2} + \frac{1}{4}\cos\Big(\frac{\pi}{p_1}\Big)(0) + \frac{1}{4}\sin\Big(\frac{\pi}{p_1}\Big)(0) = \frac{p_1}{2} \tag{A.39}$$

Subsequent to substituting (A.39) in (A.32),

$$\begin{aligned}\Sigma(\widehat{h^2}) &= \sum_{t=0}^{N-1}\big(\widehat{h_N}(t)\big)^2 = S_1^{(2)} + S_2^{(2)} + S_3^{(2)} \\ &= (3/8)p_1 + (N - 2p_1) + (3/8)p_1 = N - (5/4)p_1 \end{aligned} \tag{A.40}$$

is arrived at.

Using (A.28) and (A.40), the following results for the spectral window pertaining to the cosine taper for frequency $s = 0$ are obtained, since $p_1 \leq N/2$.

$$\widehat{\mathcal{H}}_N^{(d)}(s) = \frac{1}{\Sigma(\widehat{h^2})}\widehat{H}_N(s)\overline{\widehat{H}_N(s)} = \frac{(N - p_1)(N - p_1)}{N - (5/4)p_1} \qquad s = 0$$

$$\lim_{N\to\infty}\big(\widehat{\mathcal{H}}_N^{(d)}(s)\big) = \infty \qquad s = 0$$

It remains to show that $\widehat{\mathcal{H}}_N^{(d)}(s) \to 0$, $s \neq 0$, for $N \to \infty$. This result is obtained provided that $\widehat{H}_N(s)\overline{\widehat{H}_N(s)}$, $s \neq 0$, is bounded for $N \to \infty$. Both spectral windows $F_N(s)$ and $\widehat{\mathcal{H}}_N^{(d)}(s)$ are continuous in $s = 0$ and $\widehat{\mathcal{H}}_N^{(d)}(0) \leq F_N(0)$. This implies that $\widehat{H}_N(s)\overline{\widehat{H}_N(s)} = \big(N - (5/4)p_1\big)\widehat{\mathcal{H}}_N^{(d)}(s) \leq N \times F_N(s)$ within a small interval having its centre in $s = 0$. Within such an interval, $N \times F_N(s) = \big(\sin(\pi N s)/\sin(\pi s)\big)^2 \leq 1/\big(\sin(\pi s)\big)^2$ for $s \neq 0$ and $\widehat{H}_N(s)\overline{\widehat{H}_N(s)} \leq 1/\big(\sin(\pi s)\big)^2$ is obtained. Consequently, $\widehat{H}_N(s)\overline{\widehat{H}_N(s)}$ is bounded in a small interval having its centre in $s = 0$, except for $s = 0$.

In an interval as above, $\widehat{H}_N(s)\overline{\widehat{H}_N(s)}$ is larger than for any other s. Hence, $\widehat{H}_N(s)\overline{\widehat{H}_N(s)}$ is bounded for $s \neq 0$, which in turn implies that $\lim_{N\to\infty}\big(\widehat{\mathcal{H}}_N^{(d)}(s)\big) = \lim_{N\to\infty}\big(1/(N - (5/4)p_1)\big)\widehat{H}_N(s)\overline{\widehat{H}_N(s)} = 0$, for $s \neq 0$.

