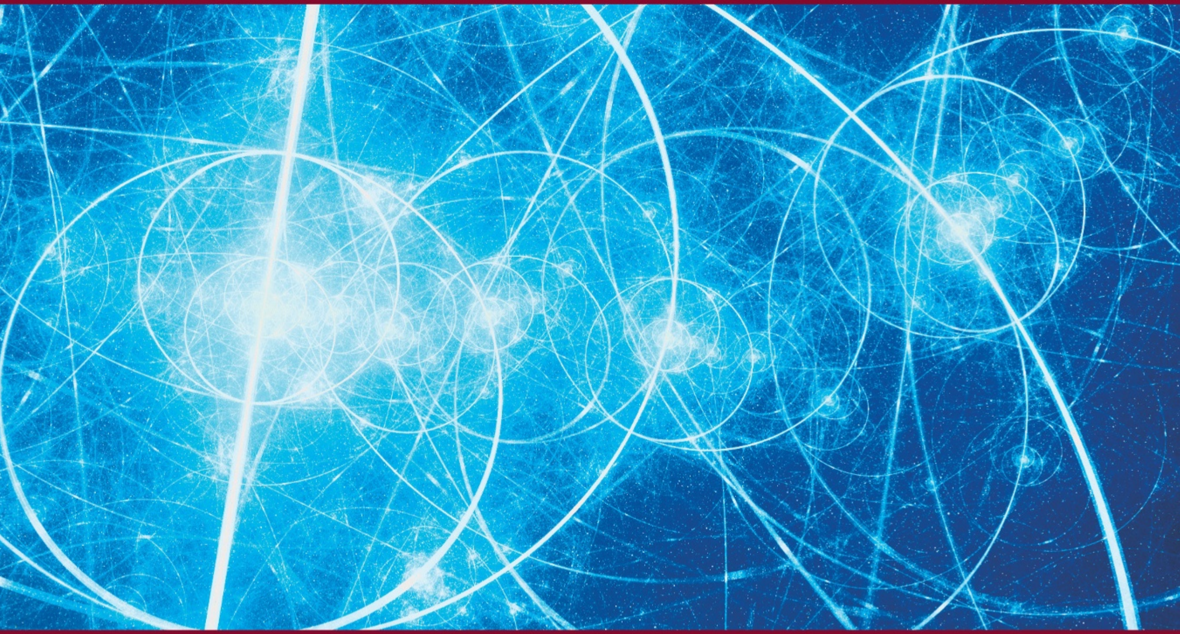
MATHEMATICS AND STATISTICS



Theory and Statistical Applications of Stochastic Processes

Yuliya Mishura
Georgiy Shevchenko

iSTE

WILEY

Theory and Statistical Applications of Stochastic Processes

Series Editor
Nikolaos Limnios

Theory and Statistical Applications of Stochastic Processes

Yuliya Mishura
Georgiy Shevchenko

iSTE

WILEY

First published 2017 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

Library of Congress Control Number: 2017953309

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN 978-1-78630-050-8

Contents

Preface

This book is concerned with both mathematical theory of stochastic processes and some theoretical aspects of statistics for stochastic processes. Our general idea was to combine classic topics of the theory of stochastic processes – measure-theoretic issues of existence, processes with independent increments, Gaussian processes, martingales, continuity and related properties of trajectories and Markov properties – with contemporary subjects – stochastic analysis, stochastic differential equations, fractional Brownian motion and parameter estimation in diffusion models. A more detailed exposition of the contents of the book is given in the Introduction.

We aimed to make the presentation of material as self-contained as possible. With this in mind, we have included several complete proofs, which are often either omitted from textbooks on stochastic processes or replaced by some informal or heuristic arguments. For this reason, we have also included some auxiliary materials, mainly related to different subjects of real analysis and probability theory, in the comprehensive appendix. However, we could not cover the full scope of the topic, so a substantial background in calculus, measure theory and probability theory is required.

The book is based on lecture courses, *Theory of stochastic processes*, *Statistics of stochastic processes*, *Stochastic analysis*, *Stochastic differential equations*, *Theory of Markov processes*, *Generalized processes of fractional Brownian motion and Diffusion processes*, taught regularly in the Mechanics and Mathematics Faculty of Taras Shevchenko National University of Kyiv and *Stochastic differential equations* lecture courses taught at the University of Verona in Spring 2016; *Fractional Brownian motion and related processes: stochastic calculus, statistical applications and modeling* taught in School in Bedlewo in March 2015; *Fractional Brownian motion and related processes* taught at Ulm University in June 2015; and a *Fractional Brownian motion in a nutshell* mini-course given at the 7th Jagna International Conference in 2014.

The book is targeted at the widest audience: students of mathematical and related programs, postgraduate students, postdoctoral researchers, lecturers, researchers, practitioners in the fields concerned with the application of stochastic processes, etc. The book would be most useful when accompanied by a problem in stochastic processes; we recommend [GUS 10] as it matches our topics best.

We would like to express our gratitude to everyone who made the creation of this book possible. In particular, we would like to thank Łukasz Stettner, Professor at the Department of Probability Theory and Mathematics of Finance, Institute of Mathematics, Polish Academy of Sciences; Luca Di Persio, Assistant Professor at the Department of Computer Science at the University of Verona; Evgeny Spodarev, Professor and Director of the Institute of Stochastics at Ulm University, for their hospitality while hosting Yuliya Mishura during lecture courses. We would also like to thank Alexander Kukush, Professor at the Department of Mathematical Analysis of Taras Shevchenko National University of Kyiv, for proofreading the statistical part of the manuscript, and Evgeniya Munchak, PhD student at the Department of Probability, Statistics, and Actuarial Mathematics of Taras Shevchenko National University of Kyiv, for her help in typesetting the manuscript.

Yuliya MISHURA
Georgiy SHEVCHENKO

September 2017

Introduction

In the world that surrounds us, a lot of events have a random (nondeterministic) structure. At molecular and subatomic levels, all natural phenomena are random. Movement of particles in the surrounding environment is accidental. Numerical characteristics of cosmic radiation and the results of monitoring the effect of ionizing radiation are random. The majority of economic factors surrounding asset prices on financial markets vary randomly. Despite efforts to mitigate risk and randomness, they cannot be completely eliminated. Moreover, in complex systems, it is often easier to reach an equilibrium state when they are not too tightly controlled. Summing-up, chance manifests itself in almost everything that surrounds us, and these manifestations vary over time. Anyone can simulate time-varying randomness by tossing a coin or rolling a dice repeatedly and recording the results of successive experiments. (If a physical random number is unavailable, one of the numerous computer algorithms to generate random numbers can be used.) In view of this ubiquity of randomness, the theory of probability and stochastic processes has a long history, despite the fact that the rigorous mathematical notion of probability was introduced less than a century ago. Let us speak more on this history.

People have perceived randomness since ancient times, for example, gambling already existed in ancient Egypt before 3000 BC. It is difficult to tell exactly when systematic attempts to understand randomness began. Probably, the most notable were those made by the prominent ancient Greek philosopher Epicurus (341–270 BC). Although his views were heavily influenced by Democritus, he attacked Democritus' materialism, which was fully deterministic. Epicurus insisted that all atoms experience some random perturbations in their dynamics. Although modern physics confirms these ideas, Epicurus himself attributed the randomness to the free will of atoms. The phenomenon of random detours of atoms was called *clinamen* (cognate to inclination) by the Roman poet Lucretius, who had brilliantly exposed Epicurus' philosophy in his poem *On the Nature of Things*.

Moving closer to present times, let us speak of the times where there was no theory of stochastic processes, physics was already a well-developed subject, but there wasn't any equipment suitable to study objects in sufficiently small microscopic detail. In 1825, botanist Robert Brown first observed a phenomenon, later called *Brownian motion*, which consisted of a chaotic movement of a pollen particle in a vessel. He could not come up with a model of this system, so just stated that the behavior is random.

A suitable model for the phenomenon arose only several decades later, in a very different problem, concerned with the pricing of financial assets traded on a stock exchange. A French mathematician Louis Bachelier (1870–1946), who aimed to find a mathematical description of stochastic fluctuations of stock prices, provided a mathematical model in his thesis "Théorie de la spéculation" [BAC 95], which was defended at the University of Paris in 1900. The model is, in modern terms, a stochastic process, which is characterized by the fact that its increments in time, in a certain statistical sense, are proportional to the square root of the time change; this "square root" phenomenon had also be observed earlier in physics; Bachelier was the first to provide a model for it. Loosely speaking, according to Bachelier, the asset price S_t at time t is modeled by

$$S_t = at + b\sqrt{t}\xi,$$

where a, b are constant coefficients, and ξ is a random variable having Gaussian distribution.

The work of Bachelier was undervalued, probably due to the fact that applied mathematics was virtually absent at the time, as well as concise probability theory. Bachelier spent his further life teaching in different universities in France and never returned to the topic of his thesis. It was only brought to the spotlight 50 years after its publication, after the death of Bachelier. Now, Bachelier is considered a precursor of mathematical finance, and the principal organization in this subject bears his name: Bachelier Finance Society.

Other works which furthered understanding towards Brownian motion were made by prominent physicists, Albert Einstein (1879–1955) and Marian Smoluchowski (1872–1917). Their articles [EIN 05] and [VON 06] explained the phenomenon of Brownian motion by thermal motion of atoms and molecules. According to this theory, the molecules of a gas are constantly moving with different speeds in different directions. If we put a particle, say of pollen which has a small surface area, inside the gas, then the forces from impacts with different molecules do not compensate each other. As a result, this *Brownian* particle will experience a chaotic movement with velocity and direction changing approximately 10^{14} times per second. This gave a physical explanation to the phenomenon observed by the botanist. It also turned out that a kinetic theory of thermal motion required a stochastic process B_t. Einstein and Smoluchowski not only described this stochastic process, but also found its important probabilistic characteristics.

Only a quarter of a century later, in 1931, Andrey Kolmogorov (1903–1987) laid the groundwork for probability theory in his pioneering works *About the Analytical Methods of Probability Theory* and *Foundations of the Theory of Probability* [KOL 31, KOL 77]. This allowed his fellow researcher Aleksandr Khinchin (1894–1859) to give a definition of stochastic process in his article [KHI 34].

There is an anecdote related to the role of Khinchin in defining a stochastic process and the origins of the "stochastic" as a synonym for randomness (the original Greek word means "guessing" and "predicting"). They say that when Khinchin defined the term "random process", it did not go well with the Soviet authorities. The reason is that the notion of random process used by Khinchin contradicted dialectical materialism (diamat). In diamat, similarly to Democritus' materialism, all processes in nature are characterized by totally deterministic development, transformation, etc., so the phrase "random process" itself sounded paradoxical. As a result, to avoid dire consequences (we recall that 1934 was the apogee of Stalin's Great Terror), Khinchin had to change the name. After some research, he came up with the term "stochastic", from *στοχαστική τέχνη*, the Greek title of *Ars Conjectandi*, a celebrated book by Jacob Bernoulli (1655–1705) published in 1713, which contains many classic results. Being popularized later by William Feller [FEL 49] and Joseph Doob [DOO 53], this became a standard notion in English and German literature. Perhaps paradoxically, in Russian literature, the term "stochastic processes" did not live for long. The 1956 Russian translation of Doob's monograph [DOO 53] of this name was entitled *Probabilistic processes*, and now the standard name is *random process*.

An alternative explanation, given, for example, in [DEL 17], attributes the term "stochastic" to Ladislaus Władysław Bortkiewicz (1868–1931), Russian economist and statistician, who in his paper, *Die Iterationen* [BOR 17], defined the term "stochastic" as "the investigation of empirical varieties, which is based on probability theory, and, therefore, on the law of large numbers. But stochastic is not simply probability theory, but above all probability theory and applications". This meaning correlates with the one given in *Ars Conjectandi* by Jacob Bernoulli, so the true origin of the term probably is somewhere between these two stories. It is also worth mentioning that Bortkiewicz is known for proving the *Poisson approximation theorem* about the convergence of binomial distributions with small parameters to the Poisson distribution, which he called *the law of small numbers*.

This historical discussion would be incomplete without mentioning Paul Lévy (1886–1971), a French mathematician who made many important contributions to the theory of stochastic processes. Many objects and theorems now bear his name: *Lévy processes*, *Lévy-Khinchin representation*, *Lévy representation*, etc. Among other things, he wrote the first extensive monograph on the (mathematical model of) Brownian motion [LÉV 65].

Further important progress in probability theory is related to Norbert Wiener (1894–1964). He was a jack of all trades: a philosopher, a journalist, but the most important legacy that he left was as a mathematician. In mathematics, his interest was very broad, from number theory and real analysis, to probability theory and statistics. Besides many other important contributions, he defined an integral (of a deterministic function) with respect to the mathematical model of Brownian motion, which now bears his name: a *Wiener process* (and the corresponding integral is called a *Wiener integral*).

The ideas of Wiener were developed by Kiyoshi Itô (1915–2008), who introduced an integral of random functions with respect to the Wiener process in [ITÔ 44]. This lead to the emergence of a broad field of *stochastic analysis*, a probabilistic counterpart to real integro-differential calculus. In particular, he defined *stochastic differential equations* (the name is self-explanatory), which allowed us to study *diffusion processes*, which are natural generalizations of the Wiener process. As with Lévy, many objects in stochastic analysis are named after Itô: *Itô integral*, *Itô process*, *Itô representation*, *Wiener-Itô decomposition*, etc.

An important contribution to the theory of stochastic processes and stochastic differential equations was made by Ukrainian mathematicians Iosif Gihman (1918–1985) and more notably by Anatoliy Skorokhod (1930–2011). Their books [GIH 72, GIK 04a, GIK 04b, GIK 07] are now classical monographs. There are many things in stochastic analysis named after Skorokhod: *Skorokhod integral*, *Skorokhod space*, *Skorokhod representation*, etc.

Our book, of course, is not the first book on stochastic processes. They are described in many other texts, from some of which we have borrowed many ideas presented here, and we are grateful to their authors for the texts. It is impossible to mention every single book here, so we cite only few texts of our selection. We apologize to the authors of many other wonderful texts which we are not able to cite here.

The extensive treatment of probability theory with all necessary context is available in the books of P. Billingsley [BIL 95], K.-L. Chung [CHU 79], O. Kallenberg [KAL 02], L. Koralov and Y. Sinai [KOR 07], M. Loève [LOÈ 77, LOÈ 78], D. Williams [WIL 91]. It is also worth mentioning the classic monograph of P. Billingsley [BIL 99] concerned with different kinds of convergence concepts in probability theory.

For books which describe the theory of stochastic processes in general, we recommend that the reader looks at the monograph by J. Doob [DOO 53], the extensive three-volume monograph by I. Gikhman and A. Skorokhod

[GIK 04a, GIK 04b, GIK 07], the textbooks of Z. Brzezniak and T. Zastawniak [BRZ 99], K.-L. Chung [CHU 79], G. Lawler [LAW 06], S. Resnick [RES 92], S. Ross [ROS 96], R. Schilling and L. Partzsch [SCH 14], A. Skorokhod [SKO 65], J. Zabczyk [ZAB 04]. It is also worth mentioning the book by A. Bulinskiy and A. Shiryaev [BUL 05], from which we borrowed many ideas; unfortunately, it is only available in Russian. Martingale theory is well presented in the books of R. Liptser and A. Shiryaev [LIP 89], J. Jacod and A. Shiryaev [JAC 03], L. Rogers and D. Williams [ROG 00a], and the classic monograph of D. Revuz and M. Yor [REV 99]. There are many excellent texts related to different aspects of Lévy processes, including the books of D. Applebaum [APP 09], K. Sato [SAT 13], W. Schoutens [SCH 03], and the collection [BAR 01].

Stochastic analysis now stands as an independent subject, so there are many books covering different aspects of it. The books of K.-L. Chung and D. Williams [CHU 90], I. Karatzas and S. Shreve [KAR 91], H. McKean [MCK 69], J.-F. Le Gall [LEG 16], L. Rogers and D. Williams [ROG 00b] cover stochastic analysis in general, and the monograph of P. Protter [PRO 04] goes much deeper into integration issues. Stochastic differential equations and diffusion processes are the subject of the best-selling textbook of B. Øksendal [ØKS 03], and the monographs of N. Ikeda and S. Watanabe [IKE 89], K. Itô and H. McKean [ITÔ 74], A. Skorokhod [SKO 65], and D. Strook and S. Varadhan [STR 06]. The ultimate guide to Malliavin calculus is given by D. Nualart [NUA 06]. Concerning financial applications, stochastic analysis is presented in the books of T. Björk [BJÖ 04], M. Jeanblanc, M. Yor, and M. Chesney [JEA 09], A. Shiryaev [SHI 99], and S. Shreve [SHR 04].

Different aspects of statistical methods for stochastic processes are covered by the books of P. Brockwell and R. Davis [BRO 06], C. Heyde [HEY 97], Y. Kutoyants [KUT 04], G. Seber and A. Lee [SEB 03].

Finally, fractional Brownian motion, one of the main research interests of the authors of this book, is covered by the books of F. Biagini *et al.* [BIA 08], Y. Mishura [MIS 08], I. Nourdin [NOU 12], D. Nualart [NUA 06], and by lecture notes of G. Shevchenko [SHE 15].

Our book consists of two parts: the first is concerned with the theory of stochastic processes and the second with statistical aspects.

In the first chapter, we define the main subjects: stochastic process, trajectory and finite-dimensional distributions. We discuss the fundamental issues: existence and construction of a stochastic process, measurability and other essential properties, and sigma-algebras generated by stochastic processes.

The second chapter is devoted to stochastic processes with independent increments. A definition is given and simple criteria which provide the existence are discussed. We also provide numerous important examples of processes with independent increments, including Lévy processes, and study their properties.

The third chapter is concerned with a subclass of stochastic processes, arguably the most important for applications: Gaussian processes. First, we discuss Gaussian random variables and vectors, and then we give a definition of Gaussian processes. Furthermore, we give several important examples of Gaussian processes and discuss their properties. Then, we discuss integration with respect to Gaussian processes and related topics. Particular attention is given to fractional Brownian motion and Wiener processes, with discussion of several integral representations of fractional Brownian motion.

The fourth chapter focuses on some delicate properties of two Gaussian processes, which are of particular interest for applications: the Wiener process and fractional Brownian motion. In particular, an explicit construction of the Wiener process is provided and nowhere differentiability of its trajectories is shown. Having in mind the question of parameter estimation for stochastic processes, we also discuss the asymptotic behavior of power variations for the Wiener process and fractional Brownian motion in this chapter.

In the fifth chapter, we attempted to cover the main topics in the martingale theory. The main focus is on the discrete time case; however, we also give several results for stochastic processes. In particular, we discuss the notions of stochastic basis with filtration and stopping times, limit behavior of martingales, optional stopping theorem, Doob decomposition, quadratic variations, maximal inequalities by Doob and Burkholder-Davis-Gundy, and the strong law of large numbers.

The sixth chapter is devoted to properties of trajectories of a stochastic process. We introduce different notions of continuity as well as important concepts of separability, indistinguishability and stochastic equivalence, and establish several sufficient conditions for continuity of trajectories and for absence of discontinuities of the second kind. To the best of our knowledge, this is the first time that the different aspects of regularity and continuity are comprehensively discussed and compared.

The seventh chapter discusses Markov processes. The definition, together with several important examples, is followed by analytical theory of Markov semigroups. The chapter is concluded by the investigation of diffusion processes, which serves as a bridge to stochastic analysis discussed in the following chapters. We provide a definition and establish important criteria and characterization of diffusion processes. We pay particular attention to the forward and backward Kolmogorov equations, which are of great importance for applications.

In the eighth chapter, we give the classical introduction to stochastic integration theory, which includes the definition and properties of Itô integral, Itô formula, multivariate stochastic calculus, maximal inequalities for stochastic integrals, Girsanov theorem and Itô representation.

The ninth chapter, which closes the theoretical part of the book, is concerned with stochastic differential equations. We give a definition of stochastic differential equations and establish the existence and uniqueness of its solution. Several properties of the solution are established, including integrability, continuous dependence of the solution on the initial data and on the coefficients of the equation. Furthermore, we prove that solutions to stochastic differential equations are diffusion processes and provide a link to partial differential equations, the Feynman-Kac formula. Finally, we discuss the diffusion model of a financial market, giving notions of arbitrage, equivalent martingale measure, pricing and hedging of contingent claims.

The tenth chapter opens the second part of the book, which is devoted to statistical aspects. It studies the estimation of parameters of stochastic processes in different scenarios: in a linear regression model with discrete time, in a continuous time linear model driven by Wiener process, in models with fractional Brownian motions, in a linear autoregressive model and in homogeneous diffusion models.

In the eleventh chapter, the classic problem of optimal filtering is studied. A statistical setting is described, then a representation of optimal filter is given as an integral with respect to an observable process. Finally, the integral Wiener-Hopf equation is derived, a linear stochastic differential equation for the optimal filter is derived, and the error of the optimal filter is identified in terms of solution of the Riccati equation. In the case of constant coefficients, the explicit solutions of these equations are found.

Auxiliary results, which are referred to in the book, are collected in Appendices 1 and 2. In Appendix 1, we give essential facts from calculus, measure theory and theory of operators. Appendix 2 contains important facts from probability theory.

PART 1

Theory of Stochastic Processes

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

1

Stochastic Processes. General Properties. Trajectories, Finite-dimensional Distributions

1.1. Definition of a stochastic process

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. Here, Ω is a sample space, i.e. a collection of all possible outcomes or results of the experiment, and $\mathcal{F}$ is a σ-field; in other words, $(\Omega, \mathcal{F})$ is a measurable space, and P is a probability measure on $\mathcal{F}$. Let $(\mathcal{S}, \Sigma)$ be another measurable space with σ-field Σ, and let us consider the functions defined on the space $(\Omega, \mathcal{F})$ and taking their values in $(\mathcal{S}, \Sigma)$. Recall the notion of random variable.

DEFINITION 1.1.– *A* random variable *on the probability space* $(\Omega, \mathcal{F})$ *with the values in the measurable space* $(\mathcal{S}, \Sigma)$ *is a measurable map* $\Omega \xrightarrow{\xi} \mathcal{S}$*, i.e. a map for which the following condition holds: the pre-image* $\xi^{-1}(B)$ *of any set* $B \in \Sigma$ *belongs to* $\mathcal{F}$*.* Equivalent forms of this definition are: for any $B \in \Sigma$, we have that

$$\xi^{-1}(B) \in \mathcal{F},$$

or, for any $B \in \Sigma$, we have that

$$\{\omega : \xi(\omega) \in B\} \in \mathcal{F}.$$

Consider examples of random variables.

1) The number shown by rolling a fair die. Here,

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}, \mathcal{F} = 2^{\Omega}, \mathcal{S} = 1, 2, 3, 4, 5, 6, \Sigma = 2^{\mathcal{S}}.$$

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

2) The price of certain assets on a financial market. Here, $(\Omega, \mathcal{F})$ can depend on the model of the market, and the space $\mathcal{S}$, as a rule, coincides with $\mathbb{R}_+ = [0, +\infty)$.

3) Coordinates of a moving airplane at some moment of time. People use different coordinate systems to determine the coordinates of the airplane that has three coordinates at any time. The coordinates are time dependent and random, to some extent, because they are under the influence of many factors, some of which are random. Here, $\mathcal{S} = \mathbb{R}^3$ for the Cartesian system, or $\mathcal{S} = \mathbb{R}^2 \times [0, 2\pi]$ for the cylindrical system, or $\mathcal{S} = \mathbb{R} \times [0, \pi] \times [0, 2\pi]$ for the spherical system.

Now, we formalize the notion of a stochastic (random) process, defined on $(\Omega, \mathcal{F}, \mathrm{P})$. We will treat a random process as a set of random variables. That said, introduce the parameter set $\mathbb{T}$ with elements $t : t \in \mathbb{T}$.

DEFINITION 1.2.– *Stochastic process on the probability space* $(\Omega, \mathcal{F}, \mathrm{P})$*, parameterized by the set* $\mathbb{T}$ *and taking values in the measurable space* $(\mathcal{S}, \Sigma)$*, is a set of random variables of the form*

$$X_t = \{X_t(\omega), t \in \mathbb{T}, \omega \in \Omega\},$$

where $X_t(\omega) : \mathbb{T} \times \Omega \to S$*.*

Thus, each parameter value $t \in \mathbb{T}$ is associated with the random variable X_t taking its value in $\mathcal{S}$. Sometimes, we call $\mathcal{S}$ a phase space. The origin of the term comes from the physical applications of stochastic processes, rather than from the physical problems which stimulated the development of the theory of stochastic processes to a large extent.

Here are other common designations of stochastic processes:

$$X(t),\ \xi(t),\ \xi_t,\ X = \{X_t, t \in \mathbb{T}\}.$$

The last designation is the best in the sense that it describes the entire process as a set of the random variables. The definition of a random process can be rewritten as follows: for any $t \in \mathbb{T}$ and any set $B \in \Sigma$

$$X_t^{-1}(B) \in \mathcal{F}.$$

Another form: for any $t \in \mathbb{T}$ and any set $B \in \Sigma$

$$\{\omega : X_t(\omega) \in B\} \in \mathcal{F}.$$

In general, the space $\mathcal{S}$ can depend on the value of t, $\mathcal{S} = \mathcal{S}_t$, but, in this book, space $\mathcal{S}$ will be fixed for any fixed stochastic process $X = \{X_t, t \in \mathbb{T}\}$. If $\mathcal{S} = \mathbb{R}$, then the process is called real or real-valued. Additionally, we assume in this case that $\Sigma = \mathcal{B}(\mathbb{R})$, i.e. $(\mathcal{S}, \Sigma) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where $\mathcal{B}(\mathcal{S})$ is a Borel σ-field on $\mathcal{S}$. If

$\mathcal{S} = \mathbb{C}$, the process is called complex or complex-valued, and if $\mathcal{S} = \mathbb{R}^d, d > 1$, the process is called vector or vector-valued. In this case, $(\mathcal{S}, \Sigma) = (\mathbb{C}, \mathcal{B}(\mathbb{C}))$ and $(\mathcal{S}, \Sigma) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, respectively.

Concerning the parameter set $\mathbb{T}$, as a rule, it is interpreted as a time set. If the time parameter is continuous, then usually either $\mathbb{T} = [a, b]$, or $[a, +\infty)$ or $\mathbb{R}$. If the time parameter is discrete, then usually either $\mathbb{T} = \mathbb{N} = 1, 2, 3, \ldots$, or $\mathbb{T} = \mathbb{Z}^+ = \mathbb{N} \cup 0$ or $\mathbb{T} = \mathbb{Z}$.

The parameter set can be multidimensional, e.g. $\mathbb{T} = \mathbb{R}^m, m > 1$. In this case, we call the process a random field. The parameter set can also be mixed, the so-called time–space set, because we can consider the processes of the form $X(t, x) = X(t, x, \omega)$, where $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$. In this case, we interpret t as time and $x \in \mathbb{R}^d$ as the coordinate in the space $\mathbb{R}^d$.

There can be more involved cases, e.g. it is possible to consider random measures $\mu(t, A, \omega)$, where $t \in \mathbb{R}_+, A \in \mathcal{B}(\mathbb{R}^d)$, or random processes defined on the groups, whose origin comes from physics. We will not consider in detail the theory of such processes.

In what follows, we consider the real-valued parameter, i.e. $\mathbb{T} \subset \mathbb{R}$, so that we can regard the parameter as time, as described above.

1.2. Trajectories of a stochastic process. Some examples of stochastic processes

1.2.1. *Definition of trajectory and some examples*

A stochastic process $X = \{X_t(\omega), t \in \mathbb{T}, \omega \in \Omega\}$ is a function of two variables, one of them being a time variable $t \in \mathbb{T}$ and the other one a sample point (elementary event) $\omega \in \Omega$. As mentioned earlier, fixing $t \in \mathbb{T}$, we get a random variable $X_t(\cdot)$. In contrast, fixing $\omega \in \Omega$ and following the values that $X_\cdot(\omega)$ takes as the function of parameter $t \in \mathbb{T}$, we get a trajectory (path, sample path) of the stochastic process. The trajectory is a function of $t \in \mathbb{T}$ and, for any t, it takes its value in $\mathcal{S}$. Changing the value of ω, we get a set of paths. They are schematically depicted in Figure 1.1.

Let us consider some examples of random processes and draw their trajectories. First, we recall the concept of independence of random variables.

DEFINITION 1.3.– *Random variables $\{\xi_\alpha, \alpha \in \mathcal{A}\}$, where $\mathcal{A}$ is some parameter set, are called* mutually independent *if for any finite subset of indices $\{\alpha_1, \ldots, \alpha_k\} \subset \mathcal{A}$ and, for any measurable sets $A_1, \ldots, A_k$, we have that*

$$\mathrm{P}\{\xi_{\alpha_1} \in A_1, \ldots, \xi_{\alpha_k} \in A_k\} = \Pi_{i=1}^k \mathrm{P}\{\xi_{\alpha_i} \in A_i\}.$$

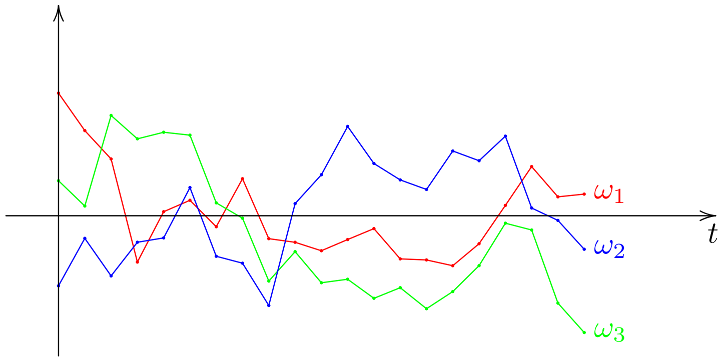


Figure 1.1. *Trajectories of a stochastic process. For a color version of the figure, see www.iste.co.uk/mishura/stochasticprocesses.zip*

1.2.1.1. *Random walks*

A *random walk* is a process with discrete time, e.g. we can put $\mathbb{T} = \mathbb{Z}^+$. Let $\{\xi_n, n \in \mathbb{Z}^+\}$ be a family of random variables taking values in $\mathbb{R}^d, d \geq 1$. Put $X_n = \sum_{i=0}^n \xi_i$. Stochastic process $X = \{X_n, n \in \mathbb{Z}_+\}$ is called a random walk in $\mathbb{R}^d$. In the case where $d = 1$, we have a random walk in the real line. In general, the random variables ξ_i can have arbitrary dependence between them, but the most developed theory is in the case of random walks with mutually independent and identically distributed variables $\{\xi_n, n \in \mathbb{Z}^+\}$. If, additionally, any random variable ξ_n takes only two values a and b with respective probabilities $\mathrm{P}\{\xi_n = a\} = p$ and $\mathrm{P}\{\xi_n = b\} = q = 1 - p \in (0,1)$, then we have a Bernoulli random walk. If $a = -b$ and $p = q = \frac{1}{2}$, then we have a symmetric Bernoulli random walk. The trajectory of the random walk consists of individual points, and is shown in Figure 1.2.

1.2.1.2. *Renewal process*

Let $\{\xi_n, n \in \mathbb{Z}^+\}$ be a family of random variables taking positive values with probability 1. Stochastic process $N = \{N_t, t \geq 0\}$ can be defined by the following formula:

$$N_t = \begin{cases} 0, t < \xi_1; \\ \sup\{n \geq 1 : \sum_{i=1}^n \xi_i \leq t\}, t \geq \xi_1. \end{cases}$$

Stochastic process $N = \{N_t, t \geq 0\}$ is called a *renewal process*. Trajectories of a renewal process are step-wise with step 1. The example of the trajectory is represented in Figure 1.3.