References

1. H. Akaike. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60:255–265, 1973.
2. H. Akima. A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. *ACM Transactions on Mathematical Software*, 4:148–159, 1978.
3. A. Arking. Absorption of solar energy in the atmosphere: Discrepancy between model and observations. *Science*, 273:779–782, 1996.
4. S. Asmussen. *Applied Probability and Queus, 2nd Ed.* Springer, New York and Berlin, 2003.
5. H. Aulitzky and H. Turner. Bioklimatische Grundlagen einer standortsgemässen Bewirtschaftung des subalpinen Lärchen-Zirbenwaldes. *Eidg.Anst. forstl.Versuchswes.Mitt.*, 58:325–580, 1982.
6. F. Aurenhammer and R. Klein. Voronoi diagrams. In J. Sack and G. Urrutia, editors, *Handbook of Computational Geometry*, pages 201–290. Elsevier Science Publishing, 2000.
7. B.R. Barkstrom et al. Results from the earth radiation budget experiment ERBE. *Adv. Space Res.*, 9:775–782, 1989.
8. B.R. Barkstrom and G.L. Smith. The earth radiation budget experiment: science and implementation. *Rev. of Geophysics*, 24:379–390, 1986.
9. M.S. Bartlett. Periodogram analysis and continuous spectra. *Biometrika*, 37:1–16, 1950.
10. N. Beamish and M.B. Priestley. Autoregressive spectral estimation. *Applied Statistics*, 30:13–25, 1981.
11. R.A. Becker et al. *The New S Language. A Programming Environment for Data Analysis and Graphics.* Wadsworth and Brooks, Pacific Grove, CA., 1988.
12. Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approch to multiple testing. *J. of the Royal Stat. Soc. B*, 57:289–300, 1995.
13. M. Bider et al. Die Reduktion der 200jährigen Basler Temperaturreihe. *Archiv Met. Geophys. Biokl. B*, 9:360–405, 1959.
14. P. Bloomfield. *Fourier Analysis of Time Series, 2nd ed.* Wiley, New York, 2000.
15. S. Bose. *An Introduction to Queuing Systems.* Kluwer Academic/Plenum Publishers (now merged with Springer), 2001. The author's web site for this book is *http://home.iitk.ac.in/~skb/ee679/ee679.html*.
16. G.E.P. Box and G.M. Jenkins. *Time Series Analysis: Forecasting and Control, rev. Ed.* Holden-Day, San Francisco, 1976.

17. G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis: Forecasting and Control, 3rd Ed.* Prentice-Hall, Englewood Cliffs, NJ., 1994.
18. R.N. Bracewell. *The Fourier Transform and its Applications, 2nd Ed.* McGraw-Hill, New York, 1978.
19. D.R. Brillinger. The key role of tapering in spectrum estimation. *IEEE transactions on Acoustics, Speech, and Signal Processing*, 29:1075–1076, 1981.
20. D.R. Brillinger. *Time Series Analysis: Data Analysis and Theory.* Holt, Rinehart and Winston, New York, 1981.
21. P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods, 2nd Ed.* Springer, New York and Berlin, 1991.
22. P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting, 2nd Ed.* Springer, New York and Berlin, 2002.
23. P.D. Bromirski. Vibrations from the "Perfect Storm". *Geochem.Geophys. Geosyst.*, 2:2000GC000119, 2001. *(http://www-ccs.ucsd.edu/~peter/Bromirski_G3_2001.pdf).*
24. S. Brönnimann et al. Extreme climate of the global troposphere and stratosphere in 1940–42 related to El Niño. *Nature*, 431:971–974, 2004.
25. N. Bukowiecki et al. Fine and ultrafine particles in the Zurich (Switzerland) area measured with a mobile laboratory: an assessment of the seasonal and regional variation throughout a year. *Atmos. Chem. Phys.*, 3:1477–1494, 2003.
26. J.P. Burg. A new analysis technique for time series data. In D.G. Childers, editor, *Modern Spectrum Analyses.* IEEE Press, New York, 1968.
27. P. Calanca et al. Gridded temperature and accumulation distributions for Greenland for use in cryospheric models. *Annals of Glaciology*, 31:118–120, 2000.
28. M.J. Campbell and A.M. Walker. A survey of statistical work on the Mackenzie River series of annual Canadian lynx trappings for the years 1821 - 1934 and a new analysis. *J. of the Royal Statistical Society A*, 140:411–431, 1977.
29. G.C. Carter. Coherence and time delay estimation. *Proc. IEEE*, 75:235–255, 1987.
30. J.M. Chambers et al. *Graphical Methods for Data Analysis.* Wadsworth International Group, Belmont, CA., 1983.
31. J.M. Chambers and T.J. Hastie, editors. *Statistical Models in S.* Wadsworth and Brooks, Pacific Grove, CA., 1992.
32. D.C. Champenay. *Fourier Transforms and their Physical Applications.* Academic Press, London, 1973.
33. C. Chatfield. *Introduction to Time-series Analysis, 5th Ed.* Chapman and Hall, London, 1996.
34. A.D. Chave et al. On the robust estimation of power spectra, coherences, and transfer functions. *Journal of Geophysical Reasearch*, 92:633–648, 1987.
35. J. Chilès and P. Delfiner. *Geostatistics: Modeling Spatial Uncertainty.* Wiley, New York, 1999.
36. B.S. Choi. *ARMA model identification.* Springer, New York and Berlin, 1992.
37. A. Christen et al. Ultra sonic anemometer instrumentation at steep slopes: wind tunnel study – field intercomparison – measurements. MAP newsletter no. 15, pp. 164-167, Mesoscale Alpine Program (MAP), Programme Office, c/o MeteoSwiss, Krähbühlstrasse 58, CH-8044 Zürich, Switzerland, 2001.
38. R.D. Cook and D.V. Hinkley. *Residuals and Influence in Regression.* Chapman and Hall, New York, 1982.