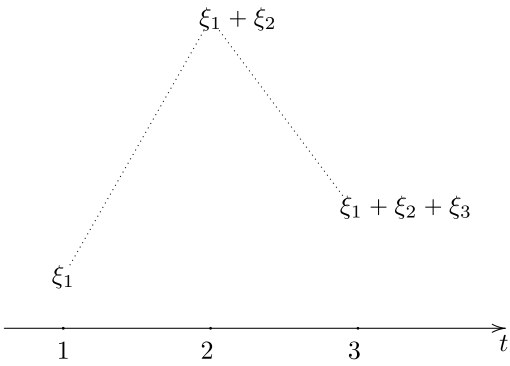


Figure 1.2. *Trajectories of a random walk*

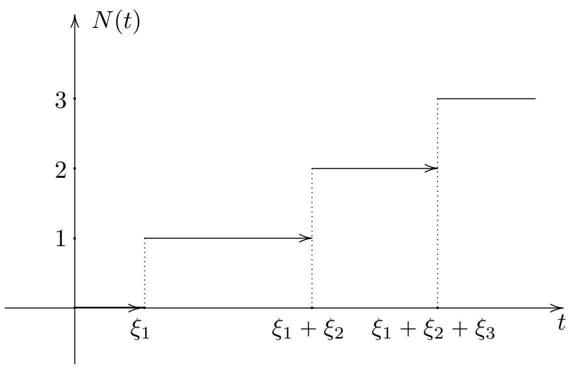


Figure 1.3. *Trajectories of a renewal process*

Random variables $T_1 = \xi_1, T_2 = \xi_1 + \xi_2, \ldots$ are called jump times, arrival times or renewal times of the renewal process. The latter name comes from the fact that the renewal processes were considered in applied problems related to moments of failure and replacement of equipment. Intervals $[0, T_1]$ and $[T_n, T_{n+1}], n \geq 1$ are called renewal intervals.

1.2.1.3. *Stochastic processes with independent values and those with independent increments*

DEFINITION 1.4.– *A stochastic process $X = \{X_t, t \geq 0\}$ is called a* process with independent values *if the random variables $\{X_t, t \geq 0\}$ are mutually independent.*

It will be shown later, in Example 6.1, that the trajectories of processes with independent values are quite irregular and, for this reason, the processes with independent values are relatively rarely used to model phenomena in nature, economics, technics, society, etc.

DEFINITION 1.5.– *A stochastic process* $X = \{X_t, t \geq 0\}$ *is called a* process with independent increments*, if, for any set of points* $0 \leq t_1 < t_2 < \ldots < t_n$*, the random variables* $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ *are mutually independent.*

Here is an example of a random process with discrete time and independent increments.

Let $X = \{X_n, n \in \mathbb{Z}^+\}$ be a random walk, $X_n = \sum_{i=0}^{n} \xi_i$, and the random variables $\{\xi_i, i \geq 0\}$ be mutually independent. Evidently, for any $0 \leq n_1 < n_2 < \ldots < n_k$, the random variables

$$X_{n_1} = \sum_{i=0}^{n_1} \xi_i, \quad X_{n_2} - X_{n_1} = \sum_{i=n_1+1}^{n_2} \xi_i, \ldots, \quad X_{n_k} - X_{n_{k-1}} = \sum_{i=n_{k-1}+1}^{n_k} \xi_i$$

are mutually independent; therefore, X is a process with discrete time and independent increments. Random processes with continuous time and independent increments are considered in detail in Chapter 2.

1.2.2. *Trajectory of a stochastic process as a random element*

Let $\{X_t, t \in \mathbb{T}\}$ be a stochastic process with the values in some set $\mathcal{S}$. Introduce the notation $\mathcal{S}^{\mathbb{T}} = \{y = y(t),\ t \in \mathbb{T}\}$ for the family of all functions defined on $\mathbb{T}$ and taking values in $\mathcal{S}$. Another notation can be $\mathcal{S}^{\mathbb{T}} = \times_{t \in \mathbb{T}} \mathcal{S}_t$, with all $\mathcal{S}_t = \mathcal{S}$ or simply $\mathcal{S}^{\mathbb{T}} = \times_{t \in \mathbb{T}} S$, which emphasizes that any element from $\mathcal{S}^{\mathbb{T}}$ is created in such a way that we take all points from $\mathbb{T}$, assigning a point from $\mathcal{S}$ to each of them. For example, we can consider $\mathcal{S}^{[0,\infty)}$ or $\mathcal{S}^{[0,T]}$ for any $T > 0$. Now, the trajectories of a random process X belong to the set $\mathcal{S}^{\mathbb{T}}$. Thus, considering the trajectories as elements of the set $\mathcal{S}^{\mathbb{T}}$, we get the mapping $X : \Omega \to \mathcal{S}^{\mathbb{T}}$, that transforms any element of Ω into some element of $\mathcal{S}^{\mathbb{T}}$. We would like to address the question of the measurability of this mapping. To this end, we need to find a σ-field $\Sigma^{\mathbb{T}}$ of subsets of $\mathcal{S}^{\mathbb{T}}$ such that the mapping X is $\mathcal{F}$-$\Sigma^{\mathbb{T}}$-measurable, and this σ-field should be the smallest possible. First, let us prove an auxiliary lemma.

LEMMA 1.1.– *Let* $\mathcal{Q}$ *and* $\mathcal{R}$ *be two spaces. Assume that* $\mathcal{Q}$ *is equipped with* σ*-field* $\mathcal{F}$*, and* $\mathcal{R}$ *is equipped with* σ*-field* $\mathcal{G}$*, where* $\mathcal{G}$ *is generated by some class* K*, i.e.* $\mathcal{G} = \sigma(K)$*. Then, the mapping* $f : \mathcal{Q} \to \mathcal{R}$ *is* $\mathcal{F}$*-*$\mathcal{G}$*-measurable if and only if it is* $\mathcal{F}$*-*K*-measurable, i.e. for any* $A \in K$*, the pre-image is* $f^{-1}(A) \in \mathcal{F}$*.*

PROOF.– Necessity is evident. To prove sufficiency, we should check that, in the case where the pre-images of all sets from K under mapping f belong to $\mathcal{F}$, the pre-images of all sets from $\mathcal{G}$ under mapping f belong to $\mathcal{F}$ as well. Introduce the family of sets

$$K_1 = \{B \in \mathcal{G} : f^{-1}(B) \in \mathcal{F}\}.$$

The properties of pre-images imply that K_1 is a σ-field. Indeed,

$$f^{-1}\left(\bigcup_{n=1}^{\infty} B_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(B_n) \in \mathcal{F},$$

if $f^{-1}(B_n) \in \mathcal{F}$,

$$f^{-1}(C_2 \setminus C_1) = f^{-1}(C_2) \setminus f^{-1}(C_1) \in \mathcal{F},$$

if $f^{-1}(C_1) \in \mathcal{F}, i = 1, 2$, and $f^{-1}(\mathcal{R}) = \mathcal{Q} \in \mathcal{F}$. It means that $K_1 \supset \sigma(K) = \mathcal{G}$, whence the proof follows. □

Therefore, to characterize the measurability of the trajectories, we must find a "reasonable" subclass of sets of $\mathcal{S}^{\mathbb{T}}$, the inverse images of which belong to $\mathcal{F}$.

DEFINITION 1.6.– *Let the point $t_0 \in \mathbb{T}$ and the set $A \subset \mathcal{S}, A \in \Sigma$ be fixed.* Elementary cylinder *with base A over point t_0 is the following set from $\mathcal{S}^{\mathbb{T}}$:*

$$C(t_0, A) = \{y = y(t) \in \mathcal{S}^{\mathbb{T}} : y(t_0) \in A\}.$$

If $\mathcal{S} = \mathbb{R}$ and A is some interval, then $C(t_0, A)$ is represented schematically in Figure 1.4. Elementary cylinder consists of the functions whose values at point t_0 belong to the set A.

Let K_{el} be the class of elementary cylinders, and $\mathbb{K}_{el} = \sigma(K_{el})$, with the σ-field being generated by the elementary cylinders.

THEOREM 1.1.– *For any stochastic process $X = \{X_t, t \in \mathbb{T}\}$, the mapping $X : \Omega \to S^{\mathbb{T}}$, which assigns to any element $\omega \in \Omega$ the corresponding trajectory $X(\cdot, \omega)$, is $\mathcal{F}$–$\mathbb{K}_{el}$-measurable.*

PROOF.– According to lemma 1.1, it is sufficient to check that the mapping X is $\mathcal{F}$–K_{el}-measurable. Let the set $C(t_0, A) \in K_{el}$. Then, the pre-image $X^{-1}(C(t_0, A)) = \{\omega \in \Omega : X(t_0, \omega) \in A\} \in \mathcal{F}$, and the theorem is proved. □

COROLLARY 1.1.– *The σ-field $\mathbb{K}_{el}$, generated by the elementary cylinders, is the smallest σ-field $\Sigma^{\mathbb{T}}$ such that for any stochastic process X, the mapping $\omega \mapsto \{X_t(\Omega), t \in \mathbb{T}\}$ is $\mathcal{F}$-$\Sigma^{\mathbb{T}}$-measurable.*

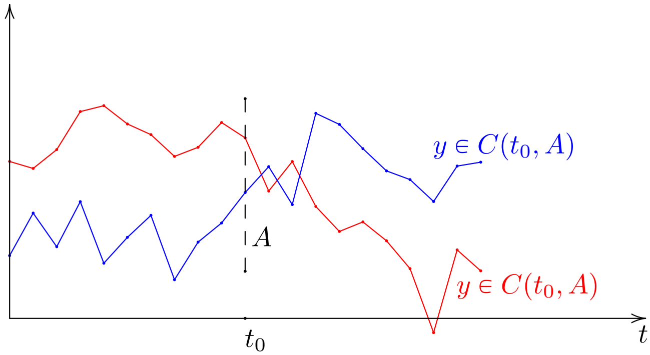


Figure 1.4. *Trajectories that belong to elementary cylinder. For a color version of the figure, see www.iste.co.uk/mishura/stochasticprocesses.zip*

1.3. Finite-dimensional distributions of stochastic processes: consistency conditions

There are two main approaches to characterizing a stochastic process: by the properties of its trajectories and by some number-valued characteristics, e.g. by finite-dimensional distributions of the values of the process. Of course, these approaches are closely related; however, any of them has its own specifics. Now we shall consider finite-dimensional distributions.

1.3.1. *Definition and properties of finite-dimensional distributions*

Let $X = \{X_t, t \in \mathbb{T}\}$ be a stochastic process taking its values in the measurable space $(\mathcal{S}, \Sigma)$. For any $k \geq 0$, consider the space $\mathcal{S}^{(k)}$, that is, a Cartesian product of $\mathcal{S}$:

$$\mathcal{S}^{(k)} = \underbrace{\mathcal{S} \times \mathcal{S} \times \ldots \times \mathcal{S}}_{k} = \times_{i=1}^{k} \mathcal{S}.$$

Let the σ-field $\Sigma^{(k)}$ of measurable sets on $\mathcal{S}^{(k)}$ be generated by all products of measurable sets from Σ.

DEFINITION 1.7.– *Finite-dimensional distributions of the process X is a family of probabilities of the form*

$$\mathbf{P} = \{\mathrm{P}\{(X_{t_1}, X_{t_2}, \ldots, X_{t_k}) \in A^{(k)}\}, k \geq 1, t_i \in \mathbb{T}, 1 \leq i \leq k, A^{(k)} \in \Sigma^{(k)}\}.$$

REMARK 1.1.– Often, especially in applied problems, *finite-dimensional distributions* are defined as the following probabilities:

$$\mathbf{P}_1 = \{\mathrm{P}\{X_{t_1} \in A_1, \ldots, X_{t_k} \in A_k\}, k \geq 1, t_i \in \mathbb{T}, A_i \in \Sigma, 1 \leq i \leq k\}.$$

Since we can write

$$\mathrm{P}\{X_{t_1} \in A_1, \ldots, X_{t_k} \in A_k\} = \mathrm{P}\left\{(X_{t_1}, \ldots, X_{t_k}) \in \times_{i=1}^{k} A_i\right\},$$

and $\times_{i=1}^{k} A_i \in \Sigma^{(k)}$, the following inclusion is evident: $\mathbf{P}_1 \subset \mathbf{P}$. The inclusion is strict, because the sets of the form $\times_{i=1}^{k} A^{(i)}$ do not exhaust $\Sigma^{(k)}$ unless $k = 1$. However, below we give a result where checking some properties for $\mathbf{P}$ is equivalent to checking them for $\mathbf{P}_1$.

1.3.2. *Consistency conditions*

Let $\pi = \{l_1, \ldots, l_k\}$ be a permutation of the coordinates $\{1, \ldots, k\}$, i.e. l_i are distinct indices from 1 to k. Denote for $A^{(k)} \in \Sigma^{(k)}$ by $\pi(A^{(k)})$ the set obtained from $A^{(k)}$ by the corresponding permutation of coordinates, e.g.

$$\pi\left(\times_{i=1}^{k} A_i\right) = \times_{i=1}^{k} A_{l_i}.$$

Denote also $\pi(X_{t_1}, \ldots, X_{t_k}) = (X_{t_{i_1}}, \ldots, X_{t_{i_k}})$ the respective permutation of vector coordinates $(X_{t_1}, \ldots, X_{t_k})$. Consider several consistency conditions which finite-dimensional distributions of random processes and the corresponding characteristic functions satisfy.

Consistency conditions (A):

1) For any $1 \leq k \leq l$, any points $t_i \in \mathbb{T}$, $1 \leq i \leq l$, and any set $A^{(k)} \in \Sigma^{(k)}$

$$\mathrm{P}\{(X_{t_1}, \ldots, X_{t_k}, X_{t_{k+1}}, \ldots, X_{t_l}) \in A^{(k)} \times \mathcal{S}^{(l-k)}\}$$
$$= \mathrm{P}\{(X_{t_1}, \ldots, X_{t_k}) \in A^{(k)}\}.$$

2) For any permutation π

$$\mathrm{P}\{\pi(X_{t_1}, \ldots, X_{t_k}) \in \pi(A^{(k)})\} = \mathrm{P}\{(X_{t_1}, \ldots, X_{t_k}) \in A^{(k)}\}. \qquad [1.1]$$

REMARK 1.2.– Assume now that $\mathcal{S} = \mathbb{R}$ and consider the characteristic functions that correspond to the finite-dimensional distributions of stochastic process X. Denote

$$\psi(\lambda_1, \ldots, \lambda_k; t_1, \ldots, t_k) = \mathrm{E} \exp\left\{ i \sum_{j=1}^{k} \lambda_j X_{t_j} \right\},$$

$\lambda_j \in \mathbb{R}$, $t_j \in \mathbb{T}$. Evidently, for $\psi(\lambda_1, \ldots, \lambda_k; t_1, \ldots, t_k)$, consistency conditions can be formulated as follows.

Consistency conditions (B):

1) For any $1 \le k \le l$ and any points $t_i \in \mathbb{T}$, $1 \le i \le l$, $\lambda_i \in \mathbb{R}$, $1 \le i \le k$

$$\psi(\lambda_1, \ldots, \lambda_k, \underbrace{0, \ldots, 0}_{l-k}; t_1, \ldots, t_k, t_{k+1}, \ldots, t_l) = \psi(\lambda_1, \ldots, \lambda_k; t_1, \ldots, t_k).$$

2) For any $k \ge 1, \lambda_i \in \mathbb{R}, t_i \in \mathbb{T}, 1 \le i \le k$

$$\psi(\pi(\overline{\lambda}); \pi(\overline{t})) = \psi(\lambda_1, \ldots, \lambda_k; t_1, \ldots, t_k),$$

where $\pi(\overline{\lambda}) = (\lambda_{i_1}, \ldots, \lambda_{i_k}), \pi(\overline{t}) = (t_{i_1}, \ldots, t_{i_k})$.

From now on, we assume that $\mathcal{S}$ is a metric space with the metric ρ, and Σ is a σ-field of Borel sets of S, generated by the metric ρ. We shall use the notation $(\mathcal{S}, \rho, \Sigma)$. Sometimes, we shall omit notations Σ and ρ yet assuming that they are fixed. Note that, for any $k > 1$, the space $\mathcal{S}^{(k)}$ is a metric space, where the metric ρ_k on the space $\mathcal{S}^{(k)}$ is defined by the formula

$$\rho_k(x, y) = \sum_{i=1}^{k} \rho(x_i, y_i), \qquad [1.2]$$

and $x = (x_1, \ldots, x_k) \in \mathcal{S}^{(k)}$, $y = (y_1, \ldots, y_k) \in \mathcal{S}^{(k)}$. Moreover, we can define the σ-field $\Sigma^{(k)}$ of the Borel sets on $\mathcal{S}^{(k)}$, generated by the metric ρ_k. (Note that it coincides with the σ-field generated by products of Borel sets from $\mathcal{S}$.)

LEMMA 1.2.– Let the metric space $(\mathcal{S}, \rho, \Sigma)$ be separable and let the finite-dimensional distributions of the process X satisfy the following version of consistency conditions.

Consistency conditions $(A1)$

1) For any $1 \le k \le l$, any points $t_i \in \mathbb{T}$, $1 \le i \le l$ and any set $A^{(k)} = \times_{i=1}^{k} A_i, A_i \in \Sigma$, the following equality holds

$$\begin{aligned} &\mathrm{P}\{X_{t_1} \in A_1, \ldots, X_{t_k} \in A_k, X_{t_{k+1}} \in \mathcal{S}, \ldots, X_{t_l} \in \mathcal{S}\} \\ &\quad = \mathrm{P}\{X_{t_1} \in A_1, \ldots, X_{t_k} \in A_k\}. \end{aligned}$$

2) For any permutation $\pi = (i_1, \ldots, i_k)$,

$$\mathrm{P}\{X_{t_{i_1}} \in A_{i_1}, \ldots, X_{t_{i_k}} \in A_{i_k}\} = \mathrm{P}\{X_{t_1} \in A_1, \ldots, X_{t_k} \in A_k\}.$$

Then the finite-dimensional distributions of the process X satisfy consistency conditions (A), where $\Sigma^{(k)}$ is a σ-field of Borel sets of $\mathcal{S}^{(k)}$. Therefore, for the stochastic process with the values in a metric separable space $(\mathcal{S}, \Sigma)$, consistency conditions for the families of sets $\mathbf{P}$ and $\mathbf{P}_1$ are fulfilled simultaneously.

PROOF.– The statement follows immediately from theorem A2.2 by noting that both sides of [1.1] are probability measures, and the sets of the form $\times_{i=1}^{k} A_i$, $A_i \in \Sigma$ form a π-system generating the σ-field $\Sigma^{(k)}$. □

1.3.3. *Cylinder sets and generated σ-algebra*

DEFINITION 1.8.– *Let* $\{t_1, \ldots, t_k\} \subset \mathbb{T}$*, the set* $A^{(k)} \in \Sigma^{(k)}$. Cylinder set *with base* $A^{(k)}$ *over the points* $\{t_1, \ldots, t_k\}$ *is the set of the form*

$$C(t_1, \ldots, t_k, A^{(k)}) = \{y = y(t) \in S^{\mathbb{T}} : (y(t_1), \ldots, y(t_k)) \in A^{(k)}\}.$$

REMARK 1.3.– If $A^{(k)}$ is a rectangle in $\mathcal{S}^{(k)}$ of the form $A^{(k)} = \times_{i=1}^{k} A_i$, then $C(t_1, \ldots, t_k, A^{(k)})$ is the intersection of the corresponding elementary cylinders:

$$C(t_1, \ldots, t_k, A^{(k)}) = \{y = y(t) \in \mathcal{S}^{\mathbb{T}} : y(t_i) \in A_i, 1 \le i \le k\} = \bigcap_{i=1}^{k} C(t_i, A_i).$$

Denote by K_{cyl} the family of all cylinder sets.

LEMMA 1.3.–

1) The family of all cylinder sets K_{cyl} is an algebra on the space $\mathcal{S}^{\mathbb{T}}$.

2) If the set $\mathcal{S}$ contains at least two points, and the set $\mathbb{T}$ is infinite, then the family of all cylinder sets is not a σ-algebra.

PROOF.– 1) Let $C(t_1^1, \ldots, t_k^1, A^{(k)})$ and $C(t_1^2, \ldots, t_m^2, B^{(m)})$ be two cylinder sets, possibly with different bases and over different sets of points. We write them as cylinder sets with different bases but over the same set of points, namely over the set $\{t_1, \ldots, t_l\} = \{t_1^1, \ldots, t_k^1\} \cup \{t_1^2, \ldots, t_m^2\}$. Specifically, define projections

$$p_1(x_1, \ldots, x_l) = (x_i, t_i \in \{t_1^1, \ldots, t_k^1\}), p_2(x_1, \ldots, x_l) = (x_i, t_i \in \{t_1^2, \ldots, t_m^2\}).$$

Then

$$C\left(t_1^1,\ldots,t_k^1,A^{(k)}\right)=C\left(t_1,\ldots,t_l,p_1^{-1}(A^{(k)})\right),$$
$$C\left(t_1^2,\ldots,t_m^2,B^{(m)}\right)=C\left(t_1,\ldots,t_l,p_2^{-1}(B^{(m)})\right),$$

so the set

$$C\left(t_1^1,\ldots,t_k^1,A^{(k)}\right)\cup C\left(t_1^2,\ldots,t_m^2,B^{(m)}\right)$$
$$=C\left(t_1,\ldots,t_l,p_1^{-1}(A^{(k)})\cup p_2^{-1}(B^{(m)})\right)$$

belongs to K_{cyl}, because

$$p_1^{-1}(A^{(k)})\cup p_2^{-1}(B^{(m)})\in\Sigma^{(l)}.$$

Similarly, the set

$$C\left(t_1^1,\ldots,t_k^1,A^{(k)}\right)\setminus C\left(t_1^2,\ldots,t_m^2,B^{(m)}\right)$$
$$=C\left(t_1,\ldots,t_l,p_1^{-1}(A^{(k)})\setminus p_2^{-1}(B^{(m)})\right)$$

belongs to K_{cyl}. Finally, for any $t_0\in\mathbb{T}$

$$\mathcal{S}^{\mathbb{T}}=\{y=y(t):y(t_0)\in\mathcal{S}\}\in K_{cyl},$$

whence it follows that the family of cylinder sets K_{cyl} is an algebra on the space $\mathcal{S}^{\mathbb{T}}$.

2) Let $\mathcal{S}$ contain at least two different points, say, s_1 and s_2, and let $\mathbb{T}$ be infinite. Then $\mathbb{T}$ contains a countable set of points $\{t_n, n\geq 1\}$. The set

$$\left(\bigcup_{i=1}^{\infty}C(t_{2i},\{s_1\})\right)\cup\left(\bigcup_{i=1}^{\infty}C(t_{2i+1},\{s_2\})\right)$$

is not a cylinder set because it cannot be described in terms of any finite set of points from $\mathbb{T}$. It means that, in this case, the family of cylinder sets K_{cyl} is not a σ-field on the space $\mathcal{S}^{\mathbb{T}}$. □

Denote by $\mathbb{K}_{cyl}$ the σ-algebra generated by the family K_{cyl} of cylinder sets: $\mathbb{K}_{cyl}=\sigma(K_{cyl})$.

LEMMA 1.4.– *For any* $k \geq 1$, $\mathbb{K}_{cyl} = \mathbb{K}_{el}$.

PROOF.– Evidently, σ-algebra $\mathbb{K}_{el} = \sigma(K_{el}) \subset \mathbb{K}_{cyl} = \sigma(K_{cyl})$, because any elementary cylinder is a cylinder set. Vice versa, for a fixed subset $\{t_1, \ldots, t_k\} \subset \mathbb{T}$, define the family

$$\mathcal{K} = \left\{ B \in \Sigma^{(k)} : C(t_1, \ldots, t_k, B) \in \mathbb{K}_{el} \right\}.$$

This is clearly a σ-algebra, which contains sets of the form $A_1 \times \cdots \times A_k$, $A_i \in \Sigma$, and therefore, $\mathcal{K} = \Sigma^{(k)}$. Consequently, we have $\mathbb{K}_{el} \supset K_{cyl}$, whence $\mathbb{K}_{el} \supset \mathbb{K}_{cyl}$, as required. □

1.3.4. *Kolmogorov theorem on the construction of a stochastic process by the family of probability distributions*

If some stochastic process is defined, then we know in particular its finite-dimensional distributions. We can say that a family of finite-dimensional distributions corresponds to a stochastic process. Consider now the opposite question. Namely, let us have a family of probability distributions. Is it possible to construct a probability space and a stochastic process on this space so that the family of probability distributions is a family of finite-dimensional distributions of the constructed process? Let us formulate this problem more precisely.

Let $(\mathcal{S}, \rho, \Sigma)$ be a metric space with Borel σ-field and $\mathbb{T}$ be a parameter set, and consider the family of functions

$$\left(\mathrm{P}\{t_1, \ldots, t_n, B^{(n)}\}, n \geq 1, t_i \in \mathbb{T}, 1 \leq i \leq n, B^{(n)} \in \Sigma^{(n)} \right), \qquad [1.3]$$

where $\Sigma^{(n)}$ is a Borel σ-field on $S^{(n)}$. Assume that, for any $t_1, \ldots, t_n \in \mathbb{T}$, the function $\mathrm{P}\{t_1, \ldots, t_n, \cdot\}$ is a probability measure on $\Sigma^{(n)}$.

THEOREM 1.2.– *[A.N. Kolmogorov] If* $(\mathcal{S}, \Sigma)$ *is a complete separable metric space with Borel* σ*-field* Σ*, and family* [1.3] *satisfies consistency conditions* $(A1)$*, then there exists a probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ *and stochastic process* $X = \{X_t, t \in \mathbb{T}\}$ *on this space and with the phase space* $(\mathcal{S}, \Sigma)$*, for which* [1.3] *is the family of its finite-dimensional distributions.*

PROOF.– We divide the proof into several steps. *Step 1.* At first, recall that according to lemma 1.2, for the separable metric space $(\mathcal{S}, \rho, \Sigma)$ conditions (A) and $(A1)$ are equivalent and continue to deal with the condition (A). Put $\Omega = \mathcal{S}^{\mathbb{T}}$, $\mathcal{F} = \mathbb{K}_{cyl}$.

Recall also that $\mathbb{K}_{cyl} = \sigma(K_{cyl})$, where K_{cyl} is the algebra of cylinder sets. Let C be the arbitrary cylinder set, and let it be represented as

$$C = C(t_1, \dots, t_n, A^{(n)}).$$

Construct the following function defined on the sets of K_{cyl}:

$$\mathrm{P}'\{C\} = \mathrm{P}\{t_1, \dots, t_n, A^{(n)}\}.$$

Note that, generally speaking, the cylinder set C admits non-unique representation. In particular, it is possible to rearrange points $t_1, \dots, t_n$ and to "turn" the base $A^{(n)}$ accordingly. Moreover, it is possible to append any finite number of points $s_1, \dots, s_m$ and replace $A^{(n)}$ with $A^{(n)} \times \mathcal{S}^{(m)}$. However, consistency conditions guarantee that $\mathrm{P}'\{C\}$ will not change under these transformations; therefore, function $\mathrm{P}'\{\cdot\}$ on K_{cyl} is well defined.

Step 2. Now we want to prove that $\mathrm{P}'\{\cdot\}$ is an additive function on K_{cyl}. To this end, consider two disjoint sets

$$C_1 = C\left(t_1^1, \dots, t_n^1, A^{(n)}\right) \text{ and } C_2 = C\left(t_1^2, \dots, t_m^2, B^{(m)}\right),$$

and let $\{t_1, \dots, t_l\} = \{t_1^1, \dots, t_n^1\} \cup \{t_1^2, \dots, t_m^2\}$. Defining projection operators $p_1(x_1, \dots, x_l) = (x_i, t_i \in \{t_1^1, \dots, t_n^1\})$, $p_2(x_1, \dots, x_l) = (x_i, t_i \in \{t_1^2, \dots, t_m^2\})$, we have

$$\begin{aligned} &C\left(t_1^1, \dots, t_n^1, A^{(n)}\right) \cup C\left(t_1^2, \dots, t_m^2, B^{(m)}\right) \\ &= C\left(t_1, \dots, t_l, p_1^{-1}(A^{(n)}) \cup p_2^{-1}(B^{(m)})\right). \end{aligned}$$

The bases $p_1^{-1}(A^{(n)})$ and $p_2^{-1}(B^{(m)})$ are disjoint, since the sets C_1 and C_2 are, so it follows from the fact that $\mathrm{P}\{t_1^1, \dots, t_n^1, \cdot\}$ is a measure with respect to the sets $A^{(n)}$ and also from consistency conditions that the following equalities hold:

$$\begin{aligned} &\mathrm{P}\{C_1 \cup C_2\} = \mathrm{P}\{t_1, \dots, t_l, p_1^{-1}(A^{(n)}) \cup p_2^{-1}(B^{(m)})\} \\ &= \mathrm{P}\{t_1, \dots, t_l, p_1^{-1}(A^{(n)})\} + \mathrm{P}\{t_1, \dots, t_l, p_2^{-1}(B^{(m)})\} \\ &= \mathrm{P}\{t_1^1, \dots, t_n^1, A^{(n)}\} + \mathrm{P}\{t_1^2, \dots, t_m^2, B^{(m)}\} = \mathrm{P}'\{C_1\} + \mathrm{P}'\{C_2\}. \end{aligned}$$

Step 3. Now we shall prove that P' is a countably additive function on K_{cyl}. Let the sets $\{C, C_n, n \geq 1\} \subset K_{cyl}, C_n \cap C_k = \emptyset$ for any $n \neq k$, and moreover, let $C = \bigcup_{n=1}^{\infty} C_n$. It is sufficient to prove that

$$\mathrm{P}'\{C\} = \sum_{n=1}^{\infty} \mathrm{P}'\{C_n\}. \quad [1.4]$$

Let us establish [1.4] in the following equivalent form. Denote $D_n = \bigcup_{k=n}^{\infty} C_k$. Then $D_1 \supset D_2 \supset \ldots$, and $\bigcap_{n=1}^{\infty} D_n = \emptyset$. Besides this, it follows from the additivity of P' on K_{cyl} that

$$\mathrm{P}'\{C\} = \sum_{k=1}^{n-1} \mathrm{P}'\{C_k\} + \mathrm{P}'\{D_n\}.$$

Therefore, in order to prove [1.4], it is sufficient to establish that

$$\lim_{n\to\infty} \mathrm{P}'\{D_n\} = 0.$$

Since the sets D_n do not increase, this limit exists. By contradiction, let

$$\lim_{n\to\infty} \mathrm{P}'\{D_n\} = \alpha > 0.$$

Without any loss of generality, we can assume that the set of points, over which D_n is defined, is growing with n. Let the points be $\{t_1, \ldots, t_{k_n}\}$, and $B_n \in \Sigma^{(k_n)}$ be the base of D_n. In other words, let $D_n = C(t_1, \ldots, t_{k_n}, B_n)$. Taking into account the fact that $\mathcal{S}$ is a completely separable metric space, we get from theorem A1.1 that the space $\mathcal{S}^{(k_n)}$ is also a completely separable metric space. Therefore, according to theorem A1.2, there exists a compact set $K_n \in \Sigma^{(k_n)}$, such that $K_n \subseteq B_n$ and

$$\mathrm{P}\{t_1, \ldots, t_{k_n}, B_n \backslash K_n) < \frac{\alpha}{2^{n+1}}.$$

Now construct a cylinder set Q_n with the base K_n over the points $t_1, \ldots, t_{k_n}$ and consider the intersection $G_n = \bigcap_{i=1}^{n} Q_i$. Let M_n be the base of the set G_n. The sets G_n are non-increasing, and their bases M_n are compact. Indeed, the set M_n is an intersection of the bases of the sets $Q_i,\ 1 \le i \le n$, but in the case where all of them are presented as the cylinder sets over the points $\{t_1, \ldots, t_{k_n}\}$. With such a record, the initial bases K_i of the sets Q_i take the form $K_i \times \mathcal{S}^{(k_n - k_i)}$ and thus remain closed although perhaps no longer compact, while the set K_n is compact. An intersection of closed sets, one of which is compact, is a compact set as well; therefore, M_n is a compact set. The fact that G_n are non-increasing means that any element of $G_{n+p}, p > 0$ belongs to G_n. Their bases M_n are non-increasing in the sense that, for any point $(y(t_1), \ldots, y(t_{k_{n+p}})) \in M_{n+p}$, its "beginning" $(y(t_1), \ldots, y(t_{k_n})) \in M_n$. Now let us prove that the sets G_n and consequently M_n are non-empty. Indeed, it follows from the additivity of P' that

$$\mathrm{P}'\{D_n \backslash G_n\} = \mathrm{P}'\{D_n \setminus \bigcap_{i=1}^{n} Q_i\} = \mathrm{P}'\left\{\bigcup_{i=1}^{n} (D_n \setminus Q_i)\right\} \le \sum_{i=1}^{n} \mathrm{P}'\left\{D_n \setminus Q_i\right\}$$

$$\le \sum_{i=1}^{n} \mathrm{P}'\{D_i \setminus Q_i\} = \sum_{i=1}^{n} \mathrm{P}\{t_1, \ldots, t_{k_i}, B_i \setminus K_i\} \le \sum_{i=1}^{\infty} \frac{\alpha}{2^{i+1}} = \frac{\alpha}{2}.$$

It means that $P'\{G_n\} \geq \frac{\alpha}{2}$, whence the sets G_n are non-empty. In turn, it means that $M_n \neq \emptyset$, and we can choose the points $(y_1^{(n)}, \ldots, y_{l_n}^{(n)}) \in M_n$, and moreover, l_n is non-decreasing in n. Take the sequence $(y_1^{(n)}, \ldots, y_{l_n}^{(n)})$ and consider its "beginning" $(y_1^{(n)}, \ldots, y_{l_1}^{(n)})$. As has just been said, the sequence $(y_1^{(n)}, \ldots, y_{l_1}^{(n)}) \in M_1$. Therefore, it contains a convergent subsequence, and then any sequence $\{y_k^{(n)}, n \geq 1\}$ for $1 \leq k \leq l_1$ contains a convergent subsequence. Take $(y_1^{(n)}, \ldots, y_{l_2}^{(n)}) \in M_2$; at once, any "column" $\{y_k^{(n)}, n \geq 1\}$ for $1 \leq k \leq l_2$ contains a convergent subsequence. Finally, any "column" $\{y_k^{(n)}, n \geq 1\}$ for $k \geq 1$ contains a convergent subsequence. Denote by $y_k^{(0)}$ the limit of convergent subsequence $\{y_k^{(n_j)}, j \geq 1\}$. Applying the diagonal method, we can choose, for any n, a convergent subsequence of vectors

$$(y_1^{(n_j)}, y_2^{(n_j)}, \ldots, y_{k_n}^{(n_j)}) \to (y_1^{(0)}, y_2^{(0)}, \ldots, y_{k_n}^{(0)}).$$

Since all the points $(y_1^{(n_j)}, y_2^{(n_j)}, \ldots, y_{k_n}^{(n_j)}) \in M_n$ and the sets M_n are closed, we get that $(y_1^{(0)}, \ldots, y_{k_n}^{(0)}) \in M_n$. Now, define a function $y = y(t) \in \mathcal{S}^{\mathbb{T}}$ by the formula $y(t_k) = y_k^{(0)}, k \geq 1$ and define $y(t)$ in an arbitrary manner in the remaining points from $\mathbb{T}$. Then arbitrary vector $(y(t_1), \ldots, y(t_{k_n})) \in M_n$; therefore, for any $n \geq 1$, the function $y \in G_n \subset D_n$. This means that $\bigcap_{n=1}^{\infty} D_n \neq \emptyset$, which contradicts to the construction of sets D_n. It means that the assumption $\lim_{n\to\infty} P'(D_n) > 0$ leads to the contradiction and so is false. Therefore, P' is a countably additive function on K_{cyl}, and consequently, P' is a probability measure on the algebra K_{cyl}.

Step 4. According to the theorem on the extension of the measure from algebra to generated σ-algebra, there exists the unique probability measure P, that is, the extension of the measure P' from K_{cyl} to $\mathbb{K}_{cyl}$. Construct a stochastic process $X = \{X_t, t \in \mathbb{T}\}$ on $(\Omega, \mathcal{F}, P)$ in the following way (recall that the elements $\omega \in \Omega$ are presented by functions $y \in \mathcal{S}^{\mathbb{T}}$):

$$X_t(\omega) = \omega(t) := y(t).$$

We first check that $X = \{X_t, t \in \mathbb{T}\}$ is indeed a stochastic process. For any set $A \in \Sigma$ and for any $t_0 \in \mathbb{T}$, we have that

$$\begin{aligned} X_{t_0}^{-1}(A) &= \{\omega : X_{t_0}(\omega) \in A\} \\ &= \{y = y(t) : y(t_0) \in A\} = C(t_0, A) \in K_{cyl} \subset \mathbb{K}_{cyl} = \mathcal{F}. \end{aligned}$$

Further,

$$\begin{aligned} &P\{(X_{t_1}, \ldots, X_{t_k}) \in A^{(k)}\} = P\{(y(t_1), \ldots, y(t_k)) \in A^{(k)}\} \\ &= P\{C(t_1, \ldots, t_k, A^{(k)})\} = P'\{C(t_1, \ldots, t_k, A^{(k)})\} = P\{t_1, \ldots, t_k, A^{(k)}\}. \end{aligned}$$

Therefore, $X = \{X_t, t \in \mathbb{T}\}$ has the prescribed finite-dimensional distributions. The theorem is proved. □

In the case where $\mathcal{S} = \mathbb{R}$, the Kolmogorov theorem can be formulated as follows.

THEOREM 1.3.– *Let a family* $\psi(\lambda_1, \ldots, \lambda_k; t_1, \ldots, t_k), k \geq 1, \lambda_j \in \mathbb{R}, t_j \geq 0$ *of characteristic functions satisfy consistency conditions* (B)*. Then there exists a probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ *and a real-valued stochastic process* $X = \{X_t, t \geq 0\}$ *for which* $\mathrm{E}\exp\{i\sum_{j=1}^{k}\lambda_j X_{t_j}\} = \psi(\lambda_1, \ldots, \lambda_k; t_1, \ldots, t_k)$.

1.4. Properties of σ-algebra generated by cylinder sets. The notion of σ-algebra generated by a stochastic process

Let $\mathbb{T} = \mathbb{R}_+ = [0, +\infty)$, $(\mathcal{S}, \Sigma)$ be a measurable space and $X = \{X_t, t \in \mathbb{T}\}$ be an $\mathcal{S}$-valued stochastic process. Consider the standard σ-algebra $\mathbb{K}_{cyl}$ generated by cylinder sets, and for any finite or countable set of points, $\mathbb{T}' = \{t_n, n \geq 1\} \subset \mathbb{T}$ forms the algebra $K_{cyl}(\{t_n, n \geq 1\})$ of cylinder sets in the following way: $A \in K_{cyl}(\{t_n, n \geq 1\})$ if and only if there exists a subset $\{t_{n_1}, \ldots, t_{n_k}\} \subset \{t_n, n \geq 1\}$ and $B^{(k)} \in \Sigma^{(k)}$ such that

$$A = C'(t_{n_1}, \ldots, t_{n_k}, B^{(k)}) := \{y : \mathbb{T}' \to \mathcal{S} : (y(t_{n_1}), \ldots, y(t_{n_k})) \in B^{(k)}\}.$$

Consider the generated σ-algebra $\mathbb{K}_{cyl}(\{t_n, n \geq 1\})$. We shall prove the statement that describes $\mathbb{K}_{cyl}$ in terms of the countable collections of points from $\mathbb{T}$.

LEMMA 1.5.– *The set* $A \subset \mathcal{S}^{\mathbb{T}}$ *belongs to* $\mathbb{K}_{cyl}$ *if and only if there exists a sequence of points* $\{t_n, n \geq 1\} \subset \mathbb{T}$ *and a set* $B \in \mathbb{K}_{cyl}(\{t_n, n \geq 1\})$*, such that the following equality holds:*

$$A = C(\{t_n, n \geq 1\}, B) := \{y \in \mathcal{S}^{\mathbb{T}} : (y(t_n), n \geq 1) \in B\}. \qquad [1.5]$$

PROOF.– Let C be any cylinder set from algebra K_{cyl},

$$C = \{y \in \mathcal{S}^{\mathbb{T}} : (y(z_1), \ldots, y(z_m)) \in B^{(m)} \subset \mathcal{S}^{(m)}\}.$$

Then C admits the representation [1.5] if we consider the arbitrary sequence of points $\{t_n, n \geq 1\}$ such that $t_n = z_n, 1 \leq n \leq m$ and $B = A(B^{(m)})$. Therefore, if we denote by $\mathbb{K}$ the sets from $\mathbb{K}_{cyl}$ that admit the representation [1.5], then $K_{cyl} \subset \mathbb{K}$. Let us establish now that $\mathbb{K}$ is a σ-algebra. Indeed, $S^{\mathbb{T}} \in \mathbb{K}$, because we can take an arbitrary sequence

$$\mathbb{T}' = \{t_n, n \geq 1\} \text{ and } B = \mathcal{S}^{\mathbb{T}'} := \times_{n=1}^{\infty}\mathcal{S} \in \mathbb{K}_{cyl}(\mathbb{T}'),$$

and get that the set $\mathcal{S}^{\mathbb{T}}$ has a form $\mathcal{S}^{\mathbb{T}} = \{y \in S^{\mathbb{T}} : (y(t_n), n \geq 1) \in B = S^{\mathbb{T}'}\}$, admitting with evidence the representation [1.5]. Further, if $A_1, A_2 \in \mathbb{K}$, they are defined over the sequences of points $\mathbb{T}^1 = \{t_n^1, n \geq 1\}$ and $\mathbb{T}^2 = \{t_n^2, n \geq 1\}$ and have the bases B_1, B_2, correspondingly. Then, we can consider these sets as defined over the same sequence of points, setting $\mathbb{S} = \{t_n^1, n \geq 1\} \cup \{t_n^2, n \geq 1\}$ and introducing the maps $p_i : y \in \mathcal{S}^{\mathbb{S}} \mapsto y|_{\mathbb{T}^i} \in S^{\mathbb{T}^i}$, $i = 1, 2$. Then $A_i = C(\mathbb{S}, p_i^{-1}(B_i))$, $i = 1, 2$. The bases $p_i^{-1}(B_i)$, $i = 1, 2$, are measurable, since the maps p_i, $i = 1, 2$ are measurable (even continuous in the topology of pointwise convergence). Therefore,

$$A_1 \setminus A_2 = C(\mathbb{S}, p_1^{-1}(B_1) \setminus p_2^{-1}(B_2)) \in \mathbb{K}.$$

Similarly, if $\{A_r, r \geq 1\} \subset \mathbb{K}$, and they are defined over the sequences of points $\mathbb{T}^r = \{t_n^r, n \geq 1\}$ and bases $B_r, r \geq 1$, we can define $\mathbb{T}^0 = \bigcup_{r=0}^{\infty} \mathbb{T}^r$ and $p_r : y \in \mathcal{S}^{\mathbb{T}^0} \mapsto y|_{\mathbb{T}^r} \in S^{\mathbb{T}^r}$, $r \geq 1$, so that

$$\bigcup_{r=1}^{\infty} A_r = C\left(\mathbb{T}^0, \bigcup_{r=1}^{\infty} p_r^{-1}(B_r)\right).$$

Thus, we have that $\mathbb{K}$ is a σ-algebra that contains K_{cyl}, i.e. $\mathbb{K} \supset \mathbb{K}_{cyl}$, but $\mathbb{K} \subset \mathbb{K}_{cyl}$ by the definition. It means that $\mathbb{K} = \mathbb{K}_{cyl}$, whence the proof follows. □

DEFINITION 1.9.– *The σ-algebra, generated by the process X is the family of sets* $\mathcal{F}^X = \{X^{-1}(A), A \in \mathbb{K}_{cyl}\}$.

REMARK 1.4.– It follows from the properties of pre-images that for any σ-algebra $\mathcal{A}$, the family of sets $\{X^{-1}, A \in \mathcal{A}\}$ is a σ-algebra; therefore, definition 1.9 is properly formulated.

LEMMA 1.6.– $\mathcal{F}^X = \sigma\{X^{-1}(A), A \in K_{cyl}\}$

PROOF.– Denote $\mathcal{F}_1^X = \{X^{-1}(A), A \in K_{cyl}\}$. On the one hand, since $\mathcal{F}^X$ is a σ-algebra and $\mathcal{F}^X$ contains all pre-images $X^{-1}(A)$ under mapping $A \in K_{cyl}$, then $\mathcal{F}^X \supset \mathcal{F}_1^X$. On the other hand, consider $\mathcal{F}_1^X$ and note that the mapping X is $\mathcal{F}_1^X$-K_{cyl}-measurable; therefore, according to lemma 1.1, the mapping X is $\mathcal{F}_1^X$-$\mathbb{K}_{cyl}$-measurable, i.e. $X^{-1}(A) \in \mathcal{F}_1^X$ for any $A \in \mathbb{K}_{cyl}$. It means that $\mathcal{F}_1^X \supset \mathcal{F}^X$, and the lemma is proved. □

The following fact is a consequence of lemma 1.4.

COROLLARY 1.2.– *The σ-algebra generated by a stochastic process X is the smallest σ-algebra containing all the sets of the form*

$$\{\omega \in \Omega : X(t_1, \omega) \in A_1, ..., X(t_k, \omega) \in A_k\}, A_i \in \Sigma, t_i \in \mathbb{T}, 1 \leq i \leq k, k \geq 1.$$

2

Stochastic Processes with Independent Increments

Throughout this chapter, the phase space $S = \mathbb{R}$ or $\mathbb{R}^d$. Therefore, we consider real-valued or vector-valued stochastic processes. Parameter set is assumed to be $\mathbb{T} = [0, +\infty)$.

2.1. Existence of processes with independent increments in terms of incremental characteristic functions

Recall that, according to definition 1.5, stochastic process $X = \{X_t, t \geq 0\}$ is a process with independent increments if for any $n \geq 1$ and for any collection of points $0 \leq t_0 < t_1 < \ldots < t_n$, the random variables $\{X_{t_0}, X_{t_1} - X_{t_0}, \ldots, X_{t_n} - X_{t_{n-1}}\}$ are mutually independent.

Let $X = \{X_t, t \geq 0\}$ be a real-valued stochastic process with independent increments. Consider the characteristic function of the increment

$$\varphi(\lambda; s, t) := \mathrm{E} \exp\{i\lambda(X_t - X_s)\}, 0 \leq s < t, \lambda \in \mathbb{R}.$$

Evidently, the following equality holds: for any $\lambda \in \mathbb{R}, 0 \leq s < u < t$

$$\varphi(\lambda; s, u)\varphi(\lambda; u, t) = \varphi(\lambda; s, t). \qquad [2.1]$$

Now we shall prove that equality [2.1] characterizes, in some sense, stochastic processes with independent increments. Recall that the function $\psi(\lambda) : \mathbb{R} \to \mathbb{C}$ is a characteristic function of some random variable if it is continuous with $\psi(0) = 1$ and non-negatively definite, that is, for any $k \geq 1$ and any $z_j \in \mathbb{C}, \lambda_j \in \mathbb{R},\ 1 \leq j \leq k$, we have that $\sum_{j,r=1}^{k} z_j \overline{z}_r \psi(\lambda_j - \lambda_r) \geq 0$.

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

THEOREM 2.1.– *Let us be given arbitrary characteristic function* $\psi(\lambda)$, $\lambda \in \mathbb{R}$, *and a family of characteristic functions*

$$\{\varphi(\lambda; s,t), 0 \le s < t < \infty, \lambda \in \mathbb{R}\},$$

and let the latter family satisfy equality [2.1]. *Then there exists a probability space* $(\Omega, \mathcal{F}, \mathrm{P})$ *and a stochastic process* $X = \{X_t, t \ge 0\}$ *with independent increments on this space, for which* $\mathrm{E}\exp\{i\lambda X_0\} = \psi(\lambda)$, *and* $\mathrm{E}\exp\{i\lambda(X_t - X_s)\} = \varphi(\lambda; s,t)$. *Moreover, all finite-dimensional distributions of such* X *are uniquely determined by* ψ *and* φ.

REMARK 2.1.– Condition [2.1] does not imply in general that increments of a process are independent; moreover, even two increments $X_t - X_u$ and $X_u - X_s$ can be dependent and still satisfy [2.1]. The above theorem establishes the existence of a process with independent increments such that the characteristic function of them satisfies [2.1], but in general there can be other processes with the same incremental characteristic function.

PROOF.– Let $\{Y_t, t \ge 0\}$ be a process with independent increments, and $\varphi(\lambda; s,t) = \mathrm{E}\exp\{i\lambda(Y_t - Y_s)\}$. For any $0 \le t_0 < t_1 < \ldots < t_n$, the following equality holds:

$$(Y_{t_0}, \ldots, Y_{t_n})^{tr} = A(Y_{t_0}, Y_{t_1} - Y_{t_0}, \ldots, Y_{t_n} - Y_{t_{n-1}})^{tr},$$

where matrix $A = \begin{pmatrix} 1 & 0 & 0 \cdots 0 \\ 1 & 1 & 0 \cdots 0 \\ \vdots & \ddots & \vdots \ \ldots \ \vdots \\ 1 & \cdots & 1 \ \ldots \ 1 \end{pmatrix}$. Denote $\overline{\lambda} = (\lambda_0, \ldots, \lambda_n)$, $\overline{Y} = (Y_{t_0}, \ldots, Y_{t_n})$, $\overline{Z} = (Y_{t_0}, Y_{t_1} - Y_{t_0}, \ldots, Y_{t_n} - Y_{t_{n-1}})$ and write

$$\begin{aligned} &\mathrm{E}\exp\{i(\lambda_0 Y_{t_0} + \lambda_1 Y_{t_1} + \ldots + \lambda_n Y_{t_n})\} = \mathrm{E}\exp\{i(\overline{\lambda}, \overline{Y})\} \\ &= \mathrm{E}\exp\{i(\overline{\lambda}, A\overline{Z}^{tr})\} = \mathrm{E}\exp\{i(A^{tr}\overline{\lambda}^{tr}, \overline{Z})\} \\ &= \mathrm{E}\exp\{i((\lambda_0 + \ldots + \lambda_n)Y_0 + (\lambda_0 + \ldots + \lambda_n)(Y_{t_0} - Y_0) \\ &\quad + (\lambda_1 + \ldots + \lambda_n)(Y_{t_1} - Y_{t_0}) \ldots + (\lambda_{n-1} + \lambda_n)(Y_{t_{n-1}} - Y_{t_{n-2}}) \\ &\quad + \lambda_n(Y_{t_n} - Y_{t_{n-1}}))\} = \psi(\lambda_0 + \ldots + \lambda_n)\varphi(\lambda_0 + \ldots + \lambda_n; 0, t_0) \\ &\quad \times \varphi(\lambda_1 + \ldots + \lambda_n; t_0, t_1) \ldots \varphi(\lambda_{n-1} + \lambda_n; t_{n-1}, t_n)\varphi(\lambda_n; t_{n-1}, t_n). \end{aligned} \qquad [2.2]$$

Thus, equality [2.2] is necessary in order for process Y to have independent increments. Vice versa, if equality [2.2] is satisfied, then the process Y has independent increments. Now, having the family $\{\varphi(\lambda; s,t)\}$ of characteristic functions and the initial characteristic function ψ, define the characteristic functions of finite-dimensional distributions by the formula

$$\varphi(\lambda_0, \ldots, \lambda_n; t_0, \ldots, t_n) = \psi(\lambda_0 + \ldots + \lambda_n)\varphi(\lambda_0 + \ldots + \lambda_n; 0, t_0) \times \varphi(\lambda_1 + \ldots + \lambda_n; t_0, t_1) \ldots \varphi(\lambda_n; t_{n-1}, t_n). \quad [2.3]$$

Let $0 \leq t_0 < t_1 < \ldots < t_n$, and $\pi(\overline{\lambda}) = (\lambda_{i_1}, \ldots, \lambda_{i_n})$ be any permutation. For the respective permutation $\pi(\bar{t}) = (t_{i_1}, \ldots, t_{i_n})$, put $\mathrm{E}\exp\{i(\pi(\overline{\lambda}), \pi(\bar{t}))\} = \mathrm{E}\exp\{i(\overline{\lambda}, \bar{t})\}$. Then the first consistency condition from the couple (B) (section 1.3.2) of the conditions for the characteristic functions is fulfilled. Now verify the second one. Consider any $1 \leq k \leq n$ and let $0 \leq t_0 < t_1 < \ldots < t_k < t_k^* < t_{k+1} < \ldots < t_n$. Then

$$\begin{aligned}&\varphi(\lambda_0, \ldots, \underset{k+1}{0}, \ldots, \lambda_n; t_0, \ldots, t_k, t_k^*, t_{k+1}, \ldots, t_n)\\ &= \psi(\lambda_0 + \ldots + \lambda_n)\varphi(\lambda_0 + \ldots + \lambda_n; 0, t_0) \qquad [2.4]\\ &\quad \times \varphi(\lambda_1 + \ldots + \lambda_n; t_0, t_1) \ldots \varphi(0 + \lambda_{k+1} + \ldots + \lambda_n, t_k, t_k^*)\\ &\quad \times \varphi(\lambda_{k+1} + \ldots + \lambda_n; t_k^*, t_{k+1}) \ldots \varphi(\lambda_n; t_{n-1}, t_n).\end{aligned}$$

It follows immediately from equality [2.1] that

$$\varphi(0 + \lambda_{k+1} + \ldots + \lambda_n; t_k, t_k^*)\varphi(\lambda_{k+1} + \ldots + \lambda_n; t_k^*, t_{k+1}) = \varphi(\lambda_{k+1} + \ldots + \lambda_n; t_k, t_{k+1}).$$

Therefore, applying formula [6.7], we get that the right-hand side of [2.4] equals

$$\varphi(\lambda_0, \ldots, \lambda_n; t_0, \ldots, t_n).$$

If $t_{n+1} > t_n$, then $\varphi(0; t_n, t_{n+1}) = 1$. It means that

$$\begin{aligned}&\varphi(\lambda_0, \ldots, \lambda_n, 0; t_0, \ldots, t_n, t_{n+1}) = \psi(\lambda_0 + \ldots + \lambda_n)\\ &\quad \times \varphi(\lambda_0 + \ldots + \lambda_n; 0, t_0)\varphi(\lambda_1 + \ldots + \lambda_n; t_0, t_1) \ldots \varphi(\lambda_n; t_{n-1} t_n)\\ &\quad \times \varphi(0; t_n, t_{n+1}) = \varphi(\lambda_0, \ldots, \lambda_n; t_0, \ldots, t_n, t_{n+1}).\end{aligned}$$

Thus, the second condition of consistency is fulfilled. Hence, by the Kolmogorov theorem in the form of theorem 1.3, there exists a probability space $(\Omega, \mathcal{F}, \mathrm{P})$, and a stochastic process $\{X_t, t \geq 0\}$ on this space, satisfying

$$\mathrm{E}\exp\{i(\lambda_0 X_{t_0} + \ldots + \lambda_n X_{t_n})\} = \varphi(\lambda_0, \ldots, \lambda_n; t_0, \ldots, t_n).$$

Then, it follows from [2.2] and [2.3] that X is a process with independent increments. The theorem is proved. $\square$

2.2. Wiener process

2.2.1. *One-dimensional Wiener process*

DEFINITION 2.1.– *A real-valued stochastic process* $W = \{W_t, t \geq 0\}$ *is called a (standard)* Wiener process *if it satisfies the following three conditions:*

1) $W_0 = 0$.

2) The process W *has independent increments.*

3) Increments $W_t - W_s$ *for any* $0 \leq s < t$ *have the Gaussian distribution with zero mean and variance* $t - s$*. In other words,* $W_t - W_s \sim \mathcal{N}(0, t-s)$.

REMARK 2.2.– The Wiener process is often called Brownian motion. Sometimes, especially in the theory of Markov processes, it is supposed that Brownian motion starts not from 0, but from some other point $x \in \mathbb{R}$.

To prove the correctness of definition 2.1, i.e. to prove that such a process exists, we apply lemma A2.2.

THEOREM 2.2.– *Definition 2.1 is valid in the sense that the Wiener process does exist.*

PROOF.– According to theorem 2.1, it is sufficient to prove that the family $\varphi(\lambda; s, t)$ of characteristic functions of the introduced process satisfies condition [2.1]. To this end, note that according to lemma A2.2,

$$\varphi(\lambda; s, t) = \mathrm{E} \exp\{i\lambda(W_t - W_s)\} = e^{-\lambda^2(t-s)/2}.$$

Therefore,

$$\varphi(\lambda; s, u)\varphi(\lambda; u, t) = e^{-\lambda^2(u-s)/2} e^{-\lambda^2(t-u)/2} = e^{-\lambda^2(t-s)/2} = \varphi(\lambda; s, t)$$

for any $0 \leq s < u < t$ and $\lambda \in \mathbb{R}$, whence the proof follows. □

Evidently, for any $t \geq 0$, $\mathrm{E}W_t = 0$. Calculate the covariance function:

$$\begin{aligned}\mathrm{Cov}(W_s, W_t) &= \mathrm{E}\, W_s W_t = \mathrm{E}\, W_{s\wedge t} W_{s\vee t} \\ &= \mathrm{E}\, W_{s\wedge t}(W_{s\vee t} - W_{s\wedge t}) + \mathrm{E}\, W_{s\wedge t}^2 = \mathrm{E}\, W_{s\wedge t}^2 = s \wedge t.\end{aligned}$$

2.2.2. *Independent stochastic processes. Multidimensional Wiener process*

Recall the definition of independent collections of sets.

DEFINITION 2.2.– *Collections of events* $\mathcal{A}_i$*,* $1 \leq i \leq n$*, are called* independent *if for any* $A_i \in \mathcal{A}_i$*,* $1 \leq i \leq n$*, the events* $A_1, A_2, \ldots, A_n$ *are independent.*

Let $X^i = \{X_t^i, t \in \mathbb{T}\}$, $1 \leq i \leq n$, be stochastic processes taking values in measurable spaces $(\mathcal{S}_i, \Sigma_i)$. They generate σ-fields $\mathcal{F}^i$, as described in section 1.4. We apply definition 2.2 to define independent stochastic processes.

DEFINITION 2.3.– *The stochastic processes* X_i*,* $1 \leq i \leq n$*, are independent if the generated* σ*-fields* $\mathcal{F}^i$*,* $1 \leq i \leq n$*, are independent.*

Recall that a class of sets $\mathcal{P}$ is called a π-system if it is closed under intersection, i.e. for any $A, B \in \mathcal{P}$, $A \cap B \in \mathcal{P}$.

LEMMA 2.1.– *Let* σ*-fields* $\mathcal{A}_i$*,* $1 \leq i \leq n$*, be generated by* π*-systems* $\mathcal{P}_i$*,* $1 \leq i \leq n$*, correspondingly. Then the* σ*-fields* $\mathcal{A}_i$*,* $1 \leq i \leq n$*, are independent if and only if* $\mathcal{P}_i$*,* $1 \leq i \leq n$*, are independent.*

PROOF.– It is enough to show that for any $A_i \in \mathcal{A}_i$, $1 \leq i \leq n$,

$$\mathrm{P}\left\{\bigcap_{i=1}^n A_i\right\} = \prod_{i=1}^n \mathrm{P}\{A_i\}. \qquad [2.5]$$

Indeed, in that case for any $B_i \in \mathcal{A}_i$, $1 \leq i \leq n$, and any $I = \{i_1, \ldots, i_k\} \subset \{1, \ldots, k\}$, we can take $A_i = B_i$, $i \in I$, and $A_i = \Omega$, $i \notin I$, to get

$$\mathrm{P}\left\{\bigcap_{i \in I} B_i\right\} = \mathrm{P}\left\{\bigcap_{i \in I}^n A_i\right\} = \prod_{i=1}^n \mathrm{P}\{A_i\} = \prod_{i \in I} \mathrm{P}\{B_i\},$$

which implies the mutual independence of B_i, $1 \leq i \leq n$.

In turn, equation [2.5] would follow from

$$\mathrm{P}\left\{\bigcap_{i=1}^n A_i\right\} = \mathrm{P}\{A_1\}\mathrm{P}\left\{\bigcap_{i=2}^n A_i\right\}, A_1 \in \mathcal{A}_1, A_2 \in \mathcal{P}_2, \ldots, A_n \in \mathcal{P}_n \qquad [2.6]$$

by applying an inductive argument. For fixed $A_2 \in \mathcal{P}_2, \ldots, A_n \in \mathcal{P}_n$ such that $\mathrm{P}\left\{\bigcap_{i=2}^n A_i\right\} \neq 0$, consider the measure

$$\mathrm{P}_1\{A_1\} = \mathrm{P}\left\{A_1 \,\middle|\, \bigcap_{i=2}^n A_i\right\} = \frac{\mathrm{P}\left\{\bigcap_{i=1}^n A_i\right\}}{\mathrm{P}\left\{\bigcap_{i=2}^n A_i\right\}}, \ A_1 \in \mathcal{A}_1.$$

This is a probability measure, which coincides with P on $\mathcal{P}_1$. Since $\mathcal{P}_1$ is a π-system, by theorem A2.2, we get [2.6] in the case where $\mathrm{P}\left\{\bigcap_{i=2}^{n} A_i\right\} \neq 0$; the equality is evident whenever $\mathrm{P}\left\{\bigcap_{i=2}^{n} A_i\right\} = 0$. Using the aforementioned inductive argument, we arrive at the statement. □

LEMMA 2.2.– *The stochastic processes X^i, $1 \le i \le n$, are independent if and only if for any $m \ge 1$, $\{t_1, \dots, t_m\} \subset \mathbb{T}$ and sets $B_{ij} \in \Sigma_i$, $1 \le i \le n$, $1 \le j \le m$, the events $\left\{X^i_{t_1} \in B_{i1}, \dots, X^i_{t_m} \in B_{im}\right\}$, $1 \le i \le n$, are independent.*

PROOF.– Consider the classes

$$\mathcal{P}_i = \left\{\left\{X^i_{t_1} \in A_1, \dots, X^i_{t_k} \in A_k\right\} \mid k \ge 1, \{t_1, \dots, t_k\} \subset \mathbb{T}, A_1, \dots, A_k \in \Sigma_i\right\}$$

for $i = 1, \dots, n$. These are obviously π-systems. Moreover, they are independent by assumption. Indeed, for any $P_i \in \mathcal{P}_i$, $1 \le i \le n$, we can write

$$P_i = \left\{X^i_{t^i_1} \in A^i_1, \dots, X^i_{t^i_k} \in A^i_{k_i}\right\} = \left\{X^i_{t_1} \in B_{i1}, \dots, X^i_{t_m} \in B_{im}\right\},$$

where $\{t_1, \dots, t_m\} = \bigcup_{i=1}^{n} \left\{t^i_1, \dots, t^i_{k_i}\right\}$, $B_{ij} = A^i_l$ if $t_j = t^i_l$ and $B_{ij} = \mathcal{S}_i$ otherwise.