39. J.W. Cooley and J.W. Tuckey. An algorithm for the machine calculation of complex fourier series. *Math. Comp.*, 19:297–301, 1965.
40. J. G. Cramer. The transactional interpretation of Quantum Mechanics. *Reviews of Modern Physics*, 58:647–687, 1986.
41. N. Cressie. *Statistics for Spatial Data, rev. Ed.* Wiley, New York, 1998.
42. P.J. Daniell. Discussion of the papers by Bartlett, Foster, Cunningham and Hind. *Suppl. to the J. of the Royal Stat. Soc*, 8:88–90, 1946.
43. H.F. Diaz and V. Markgraf, editors. *El Niño and the Southern Oscillation: Multiscale Variability and Global and Regional Impacts*. Cambridge University Press, Cambridge, 2000.
44. N. Draper and H. Smith. *Applied Regression Analysis, 3rd Ed.* Wiley, New York, 1998.
45. J. Durbin and S.J. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2001.
46. D.R. Easterling et al. On the development and use of homogenized climate data sets. *J. of Climate*, 9:1429–1434, 1996.
47. H.H. Essen et al. Does microseisms in Hamburg (Germany) reflect the wave climate in the north atlantic? *German Journal of Hydrography*, 51:33–45, 1999.
48. H.H. Essen et al. On the generation of secondary microseisms observed in northern and central Europe. *J. of Geophysical Research*, 108B10:ESE15–1, doi:10.1029/2002JB002338, 2003.
49. Federation of Digital Broad-Band Seismograph Networks (FDSN). Station book. FDSN data are stored in the IRIS data center, which is run by Incorporated Research Institutions for Seismology, 1200 New York Ave NW, Washington DC 20005. *(http://www.fdsn.org/station_book/II/NIL/nil.html).*
50. C. Frei and C. Schär. A precipitation climatology of the alps from high-resolution rain-gauge observations. *Int. J. of Climatology*, 18:873–900, 1998.
51. U. Frisch. *Turbulence: The Legacy of A.N. Kolmogorov*. Cambridge University Press, Cambridge, 1995.
52. C. Fröhlich. History of solar radiometry and the World Radiation Reference. *Metrologia*, 28:111–115, 1991.
53. C. Fröhlich and J. Lean. Total solar irradiance variations: The construction of a composite and its comparison with models. In F.L. Deubner, editor, *IAU Symposium 185: New Eyes to See Inside the Sun and Stars*, pages 89–102. Kluwer Academic Publ., Dordrecht, 1998.
54. L.S. Gandin. *Objective Analysis of Meteorological Fields*. Israel Program for Scientific Translations, Jerusalem, 1965. English translation of *Obiektivnyi analiz meteorologicheskikh polei*. Gidrometeorologicheskoe Izdatdelstvo, Leningrad, 1963.
55. G. Gardner et al. Algorithm AS154. An algorithm for exact maximum likelihood estimation of autoregressive-moving averge models by means of Kalman filtering. *Applied Statistics*, 29:311–322, 1980.
56. M. Ghil and P. Malanotte-Rizzoli. Data assimilation in meteorology and oceanography. *Advances in Geophysics*, 33:141–266, 1991.
57. H. Gilgen et al. Technical plan for BSRN (Baseline Surface Radiation Network) data management, version 2.1 (final). WMO/TD-no. 443, WCRP/WMO, World Meteorological Organization, Geneva, 1995.

58. H. Gilgen et al. Means and trends of shortwave irradiance at the surface estimated from Global Energy Balance Archive data. *J. of Climate*, 11:2042–2061, 1998.
59. H. Gilgen and A. Ohmura. The Global Energy Balance Archive. *Bull. of the American Meteorological Society*, 80:831–850, 1999. Data are available from *http://bsrn.ethz.ch/gebastatus/*.
60. L.J. Griffiths and R. Prieto-Diaz. Spectral analysis of natural seismic events using autoregressive techniques. *IEEE transactions of Geo-Science*, GE-15:13–25, 1977.
61. E.J. Hannan. The estimation of the order of an ARMA process. *Annals of Statistics*, 8:1071–1081, 1980.
62. A.C. Harvey. *Time series models, 2nd Ed.* Harvester Wheatsheaf, London, 1993.
63. R. Hausherr et al. Unpublished data from Ph.D. project "Variability of soil chemical and yield traits in differently tilled small cereal fields" at the Institute for Plant Science ETHZ. Contact persons: R. Hausherr, P. Stamp, W. Richner, Institute for Plant Science ETHZ, Universitätsstrasse 2, CH-8092 Zurich *(http://www.ipw.ethz.ch)*.
64. W. Heisenberg. *Physics and Philosophy, the Revolution in Modern Science*. Harper and Row, New York, 1966.
65. W. Heisenberg. Über den anschaulichen Inhalt der quantenmechanischen Kinematik and Mechanik. In K. Bauman and R.U. Sexl, editors, *Die Deutungen der Quantentheorie*. Vieweg, Wiesbaden, Germany, 1992. Originally published 1927, as part of the Copenhagen interpretation of Quantum Mechanics.
66. J. Herzog and G. Müller-Westermeier. Homogenitätsprüfung und Homogenisierung klimatologischer Messreihen im Deutschen Wetterdienst. DWD Bericht Nr. 202, Deutscher Wetterdienst DWD, Offenbach, 1998.
67. T.W. Horst. HATS: Field observations to obtain spatially-filtered turbulence fields from crosswind arrays of sonic anemometers in the atmospheric boundary layer. *J. Atmos. Sci.*, 61:1566–1581, 2004.
68. J.W. Hurrel et al., editors. *The North Atlantic Oscillation: Climatic Significance and Environmental Impacts*, volume 134 of *Geophysical Monograph Series*. Americal Geophysical Union, 2003.
69. J.W. Hurrell. NAO index. Joint Institute for the Study of the Atmosphere and Ocean, Department of Atmospheric Sciences, University of Washington, 408 ATG Bldg., Seattle *(http://tao.atmos.washington.edu/data_sets/nao)*.
70. J.W. Hurrell. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science*, 269:676–679, 1995.
71. Intergovernmental Panel on Climate Change (IPCC). *Climate Change 2001: The Scientific Basis*. Cambridge University Press, Cambride, 2003.
72. IRIS. Seismic monitor and data center. Run by Incorporated Research Institutions for Seismology, 1200 New York Ave NW, Washington DC 20005 *(http://iris.washington.edu/about/iris.htm)*.
73. L. Isserlis. On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables. *Biometrika*, 12:134–138, 1918.
74. G.M. Jenkins and D.G. Watts. *Spectral Analysis and its Applications*. Holden-Day, San Francisco, 1968.