Therefore, by lemma 2.1, the σ-fields $\mathcal{A}_i = \sigma(\mathcal{P}_i)$, $1 \le i \le n$, are independent. Since $\mathcal{A}_i = \mathcal{F}^i$ thanks to corollary 1.2, we get the claim. □

Now, let $\mathbb{T} = \mathbb{R}_+$ and $(S, \Sigma) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

DEFINITION 2.4.– *Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space, and let m independent real-valued Wiener processes $\{W_i(t), t \ge 0, 1 \le i \le m\}$ be defined on $(\Omega, \mathcal{F}, \mathrm{P})$.* Multidimensional Wiener process *is a vector process*

$$W(t) = \big(W_1(t), W_2(t), ..., W_m(t)\big), t \ge 0.$$

Evidently, vector of expectations is a zero one,

$$\mathrm{E}W(t) = \big(\mathrm{E}W_1(t), ..., \mathrm{E}W_m(t)\big) = 0,$$

and the matrix of covariations has a form $\mathrm{Cov}(W(t), W(s)) = (t \wedge s) E_m$, where E_m is the identity matrix of size m. For any set $A \in \mathcal{B}(\mathbb{R}^m)$

$$\mathrm{P}\{W(t) \in A\} = (2\pi t)^{-m/2} \int_A \exp\left\{-\frac{|x|^2}{2t}\right\} dx,$$

where $|x| = \left(\sum_{i=1}^{m} x_i^2\right)^{1/2}$, $dx = dx_1 \cdots dx_m$.

2.3. Poisson process

There exist at least three different approaches on how to introduce the Poisson process. The first approach is based on the characterization theorem for the processes with independent increments and gives a general definition of the (possibly non-homogeneous) Poisson process with variable intensity. The other two define the homogeneous Poisson process with constant intensity. Namely, the second approach determines the Poisson process through the transition probabilities, and this approach has rich generalizations for processes of other kinds, whereas the third one determines the Poisson process in a certain sense in a trajectory way, as the renewal process. In principle, the second and third approaches lead to the same process, although from different points of view. These three approaches are considered below.

2.3.1. *Poisson process defined via the existence theorem*

Let m be a σ-finite measure on the σ-field $\mathcal{B}(\mathbb{R}_+)$ of Borel subsets of $\mathbb{R}_+ = [0, +\infty)$, such that for any interval $(s,t] \subset \mathbb{R}_+$, $m((s,t]) \in (0, \infty)$.

DEFINITION 2.5.– *A stochastic process $N = \{N_t, t \geq 0\}$ is called a Poisson process with intensity measure m, if it satisfies three conditions:*

1) $N_0 = 0$.

2) N is a process with independent increments.

3) The increments $N_t - N_s$ for $0 \leq s < t$ have a Poisson distribution with parameter $m((s,t])$.

To prove the validity of definition 2.5, we apply lemma A2.4.

THEOREM 2.3.– *Definition 2.5 is valid in the sense that the Poisson process does exist.*

PROOF.– According to theorem 2.1, it is sufficient to establish that the family of characteristic functions $\varphi(\lambda; s,t)$ satisfies the condition [2.1]. However, according to lemma A2.4,

$$\varphi(\lambda; s,t) = \mathrm{E}\exp\{i\lambda(N_t - N_s)\} = e^{m((s,t])(e^{i\lambda}-1)},$$

therefore,

$$\varphi(\lambda; s,u)\varphi(\lambda; u,t) = e^{m((s,u])(e^{i\lambda}-1)} e^{m((u,t])(e^{i\lambda}-1)}$$
$$= e^{m((s,t])(e^{i\lambda}-1)} = \varphi(\lambda; s,t)$$

for any $0 \leq s < u < t$ and $\lambda \in \mathbb{R}$, whence the proof follows. □

REMARK 2.3.– If $m((s,t]) = \lambda(t-s)$, i.e. the measure m is proportional to the Lebesgue measure in the real line, then the Poisson process with such intensity measure is called the homogeneous Poisson process with intensity λ. Otherwise, the Poisson process is called inhomogeneous. Note also that the phase space of the Poisson process coincides with $\mathbb{Z}^+$.

2.3.2. *Poisson process defined via the distributions of the increments*

Now our goal is to characterize the homogeneous Poisson process relying on the asymptotic behavior of its so-called transition probabilities on the vanishing time interval. For more detail on the notion of transition probabilities, see Chapter 7.

THEOREM 2.4.– *Let $N = \{N_t, t \geq 0\}$ be a process with independent increments, $N_0 = 0$, and N take non-negative integer values. Moreover, assume that, for $0 \leq s \leq t$,*

$$\mathrm{P}\{N_t - N_s = 1\} = \lambda(t-s) + o(t-s),$$
$$\mathrm{P}\{N_t - N_s = 0\} = 1 - \lambda(t-s) + o(t-s),$$

and

$$\mathrm{P}\{N_t - N_s > 1\} = o(t-s)$$

as $t \to s+$, where $\lambda > 0$ is a given number. Then $\{N_t, t \geq 0\}$ is a homogeneous Poisson process with intensity λ.

REMARK 2.4.– It is clear that the third assumption of the above theorem follows from the first and second assumptions.

PROOF.– Denote $p(s,t,k) = \mathrm{P}\{N_t - N_s = k\}, \ k \in \mathbb{Z}^+$. It follows from the independence of increments and the law of total probability that, for $0 < \Delta s < t-s$, we have that

$$p(s,t,0) = p(s,s+\Delta s,0)p(s+\Delta s,t,0) = (1-\lambda\Delta s)p(s+\Delta s,t,0) + o(\Delta s).$$

The latter equality implies right-continuity of $p(s,t,0)$ in s (left-continuity is established similarly) and leads to the equation

$$\frac{\partial p(s,t,0)}{\partial s} = \lambda\, p(s,t,0).$$

The equation for the left-hand derivative has the same form, whence $p(s,t,0) = C_t e^{\lambda s}$. Since $p(t,t,0) = 1$ for $t = s$, $C_t = e^{-\lambda t}$, and

$$p(s,t,0) = \mathrm{P}\{N_t - N_s = 0\} = e^{-\lambda(t-s)}.$$

Further, for any $k \geq 1$,

$$\begin{aligned} p(s,t,k) &= p(s,s+\Delta s,0)p(s+\Delta s,t,k) \\ &\quad + p(s,s+\Delta s,1)p(s+\Delta s,t,k-1) + o(\Delta s) \end{aligned} \qquad [2.7]$$

as $\Delta s \to 0+$. It follows from equality [2.7] and the theorem's conditions that $p(s,t,k)$ is continuous from the right in s. Continuity from the left is established similarly. Now, by substituting the value of $p(s,s+\Delta s,k), k \geq 0$ from the theorem's conditions into [2.7], we get that

$$p(s,t,k) = (1-\lambda\Delta s)p(s+\Delta s,t,k) + \lambda\Delta s\, p(s+\Delta s,t,k-1) + o(\Delta s),$$

or

$$\begin{aligned} \frac{p(s,t,k) - p(s+\Delta s,t,k)}{\Delta s} &= -\lambda\, p(s+\Delta s,t,k) \\ &\quad + \lambda\, p(s+\Delta s,t,k-1) + \frac{o(\Delta s)}{\Delta s}. \end{aligned} \qquad [2.8]$$

Taking into account the continuity of $p(s,t,k)$ in s, we get that, for $\Delta s \to 0+$, equality

$$-\frac{\partial p(s,t,k)_+}{\partial s} = -\lambda\, p(s,t,k) + \lambda\, p(s,t,k-1)$$

holds, and the equation for the left-hand derivative will be the same. Hence,

$$\frac{\partial p(s,t,k)}{\partial s} = \lambda\, p(s,t,k) - \lambda\, p(s,t,k-1).$$

Now, for $k = 1$, we have that $\frac{\partial p(s,t,1)}{\partial s} = \lambda\, p(s,t,1) - \lambda e^{-\lambda(t-s)}$, whence $p(s,t,1) = e^{-\lambda(t-s)}\lambda(t-s)$. In order to apply induction, we assume that $p(s,t,k-1) = e^{-\lambda(t-s)}\frac{(\lambda(t-s))^{k-1}}{(k-1)!}$, and get the equation

$$\frac{\partial p(s,t,k)}{\partial s} = \lambda p(s,t,k) - \lambda e^{-\lambda(t-s)}\frac{(\lambda(t-s))^{k-1}}{(k-1)!}.$$

It follows that $p(s,t,k) = e^{-\lambda(t-s)}\frac{(\lambda(t-s))^k}{k!}$, and the step of induction implies that

$$\mathrm{P}\{N_t - N_s = k\} = e^{-\lambda(t-s)}\frac{(\lambda(t-s))^k}{k!}.$$

The theorem is proved. □

REMARK 2.5.– It will be shown in the following section that a Poisson process satisfies the assumptions of theorem 2.4.

2.3.3. *Poisson process as a renewal process*

Now, let $N = \{N_t, t \geq 0\}$ be a homogeneous Poisson process with parameter λ. It has non-negative integer-valued increments. Generally speaking, the trajectories of this process can be ill-behaved, but there is a version of the process which has trajectories continuous from the right with limits from the left; more detail is available in Chapter 6. Below we will consider this version. Consider the Poisson process as a renewal process. Let $\{\tau_i, i \geq 1\}$ be a sequence of independent random variables, each of which has exponential distribution with parameter $\lambda > 0$, so

$$\mathrm{P}\{\tau_i < x\} = \begin{cases} 1 - e^{-\lambda x}, \ x \geq 0, \\ 0, \ x = 0. \end{cases} \qquad [2.9]$$

Let us construct the following renewal process:

$$N_t = \begin{cases} 0, \ t < \tau_1, \\ \sup\{n \geq 1 : \sum_{i=1}^n \tau_i \leq t\}, \ t \geq \tau_1. \end{cases} \qquad [2.10]$$

It is easy to see that N has jumps of size 1 at the points $T_n = \sum_{i=1}^n \tau_i$, $n \geq 1$, called *arrival times*; the variables τ_n, $n \geq 1$, are called *inter-arrival times*.

THEOREM 2.5.– *The stochastic process, constructed by formulas* [2.9] *and* [2.10]*, is a homogeneous Poisson process with parameter* λ.

PROOF.– It is necessary to prove that the process N_t has independent increments, and that, for any $k \geq 0$ and $0 \leq s < t$,

$$\mathrm{P}\{N_t - N_s = k\} = e^{-\lambda(t-s)}\frac{(\lambda(t-s))^k}{k!}.$$

We divide the proof into several steps.

Step 1. Prove by induction that, for all $x > 0$ and $k \geq 1$,

$$P\left\{\sum_{i=1}^{k} \tau_i > x\right\} = e^{-\lambda x} \sum_{i=0}^{k-1} \frac{(\lambda x)^i}{i!}. \qquad [2.11]$$

Evidently, equality [2.11] holds for $k = 1$, since $P\{\tau_1 > x\} = e^{-\lambda x}, \ x > 0$, according to [2.9]. Assuming that [2.11] holds for $k = n$, for $k = n + 1$, we have the relations

$$P\left\{\sum_{i=1}^{n+1} \tau_i > x\right\} = P\left\{\sum_{i=1}^{n} \tau_i > x\right\} + P\left\{\sum_{i=1}^{n} \tau_i < x, \tau_{n+1} > x - \sum_{i=1}^{n} \tau_i\right\} \qquad [2.12]$$

$$= e^{-\lambda x} \sum_{i=0}^{n-1} \frac{(\lambda x)^i}{i!} + \int_0^x p_n(u) P\{\tau_{n+1} > x - u\} du,$$

where $p_n(u)$ is the probability density function of $\sum_{i=1}^{n} \tau_i$. However, since, for $k = n$, the equality [2.11] holds,

$$p_n(u) = \lambda e^{-\lambda u} \sum_{i=0}^{n-1} \frac{(\lambda u)^i}{i!} - e^{-\lambda u} \sum_{i=0}^{n-2} \frac{u^i \lambda^{i+1}}{i!} = \lambda e^{-\lambda u} \frac{(\lambda u)^{n-1}}{(n-1)!}. \qquad [2.13]$$

Now return to [2.12]:

$$e^{-\lambda x} \sum_{i=0}^{n-1} \frac{(\lambda x)^i}{i!} + \int_0^x \lambda e^{-\lambda u} \frac{(\lambda u)^{n-1}}{(n-1)!} e^{-\lambda(x-u)} du = e^{-\lambda x} \sum_{i=0}^{n} \frac{(\lambda x)^i}{i!},$$

i.e. [2.11] holds for $k = n + 1$, so for all $k \geq 1$.

Step 2. Let us find the distribution of N_t. For any $k \geq 1$, according to [2.13], we have that

$$P\{N_t = k\} = P\left\{\sum_{i=1}^{k} \tau_i \leq t, \sum_{i=1}^{k+1} \tau_i > t\right\} = P\left\{\sum_{i=1}^{k} \tau_i \leq t, \tau_{k+1} > t - \sum_{i=1}^{k} \tau_i\right\}$$

$$= \int_0^t p_k(x) P\{\tau_{k+1} > t - x\} dx = \int_0^t \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{(k-1)!} e^{-\lambda(t-x)} dx = e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

Finally, for $k = 0$, $P\{N_t = 0\} = P\{\tau_1 > t\} = e^{-\lambda t}$.

Step 3. Calculate the probability $\mathrm{P}\{N_t - N_s = k, N_s = l\}$, where $0 < s < t, k \geq 1, l \geq 1$. Evidently, it follows from the independency of τ_i that for $k \geq 2$

$$\begin{aligned}
&\mathrm{P}\{N_t - N_s = k, N_s = l\} = \mathrm{P}\{N_t = k + l, N_s = l\} \\
&= \mathrm{P}\left\{\sum_{i=1}^{l} \tau_i \leq s, s < \sum_{i=1}^{l+1} \tau_i \leq t, \sum_{i=1}^{k+l} \tau_i \leq t, \sum_{i=1}^{k+l+1} \tau_i > t\right\} \\
&= \mathrm{P}\Big\{ \sum_{i=1}^{l} \tau_i \leq s, s - \sum_{i=1}^{l} \tau_i < \tau_{l+1} \leq t - \sum_{i=1}^{l} \tau_i, \\
&\sum_{i=l+2}^{k+l} \tau_i \leq t - \sum_{i=1}^{l} \tau_i - \tau_{l+1}, \tau_{k+l+1} > t - \sum_{i=1}^{l} \tau_i - \tau_{l+1} - \sum_{i=l+2}^{k+l} \tau_i\Big\} \qquad [2.14] \\
&= \int_0^s \lambda e^{-\lambda u} \frac{(\lambda u)^{l-1}}{(l-1)!} \int_{s-u}^{t-u} \lambda e^{-\lambda x} \int_0^{t-u-x} \lambda e^{-\lambda z} \frac{(\lambda z)^{k-2}}{(k-2)!} e^{-\lambda(t-u-x-z)} dz\, dx\, du \\
&= e^{-\lambda t} \int_0^s \lambda \frac{(\lambda u)^{l-1}}{(l-1)!} \int_{s-u}^{t-u} \lambda^k \frac{(t-u-x)^{k-1}}{(k-1)!} dx\, du \\
&= e^{-\lambda t} \int_0^s \lambda \frac{(\lambda u)^{l-1}}{(l-1)!} \lambda^k \frac{(t-s)^k}{k!} du = e^{-\lambda t} \frac{(\lambda s)^l}{l!} \frac{\lambda^k (t-s)^k}{k!}.
\end{aligned}$$

Now, let $k = 1$. Then

$$\begin{aligned}
&\mathrm{P}\{N_t - N_s = 1, N_s = l\} = \mathrm{P}\{N_t = l + 1, N_s = l\} \\
&= \mathrm{P}\left\{\sum_{i=1}^{l} \tau_i \leq s, s < \sum_{i=1}^{l+1} \tau_i \leq t, \tau_{i+2} \geq t - \sum_{i=1}^{l} \tau_i - \tau_{l+1}\right\} \qquad [2.15] \\
&= \int_0^s \lambda e^{-\lambda u} \frac{(\lambda u)^{l-1}}{(l-1)!} \int_{s-u}^{t-u} \lambda e^{-\lambda x} \int_{t-u-x}^{\infty} \lambda e^{-\lambda z} dz\, dx\, du \\
&= e^{-\lambda s} \frac{(\lambda s)^l}{l!} e^{-\lambda(t-s)} \lambda(t-s).
\end{aligned}$$

Similar to equations [2.14] and [2.15], we can establish this formula for $\min(k, l) = 0$, whence the proof follows.

Step 4. Further calculations are performed taking into account [2.14]:

$$\mathrm{P}\{N_t - N_s = k\} = \sum_{l=0}^{\infty} \mathrm{P}\{N_t - N_s = k, N_s = l\}$$

$$= \sum_{l=0}^{\infty} e^{-\lambda t} \frac{(\lambda s)^l}{l!} \frac{\lambda^k (t-s)^k}{k!} = e^{-\lambda(t-s)} \frac{\lambda^k (t-s)^k}{k!}. \qquad [2.16]$$

Comparing [2.14] with [2.16], we see that the increments $N_t - N_s$ and $N_s - N_0$ are independent. Independence of arbitrary finite families of increments can be proved in a similar way. Formula [2.16] justifies that the increments have the Poisson distribution. The theorem is proved. □

2.4. Compound Poisson process

Let $\{\tau_i, i \geq 1\}$ be a sequence of independent identically distributed random variables (in what follows, we shall use the abbreviated notation iid rv or iid random variables), $N = \{N_t, t \geq 0\}$ be a Poisson process with independent measure $m = m((s,t])$ for $0 \leq s < t$, and let the process N not depend on $\{\tau_i, i \geq 1\}$. Since N takes its values in $\mathbb{Z}^+$, we can form a sum

$$X_t = \sum_{i=1}^{N_t} \xi_i.$$

As usual, we put $\sum_{i=1}^{0} = 0$. Process X is called a *compound Poisson process* with intensity measure $m = m((s,t])$ and generating random variables $\{\xi_i, i \geq 1\}$. It can also be represented in the following form. Let $\{\tau_i, i \geq 1\}$ be the lengths of the intervals between subsequent jumps of the Poisson process (τ_1 is the moment of the first jump). Then the Poisson process itself can be represented as $N_t = \sum_{i=1}^{\infty} \mathbb{1}_{\tau_i \leq t}$, and the compound Poisson process, in turn, can be represented as

$$X_t = \sum_{i=1}^{\infty} \xi_i \mathbb{1}_{\tau_i \leq t}.$$

Denote $\psi(\lambda) = \mathrm{E}\exp\{i\lambda\xi_1\}, \lambda \in \mathbb{R}$. If we denote the cumulative distribution function of ξ_1 by $\nu = \nu(dx)$, then the characteristic function can be written as $\psi(\lambda) = \int_{\mathbb{R}} e^{i\lambda x}\nu(dx)$.

THEOREM 2.6.– *The characteristic function of the increments of the compound Poisson process has a form*

$$\psi(\lambda; s, t) = \exp\{m((s,t])(\psi(\lambda) - 1)\}$$

$$= \exp\Big\{m((s,t]) \int_{\mathbb{R}} (e^{i\lambda x} - 1)\nu(dx)\Big\}. \qquad [2.17]$$

In particular, the compound Poisson process is a process with independent increments.

PROOF.– Taking into account the independence of ξ_i and N, we get

$$\psi(\lambda; s,t) = \mathrm{E}\exp\{i\lambda(X_t - X_s)\} = \mathrm{E}\left(\exp\left\{i\lambda\Big(\sum_{j=N_s+1}^{N_t}\xi_i\Big)\right\}\mathbb{1}_{N_t>N_s}\right)$$

$$+\mathrm{P}\{N_t = N_s\} = \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}\mathrm{E}\exp\left\{i\lambda\Big(\sum_{j=k+1}^{l}\xi_i\Big)\right\}\mathrm{P}\{N_s = k, N_t = l\} + e^{m((s,t])}$$

$$= \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}(\psi(\lambda))^{l-k}\mathrm{P}\{N_s = k, N_t - N_s = l-k\} + e^{m((s,t])}$$

$$= \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}(\psi(\lambda))^{l-k}\frac{(m((0,s]))^k}{k!}e^{-m([0,s))}\frac{(m((s,t]))^{l-k}}{(l-k)!}e^{-m((s,t])} + e^{m((s,t])}$$

$$= \sum_{k=0}^{\infty}e^{-m([0,s))}\frac{(m((0,s]))^k}{k!}\sum_{j=1}^{\infty}(\psi(\lambda))^j e^{-m((s,t])}\frac{(m((s,t]))^j}{j!} + e^{m((s,t])}$$

$$= e^{-m((s,t])}\sum_{j=0}^{\infty}\frac{(\psi(\lambda)m((s,t]))^j}{j!} = e^{-m((s,t])(\psi(\lambda)-1)}$$

$$= \exp\left\{m((s,t])\int_{\mathbb{R}}(e^{i\lambda x} - 1)\nu(dx)\right\}.$$

The theorem is proved. □

REMARK 2.6.– The compound Poisson process has a lot of practical applications. It can be used as a model for accumulated claims of insurance companies, total revenues in the queuing theory, etc.

2.5. Lévy processes

Consider now processes with independent increments that contain both the Wiener and the Poisson components. Let the process under consideration be real-valued. Let us fix two numbers, $a \in \mathbb{R}$ and $\sigma \geq 0$, and a measure ν on the Borel σ-field $\mathcal{B}(\mathbb{R})$, such that $\nu(\{0\}) = 0$ and

$$\int_{\mathbb{R}}(1 \wedge y^2)\nu(dy) < \infty.$$

DEFINITION 2.6.– *Stochastic process* $X = \{X_t, t \geq 0\}$ *is called a* Lévy process *with characteristics* (a, σ, ν) *if it satisfies the following three conditions:*

1) $X_0 = 0$.

2) X *is a process with independent increments.*

3) For any $0 \leq s < t$*, the characteristic function of the increment* $X_t - X_s$ *has a form*

$$\varphi(\lambda; s, t) = \mathrm{E} \exp\{i\lambda(X_t - X_s)\} = e^{(t-s)\psi(\lambda)}, \quad [2.18]$$

where the function $\psi = \psi(\lambda)$ *has a form*

$$\psi(\lambda) = ia\lambda - \frac{1}{2}\sigma^2\lambda^2 + \int_{\mathbb{R}} (e^{i\lambda y} - 1 - i\lambda y \mathbb{1}_{0<|y|<1})\nu(dy). \quad [2.19]$$

It immediately follows from equality [2.18] that

$$\varphi(\lambda; s, u)\varphi(\lambda; u, t) = \varphi(\lambda; s, t).$$

It means that such a process with independent increments does exist and is homogeneous in the sense that $\varphi(\lambda; s, t)$ depends only on the difference $t - s$. If $s = 0$, it follows from [2.18] that

$$\mathrm{E} \exp\{i\lambda X_t\} = e^{t\psi(\lambda)},$$

where $\psi(\lambda)$ is defined by the formula [2.19]. The function $\psi(\lambda)$ is called characteristic exponent of the Lévy process.

REMARK 2.7.– It turns out that any process with homogeneous independent increments is a Lévy process in the sense of the definition given above, i.e. the characteristic function of its increments has necessarily the form [2.19], called the Lévy–Khintchine representation. For more detail on this topic, see e.g. [SAT 13].

Note also that in the case where $\sigma = 0$ and $\int_{\mathbb{R}}(|x| \wedge 1)\nu(dx) < \infty$, trajectories of the Lévy process a.s. have a bounded variation on any finite interval. If additionally the Lévy measure is concentrated on $\mathbb{R}^+$, and $a \geq 0$, the trajectories of the corresponding Lévy process are a.s. non-decreasing. In the latter case, the Lévy process is called a *subordinator*.

Consider some particular examples of Lévy processes.

2.5.1. *Wiener process with a drift*

Let $\nu \equiv 0, a \neq 0, \sigma > 0$. Then $\psi(\lambda) = ia\lambda - \frac{1}{2}\sigma^2\lambda^2$, and in turn, it follows that $\mathrm{E}\exp\{i\lambda X_t\} = e^{t(ia\lambda - \frac{1}{2}\sigma^2\lambda^2)}$. Comparing with [A2.1], we see that X_t has Gaussian distribution with mean at and variance $\sigma^2 t$. Taking into account that X is a process with independent increments, we get that X can be represented via a Wiener process as $X_t = at + \sigma W_t$. Such a process is called a Wiener process with the drift coefficient a and diffusion (volatility) coefficient σ. Since the Wiener process is symmetric in the sense that $-W$ is also a Wiener process, it is natural to restrict the coefficient σ to positive values, which implies that there is no sense in considering negative values of σ.

2.5.2. *Compound Poisson process as a Lévy process*

Let $\sigma = 0$ and measure ν to be not identically zero. If additionally the measure ν is finite, i.e. $\nu(\mathbb{R}) < \infty$, then both integrals $\int_{\mathbb{R}}(e^{i\lambda y} - 1)\nu(dy)$ and $\int_{\mathbb{R}} i\lambda y \mathbb{1}_{\{0<|y|<1\}}\nu(dy)$ are finite, and $\psi(\lambda)$ can be represented in a form

$$\psi(\lambda) = ia\lambda - i\lambda \int_{\mathbb{R}} y \mathbb{1}_{0<|y|<1}\nu(dy) + \int_{\mathbb{R}} (e^{i\lambda y} - 1)\nu(dy).$$

If we assume additionally that $a = \int_{\mathbb{R}} y \mathbb{1}_{0<|y|<1}\nu(dy)$, then

$$\psi(\lambda) = \int_{\mathbb{R}} (e^{i\lambda y} - 1)\nu(dy) = \nu(\mathbb{R}) \int_{\mathbb{R}} (e^{i\lambda y} - 1)\frac{\nu(dy)}{\nu(\mathbb{R})}. \qquad [2.20]$$

Comparing formula [2.20] with [2.17], we see that X is a compound Poisson process with intensity measure $m((s,t]) = \nu(\mathbb{R})(t-s)$ and generating random variables $\{\xi_i, i \geq 1\}$ whose distribution is described by $\mathrm{P}\{\xi_i \leq x\} = \nu((-\infty, x])/\nu(\mathbb{R})$.

2.5.3. *Sum of a Wiener process with a drift and a Poisson process*

Now, let $a = \sigma = 0$, the measure ν be concentrated at the point 1 and $\nu(\{1\}) = \gamma > 0$. Then $\psi(\lambda) = \gamma(e^{i\lambda} - 1)$ and we get the characteristic function of the form $\phi(\lambda; s.t) = e^{(t-s)\psi(\lambda)} = e^{\gamma(t-s)(e^{i\lambda}-1)}$, that is, a characteristic function of the increment of a Poisson process N with intensity γ. Further, if $a \neq 0$, $\sigma > 0$, the measure ν is concentrated at the point 1 and $\nu(1) = \gamma > 0$, then $\psi(\lambda) = ia\lambda - \frac{1}{2}\sigma^2\lambda^2 + \gamma(e^{i\lambda} - 1)$, and the process X is a sum of two independent processes: a Wiener process with a drift, $at + \sigma W_t$, and a Poisson process N with intensity γ:

$$X_t = at + \sigma W_t + N_t.$$

2.5.4. *Gamma process*

Recall that the density function of Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ has a form

$$f_{\alpha,\beta}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, \quad x > 0.$$

The characteristic function has a form

$$\varphi_{\alpha,\beta}(\lambda) = \left(1 - \frac{i\lambda}{\beta}\right)^{-\alpha}.$$

Consider the Lévy process with $a = \frac{\alpha(1-e^{-\beta})}{\beta}$, $\sigma = 0$ and

$$\nu(dx) = \alpha e^{-\beta x} x^{-1} \mathbb{1}_{x>0} dx. \qquad [2.21]$$

Such Lévy process is called a *Gamma process*; since the measure ν is concentrated on the positive half-line, it is a subordinator. Its characteristic exponent equals

$$\begin{aligned}\psi(\lambda) &= ia\lambda + \int_{\mathbb{R}} \left(e^{i\lambda y} - 1 - i\lambda y \, \mathbb{1}_{0<|y|<1}\right) \nu(dy) \\ &= \frac{i\alpha\lambda(1-e^{-\beta})}{\beta} + \\ &\quad + \alpha \int_0^\infty \left(e^{i\lambda y} - 1 - i\lambda y \, \mathbb{1}_{0<y<1}\right) e^{-\beta y} y^{-1} dy \\ &= \alpha \int_0^\infty \left(e^{i\lambda y} - 1\right) e^{-\beta y} y^{-1} dy.\end{aligned}$$

Therefore, for the Gamma process, we can put $a = 0$, $\sigma = 0$ and $\psi(\lambda) = \int_0^\infty \left(e^{i\lambda x} - 1\right) \nu(dx)$ with $\nu(dx)$ defined in [2.21].

2.5.5. *Stable Lévy motion*

For $\alpha \in (0, 2)$, a Lévy process is called a (standard) *symmetric α-stable Lévy motion* if

$$\mathrm{E} e^{i\lambda X_t} = e^{-t|\lambda|^\alpha}, \ t \geq 0.$$

This is a generalization of the Wiener process (which, up to a constant, corresponds to $\alpha = 2$) with heavy tails of distribution: its variance is infinite for any $\alpha \in (0, 2)$, and the expectation is infinite for $\alpha \in (0, 1]$.

2.5.6. *Stable Lévy subordinator with stability parameter* $\alpha \in (0,1)$

For $\alpha \in (0,1)$, a subordinator X with Lévy measure $\nu(dx) = \frac{\alpha}{\Gamma(1-\alpha)} x^{-\alpha-1}$, $x > 0$ is called a stable process with stability parameter α. The characteristic exponent equals $\psi_\alpha(\lambda) = \int_0^\infty (e^{i\lambda y} - 1)\nu(dy)$.

REMARK 2.8.– For more detail concerning Lévy processes, see e.g. books [BAR 01, SAT 13, SCH 03].

3

Gaussian Processes. Integration with Respect to Gaussian Processes

3.1. Gaussian vectors

Let $m \geq 1$, let and $\overline{\xi} = (\xi_1, \ldots, \xi_m)$ be a random vector.

DEFINITION 3.1.– *Vector $\overline{\xi}$ is called* Gaussian *if there exists a non-random vector $\overline{a} = (a_1, \ldots, a_m) \in \mathbb{R}^m$ and a non-negatively definite symmetric non-random matrix $C = \{c_{ik}\}_{i,k=1}^m$ such that, for any vector $\overline{\lambda} = (\lambda_1, \ldots, \lambda_m) \in \mathbb{R}^m$,*

$$\varphi_{\overline{\xi}}(\overline{\lambda}) := \mathrm{E} \exp\left\{i(\overline{\lambda}, \overline{\xi})\right\} = \exp\left\{i(\overline{\lambda}, \overline{a}) - \frac{1}{2}(C\overline{\lambda}, \overline{\lambda})\right\}, \qquad [3.1]$$

where $(\overline{x}, \overline{y}) = \sum_{j=1}^m x_j y_j$ stands for the inner product in $\mathbb{R}^m$.

LEMMA 3.1.–

i) Coefficients in the representation [3.1] *equal*

$$a_j = \mathrm{E}\xi_j \ \textit{ and } \ c_{jk} = \mathrm{Cov}(\xi_j, \xi_k) = \mathrm{E}(\xi_j - a_j)(\xi_k - a_k).$$

ii) Let $\det C \neq 0$. *Then the coordinates of vector $\overline{\xi} - \overline{a}$ are linearly independent. In this case, the distribution of the vector $\overline{\xi}$ is non-degenerate,* $\operatorname{supp} \overline{\xi} = \mathbb{R}^m$*, i.e. the distribution of $\overline{\xi}$ has strictly positive density on $\mathbb{R}^m$, and this density equals*

$$p_{\overline{\xi}}(\overline{x}) = (2\pi)^{-\frac{m}{2}} (\det C)^{-\frac{1}{2}} \exp\left\{-(C^{-1}(\overline{x} - \overline{a}), \overline{x} - \overline{a})\right\}, \qquad [3.2]$$

where C^{-1} is the inverse matrix to C.

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

iii) Let $\det C = 0$*. Then coordinates of* $\overline{\xi} - \overline{a}$ *are linearly dependent, the number of linearly independent coordinates equals* $m_1 := \operatorname{rank} C < m$ *and the distribution of the vector* $\overline{\xi}$ *is concentrated on* $\mathbb{R}^{m_1}$*.*

PROOF.– i) On the one hand, differentiate $\varphi_{\overline{\xi}}(\overline{\lambda})$:

$$\left(\frac{\partial}{\partial \lambda_j}\varphi_{\overline{\xi}}(\overline{\lambda})\right)\bigg|_{\overline{\lambda}=0} = \frac{\partial}{\partial \lambda_j}\exp\left\{i(\overline{\lambda},\overline{a}) - \frac{1}{2}(C\overline{\lambda},\overline{\lambda})\right\}\bigg|_{\overline{\lambda}=0} = ia_j.$$

On the other hand, it follows from the properties of the characteristic functions that

$$\frac{\partial}{\partial \lambda_j}\mathrm{E}\exp\left\{i(\overline{\lambda},\overline{\xi})\right\}\bigg|_{\overline{\lambda}=0} = i\mathrm{E}\xi_j,$$

whence $a_j = \mathrm{E}\xi_j$. Similarly, differentiating in λ_j and λ_k, we get that

$$\left(\frac{\partial^2}{\partial \lambda_j \partial \lambda_k}\varphi_{\overline{\xi}}(\overline{\lambda})\right)\bigg|_{\overline{\lambda}=0} = -a_j a_k - c_{jk},$$

and, at the same time,

$$\left(\frac{\partial^2}{\partial \lambda_j \partial \lambda_k}\mathrm{E}\exp\left\{i(\overline{\lambda},\overline{\xi})\right\}\right)\bigg|_{\overline{\lambda}=0} = -\mathrm{E}\xi_j\xi_k,$$

whence $c_{jk} = \mathrm{E}\xi_j\xi_k - a_j a_k = \mathrm{Cov}(\xi_j,\xi_k)$.

ii) Let $\det C \neq 0$ (which means that $\det C > 0$). Then the rows of matrix C are linearly independent, i.e. if $\sum_{k=1}^m \beta_k c_{lk} = 0$, $1 \le l \le m$, then $\beta_k = 0, 1 \le k \le m$. Assume that coordinates $(\xi_1 - a_1, \ldots, \xi_m - a_m)$ are linearly dependent, i.e. there exists vector $\overline{\beta} = (\beta_k, 1 \le k \le m)$, not identically zero, and such that $\sum_{k=1}^m \beta_k(\xi_k - a_k) = 0$. Then, for any $1 \le l \le m$,

$$\mathrm{E}\left(\sum_{k=1}^m \beta_k(\xi_k - a_k)(\xi_l - a_l)\right) = \sum_{k=1}^m \beta_k c_{lk} = 0,$$

and we get a contradiction. Therefore, coordinates $(\xi_1 - a_1, \ldots, \xi_m - a_m)$ are linearly independent.

In order to prove that the density of non-degenerate distribution has a form [3.2], it is sufficient to establish that, for any $\overline{\lambda} \in \mathbb{R}^m$,

$$J := \int_{\mathbb{R}^m} e^{i(\overline{\lambda},\overline{x})} p_{\overline{\xi}}(\overline{x}) d\overline{x} = \exp\left\{i(\overline{\lambda},\overline{a}) - \frac{1}{2}(C\overline{\lambda},\overline{\lambda})\right\},$$

or, equivalently, to establish that

$$(2\pi)^{-\frac{m}{2}}(\det C)^{-\frac{1}{2}}\int_{\mathbb{R}^m} e^{i(\overline{\lambda},\overline{x})}e^{-(C^{-1}(\overline{x}-\overline{a}),\overline{x}-\overline{a})}d\overline{x} = \\ = \exp\left\{i(\overline{\lambda},\overline{a})-\frac{1}{2}(C\overline{\lambda},\overline{\lambda})\right\}. \qquad [3.3]$$

Matrix C is symmetric and non-negatively definite. Then it follows from matrix theory that there exists an orthogonal matrix B, i.e. $BB^{\mathrm{tr}} = E$ (unit matrix) and such that $B^{\mathrm{tr}}CB = D$, where D is a diagonal matrix of the form $D = \begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_m \end{pmatrix}$, $d_j \geq 0$, $1 \leq j \leq m$. Since $\det C \neq 0$, then $d_j > 0$, $1 \leq j \leq m$. Changing the variables in the left-hand side of [3.3], $\overline{x} - \overline{a} = B\overline{y}$, and putting $\overline{\lambda} = B\overline{\mu}$, we get, taking into account that $C^{-1} = (C^{-1})^{\mathrm{tr}}$:

$$\begin{aligned} J &= (2\pi)^{-\frac{m}{2}}(\det C)^{-\frac{1}{2}}\int_{\mathbb{R}^m} e^{i(B\overline{\mu},\overline{a}+B\overline{y})}e^{-\frac{1}{2}(C^{-1}B\overline{y},B\overline{y})}d\overline{y} \\ &= (2\pi)^{-\frac{m}{2}}(\det C)^{-\frac{1}{2}}e^{i(\overline{\lambda},\overline{a})}\int_{\mathbb{R}^m} e^{i(B\overline{\mu})^{\mathrm{tr}}B\overline{y}-\frac{1}{2}(C^{-1}B\overline{y})^{\mathrm{tr}}B\overline{y}}d\overline{y} \\ &= (2\pi)^{-\frac{m}{2}}(\det C)^{-\frac{1}{2}}e^{i(\overline{\lambda},\overline{a})}\int_{\mathbb{R}^m} \exp\left\{i(\overline{\mu},\overline{y})-\frac{1}{2}\overline{y}^{\mathrm{tr}}B^{\mathrm{tr}}(C^{-1})^{\mathrm{tr}}B\overline{y}\right\}d\overline{y} \\ &= (2\pi)^{-\frac{m}{2}}(\det C)^{-\frac{1}{2}}e^{i(\overline{\lambda},\overline{a})}\int_{\mathbb{R}^m} \exp\left\{i(\overline{\mu},\overline{y})-\frac{1}{2}\overline{y}^{\mathrm{tr}}B^{\mathrm{tr}}C^{-1}B\overline{y}\right\}d\overline{y}. \end{aligned}$$

Calculate $B^{\mathrm{tr}}C^{-1}B$. Let $B^{\mathrm{tr}}C^{-1}B = X$. Then $C^{-1}B = BX$, $C^{-1} = BXB^{\mathrm{tr}}$, whence $CBXB^{\mathrm{tr}} = E$, or $B^{\mathrm{tr}}CBXB^{\mathrm{tr}} = B^{\mathrm{tr}}$, whence $DXB^{\mathrm{tr}} = B^{\mathrm{tr}}$, or $DX = E$. Finally, $X = D^{-1}$. Therefore,

$$\begin{aligned} J &= (2\pi)^{-\frac{m}{2}}(\det C)^{-\frac{1}{2}}e^{i(\overline{\lambda},\overline{a})}\int_{\mathbb{R}^m} \exp\left\{i(\overline{\mu},\overline{y})-\frac{1}{2}\overline{y}^{\mathrm{tr}}D^{-1}\overline{y}\right\}d\overline{y} \\ &= (2\pi)^{-\frac{m}{2}}(\det C)^{-\frac{1}{2}}e^{i(\overline{\lambda},\overline{a})}\int_{\mathbb{R}^m} \exp\left\{i\sum_{k=1}^{m}\mu_k y_k-\frac{1}{2}\sum_{k=1}^{m}d_k^{-1}y_k^2\right\}d\overline{y}. \end{aligned}$$

Note that $\det C = \det D = \prod_{k=1}^{m} d_k$. Therefore,

$$J = (2\pi)^{-\frac{m}{2}} \prod_{k=1}^{m} d_k^{-\frac{1}{2}} e^{i(\overline{\lambda},\overline{a})} \prod_{k=1}^{m} \int_{\mathbb{R}} e^{i\lambda\mu_k y_k - \frac{1}{2} d_k^{-1} y_k^2} dy = \prod_{k=1}^{m} e^{i\lambda_k a_k - \frac{\mu_k^2 d_k^2}{2}}$$

$$= \exp\left\{ i(\overline{\lambda},\overline{a}) - \frac{1}{2}\overline{\mu}^{tr} D\overline{\mu} \right\} = \exp\left\{ i(\overline{\lambda},\overline{a}) - \frac{1}{2}\overline{\mu}^{tr} B^{\mathrm{tr}} C B\overline{\mu} \right\}$$

$$= \exp\left\{ i(\overline{\lambda},\overline{a}) - \frac{1}{2}(C\overline{\lambda},\overline{\lambda}) \right\}.$$

iii) If $\det C = 0$, then the rows of matrix C are linearly dependent, so that there exist β_l, $1 \le l \le m$ such that $\overline{\beta} = (\beta_1, \dots, \beta_m) \ne 0$, and $\sum_{l=1}^{m} \beta_l c_{lk} = 0$ for $1 \le k \le m$, whence

$$\sum_{k,l=1}^{m} \beta_l \beta_k \mathrm{Cov}(\xi_l, \xi_k) = \mathrm{E}\left(\sum_{l=1}^{m} \beta_l (\xi_l - a_l) \right)^2 = 0.$$

This means that $\sum_{l=1}^{m} \beta_l (\xi_l - a_l) = 0$ with $\overline{\beta} \ne 0$, and coordinates of $\overline{\xi} - \overline{a}$ are linearly dependent. These reasons clearly demonstrate that the number m_1 of linearly independent coordinates equals $\mathrm{rank}\, C$ and other coordinates are linear combinations of linearly independent ones, and so the distribution of $\overline{\xi}$ is concentrated on some m_1-dimensional subspace. □

REMARK 3.1.– Recall that the coordinates of a Gaussian vector are independent (in standard sense, as the random variables), if and only if they are non-correlated. If some subsets of coordinates of a random vector are Gaussian, it does not mean that the vector itself is Gaussian.

3.2. Theorem of Gaussian representation (theorem on normal correlation)

Consider a Gaussian vector $(\xi, \xi_1, \dots, \xi_n)$. Introduce the σ-field $\mathcal{F}_n = \sigma\{\xi_1, \dots, \xi_n\}$.

THEOREM 3.1.– ***Theorem on normal correlation***. *There exist constants d and d_j, $1 \le j \le n$ such that*

$$\mathrm{E}(\xi | \mathcal{F}_n) = \sum_{j=1}^{n} d_j \xi_j + d.$$

The values of the constants will be specified in the proof of the theorem.

PROOF.– Denote $m = \mathrm{E}\xi$, $m_j = \mathrm{E}\xi_j$, $\widetilde{\xi} = \xi - m$ and $\widetilde{\xi}_j = \xi_j - m_j$, $c_{jk} = \mathrm{E}\widetilde{\xi}_j\widetilde{\xi}_k = \mathrm{Cov}(\xi_j, \xi_k)$, $c_j = \mathrm{E}\widetilde{\xi}\widetilde{\xi}_j = \mathrm{Cov}(\xi, \xi_j)$. Without loss of generality, we can assume that $\{\xi_1, \dots, \xi_n\}$ are linearly independent. Indeed, let $\{\xi_1, \dots, \xi_{n'}\}$ be linearly independent and $\{\xi_{n'+1}, \dots, \xi_n\}$ be the linear combinations of $\{\xi_1, \dots, \xi_{n'}\}$. Then $\mathcal{F}_n = \mathcal{F}_{n'}$. If we get that $\mathrm{E}(\xi|\mathcal{F}_{n'}) = \sum_{j=1}^{n'} d_j\xi_j + d$, then $\mathrm{E}(\xi|\mathcal{F}_n) = \sum_{j=1}^{n} d_j\xi_j + d$, with $d_{n'+1} = \dots = d_n = 0$. Therefore, let $\{\xi_1, \dots, \xi_n\}$ be linearly independent. In this case, according to lemma 3.1, $\det C \neq 0$, where $C = \{c_{jk}\}_{j,k=1}^n$ is a covariance matrix. Now, let us find constants $\{\alpha_1, \dots, \alpha_n\}$ such that the centered random variable $\widetilde{\xi} - \sum_{j=1}^n \alpha_j\widetilde{\xi}_j$ and any of $\widetilde{\xi}_j$ become orthogonal:

$$\mathrm{E}\left(\widetilde{\xi} - \sum_{r=1}^n \alpha_r\widetilde{\xi}_r\right)\widetilde{\xi}_j = 0, \;\; 1 \le j \le n.$$

The latter system of equations is equivalent to the following one:

$$c_j - \sum_{r=1}^n \alpha_r c_{jr} = 0, \;\; 1 \le j \le n.$$

This system of n linear equations has a non-zero determinant which equals $\det C$; therefore, it has the unique solution $\alpha_r = \frac{\det C_r}{\det C}$, where C_r can be obtained by replacing in C rth column with $(C_1, \dots, C_n)^{\mathrm{tr}}$. Now, $\widetilde{\xi} - \sum_{j=1}^n \alpha_j\widetilde{\xi}_j$ is independent with the vector $(\widetilde{\xi}_1, \dots, \widetilde{\xi}_n)$ because it follows from orthogonality that, for any $(\lambda, \lambda_1, \dots, \lambda_n)$,

$$\mathrm{E}\exp\left\{i\lambda\left(\widetilde{\xi} - \sum_{j=1}^n \alpha_j\widetilde{\xi}_j\right) + i\sum_{j=1}^n \lambda_j\widetilde{\xi}_j\right\}$$
$$= \exp\left\{-\frac{1}{2}\lambda^2\mathrm{E}\left(\widetilde{\xi} - \sum_{j=1}^n \alpha_j\widetilde{\xi}_j\right)^2 - \frac{1}{2}\sum_{j,k=1}^n \lambda_j\lambda_k c_{jk}\right\}$$
$$= \mathrm{E}\exp\left\{i\lambda\left(\widetilde{\xi} - \sum_{j=1}^n \alpha_j\widetilde{\xi}_j\right)\right\}\mathrm{E}\exp\left\{i\sum_{j=1}^n \lambda_j\widetilde{\xi}_j\right\}.$$

Finally, we get

$$\begin{aligned}\mathrm{E}(\xi|\mathcal{F}_n) &= \mathrm{E}(\widetilde{\xi}|\mathcal{F}_n) + m = \mathrm{E}\left(\widetilde{\xi} - \sum_{j=1}^n \alpha_j \widetilde{\xi}_j \,\middle|\, \mathcal{F}_n\right) + \sum_{j=1}^n \alpha_j \widetilde{\xi}_j + m \\ &= \mathrm{E}\left(\widetilde{\xi} - \sum_{j=1}^n \alpha_j \widetilde{\xi}_j\right) + \sum_{j=1}^n \alpha_j \widetilde{\xi}_j + m \\ &= \sum_{j=1}^n \alpha_j \widetilde{\xi}_j + m = \sum_{j=1}^n \alpha_j \xi_j + m + \sum_{j=1}^n \alpha_j m_j.\end{aligned}$$

Therefore, $d = m + \sum_{j=1}^n \alpha_j m_j$ and $d_j = \alpha_j = \frac{\det C_j}{\det C}$, and the theorem is proved. □

3.3. Gaussian processes

Let $X = \{X_t, t \in \mathbb{T}\}$ be a real-valued stochastic process.

DEFINITION 3.2.– *Stochastic process X is* Gaussian *if all its finite-dimensional distributions are Gaussian, i.e. for any $m \geq 1$ and any $t_1, \ldots, t_m \in \mathbb{T}$ random vector $(X_{t_1}, \ldots, X_{t_m})$ is Gaussian.*

REMARK 3.2.– If only one-dimensional, or one- and two-dimensional distributions are Gaussian, it does not mean that the process is Gaussian.

Definition 3.2, together with definition 3.1, means that there exist function $\{a(t), t \in \mathbb{T}\}$ and function of two variables $\{R(t,s), (t,s) \in \mathbb{T} \times \mathbb{T}\}$ such that, for any $m \geq 1$ and any $\overline{\lambda} = (\lambda_1, \ldots, \lambda_m)$,

$$\mathrm{E}\exp\left\{i \sum_{j=1}^m \lambda_j X_{t_j}\right\} = \exp\left\{i \sum_{j=1}^m \lambda_j a(t_j) - \frac{1}{2} \sum_{j,k=1}^m R(t_j, t_k)\lambda_j \lambda_k\right\}.$$

It follows from lemma 3.1 that $a(t) = \mathrm{E}X_t$ (*mean function*) and $R(t,s) = \mathrm{E}(X_t - a(t))(X_s - a(s))$ (*covariance function*). Therefore, function R has the properties:

(R)

i) $R(t,s) = R(s,t)$, $(s,t) \in \mathbb{T} \times \mathbb{T}$;

ii) for any $n \geq 1$, any $t_1, \ldots, t_n \in \mathbb{T}$ and any $b_1, \ldots, b_n \in \mathbb{R}$

$$\sum_{j,k=1}^n R(t_j, t_k) b_j b_k \geq 0.$$

Property (i) is evident. To prove (ii), note that

$$\sum_{j,k=1}^{n} R(t_j,t_k)b_j b_k = \sum_{j,k=1}^{n} \mathrm{E}(X_{t_k}-a(t_k))(X_{t_j}-a(t_j))b_j b_k$$
$$= \mathrm{E}\left(\sum_{j=1}^{n} b_j(X_{t_j}-a(t_j))\right)^2 \geq 0.$$

DEFINITION 3.3.– *Function $R = R(t,s) : \mathbb{T}\times\mathbb{T} \to \mathbb{R}$ is called symmetric and non-negative definite if it satisfies properties (R), (i) and (ii).*

The next theorem states that any real-valued function $a = a(t) : \mathbb{T} \to \mathbb{R}$ and any real-valued non-negatively definite function $R = R(t,s) : \mathbb{T}\times\mathbb{T} \to \mathbb{R}$ define some Gaussian process on $\mathbb{T}$ with finite-dimensional distributions defined uniquely.

THEOREM 3.2.– *Let us have an arbitrary function $a : \mathbb{T} \to \mathbb{R}$ and a symmetric non-negative definite function $R : \mathbb{T}\times\mathbb{T} \to \mathbb{R}$. Then there exists a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and a unique in the sense of finite-dimensional distributions Gaussian process $X = \{X_t : \mathbb{T}\times\Omega \to \mathbb{R}\}$, such that $a = a(t)$ is its mean function: $\mathrm{E}X_t = a_t$, and $R = R(t,s)$ is its covariance function: $\mathrm{E}(X_t - a(t))(X_s - a(s)) = R(t,s)$.*

PROOF.– For any $n \geq 1$ and $t_1,\dots,t_n \in \mathbb{T}$, let us define the characteristic functions

$$\psi(\lambda_1,\dots,\lambda_n;t_1,\dots,t_n) = \exp\left\{ i\sum_{j=1}^{n}\lambda_j a(t_j) - \frac{1}{2}\sum_{j,k=1}^{n}\lambda_j\lambda_k R(t_j,t_k)\right\}. \quad [3.4]$$

Then the consistency conditions (B) are fulfilled. Indeed, for any $1 \leq l \leq n$ and any points $t_j \in \mathbb{T}$, $1 \leq j \leq n$, $\lambda_j \in \mathbb{R}$, $1 \leq j \leq n$, we have that

$$\psi(\lambda_1,\dots,\lambda_l,\underbrace{0,\dots,0}_{n-k};t_1,\dots,t_l,t_{l+1},\dots,t_n)$$
$$= \exp\left\{ i\sum_{j=1}^{l}\lambda_j a(t_j) - \frac{1}{2}\sum_{j,k=1}^{l}\lambda_j\lambda_k R(t_j,t_k)\right\} = \psi(\lambda_1,\dots,\lambda_l;t_1,\dots,t_l),$$

so condition (B), 1) holds. Further, for any $n \geq 1$, $\lambda_j \in \mathbb{R}$, $t_j \in \mathbb{T}$, $1 \leq j \leq n$ and any permutation $\pi(\overline{\lambda}) = (\lambda_{i_1}, \ldots, \lambda_{i_n})$ and the corresponding permutation $\pi(\overline{t}) = (t_{i_1}, \ldots, t_{i_n})$, we have that

$$\psi(\pi(\overline{\lambda}); \pi(\overline{t})) = \exp\left\{ i\sum_{j=1}^{n} \lambda_{i_j} a(t_{i_j}) - \frac{1}{2}\sum_{j,k=1}^{n} \lambda_{i_j}\lambda_{i_k} R(t_{i_j}, t_{i_k}) \right\}$$

$$= \exp\left\{ i\sum_{j=1}^{n} \lambda_j a(t_j) - \frac{1}{2}\sum_{j,k=1}^{n} \lambda_j\lambda_k R(t_j, t_k) \right\} = \psi(\lambda_1, \ldots, \lambda_n; t_1, \ldots, t_n),$$

so condition (B), 2) holds. Applying the Kolmogorov theorem in the form of theorem 1.3, we get that there exists a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and a stochastic process $X = \{X_t : \mathbb{T} \times \Omega \to \mathbb{R}\}$ for which

$$\psi(\lambda_1, \ldots, \lambda_n; t_1, \ldots, t_n) = \mathrm{E}\exp\left\{ i\sum_{j=1}^{n} \lambda_j X_{t_j} \right\}.$$

According to [3.4], process X is Gaussian; $a = a(t)$ is its mean function and $R = R(t, s)$ is its covariance function. Since finite-dimensional distributions of the Gaussian process are uniquely determined by its mean and covariance function, such a process is unique in the sense of finite-dimensional distributions. The theorem is proved. □

3.4. Examples of Gaussian processes

In what follows, we shall consider centered Gaussian processes, i.e. $\mathrm{E}X_t = 0$ for any $t \in \mathbb{T}$. Any such process is uniquely determined by its covariance function; therefore, giving an example of the Gaussian process is equivalent to providing an example of a covariance function.

3.4.1. *Wiener process as an example of a Gaussian process*

Consider a Wiener process $W = \{W_t, t \geq 0\}$ satisfying definition 2.1. Recall that, for any $t \geq 0$, $\mathrm{E}W_t = 0$ and the covariance function equals

$$\mathrm{Cov}(W_s, W_t) = \mathrm{E}\, W_s W_t = s \wedge t.$$

According to theorem 2.2, the Wiener process does exist; therefore, the function $R(s, t) = s \wedge t$ is non-negative definite, as any covariance function. With this in mind, consider another definition of a Wiener process.

DEFINITION 3.4.– *Stochastic process* $W = \{W_t, t \geq 0\}$ *is a Wiener process if it satisfies three assumptions:*

1) W *is a Gaussian process;*

2) $\mathrm{E}\, W_t = 0$*, for any* $t \geq 0$*;*

3) $\mathrm{Cov}(W_s, W_t) = s \wedge t, \quad s, t \geq 0.$

THEOREM 3.3.– *Definitions 2.1 and 3.4 of the Wiener process are equivalent.*

PROOF.– Let the process W satisfy definition 3.4. Then, for any $0 = t_0 \leq t_1 < t_2 < \ldots < t_n$, vector $(W_{t_1}, W_{t_2} - W_{t_1}, W_{t_3} - W_{t_2}, \ldots, W_{t_n} - W_{t_{n-1}})$ is Gaussian. Moreover, for any $\overline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$,

$$\begin{aligned} &\mathrm{E}\exp\left\{i\sum_{k=1}^{n}\lambda_k(W_{t_k} - W_{t_{k-1}})\right\} \\ &= \exp\left\{-\frac{1}{2}\sum_{j,k=1}^{n}\lambda_k\lambda_j\mathrm{E}\left((W_{t_k} - W_{t_{k-1}})(W_{t_j} - W_{t_{j-1}})\right)\right\} \qquad [3.5] \\ &= \exp\left\{-\frac{1}{2}\sum_{j,k=1}^{n}\lambda_k\lambda_j(t_k \wedge t_j - t_{k-1} \wedge t_j - t_k \wedge t_{j-1} + t_{k-1} \wedge t_{j-1})\right\} \\ &= \exp\left\{-\frac{1}{2}\sum_{k=1}^{n}\lambda_k^2(t_k - t_{k-1})\right\}, \end{aligned}$$

because, if e.g. $k > j$, then $t_k \wedge t_j - t_{k-1} \wedge t_j - t_k \wedge t_{j-1} + t_{k-1} \wedge t_{j-1} = t_j - t_j - t_{j-1} + t_{j-1} = 0$. Therefore, it follows from [3.5] that

$$\mathrm{E}\exp\left\{i\sum_{k=1}^{n}\lambda_k(W_{t_k} - W_{t_{k-1}})\right\} = \prod_{k=1}^{n}\mathrm{E}\exp\left\{i\lambda_k(W_{t_k} - W_{t_{k-1}})\right\},$$

i.e. the increments $(W_{t_1}, W_{t_2} - W_{t_1}, \ldots, W_{t_n} - W_{t_{n-1}})$ are mutually independent. Any increment is a Gaussian random variable with $\mathrm{E}(W_t - W_s) = 0$ and $\mathrm{E}(W_t - W_s)^2 = t - 2(s \wedge t) + s = t - s$, if $s \leq t$. Finally, $\mathrm{E}W_0^2 = 0$, so W starts from zero, and we get that W satisfies definition 2.1. Conversely, let W satisfy definition 2.1. Then, for any $0 = t_0 \le t_1 < t_2 < \ldots < t_n$ and for any $\overline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_n) \in \mathbb{R}^n$,

$$\begin{aligned}
&\mathrm{E}\exp\left\{ i\sum_{k=1}^{n}\lambda_k W_{t_k}\right\} = \mathrm{E}\exp\left\{ i\sum_{j=1}^{n}(\lambda_j+\ldots+\lambda_n)(W_{t_j}-W_{t_{j-1}})\right\} \\
&= \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}(\lambda_j+\ldots+\lambda_n)^2(t_j-t_{j-1})\right\} \qquad [3.6] \\
&= \exp\left\{-\frac{1}{2}\sum_{j=1}^{n}\sum_{k,l=j}^{n}\lambda_k\lambda_l(t_j-t_{j-1})\right\} \\
&= \exp\left\{-\frac{1}{2}\sum_{k,l=1}^{n}\lambda_k\lambda_l\sum_{j=1}^{k\wedge l}(t_j-t_{j-1})\right\} = \exp\left\{-\frac{1}{2}\sum_{k,l=1}^{n}\lambda_k\lambda_l\, t_{k\wedge l}\right\}.
\end{aligned}$$

Equalities [3.6] mean that $\{W_t, t \ge 0\}$ is a Gaussian process with $\mathrm{E}W_t = 0$ and $\mathrm{Cov}(W_t, W_s) = t \wedge s$. The theorem is proved. □

3.4.2. *Fractional Brownian motion*

Let $H \in (0,1)$.

DEFINITION 3.5.– Fractional Brownian motion *with Hurst index* $H \in (0,1)$ *is a centered Gaussian process* $B^H = \left\{B_t^H, t \ge 0\right\}$ *with covariance function*

$$R(t,s) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right). \qquad [3.7]$$

THEOREM 3.4.– *Formula* [3.7] *defines a covariance function for any* $H \in (0,1)$.

PROOF.– We follow the lines of the proof of proposition 1.6 from [NOU 12]. The symmetry property of $R(t,s)$ is evident; therefore, we have to prove only that it is non-negatively definite. To this end, denote

$$c_H = \int_0^\infty (1 - e^{-u^2})u^{-1-2H}du,$$

and observe that, for any $x \in \mathbb{R}$,

$$\frac{1}{c_H}\int_0^\infty \frac{(1-e^{-u^2x^2})}{u^{1+2H}}du = \left|\begin{matrix} u|x| = y \\ du = \frac{dy}{|x|}\end{matrix}\right| = \frac{1}{c_H}\int_0^\infty \frac{(1-e^{-y^2})}{y^{1+2H}|x|^{-1-2H}}\frac{dy}{|x|} = |x|^{2H}.$$

Therefore,

$$\frac{1}{2}\left(t^{2H}+s^{2H}-|t-s|^{2H}\right)$$
$$=\frac{1}{2c_H}\left(\int_0^\infty\left(1-e^{-u^2s^2}+1-e^{-u^2t^2}-1+e^{-u^2(t-s)^2}\right)u^{-1-2H}du\right)$$
$$=\frac{1}{2c_H}\int_0^\infty\left(\left(1-e^{-u^2s^2}\right)\left(1-e^{-u^2t^2}\right)\right.$$
$$\left.+\left(e^{-u^2(t-s)^2}-e^{-u^2s^2-u^2t^2}\right)\right)u^{-1-2H}du$$
$$=\frac{1}{2c_H}\int_0^\infty\left(1-e^{-u^2s^2}\right)\left(1-e^{-u^2t^2}\right)u^{-1-2H}du$$
$$+\frac{1}{2c_H}\int_0^\infty\left(e^{-u^2s^2-u^2t^2}\left(e^{2u^2ts}-1\right)\right)u^{-1-2H}du:=I_1(t,s)+I_2(t,s).$$

Concerning I_1, we can state immediately that, for any $n\geq 1$, any λ_i, $1\leq i\leq n$, and any $0\leq t_1<t_2<\ldots<t_n$,

$$\sum_{j,k=1}^n\lambda_j\lambda_k I_1(t_j,t_k)$$
$$=\frac{1}{2c_H}\sum_{j,k=1}^n\lambda_j\lambda_k\int_0^\infty\left(1-e^{-u^2t_j^2}\right)\left(1-e^{-u^2t_k^2}\right)u^{-1-2H}du \qquad [3.8]$$
$$=\frac{1}{2c_H}\left(\sum_{j=1}^n\lambda_j\int_0^\infty\left(1-e^{-u^2t_j^2}\right)u^{-1-2H}du\right)^2\geq 0.$$

Furthermore, $I_2(t,s)$ can be represented as

$$I_2(t,s)=\frac{1}{2c_H}\int_0^\infty\sum_{l=1}^\infty e^{-u^2s^2-u^2t^2}\frac{(2u^2ts)^l}{l!}u^{-1-2H}du.$$

Therefore,

$$\sum_{j,k=1}^n\lambda_j\lambda_k I_2(t_j,t_k)=\frac{1}{2c_H}\sum_{l=1}^\infty 2^l\int_0^\infty\sum_{j,k=1}^n\lambda_j\lambda_k e^{-u^2t_j^2}e^{-u^2t_k^2}t_j^l t_k^l u^{-1-2H}du$$
$$=\frac{1}{2c_H}\sum_{l=1}^\infty 2^l\int_0^\infty\left(\sum_{j=1}^n\lambda_j e^{-u^2t_j^2}t_j^l\right)^2u^{-1-2H}du\geq 0. \qquad [3.9]$$

Proof follows now from [3.8] and [3.9]. □

Let $H = \frac{1}{2}$. Then $R(t,s) = \frac{1}{2}(t + s - |t - s|) = s \wedge t$, so we have the Wiener process. This means that the Wiener process is a particular case of the fractional Brownian motion. The properties of fBm are considered in detail in the book [MIS 08].