75. P.D. Jones. Global average temparatures. Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK *(http://www.cru.uea.ac.uk/data/)*.
76. P.D. Jones. NAO index. Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, UK *(http://www.cru.uea.ac.uk/data/nao.htm)*.
77. P.D. Jones et al. Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *Int. J. of Climatology*, 17:1433–1450, 1997.
78. R.H. Jones. Maximum likelihood fitting of ARMA models to times series with missing observations. *Technometrics*, 20:389–395, 1980.
79. R.H. Jones. Time series analysis – time domain. In A.H. Murphy and R.W. Katz, editors, *Probability, Statistics, and Decision Making in the Atmospheric Sciences*. Westview Press, Boulder, CO, 1985.
80. P. Jussel et al. Transport modeling in heterogeneous aquifers: 1. Stastistical discription and numerical generation. *Water Resources Research*, 30:1803–1818, 1994.
81. I. Karatsas and S. Shreve. *Brownian Motion and Stochastic Calculus, 2nd Ed.* Springer, New York and Berlin, 1997.
82. T.R. Karl et al. Detecting climate variations and change: New challenges for observing and data management systems. *J. of Climate*, 6:1481–1494, 1993.
83. S.M. Kay. *Modern Spectral Estimation: Theory and Application*. Prentice-Hall, Englewood Cliffs, NJ., 1988.
84. C.D. Keeling and T.P. Whorf. Atmospheric CO_2 records from sites in the SIO air sampling network. In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn. *(http://cdiac.esd.ornl.gov/trends/sio-mlo.htm)*, 2004.
85. H.R. Künsch. State Space and Hidden Markov Models. In O.E. Barndorff-Nielsen, C. Cox, and C. Klüppelberg, editors, *Complex Stochastic Systems*. Chapman and Hall / CRC, 2001.
86. Lamont-Doherty Earth Observatory. Southern oscillation (SO) index. Lamont-Doherty Earth Observatory at the Columbia Univesity, New York *(http://ingrid.ldeo.columbia.edu/descriptions/soi.html)*.
87. G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65:297–303, 1978.
88. R. Lund and J. Reeves. Detection of undocumented changepoints: a revision of the two-phase regression model. *J. of Climate*, 15:2547–2554, 2002.
89. S.L. Marple. *Digital Spectral Analysis with Applications*. Prentice-Hall, Englewood Cliffs, NJ., 1987.
90. J. Marshall et al. North Atlantic climate variability: phenomena, impacts and mechanisms. *Int. J. of Climatology*, 21:1863–1898, 2001.
91. Mathematics at the University of York. Collection of time series. Department of Mathematics, University of York, Heslington, York YO10 5DD *(http://www.york.ac.uk/depts/maths/data/ts/)*.
92. G. Mathéron. The intrinsic random functions and their applications. *Adv. in Applied Probability*, 5:439–468, 1973.
93. J. Mathieu and J. Scott. *An Introduction to Turbulent Flow*. Cambridge University Press, Cambridge, 2000.
94. B. McArthur. Baseline Surface Radiation Network BSRN. Operations Manual. WMO/TD-no. 879, WCRP/WMO, World Meteorological Organization, Geneva, 1998.