3.4.3. *Sub-fractional and bi-fractional Brownian motion*

Consider some generalizations of fractional Brownian motion

DEFINITION 3.6.– *Sub-fractional Brownian motion* $C^H = \{C_t^H, t \in [0,1]\}$ *is a centered Gaussian process with covariance function*

$$R_{C^H}(t,s) = t^{2H} + s^{2H} - \frac{1}{2}(|t+s|^{2H} + |t-s|^{2H}),$$

$H \in (0,1)$. This process was studied, e.g. in [BOJ 04].

DEFINITION 3.7.– *Bi-fractional Brownian motion* $B^{H,K} = \{B_t^{H,K}, t \in [0,1]\}$ *is a centered Gaussian process with covariance function*

$$R_{B^{H,K}}(t,s) = \frac{1}{2^K}\Big((t^{2H} + s^{2H})^K - |t-s|^{2HK}\Big),$$

$H \in (0,1)$, $K \in (0,1)$. This process was studied, e.g. in [RUS 06].

3.4.4. *Brownian bridge*

It is interesting that we can construct a Gaussian process that takes the prescribed values in the endpoints of an interval.

DEFINITION 3.8.– *Brownian bridge between points* 0 *and* T *in time and points* $a, b \in \mathbb{R}$ *is a Gaussian process*

$$B = \{B_t, t \in [0,T]\}$$

with

$$\mathrm{E}B_t = \frac{1}{T}(a(T-t) + bt)$$

and covariance function

$$R(t,s) = t \wedge s - \frac{ts}{T}.$$

It turns out that the Brownian bridge can be constructed using a standard Wiener process W; we provide one of the possible constructions. Define

$$B_t = a + W_t + \frac{(b - a - W_T)t}{T}.$$

It is clear that $B_0 = a$, $B_T = b$ and $\mathrm{E}B_t = \frac{1}{T}(a(T-t) + bt)$. Let us calculate the covariance:

$$\begin{aligned}\mathrm{Cov}(B_t, B_s) &= \mathrm{E}\Big(W_t - \frac{t}{T}W_T\Big)\Big(W_s - \frac{s}{T}W_T\Big)\\ &= \mathrm{E}W_tW_s - \frac{t}{T}\mathrm{E}W_sW_T - \frac{s}{T}\mathrm{E}W_tW_T - \frac{ts}{T^2}\mathrm{E}W_T^2\\ &= t \wedge s - \frac{2ts}{T} + \frac{tsT}{T^2} = t \wedge s - \frac{ts}{T},\end{aligned}$$

as required. For further discussion concerning the Brownian bridge, see section 9.1.2.

3.4.5. *Ornstein–Uhlenbeck process*

We can consider the Ornstein–Uhlenbeck process defined either on $\mathbb{R}^+$ or on $\mathbb{R}$ (the same as the Wiener process and the fractional Brownian motion, but their two-sided versions will be considered in detail later, in section 3.6).

DEFINITION 3.9.– One-sided Ornstein–Uhlenbeck process $X = \{X_t, t \geq 0\}$ *is a Gaussian process* $X = \{X_t, t \geq 0\}$ *with* $\mathrm{E}X_t = x_0 e^{\theta t}$ *and covariance function*

$$R(t,s) = \frac{\sigma^2}{2\theta} e^{\theta t + \theta s}\left(1 - e^{-2\theta(t \wedge s)}\right) = \frac{\sigma^2}{2\theta}\big(e^{\theta(t+s)} - e^{-\theta|t-s|}\big),$$

$\theta \in \mathbb{R}, t, s \geq 0$.

DEFINITION 3.10.– Two-sided Ornstein–Uhlenbeck process $X = \{X_t, t \in \mathbb{R}\}$ *is a Gaussian process with* $\mathrm{E}X_t = x_0$ *and* $R(t,s) = \frac{\sigma^2}{2\theta} e^{\theta(t \vee s - t \wedge s)} = -\frac{\sigma^2}{2\theta} e^{\theta|t-s|}$, $\theta < 0$, $t, s \in \mathbb{R}$.

An explicit construction and properties of Ornstein–Uhlenbeck processes will be considered in section 9.1.2. Note that the expectation of the two-sided Ornstein–Uhlenbeck process is constant, and the covariance function depends only on the difference between t and s. Such processes are called *wide-sense stationary*.

3.5. Integration of non-random functions with respect to Gaussian processes

3.5.1. *General approach*

Consider a centered Gaussian process $X = \{X_t, t \geq 0\}$ with covariance function $R(t,s)$, $(t,s) \in \mathbb{R}_+^2$. Consider a fixed rectangle $[0,T]^2$.

DEFINITION 3.11.– *A function* $f = f(t)\colon [0,T] \to \mathbb{R}$ *is called* elementary *if it has a form*

$$f(t) = \sum_{j=0}^{m-1} a_j \mathbb{1}_{t\in(t_j,t_{j+1}]}, f_j \in \mathbb{R},$$

where $\{0 = t_0 < t_1 < \cdots < t_m = T\}$ *is a partition of* $[0,T]$.

DEFINITION 3.12.– *Integral* $I(f,X)([0,T])$ *of elementary function* f *w.r.t. a Gaussian process* X *is defined as a sum*

$$I(f,X,[0,T]) = \int_0^T f(t)dX_t := \sum_{j=0}^{m-1} a_j(X_{t_{j+1}} - X_{t_j}).$$

It is evident that, for any elementary function f, the integral $\int_0^T f(t)dX_t$ is a Gaussian random variable with mean $\mathrm{E}\int_0^T f(t)dX_t = 0$ and variance

$$\mathrm{E}\left(\int_0^T f(t)dX_t\right)^2 = \sum_{j,k=0}^{m-1} a_j a_k R(\Delta_{jk}) \qquad [3.10]$$

where

$$R(\Delta_{jk}) = R(t_{j+1},t_{k+1}) - R(t_j,t_{k+1}) - R(t_{j+1},t_k) + R(t_j,t_k)$$

is the rectangular increment of R over the rectangle $\Delta_{jk} = (t_j,t_{j+1}] \times (t_k,t_{k+1}]$, $j,k = 0,1,\ldots,m-1$.

Assume that the covariance function R satisfies the following technical assumption:

(R) The covariance function R is absolutely continuous w.r.t. the Lebesgue measure λ_2 on $[0,T]^2$, i.e. there exists a function r, integrable w.r.t. the measure λ_2 on $[0,T]^2$ such that, for any $s_1,s_2,t_1,t_2 \in [0,T]$ with $s_1 < t_1, s_2 < t_2$,

$$R(t_1,t_2) - R(s_1,t_2) - R(t_2,s_1) + R(t_1,s_1) = \int_{s_1}^{t_1}\int_{s_2}^{t_2} r(u_1,u_2)du_2\,du_1.$$

We have from [3.10] that for any elementary function f,

$$\mathrm{E}\left(\int_0^T f(t)dX_t\right)^2 = \int_0^T \int_0^T f(t)f(s)r(t,s)dt\,ds =: K(f). \qquad [3.11]$$

This motivates the following definition. Denote by $\mathcal{L}_{2,r}([0,T])$ the class of Borel functions $f\colon [0,T] \to \mathbb{R}$ for which

$$\int_0^T \int_0^T |f(t)||f(s)|\,|r(t,s)|\,dt\,ds < \infty.$$

so that the integral $K(f) = \int_0^T \int_0^T f(t)f(s)r(t,s)dt\,ds$ is well defined.

THEOREM 3.5.– *Let $f \in \mathcal{L}_{2,r}([0,T])$. Then there exists a sequence of elementary functions $g_n = g_n(t), n \geq 1, t \in [0,T]$, such that*

$$K(f - g_n)\,ds \to 0,\ n \to \infty, \qquad [3.12]$$

and

$$K(g_m - g_n) \to 0,\ n, m \to \infty, \qquad [3.13]$$

so, by [3.11] *the limit of $\int_0^T g_n(t)dX_t$ exists in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$, and we can define the integral $\int_0^T f(t)dX_t$ as a limit of the integrals $\int_0^T g_n(t)dX_t$ in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$.*

PROOF.– Consider first a simple function of the form

$$h(t) = \sum_{k=1}^{N} c_k \mathbb{1}_{A_k}(t),$$

where $A_k \in \mathcal{B}([0,T]), k = 1, \ldots, N$. Since the Borel σ-algebra $\mathcal{B}([0,T])$ is generated by the semiring of half-open intervals of the form $(a,b]$, then by the Caratheodory approximation theorem (see [BIL 95, theorem 11.4]), for any $\varepsilon > 0$ and each $k = 1, 2, \ldots, N$, there exist disjoint intervals $(a_i, b_i]$, $i = 1, \ldots, m_k$, such that

$$\lambda_1\left(A_k \,\triangle\, \bigcup_{i=1}^{m_k} (a_i, b_i]\right) < \frac{\varepsilon}{N}.$$

Then, defining elementary functions

$$h_\varepsilon(t) = \sum_{k=1}^{N} c_k \sum_{i=1}^{m_k} \mathbb{1}_{(a_i,b_i]}(t),$$

we have

$$\lambda_1(\{t \in [0,T] : h_\varepsilon(t) \neq h(t)\}) \leq \sum_{k=1}^{N} \lambda_1\Big(A_k \triangle \bigcup_{i=1}^{m_k}(a_i, b_i]\Big) < \varepsilon.$$

Consequently, $h_\varepsilon \xrightarrow{\lambda_1} h$, $\varepsilon \to 0+$. Moreover, $|h_\varepsilon(t)| \leq \max_{1\leq k\leq N}|c_k|$. Therefore, by the Lebesgue dominated convergence theorem, we have

$$K(h - h_\varepsilon) \to 0, \ \varepsilon \to 0+$$

and $K(h_\varepsilon) \to K(h)$, $\varepsilon \to 0+$, so it follows from [3.11] that $K(h) \geq 0$ and $\sqrt{K(h)}$ is a seminorm on the set of simple functions.

Now let f_n be simple measurable functions, such that, for any $t \in [0,T]$, $|f_n(t)| \leq |f(t)|$ and $f_n(t) \to f(t)$, $n \to \infty$. Then by the Lebesgue dominated convergence theorem

$$K(f_n - f) \to 0, \ n \to \infty,$$

and $K(f_n) \to K(f)$, $n \to \infty$. As mentioned above, it follows that $K(f) \geq 0$ and $\sqrt{K(f)}$ is a seminorm on $\mathcal{L}_{2,r}([0,T])$. We can then approximate the functions f_n in probability by elementary functions g_n so that $K(f_n - g_n) \to 0$, $n \to \infty$. Then we get [3.12] and [3.13] by virtue of the triangle inequality. □

REMARK 3.3.– Obviously, $\int_0^T f(t)dX_t$ is a Gaussian random variable with zero mean and variance $K(f)$.

3.5.2. *Integration of non-random functions with respect to the Wiener process*

Let $W = \{W_t, t \geq 0\}$ be a Wiener process. Then, on the one hand, we cannot directly apply theorem 3.5, because the covariance function $R(s,t) = s \wedge t$ is not absolutely continuous w.r.t. the Lebesgue measure on any rectangle $[0,T]^2$. Indeed, consider, for $n \geq 1$, $0 = t_0^n < t_1^n < \cdots < t_n^n = T$ and note that

$$\begin{aligned}\mathbb{R}((t_k^n, t_{k+1}^n]^2) &= R(t_{k+1}^n, t_{k+1}^n) - R(t_{k+1}^n, t_k^n) - R(t_k^n, t_{k+1}^n) + R(t_k^n, t_k^n)\\ &= t_{k+1}^n - t_k^n,\end{aligned}$$

so

$$R\Big(\bigcup_{k=0}^{n-1}(t_k^n, t_{k+1}^n]^2\Big) = \sum_{k=0}^{n-1}\big(t_{k+1}^n - t_k^n\big) = T,$$

while

$$\lambda_2\left(\bigcup_{k=0}^{n-1}(t_k^n, t_{k+1}^n]^2\right) = \sum_{k=0}^{n-1}\left(t_{k+1}^n - t_k^n\right)^2 \le T \max_{0\le k\le n-1}\left(t_{k+1}^n - t_k^n\right)$$

may be arbitrarily small. On the other hand, the theory of integration of non-random functions is very simple in this case because the increments in the Wiener process are independent (non-correlated, orthogonal), and for any elementary function $f(t) = \sum_{j=0}^{n-1} f_j \mathbb{1}_{t\in(t_k^n, t_{k+1}^n]}$ we can define

$$I(f, W, [0,T]) = \sum_{j=0}^{n-1} f_j (W_{t_{j+1}} - W_{t_j})$$

that is a Gaussian r.v. with $\mathrm{E} I(f, W, [0,T]) = 0$ and $\mathrm{E}\left(I(f, W, [0,T])\right)^2 = \sum_{j=0}^{n-1} f_j^2 \Delta t_j$, where $\Delta t_j = t_{j+1} - t_j$. Now, let $f \in \mathcal{L}_2([0,T], \lambda_1)$. Then there exists a sequence $\widetilde{f}_n(t) = \sum_{j=0}^{k_n-1} \widetilde{f}_{n,j} \mathbb{1}_{t\in(t_{n,j}, t_{n,j+1}]}$ of elementary functions such that $\int_0^T |f(t) - \widetilde{f}_n(t)|^2 dt \to 0$, $n \to \infty$. Then $\widetilde{f}_n$ is a Cauchy sequence in $\mathcal{L}_2([0,T])$, and consequently, $I(\widetilde{f}_n, W, [0,T])$ is a Cauchy sequence in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ because

$$\mathrm{E}\left(I(\widetilde{f}_n, W, [0,T]) - I(\widetilde{f}_m, W, [0,T])\right)^2 = \int_0^T |\widetilde{f}_n(t) - \widetilde{f}_m(t)|^2 dt.$$

Therefore, we can define $\int_0^T f(t) dW_t$ as the limit

$$\int_0^T f(t) dW_t = \lim_{n\to\infty} I(\widetilde{f}_n, W, [0,T]) \text{ in } \mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P}).$$

THEOREM 3.6.–

1) The integral $\int_0^T f(t) dW_t$ is a Gaussian r.v. with $\mathrm{E}\int_0^T f(t) dW_t = 0$ and

$$\mathrm{E}\left(\int_0^T f(t) dW_t\right)^2 = \int_0^T f^2(t) dt,$$

or, in other words,

$$\left\|\int_0^T f(t) dW_t\right\|_{\mathcal{L}_2(\Omega,\mathcal{F},\mathrm{P})} = \|f\|_{\mathcal{L}_2([0,T],\lambda_1)}$$

(isometry property).

2) The integral $\int_0^T f(t)dW_t$ *does not depend on the approximating sequence* $\widetilde{f}_n = \left\{\widetilde{f}_n(t), t \in [0,T]\right\}$ *of elementary functions.*

3) Let $g_n = \{g_n(t), t \in [0,T]\} \in \mathcal{L}_2([0,T], \lambda_1)$ *and* $\|g_n - f\|_{\mathcal{L}_2([0,T],\lambda_1)} \to 0$ *as* $n \to \infty$*. Then*

$$\int_0^T f(t)dW_t = \lim_{n\to\infty} \int_0^T g_n(t)dW_t \ \ in \ \ \mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P}).$$

4) For any $\alpha, \beta \in \mathbb{R}$ *and any* $f, g \in \mathcal{L}_2([0,T], \lambda_1)$*,*

$$\int_0^T (\alpha f(t) + \beta g(t))dW_t = \alpha \int_0^T f(t)dW_t + \beta \int_0^T g(t)dW_t.$$

PROOF.– 1) By definition, $\int_0^T f(t)dW_t$ is a limit in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ of

$$I(\widetilde{f}_n, W, [0,T]) = \sum_{j=0}^{k_n-1} \widetilde{f}_{n,j}(W_{t_{n,j+1}} - W_{t_{n,j}}),$$

which are Gaussian r.v. with $\mathrm{E}I(\widetilde{f}_n, W, [0,T]) = 0$ and

$$\mathrm{E}(I(\widetilde{f}_n, W, [0,T]))^2 = \sum_{j=0}^{k_n-1} \widetilde{f}^2_{n,j}\Delta t_{n,j} = \int_0^T \widetilde{f}^2_n(t)dt.$$

Then it is a weak limit too, so, according to lemma A2.5, we get that $\int_0^T f(t)dW_t$ is a Gaussian r.v. with zero mean and variance $\int_0^T f^2(t)dt$.

2) Let $\widetilde{g}_n = \{\widetilde{g}_n(t), t \in [0,T]\}$ be another sequence of elementary approximating functions, i.e.

$$\int_0^T |f(t) - \widetilde{g}_n(t)|^2 dt \to 0, \ \ n \to \infty.$$

Then

$$\mathrm{E}\left|\int_0^T \widetilde{g}_n(t)dW_t - \int_0^T \widetilde{f}_n(t)dW_t\right|^2 = \int_0^T |\widetilde{f}_n(t) - \widetilde{g}_n(t)|^2 dt \to 0, \ \ n \to \infty,$$

which means that, in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$,

$$\lim_{n\to\infty} \int_0^T \widetilde{g}_n(t)dW_t = \lim_{n\to\infty} \int_0^T \widetilde{f}_n(t)dW_t = \int_0^T f(t)dW_t.$$

3) We can proceed as in the previous section. Indeed, it follows from isometry property that

$$\mathrm{E}\left|\int_0^T f(t)dW_t - \int_0^T g_n(t)dW_t\right|^2 = \int_0^T |f(t) - g_n(t)|^2 dt \to 0, \;\; n \to \infty.$$

4) Let $\widetilde{f}_n(t), \widetilde{g}_n(t)$ be a sequence of elementary functions such that $\widetilde{f}_n \to f$, $\widetilde{g}_n \to g$ in $\mathcal{L}_2([0,T], \lambda_1)$. Then $\alpha\widetilde{f}_n(t) + \beta\widetilde{g}_n(t)$ is an elementary function that admits the representation, say,

$$\alpha\widetilde{f}_n(t) + \beta\widetilde{g}_n(t) = \sum_{j,k=0}^{k_n-1} (\alpha c_k + \beta d_j)\mathbb{1}_{t\in\Delta_{n,j,k}},$$

where the intervals $\Delta_{n,j,k}$ have no common interior points and $\bigcup_{j,k=0}^{k_n-1} \Delta_{n,j,k} = [0,T]$. Therefore,

$$\begin{aligned}\int_0^T (\alpha\widetilde{f}_n(t) + \beta\widetilde{g}_n(t))dW_t &= \alpha \sum_{j,k=0}^{k_n-1} c_k \Delta W(\Delta_{n,j,k}) + \beta \sum_{j,k=0}^{k_n-1} d_j \Delta W(\Delta_{n,j,k}) \\ &= \alpha \int_0^T \widetilde{f}_n(t)dW_t + \beta \int_0^T \widetilde{g}_n(t)dW_t.\end{aligned}$$

Further,

$$\int_0^T (\alpha f_n(t) + \beta g_n(t))dW_t \to \int_0^T (\alpha f(t) + \beta g(t))dW_t,$$

$$\alpha \int_0^T f_n(t)dW_t + \beta \int_0^T g_n(t))dW_t \to \alpha \int_0^T f(t)dW_t + \beta \int_0^T g(t)dW_t,$$

whence the proof follows. □

3.5.3. *Integration w.r.t. the fractional Brownian motion*

Let $B^H = \left\{B_t^H, t \geq 0\right\}$ be the fractional Brownian motion with Hurst index $H \in \left(\frac{1}{2}, 1\right)$. We know that $R(s,t) = \frac{1}{2}(s^{2H} + t^{2H} - |s-t|^{2H})$, and in the case

$H \in \left(\frac{1}{2}, 1\right)$, for any rectangle $\Delta = [s_1, s_2] \times [t_1, t_2]$, the increment over this rectangle equals

$$\begin{aligned}R(\Delta) &= R(s_2,t_2) - R(s_1,t_2) - R(s_2,t_1) + R(s_1,t_1)\\ &= \frac{1}{2}\left(s_2^{2H} + t_2^{2H} - |s_2-t_2|^{2H} - s_1^{2H} - t_2^{2H} + |s_1-t_2|^{2H}\right.\\ &\quad \left.-s_2^{2H} - t_1^{2H} + |s_2-t_1|^{2H} + s_1^{2H} + t_1^{2H} - |s_1-t_1|^{2H}\right)\\ &= \frac{1}{2}\left(|s_1-t_2|^{2H} + |s_2-t_1|^{2H} - |s_2-t_2|^{2H} - |s_1-t_1|^{2H}\right)\\ &= H(2H-1)\int_{s_1}^{s_2}\int_{t_1}^{t_2} |u-v|^{2H-2} du\, dv \geq 0.\end{aligned}$$

Therefore, the covariance function R is increasing as a function of two variables, in the sense that its increment is positive over any rectangle, so it coincides with $|R|$, and generates the measure, which we also denote by R, on $\mathcal{B}([0,T]^2)$. This measure is absolutely continuous w.r.t. the Lebesgue measure. Note that $\frac{\partial^2 R}{\partial s \partial t} > T^{2H-2}$, and hence, it is separated from zero and we can apply remark 3.3, (ii), and state that integral $\int_0^T f(s)dB_s^H$ exists for any $f \in \mathcal{L}_2([0,T]^2, R)$. However, for technical simplicity, we restrict the class of integrable functions. To be more precise, let us formulate the following Hardy–Littlewood theorem (see e.g. [SAM 93]).

THEOREM 3.7.– *Let $0 < \alpha < 1$. Then, for any $1 < p < \frac{1}{\alpha}$ and $q = \frac{p}{1-\alpha p}$, there exists a constant $C_{p,q,\alpha}$ such that*

$$\left(\int_{[0,T]} \left(\int_{[0,T]} |f(u)||x-u|^{\alpha-1} du\right)^q dx\right)^{\frac{1}{q}} \leq C_{p,q,\alpha} \|f\|_{L_p([0,T],\lambda_1)}. \quad [3.14]$$

THEOREM 3.8.– *Let function $f \in \mathcal{L}_{\frac{1}{H}}([0,T], \lambda_1)$. Then*

i) $f \in \mathcal{L}_2([0,T]^2, R)$.

ii) We can define integral $\int_0^T f(s)dB_s^H$ as the limit in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ of the integrals of elementary functions, and $\int_0^T f(s)dB_s^H$ is a Gaussian random variable with $\mathrm{E}\int_0^T f(s)dB_s^H = 0$ and

$$\mathrm{E}\left(\int_0^T f(s)dB_s^H\right)^2 = H(2H-1)\int_0^T\int_0^T f(s)f(t)|s-t|^{2H-2} dsdt.$$

iii) Moreover, $\int_0^T f(s)dB_s^H$ is the limit in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ of the integrals of any sequence f_n of elementary functions such that $\|f - f_n\|_{\mathcal{L}_{\frac{1}{H}}([0,T],\lambda_1)} \to 0$ as $n \to \infty$.

PROOF.– To prove (i), it is enough to prove that any function $f \in \mathcal{L}_{\frac{1}{H}}([0,T],\lambda_1)$ belongs to $\mathcal{L}_2([0,T]^2, R)$. In turn, it is sufficient to establish that the iterated integral is finite,

$$I := \int_{[0,T]} |f(u)| \left(\int_{[0,T]} |f(v)||u-v|^{2H-2} dv \right) du < \infty.$$

Applying theorem 3.7 with $\alpha = 2H-1$, $p = \frac{1}{H}$ and $q = \frac{p}{1-2\alpha p} = \frac{1}{1-H}$, we obtain that

$$\begin{aligned} I &\le \left(\int_{[0,T]} |f(u)|^{\frac{1}{H}} du \right)^H \left(\int_{[0,T]} \left(\int_{[0,T]} |f(v)||u-v|^{2H-1} dv \right)^{\frac{1}{1-H}} du \right)^{1-H} \\ &\le \|f\|_{L_{\frac{1}{H}}([0,T],\lambda_1)} C_H \|f\|_{L_{\frac{1}{H}}([0,T],\lambda_1)} = C_H \|f\|^2_{L_{\frac{1}{H}}([0,T],\lambda_1)}, \qquad [3.15] \end{aligned}$$

where we have denoted $C_H = C_{1/H,1/(1-H),2H-1}$, for brevity. So, (i) is proved, and (ii) follows immediately from (i) and theorem 3.5. Now, consider arbitrary sequence f_n of elementary functions such that $\|f - f_n\|_{\mathcal{L}_{\frac{1}{H}}([0,T],\lambda_1)} \to 0, \ n \to \infty$. Then, similar to [3.15], we can get that

$$\begin{aligned} &\mathrm{E}\left(\int_0^T f(s) dB_s^H - \int_0^T f_n(s) dB_s^H \right)^2 \\ &\le \int_{[0,T]} |f(u) - f_n(u)| \left(\int_{[0,T]} |f(v) - f_n(v)||u-v|^{2H-2} dv \right) du \\ &\le \left(\int_{[0,T]} |f(u) - f_n(u)|^{\frac{1}{H}} du \right)^H \\ &\times \left(\int_{[0,T]} \left(\int_{[0,T]} |f(v) - f_n(v)||u-v|^{2H-1} dv \right)^{\frac{1}{1-H}} du \right)^{1-H} \\ &\le \|f - f_n\|_{L_{\frac{1}{H}}([0,T],\lambda_1)} C_H \|f - f_n\|_{L_{\frac{1}{H}}([0,T],\lambda_1)} \\ &= C_H \|f - f_n\|^2_{L_{\frac{1}{H}}([0,T],\lambda_1)} \to 0, \ n \to \infty. \end{aligned}$$

Therefore, (iii) is established, and the theorem is proved. □

REMARK 3.4.– Let a function f be continuous, therefore bounded by some constant, on $[0,T]$. Then evidently $\int_0^T f(s) dB_s^H$ exists. Consider any sequence of partitions with vanishing diameter and the sequence of elementary functions of the form $f_n(t) = \sum_{k=1}^{k_n} f(t_k^n) \mathbb{1}_{(t_k^n, t_{k+1}^n]}$. Being bounded by the same constant, f_n tends to f pointwise, and therefore, in $\mathcal{L}_{\frac{1}{H}}([0,T],\lambda_1)$. This means that $\int_0^T f(s) dB_s^H = \lim \int_0^T f_n(s) dB_s^H$ in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$.

3.6. Two-sided Wiener process and fractional Brownian motion: Mandelbrot–van Ness representation of fractional Brownian motion

First, construct a *two-sided Wiener process.*

DEFINITION 3.13.– *Two-sided Wiener process* $W = \{W_t, t \in \mathbb{R}\}$ *is a Gaussian process with* $\mathrm{E}W_t = 0$ *and*

$$\mathrm{E}W_tW_s = \begin{cases} |t| \wedge |s|, & s \cdot t \geq 0, \\ 0, & s \cdot t < 0. \end{cases}$$

A two-sided Wiener process can also be characterized as a process with independent increments, such that $W_0 = 0$ and $W_t - W_s \sim \mathcal{N}(0, t-s)$ for $-\infty < s < t < +\infty$. It can be constructed explicitly in the following way: let $W^1 = \left\{W_t^1, t \geq 0\right\}$ be a Wiener process, and $W^2 = \left\{W_t^2, t \geq 0\right\}$ be a Wiener process independent of W^1. Then $W_t = \begin{cases} W_t^1, & t \geq 0, \\ W_{-t}^2, & t < 0 \end{cases}$ is a two-sided Wiener process.

Now we are in a position to construct the so-called *Mandelbrot–Van Ness representation* of fractional Brownian motion ([MAN 68]) and to introduce two-sided fractional Brownian motion. For any $H \in (0,1)$, define the non-random kernel

$$k_H^+(t,u) = (t-u)_+^{H-\frac{1}{2}} - (-u)_+^{H-\frac{1}{2}}, \quad -\infty < u < t < +\infty,$$

where we use the notation $a_+ = a\mathbb{1}_{a>0}$. Moreover, denote the constant

$$C_H^{(1)} = \left(\int_0^\infty \left((1+x)^{H-\frac{1}{2}} - x^{H-\frac{1}{2}}\right)^2 dx + \frac{1}{2H}\right) = \frac{(2H\sin(\pi H)\Gamma(2H))^{\frac{1}{2}}}{\Gamma\left(H+\frac{1}{2}\right)},$$

whose value was calculated e.g. in [MIS 08]. Now, consider the Wiener integral

$$B_t^H := C_H^{(1)} \int_{\mathbb{R}} k_H^+(t,u)dW_u, \quad t \in \mathbb{R}. \qquad [3.16]$$

Note that the explicit representation of B_t^H is a little bit different for $t > 0$ and $t \leq 0$:

$$B_t^H = C_H^{(1)} \left(\int_{-\infty}^0 \left((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}\right) dW_u + \int_0^t (t-u)^{H-\frac{1}{2}} dW_u\right), t > 0,$$

and

$$B_t^H = C_H^{(1)} \left(\int_{-\infty}^t \left((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}}\right) dW_u + \int_t^0 (-u)^{H-\frac{1}{2}} dW_u\right), t < 0.$$

However, for any $t \in \mathbb{R}$, $\mathrm{E}(B_t^H)^2 = t^{2H}$. Indeed, for $t > 0$,

$$\mathrm{E}(B_t^H)^2 = (C_H^{(1)})^2 \left(\int_{-\infty}^0 \left((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}} \right)^2 du + \int_0^t (t-u)^{2H-1} du \right)$$
$$= (C_H^{(1)})^2 t^{2H} \left(\int_0^\infty \left((1+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}} \right)^2 du + \frac{1}{2H} \right) = 1,$$

since

$$\int_0^\infty \left((1+u)^{H-\frac{1}{2}} - u^{H-\frac{1}{2}} \right)^2 du + \frac{1}{2H} = \frac{\Gamma^2(H+\frac{1}{2})}{2H \sin(\pi H)\Gamma(2H)},$$

according to [MIS 08]. For $t < 0$,

$$\mathrm{E}(B_t^H)^2 = (C_H^{(1)})^2 \left(\int_{-\infty}^t \left((t-u)^{H-\frac{1}{2}} - (-u)^{H-\frac{1}{2}} \right)^2 du + \int_t^0 (-u)^{2H-1} du \right)$$
$$= (C_H^{(1)})^2 |t|^{2H} \left(\int_{-\infty}^1 \left((-1-z)^{H-\frac{1}{2}} - (-z)^{H-\frac{1}{2}} \right)^2 dz + \frac{1}{2H} \right)$$
$$= (C_H^{(1)})^2 |t|^{2H} \left(\int_1^\infty \left((z-1)^{H-\frac{1}{2}} - z^{H-\frac{1}{2}} \right)^2 dz + \frac{1}{2H} \right)$$
$$= (C_H^{(1)})^2 |t|^{2H} \left(\int_0^\infty \left(z^{H-\frac{1}{2}} - (z+1)^{H-\frac{1}{2}} \right)^2 dz + \frac{1}{2H} \right) = |t|^{2H}.$$

Furthermore, for $h > 0$, it holds that

$$B_{s+h}^H - B_s^H = C_H^{(1)} \left(\int_{\mathbb{R}} \left(k_H^+(s+h,u) - k^+(s,u) \right) dW_u \right),$$

and, for $0 < s < s+h$, we have that

$$\mathrm{E}\left(B_{s+h}^H - B_s^H\right)^2 = (C_H^{(1)})^2 \left(\int_{-\infty}^s \left((s+h-u)^{H-\frac{1}{2}} - (s-u)^{H-\frac{1}{2}} \right)^2 du \right.$$
$$\left. + \int_s^{s+h} (s+h-u)^{2H-1} du \right)$$
$$= (C_H^{(1)})^2 \left(\int_0^\infty \left((h+z)^{H-\frac{1}{2}} - z^{H-\frac{1}{2}} \right)^2 dz + \frac{h^{2H}}{2H} \right)$$
$$= (C_H^{(1)})^2 h^{2H} \left(\int_0^\infty \left((1+z)^{H-\frac{1}{2}} - z^{H-\frac{1}{2}} \right)^2 dz + \frac{1}{2H} \right) = h^{2H},$$

while for $-\infty < s < s+h < 0$, we have that

$$\mathrm{E}\left(B^H_{s+h} - B^H_s\right)^2 = (C^{(1)}_H)^2 \left(\int_{-\infty}^{s} \left((s+h-u)^{H-\frac{1}{2}} - (s-u)^{H-\frac{1}{2}} \right)^2 du \right.$$

$$\left. + \int_s^{s+h} \left((s+h-u)^{2H-1} \right)^2 du \right)$$

$$= (C^{(1)}_H)^2 \left(\int_0^{\infty} \left((h+z)^{H-\frac{1}{2}} - z^{H-\frac{1}{2}} \right)^2 dz + \frac{h^{2H}}{2H} \right) = h^{2H}.$$

The case $-\infty < s < 0 < s+h < \infty$ can be considered similarly. Finally, we get that the Gaussian process B^H introduced by the relation [3.16] has zero expectation and covariance function

$$\mathrm{E}B^H_t B^H_s = \frac{1}{2}\left(\mathrm{E}(B^H_t)^2 + \mathrm{E}(B^H_s)^2 - \mathrm{E}(B^H_t - B^H_s)^2 \right)$$

$$= \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right).$$

In connection with the above, we can introduce the following definition.

DEFINITION 3.14.– A two-sided Brownian motion $B^H = \left\{ B^H_t, t \in \mathbb{R} \right\}$ *is a centered Gaussian process with the covariance function*

$$\mathrm{E}B^H_t B^H_s = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right). \qquad [3.17]$$

The above calculations give us the following result.

THEOREM 3.9.– *The Mandelbrot–van Ness representation of the form*

$$B^H_t = \int_{\mathbb{R}} k^+_H(t,u) dW_u, \quad t \in \mathbb{R}, \qquad [3.18]$$

where $W = \{W_t, t \geq 0\}$ *is a two-sided Wiener process, gives us a two-sided fractional Brownian motion* B^H.

REMARK 3.5.–

i) Obviously, one-sided fBm admits representation [3.18] for $t \geq 0$.

ii) Principally, it is possible to check that [3.17] is indeed a covariance function similarly to the verification made in theorem 3.4. However, since we have an explicit representation [3.18] of fBm, it is not necessary to do so.

iii) It is possible to prove that, for any two-sided fractional Brownian motion B^H, there exists a two-sided Wiener process W for which [3.18] holds, see also remark 3.6.

3.7. Representation of fractional Brownian motion as the Wiener integral on the compact integral

Consider the so-called Molchan–Golosov representation of fBm as the Wiener integral on the compact integral ([MOL 69a, NOR 99, JOS 06]).

Introduce the kernel

$$\kappa_H(t,s) = c_H \left[t^{H-\frac{1}{2}} s^{\frac{1}{2}-H} (t-s)^{H-\frac{1}{2}} \right.$$
$$\left. - \left(H - \frac{1}{2}\right) s^{\frac{1}{2}-H} \int_s^t u^{H-\frac{3}{2}} (u-s)^{H-\frac{1}{2}} du \right],$$

where $c_H = \left(\frac{2H\Gamma\left(\frac{3}{2}-H\right)}{\Gamma(2-2H)\Gamma(H+\frac{1}{2})} \right)^{\frac{1}{2}}$.

For $H > \frac{1}{2}$, we can integrate by parts and reduce the kernel $\kappa_H(t,s)$ to

$$\kappa_H(t,s) = \left(H - \frac{1}{2}\right) c_H s^{\frac{1}{2}-H} \int_s^t u^{H-\frac{1}{2}} (u-s)^{H-\frac{3}{2}} du.$$

THEOREM 3.10.– *Let* $W = \{W_t, t \geq 0\}$ *be a Wiener process. Then the stochastic process*

$$B_t^H = \int_0^t \kappa_H(t,s) dW_s \qquad [3.19]$$

is a fractional Brownian motion.

PROOF.– Consider, for technical simplicity, the case $H > \frac{1}{2}$ and for any $t > 0$ denote $C_H = (H - \frac{1}{2}) c_H$. Then we can transform $\int_0^t \kappa_H^2(t,s) ds$:

$$\int_0^t \kappa_H^2(t,s) ds = C_H^2 \int_0^t s^{1-2H} \left(\int_s^t u^{H-\frac{1}{2}} (u-s)^{H-\frac{3}{2}} du \right)^2 ds$$
$$= C_H^2 \int_0^t s^{1-2H} \int_s^t u^{H-\frac{1}{2}} (u-s)^{H-\frac{3}{2}} du \int_s^t v^{H-\frac{1}{2}} (v-s)^{H-\frac{3}{2}} dv ds$$
$$= C_H^2 \int_0^t \int_0^t u^{H-\frac{1}{2}} v^{H-\frac{1}{2}} \int_0^{u \wedge v} s^{1-2H} (u-s)^{H-\frac{3}{2}} (v-s)^{H-\frac{3}{2}} ds\, du\, dv.$$

According to lemma 2.2 (i) from [NOR 99], for $\mu > 0, \nu > 0, c > 1$

$$\int_0^1 t^{\mu-1}(1-t)^{\nu-1}(c-t)^{-\mu-\nu}dt = c^{-\nu}(c-1)^{-\mu}B(\mu,\nu).$$

In this case, $\mu - 1 = 1-2H, \nu - 1 = H - \frac{3}{2}$, so that $-\nu = \frac{1}{2} - H, -\mu = 2H-2$. Therefore,

$$\begin{aligned}
&\int_0^{u\wedge v} s^{1-2H}(u-s)^{H-\frac{3}{2}}(v-s)^{H-\frac{3}{2}}ds \\
&= (u\wedge v)^{-1}\int_0^1 s^{1-2H}(1-s)^{H-\frac{3}{2}}\left(\frac{u\vee v}{u\wedge v} - s\right)^{H-\frac{3}{2}} ds \\
&= (u\wedge v)^{-1}\left(\frac{u\vee v}{u\wedge v}\right)^{\frac{1}{2}-H}\left(\frac{u\vee v}{u\wedge v} - 1\right)^{2-2H} B\left(2-2H, H-\frac{1}{2}\right).
\end{aligned}$$

We can continue with $\int_0^t \kappa_H^2(t,s)ds$:

$$\begin{aligned}
&\int_0^t \kappa_H^2(t,s)ds = C_H^2 B\left(2-2H, H-\frac{1}{2}\right)\times \\
&\times\left(\int_0^t u^{H-\frac{1}{2}}\int_0^u v^{H-\frac{1}{2}}v^{-1}\left(\frac{u}{v}\right)^{\frac{1}{2}-H}\left(\frac{u}{v}-1\right)^{2H-2} dv\, du\right. \\
&\left.+\int_0^t u^{H-\frac{1}{2}}\int_u^t v^{H-\frac{1}{2}}u^{-1}\left(\frac{v}{u}\right)^{\frac{1}{2}-H}\left(\frac{v}{u}-1\right)^{2H-2} du\, dv\right) \\
&= C_H^2 B\left(2-2H, H-\frac{1}{2}\right)\left(\int_0^t\int_0^u (u-v)^{2H-2}dv\, du\right. \\
&\left.+\int_0^t\int_u^t (v-u)^{2H-2}dv\, du\right) \\
&= \frac{C_H^2 B\left(2-2H, H-\frac{1}{2}\right)}{H(2H-1)}t^{2H} < \infty.
\end{aligned}$$

Therefore, integral $B_t^H := \int_0^t \kappa_H(t,s)dW_s$ exists, and so $\mathrm{E}B_t^H = 0$. Moreover,

$$\frac{C_H^2 B\left(2-2H, H-\frac{1}{2}\right)}{H(2H-1)} = \frac{2H\Gamma\left(\frac{3}{2}-H\right)\Gamma(2-2H)\Gamma\left(H-\frac{1}{2}\right)\left(H-\frac{1}{2}\right)^2}{\Gamma(2-2H)\Gamma\left(H+\frac{1}{2}\right)H(2H-1)\Gamma\left(\frac{3}{2}-H\right)} = 1,$$

so, $\mathrm{E}(B_t^H)^2 = t^{2H}$. Applying the same transformations, we can calculate the covariance function. Let $t > s$, then

$$\mathrm{E}B_t^H B_s^H = C_H^2 \int_0^s \kappa(t,u)\kappa(s,u)du$$
$$= C_H^2 \int_0^s u^{1-2H} \int_u^t v^{H-\frac{1}{2}}(v-u)^{H-\frac{3}{2}}dv \int_u^s z^{H-\frac{1}{2}}(z-u)^{H-\frac{3}{2}}dz\, du$$
$$= C_H^2 \int_0^t \int_0^s v^{H-\frac{1}{2}} z^{H-\frac{1}{2}} \int_0^{z\wedge v} u^{1-2H}(v-u)^{H-\frac{3}{2}}(z-u)^{H-\frac{3}{2}}dz\, du\, dv$$
$$= C_H^2 \left(\int_0^s \int_0^s + \int_s^t \int_0^s\right) = s^{2H} + C_H^2 \int_s^t \int_0^s = s^{2H}$$
$$+C_H^2 B\left(2-2H, H-\frac{1}{2}\right) \int_s^t \int_0^s v^{H-\frac{1}{2}} z^{H-\frac{1}{2}} z^{-1} \left(\frac{v}{z}\right)^{\frac{1}{2}-H} \left(\frac{v}{z}-1\right)^{2H-2} dzdv$$
$$= s^{2H} + \frac{1}{2H(2H-1)} \int_s^t \int_0^s (v-z)^{2H-2}dzdv$$
$$= s^{2H} + \frac{1}{2}(t^{2H} - s^{2H} - (t-s)^{2H}) = \frac{1}{2}(s^{2H} + t^{2H} - |t-s|^{2H}).$$

The case $s > t$ can be considered similarly, and the proof follows. □

REMARK 3.6.– Any fractional Brownian motion admits representation [3.19] with some Wiener process. Indeed, let F be the Gauss hypergeometric function of parameters a, b, c and variable $z \in \mathbb{R}$ defined as

$$F(a,b,c,z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)} \int_0^1 v^{b-1}(1-v)^{c-b-1}(1-vz)^{-a}dv,$$

and $\widetilde{C}_H = \frac{(\Gamma(2-2H))^{\frac{1}{2}}}{2H\Gamma(H+\frac{1}{2})^{\frac{1}{2}}\Gamma(\frac{3}{2}-H)^{\frac{3}{2}}}$. It was established, e.g. in [JOS 06], that if we have the fractional Brownian motion $B^H = \{B_t^H, t \geq 0\}$, then for any $t > 0$ the integral

$$W_t = \widetilde{C}_H \int_0^t (t-s)^{\frac{1}{2}-H} F\left(\frac{1}{2}-H, \frac{1}{2}-H, \frac{3}{2}-H, \frac{s-t}{s}\right) dB_s^H,$$

exists and $W = \{W_t, t \geq 0\}$ is a Wiener process, according to which B^H admits the representation [3.19].

4

Construction, Properties and Some Functionals of the Wiener Process and Fractional Brownian Motion

4.1. Construction of a Wiener process on the interval $[0, 1]$

Consider interval $[0, 1]$ and the sequences of Haar and Schauder functions on this interval. Haar functions are constructed as follows:

$$H_0(t) = 1, \ \ t \in [0, 1]; \ \ H_1(t) = \mathbb{1}_{\left[0, \frac{1}{2}\right)}(t) - \mathbb{1}_{\left[\frac{1}{2}, 1\right)}(t);$$

$$H_2(t) = \sqrt{2}\left(\mathbb{1}_{\left[0, \frac{1}{4}\right)}(t) - \mathbb{1}_{\left[\frac{1}{4}, \frac{1}{2}\right)}(t)\right);$$

$$H_3(t) = \sqrt{2}\left(\mathbb{1}_{\left[\frac{1}{2}, \frac{3}{4}\right)}(t) - \mathbb{1}_{\left[\frac{3}{4}, 1\right)}(t)\right),$$

and, in general, we can divide the function into the groups, with functions $H_{2^{n-1}}(t), \dots, H_{2^n-1}(t)$ in the nth group, $n \geq 1$, and for $2^{n-1} \leq j \leq 2^n - 1$

$$H_j(t) = 2^{\frac{n-1}{2}}\left(\mathbb{1}_{\left[\frac{2j-2^n}{2^n}, \frac{2j+1-2^n}{2^n}\right)}(t) - \mathbb{1}_{\left[\frac{2j+1-2^n}{2^n}, \frac{2j+2-2^n}{2^n}\right)}(t)\right).$$

The functions are depicted in Figure 4.1.

Denote for brevity $\mathcal{L}_2([0, 1]) = \mathcal{L}_2([0, 1], \lambda_1)$, where λ_1 is the Lebesgue measure on $[0, 1]$. Evidently, for any $n \geq 1$,

$$H_n \in \mathcal{L}_2([0, 1]), \|H_n\|_{\mathcal{L}_2([0,1])} = 1 \text{ and } \langle H_n, H_k\rangle_{\mathcal{L}_2([0,1])} = 0 \text{ for } n \neq k.$$

Furthermore, Haar functions create an orthonormal basis in $\mathcal{L}_2([0, 1])$. Indeed, any indicator of dyadic interval $\left[\frac{k}{2^n}, \frac{l}{2^n}\right)$ can be presented as a linear combination of

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

Haar functions from nth series, and the family of such indicators is tight in $\mathcal{L}_2([0,1])$. Therefore, any function $f \in \mathcal{L}_2([0,1])$ can be presented as

$$f(t) = \sum_{n=0}^{\infty} \langle f, H_n \rangle_{\mathcal{L}_2([0,1])} H_n(t),$$

where

$$\langle f, g \rangle_{\mathcal{L}_2([0,1])} = \int_0^1 f(t)g(t)dt = \sum_{n=0}^{\infty} \langle f, H_n \rangle_{\mathcal{L}_2([0,1])} \langle g, H_n \rangle_{\mathcal{L}_2([0,1])}.$$

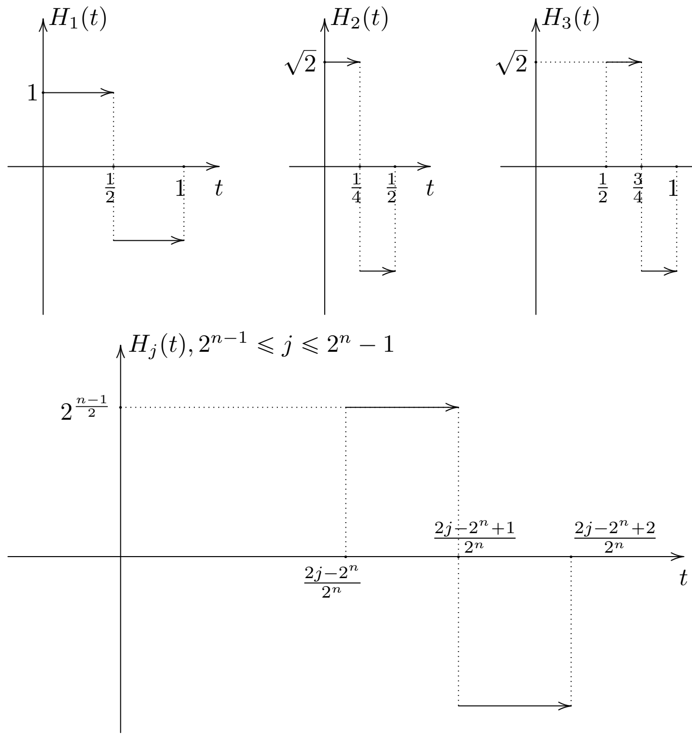


Figure 4.1. *Haar functions*

Schauder functions are defined as follows:

$$S_j(t) = \int_0^t H_j(s)ds = \left\langle H_j, \mathbb{1}_{[0,t]} \right\rangle_{\mathcal{L}_2([0,1])}, \ \ t \in [0,1].$$

However, a simpler description is as follows:

$$S_0(t) = t; \;\; S_1(t) = t \cdot \mathbb{1}_{0 \le t \le \frac{1}{2}} + (1-t)\mathbb{1}_{\frac{1}{2} \le t \le 1};$$

$$S_j(t) = 2^{-\frac{n+1}{2}} S_1(2^{n-1}t - j + 2^{n-1}), \;\; 2^{n-1} \le j \le 2^n - 1.$$

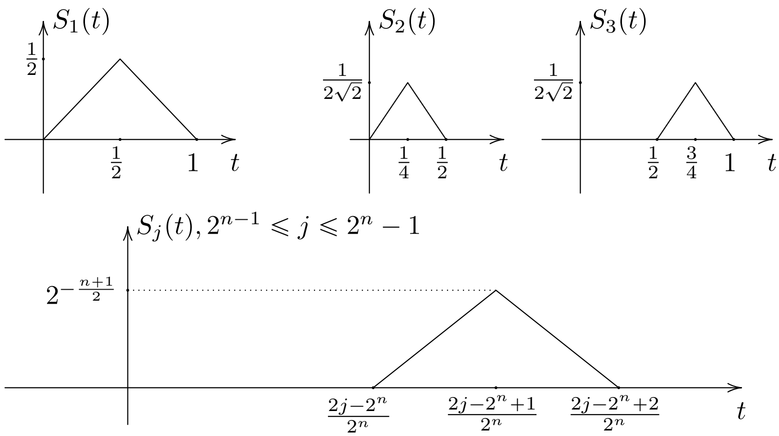


Figure 4.2. *Schauder functions*

In the ensuing considerations, we need the following two properties of Schauder functions:

i) For any $t \in [0,1]$, only one function $S_j(t)$ in the nth series, i.e. with some index $2^{n-1} \le j \le 2^n - 1$, is non-zero.

ii) $\max_{t\in[0,1]} S_j(t) = 2^{-\frac{n+1}{2}}$, $2^{n-1} \le j \le 2^n - 1$, $n \ge 1$.

Now, we can establish the following result.

THEOREM 4.1.– *Stochastic process $W = \{W_t, t \in [0,1]\}$, of the form*

$$W_t = \sum_{k=0}^{\infty} S_k(t)\xi_k, \;\; t \in [0,1], \qquad [4.1]$$

where $S_k = \{S_k(t), t \in [0,1]\}$ is a sequence of Schauder functions and $\{\xi_k, k \ge 0\}$ is the sequence of iid $\mathcal{N}(0,1)$ random variables, is a Wiener process. The series [4.1] *converges uniformly on* $[0,1]$ *with probability 1 and for any* $t \in [0,1]$ *in* $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$. *The trajectories of* W *are continuous a.s.*

PROOF.– We divide the proof into several steps.

1) First, establish convergence in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$. To this end, observe that for $N > M \geq 1$

$$\mathrm{E}\left|\sum_{k=M+1}^{N} S_k(t)\xi_k\right|^2 = \sum_{k=M+1}^{N} S_k^2(t) = \sum_{k=M+1}^{N} \left\langle H_k, \mathbb{1}_{[0,t]}\right\rangle^2 \to 0$$

$$\text{as } M, N \to \infty \text{ because } \sum_{k=1}^{\infty} \left\langle H_k, \mathbb{1}_{[0,t]}\right\rangle^2 = \left\|\mathbb{1}_{[0,t]}\right\|^2_{\mathcal{L}_2([0,1])} = t.$$

It means that the sequence $\sum_{k=1}^{K} S_k(t)\xi_k, \quad K \geq 1$ is a Cauchy sequence in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$; therefore, it converges in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$, for any $t \in [0,1]$. Denote by W_t its limit in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$. Recall that the convergence in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ implies the weak convergence of finite-dimensional distributions which in turn is equivalent to the convergence of the characteristic functions. Therefore, for any $0 = t_0 < t_1 < \ldots < t_K$ and any vector $\overline{\lambda} = (\lambda_1, \ldots, \lambda_K)$,

$$\mathrm{E}\exp\left\{i\sum_{k=1}^{K} \lambda_k (W_{t_k} - W_{t_{k-1}})\right\}$$

$$= \lim_{M\to\infty} \mathrm{E}\exp\left\{i\sum_{k=1}^{K} \lambda_k \sum_{j=1}^{M} \xi_j \left\langle H_j, \mathbb{1}_{[t_{k-1},t_k)}\right\rangle_{\mathcal{L}_2([0,1])}\right\}$$

$$= \lim_{M\to\infty} \mathrm{E}\exp\left\{i\sum_{j=1}^{M} \xi_j \left(\sum_{k=1}^{K} \lambda_k \left\langle H_j, \mathbb{1}_{[t_{k-1},t_k)}\right\rangle_{\mathcal{L}_2([0,1])}\right)\right\}$$

$$= \lim_{M\to\infty} \exp\left\{-\frac{1}{2}\sum_{j=1}^{M} \left(\sum_{k=1}^{K} \lambda_k \left\langle H_j, \mathbb{1}_{[t_{k-1},t_k)}\right\rangle_{\mathcal{L}_2([0,1])}\right)^2\right\}$$

$$= \lim_{M\to\infty} \exp\left\{-\frac{1}{2}\sum_{k,r=1}^{K} \lambda_k\lambda_r \sum_{j=1}^{M} \left\langle H_j, \mathbb{1}_{[t_{k-1},t_k)}\right\rangle_{\mathcal{L}_2([0,1])}\right.$$

$$\left.\times \left\langle H_j, \mathbb{1}_{[t_{r-1},t_r)}\right\rangle_{\mathcal{L}_2([0,1])}\right\}$$

$$= \exp\left\{-\frac{1}{2}\sum_{k,r=1}^{K}\lambda_k\lambda_r\sum_{j=1}^{\infty}\left\langle H_j, \mathbb{1}_{[t_{k-1},t_k)}\right\rangle_{\mathcal{L}_2([0,1])}\right.$$

$$\left.\times\left\langle H_j, \mathbb{1}_{[t_{r-1},t_r)}\right\rangle_{\mathcal{L}_2([0,1])}\right\}$$

$$= \exp\left\{-\frac{1}{2}\sum_{k,r=1}^{K}\lambda_k\lambda_r\left\langle \mathbb{1}_{[t_{k-1},t_k)}, \mathbb{1}_{[t_{r-1},t_r)}\right\rangle_{\mathcal{L}_2([0,1])}\right\}$$

$$= \exp\left\{-\frac{1}{2}\sum_{k=1}^{K}\lambda_k^2(t_k - t_{k-1})\right\}.$$

These equalities can be read as

$$\mathrm{E}\exp\left\{i\sum_{k=1}^{K}\lambda_k(W_{t_k} - W_{t_{k-1}})\right\} = \prod_{k=1}^{K}\mathrm{E}\exp\left\{i\lambda_k(W_{t_k} - W_{t_{k-1}})\right\}$$

$$= \prod_{k=1}^{K}\exp\left\{-\frac{1}{2}\lambda_k^2(t_k - t_{k-1})\right\},$$

and it means that W is a process with independent increments and the increment $W_t - W_s = \mathcal{N}(0, t-s)$. Therefore, W is a Wiener process.

2) Second, prove that the series converges uniformly a.s. Indeed, for any $x > 0$ and $\xi = \mathcal{N}(0,1)$, we have that

$$\mathrm{P}\{|\xi| > x\} = 2\mathrm{P}\{\xi > x\} = 2\frac{1}{\sqrt{2\pi}}\int_x^{\infty} e^{-\frac{y^2}{2}}\,dy$$

$$= \sqrt{\frac{2}{\pi}}\left(-\left.\frac{e^{-\frac{y^2}{2}}}{y}\right|_x^{\infty} - \int_x^{\infty}\frac{e^{-\frac{y^2}{2}}}{y^2}\,dy\right) \le \sqrt{\frac{2}{\pi}}\frac{e^{-\frac{x^2}{2}}}{x}.$$

Therefore, for any $k \ge 1$ and for $\{\xi_j, j \ge 1\}$ consisting of $\mathcal{N}(0,1)$ random variables,

$$\mathrm{P}\left\{\max_{1\le j\le k}|\xi_j| > x\right\} \le \sum_{j=1}^{k}\mathrm{P}\{|\xi_j| > x\} \le k\sqrt{\frac{2}{\pi}}\frac{\exp\{-\frac{x^2}{2}\}}{x}.$$

In particular,

$$\mathrm{P}\left\{\max_{1\le j\le k}|\xi_j|\ge k^{\frac{1}{3}}\right\}\le k^{\frac{2}{3}}\sqrt{\frac{2}{\pi}}\exp\left\{-\frac{k^{\frac{2}{3}}}{2}\right\}.$$

As $\sum_{k=1}^{\infty} k^{\frac{2}{3}}\exp\left\{-\frac{k^{\frac{2}{3}}}{2}\right\}<\infty$, we get from the Borel–Cantelli lemma that, for any $\omega\in\Omega^{'}$, $\mathrm{P}\{\Omega^{'}\}=1$, there exists $k=k(\omega)$ that, for $k\ge k(\omega)$,

$$\max_{1\le j\le k}|\xi_j|\le k^{\frac{1}{3}}.$$

Therefore, for $N\ge 1+\log_2 k(\omega)$, we have that $2^{N-1}\ge k(\omega)$ and, for $2^{N-1}\le j\le 2^N-1$, we have that $|\xi_j|\le\left(2^N-1\right)^{\frac{1}{3}}\le 2^{\frac{N}{3}}$. Moreover, for $2^{N-1}\le j\le 2^N-1$ and for any $0\le t\le 1$, only one Schauder function with such index j is non-zero, and additionally, it does not exceed $2^{-\frac{N+1}{2}}$. Finally, for $2^{N-1}\ge k(\omega)$, we get the bound

$$\sum_{j=2^{N-1}}^{2^N-1}S_j(t)|\xi_j|\le 2^{\frac{N}{3}}2^{-\frac{N+1}{2}}\le 2^{-\frac{N}{6}}.$$

The latter inequality implies that

$$\sum_{j=2^{N-1}}^{\infty}S_j(t)|\xi_j|\le\frac{2^{-\frac{N}{6}}}{1-2^{-\frac{N}{6}}}$$

and this upper bound does not depend on t. In turn, it means that the series $\sum_{j=1}^{\infty}S_j(t)\xi_j$ converges uniformly on $[0,1]$ and, consequently, the trajectories of W are continuous a.s. The theorem is proved. □

4.2. Construction of a Wiener process on $\mathbb{R}^+$

Let $W_0=\{W_0(t),t\in[0,1]\}$ be a Wiener process with a.s. continuous trajectories on $[0,1]$. Consider a sequence $\{W_n=\{W_n(t),t\in[0,1]\}\,n\ge 1\}$ of independent copies of W (the notion of independent processes was discussed in section 2.2.2). Define a stochastic process

$$W_t=W_0(t)\,\mathbb{1}_{t\in[0,1)}+\sum_{k=1}^{\infty}\left(W_0(1)+\ldots+W_{k-1}(1)+W_k(t-k)\right)\,\mathbb{1}_{t\in[k,k+1)}.$$

THEOREM 4.2.– *Stochastic process $W=\{W_t,t\ge 0\}$ is a Wiener process with a.s. continuous trajectories.*

PROOF.– Consider the points $0 = t_0 < t_1 < \ldots < t_N$ and suppose that

$$t_0, t_1, \ldots, t_{n_1} \in [0,1),\ t_{n_1+1}, \ldots, t_{n_2} \in [1,2),$$
$$\ldots,\ t_{n_N+1}, \ldots, t_{n_{N+1}} \in [N-1, N].$$

(The case where some unit intervals between 0 and N do not contain the points can be considered similarly). Then, for any $n_k + 1 \le p \le n_{k+1} - 1$,

$$W_{t_{p+1}} - W_{t_p} = W_k(t_{p+1} - k) - W_k(t_p - k),$$

therefore, these increments are mutually independent and are independent with any other increments that belong to the interiors of the intervals $[k, k+1)$. Furthermore, the increments between two neighbor points from different intervals have a form

$$W_{k+1}(t_{n_{k+1}+1} - k + 1) + W_k(1) - W_k(t_{n_k} - k)$$

and they are independent with any other increments of W_{k+1} and W_k. The situation when some neighbor points belong to different but not neighboring intervals is considered similarly. Therefore, W has independent increments and starts from 0.

Consider $0 \le s < t$ and let $s \in [k, k+1)$, $t \in [l, l+1)$, $k < l$. Then

$$\begin{aligned} W_t - W_s &= W_0(1) + \ldots + W_{l-1}(1) + W_l(t-l) \\ &\quad - (W_0(1) + \ldots + W_{k-1}(1) + W_k(s-k)) \\ &= (W_k(1) + \ldots + W_{l-1}(1)) + (W_l(t-l) - W_k(s-k)) \end{aligned}$$

and $\mathrm{E}(W_t - W_s) = 0$,

$$\begin{aligned} \mathrm{E}(W_t - W_s)^2 &= \mathrm{E}\left(W_k(1) - W_k(s-k)\right)^2 + \mathrm{E}\left(W_{k+1}(1)\right)^2 + \ldots + \mathrm{E}\left(W_{l-1}(1)\right)^2 \\ &\quad + \mathrm{E}\left(W_l(t-l)\right)^2 = (k+1-s) + (l-1-k) + (t-l) = t-s. \end{aligned}$$

Moreover, $W_t - W_s$ is a sum of independent Gaussian variables

$$(W_k(1) - W_k(s-k)), W_{k-1}(1), \ldots, W_{l+1}(1), W_l(t-l),$$

therefore, it is a Gaussian random variable. It means that W is a Wiener process. Continuity of the trajectories follows directly from its construction. The theorem is proved. □

4.3. Nowhere differentiability of the trajectories of a Wiener process

In the previous section, we established that there exists a Wiener process with continuous trajectories. However, the trajectories are irregular in the sense that almost all trajectories have no derivative at any point $t \in \mathbb{R}^+$. It is hard to depict such trajectories since there is no fixed direction of the trajectory at any point. The approximate form of such a trajectory is shown in Figure 4.3.

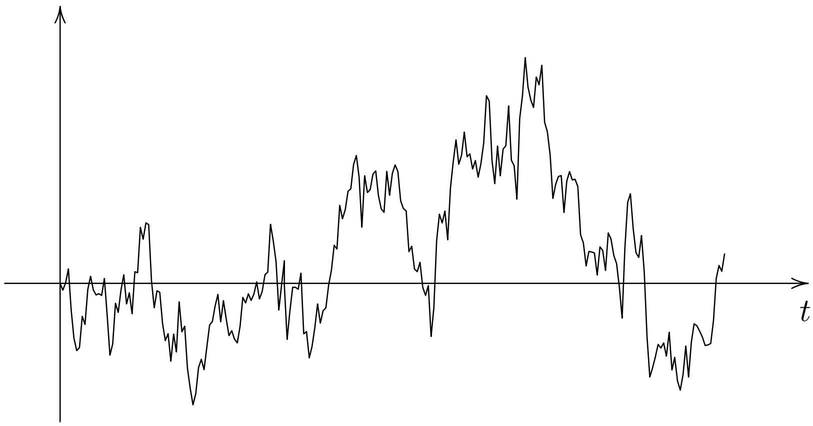


Figure 4.3. *Wiener process*

Sometimes the absence of derivative is argued in the following way. For any $t > 0$,

$$\mathrm{E}\left|\frac{W_{t+h} - W_t}{h}\right|^2 = \frac{1}{h} \to \infty, \quad \text{as} \;\; h \to 0.$$

It means that the Wiener process is not differentiable in the so-called mean-square sense; in other words, it means that the derivative, even if exists, is not square integrable. To establish non-differentiability of the trajectories, we need more subtle arguments. The result was originally proved in [PAL 33] by Paley, Wiener and Zygmund, and then proved in a more simple form in [DVO 61] by Dvoretzky, Erdös and Kakutani.