95. A.I. McLeod and C. Jimenez. Nonnegative definiteness of the sample autocovariance function. *The American Statistician*, 39:237–238, 1984.
96. P. Moin and J. Kim. Tackling turbulence with supercomputers. *Scientific American*, Jan. 1997. *(http://www.sciam.com/0197issue/0197moin.html).*
97. A.S. Monin and A.M. Yaglom. *Statistical Fluid Mechanics, Vols. I and II, 3rd and 5th printing.* The MIT Press, Cambridge, MA., 1987. Translation of *Statisticheskaya gidromekhanika - Mekhanika Turbulenosti,* Nauka Press, Moscow, 1965.
98. R.B. Montgomery. Report on the work of G.T. Walker. *Monthly Weather Review*, supplement no. 39:1–22, 1940.
99. R.B. Montgomery. Verification of three of Walker's seasonal forecasting formulae for India rain. *Monthly Weather Review*, supplement no. 39:23–34, 1940.
100. National Geophysical Data Center. Sunspot numbers. The National Geophysical Data Center is run by the National Oceanic and Atmospheric Administration of the US Department of Commerce and located in Boulder, CO. *(http://www.ngdc.noaa.gov/stp/SOLAR/SSN/ssn.html).*
101. A.A. Nowroozi. Table for Fisher's test of significance in harmonic analysis. *Geophysical Journal of the Royal Astronomical Society*, 12:517–520, 1967.
102. A.H. Nutall and G.C. Carter. Spectral estimation using combined time and lag weighting. *Proc. IEEE*, 70:1115–1125, 1982.
103. A. Ohmura et al. Baseline Surface Radiation Network (BSRN/WCRP), a new precision radiometry for climate research. *Bull. of the American Meteorological Society*, 79:2115–2136, 1998. Data are available from *http://bsrn.ethz.ch.*
104. A. Papoulis. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, New York, 1981.
105. E.J. Pebesma. Gstat user's manual. Available from *http://www.gstat.org*, a domain hosted by the Department of Geography at Utrecht University, the Netherlands.
106. E.J. Pebesma. Multivariable Geostatistics in S: the Gstat package. *Computers and Geosciences*, 30:683–691, 2004.
107. E.J. Pebesma and C.G. Wesseling. Gstat, a program for geostatistical modelling, prediction und simulation. *Computers and Geosciences*, 24:17–31, 1998.
108. D.B. Percival and A.T. Walden. *Spectral Analysis of Time Series for Physical Applications.* Cambridge University Press, Cambridge, 1993.
109. T.C. Peterson et al. Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. of Climatology*, 18:1493–1517, 1998.
110. S.G.H. Philander. *El Niño, La Niña, and the Southern Oscillation.* Academic Press, San Diego, 1990.
111. D.C. Powell and C.E. Elderkin. An investigation of the application of Taylor's hypothesis to atmospheric boundary layer turbulence. *J. Atmos. Sci.*, 31:990–1002, 1974.
112. M.B. Priestley. Basic considerations in the estimation of spectra. *Technometrics*, 4:551–564, 1962.
113. M.B. Priestley. *Spectral Analysis and Time Series, Vols. I and II.* Academic Press, London, 1981.
114. R Foundation for Statistical Computing. The R project for statistical computing. The R Foundation, c/o Institut für Statistik und Wahrscheinlichkeitstheorie, Technische Universität Wien, Wiedner Hauptstrasse 8, 1040 Wien, Austria *(http://www.r-project.org).*

115. C.F. Ropelewski. SO index. Climatic Research Unit, University of East Anglia, Norwich NR4 7TJ, Uk *(http://www.cru.uea.ac.uk/data/soi.htm)*.
116. C.F. Ropelewski and P.D. Jones. An extension of the Tahiti-Darwin Southern Oscillation index. *Monthly Weather Review*, 115:2161–2165, 1987.
117. M. Rosenblatt. *Stationary Sequences and Random Fields*. Birkhäuser, Boston, 1985.
118. A.J. Rossman. *Workshop Statistics, 2nd Ed.* Springer, New York and Berlin, 2000.
119. M.W. Rotach. Turbulence close to a rough urban surface part I: Reynolds stress. *Boundary-Layer Meteorol.*, 65:1–28, 1993.
120. M.W. Rotach et al. Turbulence structure and exchange processes in an Alpine valley: The Riviera project. *Bull. of the American Meteorological Society*, 85:1367–1385, 2004.
121. B. Rudolf. Die Bestimmung der zeitlich-räumlichen Struktur des globalen Niederschlags. DWD Bericht Nr. 196, Deutscher Wetterdienst DWD, Offenbach, 1995.
122. G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.
123. D.S. Shepard. Computer mapping: the SYMAP interpolation algorithm. In G.L. Gaile and C.J. Willmott, editors, *Spatial Statistics and Models*. Reidel, Dordrecht, Netherlands, 1984.
124. A.F. Siegel. Periodogram tests. *J. of Am. Stat. Ass.*, 75:345, 1980.
125. E. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty – V: the discrete case. *Bell System Tech. J.*, 57:1371–1430, 1978.
126. W.I. Smirnov. *Lehrgang der höheren Mathematik, Teil V*. Deutscher Verlag der Wissenschaften, Berlin, 1974.
127. G.L. Smith et al. Clouds and earth radiant energy system (CERES): an overview. *Adv. Space Res.*, 24:1125–1131, 2004.
128. W. Stahel. *Statistische Datenanalyse. Eine Einführung für Naturwissenschaftler, 3. Aufl.* Vieweg, Braunschweig, 2000.
129. N.C. Stenseth et al. Common dynamic structure of canada lynx populations within three climatic regions. *Science*, 285:1071–1073, 1999.
130. H.H. Storrer. *Einführung in die mathematische Behandlung der Naturwissenschaften II.* Birkhäuser Skripten Band 8. Birkhäuser, Berlin, 1995.
131. Swiss Federal Institute for Snow and Avalanche Research (SLF). Snow and avalanche information. The SLF is located in Davos, Switzerand. In the winter season, it makes daily snow and avalanche information available at *http://www.slf.ch/*.
132. Swiss Hydrological Survey. Data Processing. Papiermühlestrasse 72, CH-3063 Ittigen, Switzerland *(http://www.bwg.admin.ch/bwg/abteil/e/abthyd.htm)*.
133. Swiss Seismological Service. Verification of Nuclear Explosions. The Swiss Seismological Service is run by the Institute of Geophysics, ETH Zurich *(http://www.seismo.ethz.ch/bsv)*.
134. D.J. Thomson. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70:1055–1096, 1982.
135. D.J. Thomson. Quadratic-inverse spectrum estimates: application to palaeoclimatology. *Phil.Trans.Roy.Soc.A*, 332:539–597, 1990.
136. W.N. Venables and B.D. Ripley. *Modern Applied Statistics with Splus, 3rd Ed.* Springer, New York and Berlin, 1999.

137. W.N. Venables and B.D. Ripley. *S Programming*. Springer, New York and Berlin, 2000.
138. E.O. Waagen. Personal communication, May 2005. Elizabeth O. Waagen, Senior Technical Assistant, American Association of Variable Star Observers (AAVSO), 25 Birch Street, Cambridge, MA 02138. *(http://www.aavso.org/)*.
139. A.T. Walden and R.E. White. Estimating the statistical bandwidth of a time series. *Biometrika*, 77:699–707, 1990.
140. X.L. Wang. Comments on "Detection of undocumented changepoints: A revision of the two-phase regression model". *J. of Climate*, 16:3383–3385, 2003.
141. H. Wanner et al. North Atlantic oscillation – Concepts and studies. *Surveys in Geophysics*, 22:321–382, 2001.
142. E.C. Weatherhead et al. Factors affecting the detection of trends: Statistical considerations and applications to environmental data. *Journal of Geophysical Research*, 103:17149–17161, 1998.
143. N. Weiss. *Introductory Statistics, 6th Ed.* Pearson Addison Wesley, New York, 2001.
144. P.D. Welch. The use of Fast Fourier Transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, 15:70–73, 1967.
145. P.H. Westfall and S.S. Young. *Resampling Based Multiple Testing*. Wiley, New York, 1993.
146. E.T. Whittaker and G. Robinson. *The Calculus of Observations: A Treatise on Numerical Mathematics, 4th ed*. Blackie, London, 1944.
147. S. Wichert et al. Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics*, 20:5–20, 2004.
148. B.A. Wielicki et al. Changes in Earth's albedo measured by satellite. *Science*, 308:825, 2005.
149. M. Wild et al. Validation of general circulation model radiative fluxes using surface observations from the global energy balance archive. *J. of Climate*, 8:1309–1324, 1995.
150. M. Wild et al. The disposition of radiative energy in the global climate system: GCM versus observational estimates. *Climate Dynamics*, 14:853–869, 1998.
151. M. Wild et al. On the consistency of trends in radiation and temperature records and implications for the global hydrological cycle. *Geophysical Research Letters*, 31:L11201,doi:10.1029/2003GL019188, 2004.
152. M. Wild et al. From dimming to brightening: Decadal changes in solar radiation at the Earth's surface. *Science*, 308:847–850, 2005.
153. W.B. Wu. Fourier transforms of stationary processes. *Proc.Amer.Math.Soc.*, 133:285–293, 2005. *(http://www.ams.org/proc/2005-133-01/S0002-9939-04-07528-8/)*.
154. C. Wunsch. The interpretation of short climate records, with comments on the North Atlantic and Southern Oscillations. *Bull. of the American Meteorological Society*, 80:245–255, 1999.
155. A.M. Yaglom. *Einführung in die Theorie der stationären Zufallsfunktionen*. Akademie Verlag, Berlin, 1959.

Index