THEOREM 4.3.– *[Paley–Wiener–Zygmund–Dvoretzky–Erdös–Kakutani] Almost all trajectories of a Wiener process have everywhere a lower derivative $-\infty$ and upper derivative $+\infty$, i.e.*

$$\mathrm{P}\left\{\liminf_{h\to 0} \frac{W_{t+h} - W_t}{h} = -\infty \textit{ and } \limsup_{h\to 0} \frac{W_{t+h} - W_t}{h} = +\infty \textit{ for all } t \geq 0\right\} = 1.$$

REMARK 4.1.– For $t = 0$, we consider right-hand upper and lower derivatives.

PROOF.– For technical simplicity, we prove only the weaker result:

$$\mathrm{P}\left\{\limsup_{h\to 0}\left|\frac{W_{t+h}-W_t}{h}\right| = +\infty \text{ for all } t \geq 0\right\} = 1. \qquad [4.2]$$

Relation [4.2] is equivalent to the following one:

$$\mathrm{P}\left\{\liminf_{h\to 0}\frac{W_{t+h}-W_t}{h} = -\infty \text{ or } \limsup_{h\to 0}\frac{W_{t+h}-W_t}{h} = +\infty \text{ for all } t \geq 0\right\} = 1.$$

Introduce the event

$$B = \left\{\omega : \text{ there exists } t \geq 0 \text{ for which } \limsup_{h\to 0+}\left|\frac{W_{t+h}-W_t}{h}\right| < \infty\right\}.$$

Consider any $t \geq 0$. Let $t \in [m, m+1)$ for some $m \geq 0$. Assume that, for some $\omega \in B$,

$$\limsup_{h\to 0}\frac{|W_{t+h}-W_t|}{h} < \infty.$$

Then there exists $N \in \mathbb{N}$ such that for this $\omega \in B$

$$\limsup_{h\to 0}\frac{|W_{t+h}-W_t|}{h} \leq N.$$

Assume that

$$\limsup_{h\to 0+}\frac{|W_{t+h}-W_t|}{h} \leq N$$

(the case where $h \uparrow 0$ is considered similarly). Then there exists $j \geq 1$ such that $t + \frac{1}{j} \leq m + 1$ and for any $0 < h < \frac{1}{j}$ $|W_{t+h} - W_t| \leq Nh$.

Now, let $n \geq 1$ be such that $\frac{4}{n} \leq \frac{1}{j}$ and let $1 \leq k \leq n$ be such that $m + \frac{k-1}{n} \leq t \leq m + \frac{k}{n}$. Then $t \leq m + \frac{k+3}{n} \leq t + \frac{1}{j}$ and

$$\begin{aligned}
&|W_{m+\frac{k+1}{n}} - W_{m+\frac{k}{n}}| \leq |W_{m+\frac{k+1}{n}} - W_t| + |W_{m+\frac{k}{n}} - W_t| \\
&\leq N|m + \frac{k+1}{n} - t| + N|m + \frac{k}{n} - t| \leq N\frac{2}{n} + N\frac{1}{n} = \frac{3N}{n}, \\
&|W_{m+\frac{k+2}{n}} - W_{m+\frac{k+1}{n}}| \leq |W_{m+\frac{k+2}{n}} - W_t| + |W_{m+\frac{k+1}{n}} - W_t| \\
&\leq N|m + \frac{k+2}{n} - t| + N|m + \frac{k+1}{n} - t| \leq N\frac{3}{n} + N\frac{2}{n} = \frac{5N}{n},
\end{aligned}$$

and similarly,

$$\begin{aligned}
&|W_{m+\frac{k+3}{n}} - W_{m+\frac{k+2}{n}}| \le |W_{m+\frac{k+3}{n}} - W_t| + |W_{m+\frac{k+2}{n}} - W_t| \\
&\le N|m + \frac{k+3}{n} - t| + N|m + \frac{k+2}{n} - t| \le N\frac{4}{n} + N\frac{3}{n} = \frac{7N}{n}.
\end{aligned}$$

Introduce the events

$$\begin{aligned}
B_{N,k,n,m} = \Big\{\omega : &|W_{m+\frac{k+1}{n}} - W_{m+\frac{k}{n}}| \le \frac{3N}{n}, |W_{m+\frac{k+2}{n}} - W_{m+\frac{k+1}{n}}| \le \frac{5N}{n}, \\
&|W_{m+\frac{k+3}{n}} - W_{m+\frac{k+2}{n}}| \le \frac{7N}{n}\Big\}.
\end{aligned}$$

Then

$$B \subseteq \bigcup_{m=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcup_{j=1}^{\infty} \bigcap_{n=4j}^{\infty} \bigcup_{k=1}^{n} B_{N,k,n,m}.$$

Taking into account the independence of increments in the Wiener process W that are included in the event $B_{N,k,n,m}$, consider

$$\begin{aligned}
&\mathrm{P}\left\{\bigcap_{n=4j}^{\infty} \bigcup_{k=1}^{n} B_{N,k,n,m}\right\} \le \liminf_{n\to\infty} \mathrm{P}\left\{\bigcup_{k=1}^{n} B_{N,k,n,m}\right\} \le \liminf_{n\to\infty} \sum_{k=1}^{n} \mathrm{P}\left\{B_{N,k,n,m}\right\} \\
&\le \liminf_{n\to\infty} \sum_{k=1}^{n} \mathrm{P}\left\{|W_{m+\frac{k+1}{n}} - W_{m+\frac{k}{n}}| \le \frac{3N}{n}\right\} \mathrm{P}\left\{|W_{m+\frac{k+2}{n}} - W_{m+\frac{k+1}{n}}| \le \frac{5N}{n}\right\} \\
&\times \mathrm{P}\left\{|W_{m+\frac{k+3}{n}} - W_{m+\frac{k+2}{n}}| \le \frac{7N}{n}\right\} \\
&= \liminf_{n\to\infty} \sum_{k=1}^{n} \mathrm{P}\left\{\left|\mathcal{N}\left(0, \frac{1}{n}\right)\right| \le \frac{3N}{n}\right\} \mathrm{P}\left\{\left|\mathcal{N}\left(0, \frac{1}{n}\right)\right| \le \frac{5N}{n}\right\} \\
&\times \mathrm{P}\left\{\left|\mathcal{N}\left(0, \frac{1}{n}\right)\right| \le \frac{7N}{n}\right\} \\
&\le \liminf_{n\to\infty} n \cdot \mathrm{P}\left\{\left|\mathcal{N}\left(0, \frac{1}{n}\right)\right| \le \frac{7N}{n}\right\}^3 = \liminf_{n\to\infty} n \cdot \left(\frac{\sqrt{n}}{\sqrt{2\pi}} \int_{-\frac{7N}{n}}^{\frac{7N}{n}} e^{-\frac{x^2 n}{2}} dx\right)^3 \\
&= \left|x\sqrt{n} = y\right| = \liminf_{n\to\infty} n \cdot \left(\frac{1}{\sqrt{2\pi}} \int_{-\frac{7N}{\sqrt{n}}}^{\frac{7N}{\sqrt{n}}} e^{-\frac{y^2}{2}} dy\right)^3 \le \lim_{n\to\infty} n \cdot \frac{1}{2\pi^{\frac{3}{2}}} \frac{(14N)^3}{n^{\frac{3}{2}}} = 0.
\end{aligned}$$

Therefore,

$$\mathrm{P}(B) \leq \sum_{m=1}^{\infty} \sum_{N=1}^{\infty} \sum_{j=1}^{\infty} \mathrm{P}\left\{ \bigcap_{n=4j}^{\infty} \bigcup_{k=1}^{n} B_{N,k,n,m} \right\} = 0,$$

and the theorem is proved. □

4.4. Power variation of the Wiener process and of the fractional Brownian motion

4.4.1. *Ergodic theorem for power variations*

Let $X = \{X_t, t \geq 0\}$ be a stochastic process. Denote for any finite set of points $\pi = \{0 \leq t_1 < t_2 < \ldots < t_n\}$ and $p > 0$

$$\mathrm{Var}^{(p)}(X, \pi) = \sum_{k=0}^{n-1} \left| X_{t_{k+1}} - X_{t_k} \right|^p .$$

First, consider the case where $\pi = \pi_n = \{0, 1, \ldots, n\}$.

LEMMA 4.1.– *For any* $H \in (0,1)$,

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{Var}^{(p)}(B^H, \pi_n) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \left| B^H_{k+1} - B^H_k \right|^p = \frac{p}{2} \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{p}{2}\right)$$

with probability 1.

PROOF.– At first, note that the fractional Brownian motion with any Hurst index $H \in (0,1)$ has stationary increments (being not stationary itself). Indeed, for any $0 \leq t_1 < t_2 \leq t_2 \leq t_3 < t_4$ and $h > 0$

$$\begin{aligned}
&\mathrm{E}(B^H_{t_2+h} - B^H_{t_1+h})(B^H_{t_4+h} - B^H_{t_3+h}) \\
&= \frac{1}{2} \big((t_2+h)^{2H} + (t_4+h)^{2H} - |t_4 - t_2|^{2H} - (t_1+h)^{2H} \\
&- (t_4+h)^{2H} + |t_4 - t_1|^{2H} - (t_2+h)^{2H} - (t_3+h)^{2H} \\
&+ |t_3 - t_2|^{2H} + (t_1+h)^{2H} + (t_3+h)^{2H} - |t_3 - t_1|^{2H} \big) \\
&= \frac{1}{2} \left((t_4 - t_1)^{2H} + (t_3 - t_2)^{2H} - (t_4 - t_2)^{2H} - (t_3 - t_1)^{2H} \right),
\end{aligned}$$

and the last value does not depend on h. Since, for any $k > 1$, the common distribution of $(B^H_{t_2+h} - B^H_{t_1+h}, B^H_{t_3+h} - B^H_{t_2+h}, \ldots, B^H_{t_k+h} - B^H_{t_{k-1}+h})$ depends only on the covariance matrix

$$C_{ij}(h) = \mathrm{E}(B^H_{t_j+h} - B^H_{t_{j-1}+h})(B^H_{t_k+h} - B^H_{t_{k-1}+h})$$
$$= \mathrm{E}(B^H_{t_j} - B^H_{t_{j-1}})(B^H_{t_k} - B^H_{t_{k-1}}) = C_{ij}(0),$$

we get the equality of the distributions:

$$(B^H_{t_2+h} - B^H_{t_1+h}, B^H_{t_3+h} - B^H_{t_2+h}, \ldots, B^H_{t_k+h} - B^H_{t_{k-1}+h})$$
$$\overset{\mathrm{d}}{=} (B^H_{t_2} - B^H_{t_1}, B^H_{t_3} - B^H_{t_2}, \ldots, B^H_{t_k} - B^H_{t_{k-1}}).$$

Now, consider the stationary sequence $(B^H_1, B^H_2 - B^H_1, \ldots, B^H_n - B^H_{n-1}, \ldots)$. Using the generalized formula for the power binomial function,

$$(1+x)^\alpha = 1 + \alpha x + \frac{\alpha(\alpha-1)}{2}x^2 + o(x^2),$$

we can calculate the following value of the covariance function:

$$R(n) := \mathrm{E}B^H_1(B^H_n - B^H_{n-1}) = \frac{1}{2}(1 + n^{2H} - (n-1)^{2H} - 1 - (n-1)^{2H}$$
$$+(n-2)^{2H}) = \frac{1}{2}(n^{2H} + (n-2)^{2H} - 2(n-1)^{2H})$$
$$= \frac{1}{2}n^{2H}\left(1 + \left(1 - \frac{2}{n}\right)^{2H} - 2\left(1 - \frac{1}{n}\right)^{2H}\right)$$
$$= \frac{1}{2}n^{2H}\left(1 + 1 - 2H\frac{2}{n} + \frac{2H(2H-1)}{2}\frac{4}{n^2} + o\left(\frac{1}{n^2}\right)\right.$$
$$\left. -2\left(1 - 2H\frac{1}{n} + \frac{2H(2H-1)}{2}\frac{1}{n^2} + o\left(\frac{1}{n^2}\right)\right)\right)$$
$$= \frac{1}{2}n^{2H}\left(-\frac{4H}{n} + \frac{4H}{n} + 4H(2H-1)\frac{1}{n^2} - 2H(2H-1)\frac{1}{n^2} + o\left(\frac{1}{n^2}\right)\right)$$
$$= \frac{1}{2}n^{2H}2H(2H-1)\frac{1}{n^2} + o\left(\frac{1}{n^{2-2H}}\right) \to 0 \text{ as } n \to \infty.$$

Therefore, according to theorem A2.15,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \left|B^H_{k+1} - B^H_k\right|^p \to \mathrm{E}|B^H_1|,$$

and the proof follows from lemma A2.3. □

4.5. Self-similar stochastic processes

4.5.1. *Definition of self-similarity and some examples*

Let $X = \{X_t, t \geq 0\}$ and $Y = \{Y_t, t \geq 0\}$ be two real-valued stochastic processes. We shall write $\{X_t\} \overset{\mathrm{d}}{=} \{Y_t\}$ if they have the same finite-dimensional distributions.

DEFINITION 4.1.– *We say that the process X is* self-similar *if, for any $a > 0$, there exists $b > 0$, such that*

$$\{X_{at}\} \overset{\mathrm{d}}{=} \{bX_t\}. \qquad [4.3]$$

THEOREM 4.4.– *Let $X = \{X_t, t \geq 0\}$ be a non-trivial stochastically continuous at zero and self-similar stochastic process. Then there exists a unique $H \geq 0$, such that b in* [4.3] *has a form $b = a^H$.*

PROOF.– Let $a > 0$ be given and for some $t > 0$ $X_{at} \overset{\mathrm{d}}{=} bX_t$. As X_t is non-trivial, b is uniquely determined by a, so is a function of a, denoted as $b(a)$. Then

$$X_{aa_1t} \overset{\mathrm{d}}{=} b(a)X_{a_1t} \overset{\mathrm{d}}{=} b(a)b(a_1)X_t.$$

Therefore,

$$b(aa_1) = b(a)b(a_1). \qquad [4.4]$$

Now, let $a < 1$. We have that $X_{a^nt} \overset{\mathrm{d}}{=} (b(a))^nX_t$, while $a^nt \to 0+$. As X is stochastically continuous at zero, we conclude that $b(a) \leq 1$. Further, $b\left(\frac{a_1}{a_2}\right) = \frac{b(a_1)}{b(a_2)}$ and $\frac{b(a_1)}{b(a_2)} \leq 1$ for $a_1 \leq a_2$. Therefore, $b(a)$ is a non-decreasing function satisfying [4.4]. Therefore, $b(a)$ is a power function, $b(a) = a^H$ for some $H \geq 0$. □

THEOREM 4.5.– *Fractional Brownian motion with Hurst index $H \in (0, 1)$ is a self-similar process with $b = a^H$, according to definition 4.1.*

PROOF.– For any $a > 0$ and $0 \leq t_1 < t_2$,

$$\begin{aligned}\mathrm{E}B^H_{at_1}B^H_{at_2} &= \frac{1}{2}\left((at_1)^{2H} + (at_2)^{2H} - |at_1 - at_2|^{2H}\right)\\ &= \frac{1}{2}a^{2H}\left(t_1^{2H} + t_2^{2H} - (t_2 - t_1)^{2H}\right) = a^{2H}\mathrm{E}B^H_{t_1}B^H_{t_2}.\end{aligned}$$

Therefore, the covariance matrix for vector $(B^H_{at_1}, B^H_{at_2}, \dots, B^H_{at_k})$ equals the covariance matrix for $(B^H_{t_1}, B^H_{t_2}, \dots, B^H_{t_k})$, multiplied by a^{2H}, and

$$\mathrm{E}\exp\left\{i\sum_{k=1}^{n}\lambda_k B^H_{at_k}\right\} = \mathrm{E}\exp\left\{i\sum_{k=1}^{n}\lambda_k a^H B^H_{t_k}\right\}. \qquad [4.5]$$

As the characteristic functions uniquely determine the distribution, the proof follows immediately from [4.5]. □

4.5.2. *Power variations of self-similar processes on finite intervals*

Now, consider any $p > 0$, fix interval $[0, T]$, introduce the sequence of partitions

$$\overline{\pi}_n = \left\{T\delta_k, \delta_k = \frac{k}{2^n}, 0 \le k \le 2^n\right\},$$

and let

$$\mathrm{Var}^{(p)}(B^H, \overline{\pi}_n, [0,T]) := \sum_{k=0}^{2^n-1}\left|B^H_{T\delta_{k+1}} - B^H_{T\delta_k}\right|^p.$$

Note that according to the self-similarity of fractional Brownian motion

$$\sum_{k=0}^{2^n-1}\left|B^H_{T\delta_{k+1}} - B^H_{T\delta_k}\right|^p \overset{\mathrm{d}}{=} \left(\frac{T}{2^n}\right)^{pH}\sum_{k=0}^{2^n-1}|B^H_{k+1} - B^H_k|^p.$$

consequently,

$$\frac{2^{n(pH-1)}}{T^{pH}}\sum_{k=0}^{2^n-1}\left|B^H_{T\delta_{k+1}} - B^H_{T\delta_k}\right|^p \overset{\mathrm{d}}{=} \frac{1}{2^n}\sum_{k=0}^{2^n-1}|B^H_{k+1} - B^H_k|^p.$$

Applying lemma 4.1, we get that

$$\frac{1}{2^n}\sum_{k=0}^{2^n-1}|B^H_{k+1} - B^H_k|^p \to \mathrm{E}|B^H_1|^p = \left(\frac{2^p}{\pi}\right)^{\frac{1}{2}}\Gamma\left(\frac{p}{2}\right)$$

in probability. The almost sure convergence can be shown as in proposition 2.1 of [DOZ 14], so we have the following result.

THEOREM 4.6.– *For any $T > 0$, $p > 0$ and $H \in (0,1)$,*

$$\frac{2^{n(pH-1)}}{T^{pH}} \mathrm{Var}^{(p)}(B^H, \overline{\pi}_n, [0,T]) \to \left(\frac{2^p}{\pi}\right)^{\frac{1}{2}} \Gamma\left(\frac{p}{2}\right) \text{ a.s. as } n \to \infty.$$

COROLLARY 4.1.–

i) Let $p > \frac{1}{H}$. Then

$$\mathrm{Var}^{(p)}(B^H, \overline{\pi}_n, [0,T]) \to 0 \text{ a.s. as } n \to \infty.$$

Let $p < \frac{1}{H}$. Then

$$\mathrm{Var}^{(p)}(B^H, \overline{\pi}_n, [0,T]) \to \infty \text{ a.s. as } n \to \infty.$$

Let $p = \frac{1}{H}$. Then

$$\mathrm{Var}^{\left(\frac{1}{H}\right)}(B^H, \overline{\pi}_n, [0,T]) \to T \text{ as } n \to \infty.$$

ii) Let $H = \frac{1}{2}$, $p = 2$. Then $B^H = W$, a Wiener process, and we have that

$$\sum_{k=0}^{2^n-1} \left(W_{T\delta_{k+1}} - W_{T\delta_k}\right)^2 \to T \text{ a.s. as } n \to \infty.$$

What about non-uniform partitions? Consider any sequence of partitions

$$\pi_n = \left\{0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_{k_n}^{(n)} = T\right\}$$

such that

$$|\pi_n| = \max_{1 \le k \le k_n} |t_k^{(n)} - t_{k-1}^{(n)}| \to 0.$$

The ergodic theorem does not work in this case. Consider the Wiener process and establish at first the convergence of quadratic variations in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$.

THEOREM 4.7.– *The sequence of quadratic variations*

$$S_n := \sum_{k=0}^{n-1} \left(W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}}\right)^2 \to T \text{ in } \mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P}).$$

PROOF.– Consider

$$E(S_n - T)^2 = E\left(\sum_{k=0}^{n-1}\left((\Delta W_k)^2 - \Delta t_k\right)\right)^2,$$

where

$$\Delta W_k = W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}}, \quad \Delta t_k = t_k^{(n)} - t_{k-1}^{(n)}.$$

We have that, for $k \neq j$,

$$E\left((\Delta W_k)^2 - \Delta t_k\right)\left((\Delta W_j)^2 - \Delta t_j\right) = 0,$$

therefore,

$$\begin{aligned}
E(S_n - T)^2 &= E\left(\sum_{k=0}^{n-1}\left((\Delta W_k)^2 - \Delta t_k\right)\right)^2 = \sum_{k=0}^{n-1} E\left((\Delta W_k)^2 - \Delta t_k\right)^2 \\
&= \sum_{k=0}^{n-1} E(\Delta W_k)^4 - 2\sum_{k=0}^{n-1} \Delta t_k E(\Delta W_k)^2 + \sum_{k=0}^{n-1}(\Delta t_k)^2 \\
&= 3\sum_{k=0}^{n-1}(\Delta t_k)^2 - 2\sum_{k=0}^{n-1}\Delta t_k \cdot \Delta t_k + \sum_{k=0}^{n-1}(\Delta t_k)^2 \\
&= 2\sum_{k=0}^{n-1}(\Delta t_k)^2 \le 2|\pi_n| \cdot T \to 0, \; n \to \infty.
\end{aligned}$$

□

REMARK 4.2.– By using similar calculations, it is easy to prove that, for any $p \in \mathbb{N}$, there exists $C_p > 0$ such that $E(S_n - T)^{2p} \le C_p|\pi_n|^p$. Now, let $|\pi_n| = O(n^{-\lambda})$ for some $\lambda > 0$. Then it is possible to prove that $S_n \to T$ a.s. as $n \to \infty$. Indeed, for any $\varepsilon > 0$ and any $\alpha > 0$, there exists $p \in \mathbb{N}$ such that $|\pi_n|^p = O(n^{-1-\alpha})$ for some $\alpha > 0$. Then the series

$$\begin{aligned}
\sum_{n=1}^{\infty} P|S_n - T| > \varepsilon) &\le \varepsilon^{-2}\sum_{n=1}^{\infty} E(S_n - T)^{2p} \le C_p\varepsilon^{-2}\sum_{n=1}^{\infty}|\pi_n|^p \\
&\le C_p\varepsilon^{-2}\sum_{n=1}^{\infty} n^{-1-\alpha}
\end{aligned}$$

converges, and the almost sure convergence follows from the Borel–Cantelli lemma.

REMARK 4.3.– If for some stochastic process $X = \{X_t, t \in [0,T]\}$, for any $t \in [0,T]$ and for any sequence $\left\{\pi_n(t) = 0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_{k_n}^{(n)} = t\right\}$ of partitions such that $|\pi_n(t)| \to 0$, we have that $\sum_{k=0}^{n-1}\left(X_{t_k^{(n)}} - X_{t_{k-1}^{(n)}}\right)^2$ has a limit $[X]_t$ in probability, then we say that X has a quadratic variation $[X] = \{[X]_t, t \in [0,T]\}$. Evidently, $[X]_t$ is a non-decreasing process on $[0,T]$. Theorem 4.7 states that the Wiener process W has quadratic variation $[W]_t = t$, for any $t \geq 0$. Using theorem 4.6, it is possible to prove similarly to Theorem 4.7 that, for any $H \in \left(0, \frac{1}{2}\right)$, the quadratic variation of B^H is infinite and, for $H \in \left(\frac{1}{2}, 1\right)$, it equals zero.

5

Martingales and Related Processes

5.1. Notion of stochastic basis with filtration

Consider a probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Let a family $\{\mathcal{F}_t, t \geq 0\}$ of σ-fields satisfy the following assumptions:

A) i) For any $0 \leq s < t$

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}.$$

ii) For any $t \geq 0$

$$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s \quad \text{(continuity “from the right”).}$$

iii) $\mathcal{F}_0$ contains all the sets from $\mathcal{F}$ of zero P-measure.

DEFINITION 5.1.– *The family* $\{\mathcal{F}_t, t \geq 0\}$ *satisfying assumptions* (A)*, is called a flow of* σ*-fields, or a* filtration.

REMARK 5.1.– We can define filtration for the discrete time: the family $\{\mathcal{F}_n, n \geq 0\}$ of σ-fields is called a filtration if, for any $0 \leq m < n$, $\mathcal{F}_m \subset \mathcal{F}_n \subset \mathcal{F}$ and $\mathcal{F}_0$ contain all the sets from $\mathcal{F}$ of zero P-measure.

REMARK 5.2.– The notion of filtration reflects the fact that information is increasing in time: the more time passed, the more events we could observe, and the richer the corresponding σ-field. Continuity “from the right” means that each σ-field $\mathcal{F}_t$ is sufficiently rich to contain all “future sprouts”, and condition (iii) means the completeness of all σ-fields.

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

Sometimes, the collection $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathrm{P})$ or $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 0}, \mathrm{P})$ is called a *stochastic basis* with filtration.

DEFINITION 5.2.– *Stochastic process* $X = \{X_t, t \geq 0\}$ $(X = \{X_n, n \geq 0\})$ *is said to be* adapted *to the filtration* $\{\mathcal{F}_t\}_{t\geq 0}$ $(\{\mathcal{F}_n\}_{n\geq 0})$ *if, for any* $t \geq 0$*,* X_t *is* $\mathcal{F}_t$*-measurable* (*for any* $n \geq 0$*,* X_n *is* $\mathcal{F}_n$*-measurable*).

If we write $\{X_t, \mathcal{F}_t, t \geq 0\}$, then it means that X is $\mathcal{F}$-adapted.

REMARK 5.3.– Adaptedness of a stochastic process means that, for any moment of time, the values of the process "agree" with the information available up to this moment of time.

Consider also the notion of predictability, but only for discrete-time process.

DEFINITION 5.3.– *Let* $\{\mathcal{F}_n\}_{n\geq 0}$ *be a filtration. A stochastic process* $X = \{X_n, n \geq 0\}$ *is called predictable w.r.t. this filtration if* X_0 *is a constant, and for any* $n \geq 1$*,* X_n *is a* $\mathcal{F}_{n-1}$*-measurable random variable.*

Throughout this chapter, we consider the phase spaces $\mathcal{S} = \mathbb{R}$ or $\mathbb{R}^d, d > 1$.

REMARK 5.4.– Let $X = \{X_t, t \geq 0\}$ be a stochastic process. Similarly to definition 1.9, we can define σ-algebra $\mathcal{F}_t^X$ generated by the process X restricted to the interval $[0, t]$. According to corollary 1.2, σ-algebra $\mathcal{F}_t^X$ is the smallest σ-algebra containing the sets $\{\omega \in \Omega : X(t_1, \omega) \in A_1, \ldots, X(t_k, \omega) \in A_k\}$, $A_i \subset \mathbb{R}$, $A_i \in \mathcal{B}(\mathbb{R})$, $t_i \leq t, 1 \leq i \leq k$. We denote it by $\mathcal{F}_t^X = \sigma\{X_s, s \leq t\}$ and say that $\left\{\mathcal{F}_t^X\right\}_{t\geq 0}$ is a *natural filtration* generated by process X. Any stochastic process is adapted to its natural filtration. Moreover, if X is adapted to $\{\mathcal{F}_t\}_{t\geq 0}$, then $\mathcal{F}_t^X \subset \mathcal{F}_t$ for $t \geq 0$.

5.2. Notion of (sub-, super-) martingale: elementary properties

Let $\mathbb{T}$ be a set with a linear order. It can be $\mathbb{R}^+$ or $\mathbb{Z}^+ = \mathbb{N} \cup \{0\}$. Let $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{T}}, \mathrm{P}\right)$ be a stochastic basis with filtration.

DEFINITION 5.4.– *A stochastic process* $\{X_t, t \in \mathbb{T}\}$ *is said to be a* martingale *w.r.t. a filtration* $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ *if it satisfies the following three conditions.*

i) For any $t \in \mathbb{T}$*, the random variable* $X_t \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$ *(this means that the process* X *is integrable on* $\mathbb{T}$*).*

ii) For any $t \in \mathbb{T}$*,* X_t *is* $\mathcal{F}_t$*-measurable, so the process* X *is* $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$*-adapted.*

iii) For any $s, t \in \mathbb{T}$ *such that* $s \leq t$*, it holds that* $\mathrm{E}(X_t|\mathcal{F}_s) = X_s$ P*-a.s.*

If we change in condition (iii) the sign $=$ for $\geq$ and obtain $\mathrm{E}(X_t|\mathcal{F}_s) \geq X_s$ P-a.s. for any $s \leq t$, we get the definition of a *submartingale*; if $\mathrm{E}(X_t|\mathcal{F}_s) \leq X_s$ P-a.s. for any $s \leq t$, $s,t \in \mathbb{T}$, then we have a *supermartingale*. A vector process is called (sub-, super-) martingale if the corresponding property has each of its components. Evidently, any martingale is a (sub-, super) martingale. If X is a submartingale, then $-X$ is a supermartingale and vice versa.

LEMMA 5.1.–

1) Each (sub-, super-) martingale has the same property w.r.t. its natural filtration.

2) Property (iii) is equivalent to the following one: for any $s \leq t$, $s,t \in \mathbb{T}$ $\mathrm{E}(X_t - X_s|\mathcal{F}_s) = 0$ (≥ 0, ≤ 0 for (sub-, super) martingales).

3) If $\mathbb{T} = \mathbb{Z}^+$, then property (iii) is equivalent to the following one: for any $n \geq 0$ $\mathrm{E}(X_{n+1}|\mathcal{F}_n) = X_n$ or, that is the same, $\mathrm{E}(X_{n+1} - X_n|\mathcal{F}_n) = 0$.

PROOF.– 1) Let $\{X_t, \mathcal{F}_t, t \in \mathbb{T}\}$ be a martingale. (Sub- and supermartingales can be considered similarly.) Then X_t is $\mathcal{F}_t^X$-measurable for any $t \in \mathbb{T}$, and $\mathrm{E}(X_t|\mathcal{F}_s^X) = \mathrm{E}(\mathrm{E}(X_t|\mathcal{F}_s)|\mathcal{F}_s^X) = \mathrm{E}(X_s|\mathcal{F}_s^X) = X_s$, because $\mathcal{F}_t^X \subset \mathcal{F}_t$, as it was mentioned in remark 5.4.

Statement 2) is evident, and to establish 3), we only need to prove that if $\mathrm{E}(X_{n+1}|\mathcal{F}_n) = X_n$, then $\{X_n, \mathcal{F}_n, n \geq 0\}$ is a martingale. However, in this case, for any $n > m$

$$\mathrm{E}(X_n|\mathcal{F}_m) = \mathrm{E}(\mathrm{E}(X_n|\mathcal{F}_{n-1})|\mathcal{F}_m) = \mathrm{E}(X_{n-1}|\mathcal{F}_m) = \ldots = \mathrm{E}(X_{m+1}|\mathcal{F}_m) = X_m.$$

□

REMARK 5.5.–

i) It is very easy to check that $\mathrm{E}X_s = c$, if X is a martingale and $\mathrm{E}X_s$ increases (decreases) if X is a submartingale (supermartingale).

ii) Let $\{\xi_n, \mathcal{F}_n, n \geq 1\}$ be a sequence of integrable random variables for which $\mathrm{E}\{\xi_n|\mathcal{F}_{n-1}\} = 0, n \geq 1$. We say that the sequence is a *martingale difference*, or forms a martingale difference. Obviously, a stochastic process $\{X_n, \mathcal{F}_n, n \geq 0\}$, is a martingale if and only if $\{X_n - X_{n-1}, \mathcal{F}_n, n \geq 1\}$ is a martingale difference.

5.3. Examples of (sub-, super-) martingales

EXAMPLE 5.1.– **(Random walk).** *Let $\{\xi_i, i \geq 0\}$ be a sequence of integrable independent random variables. Consider $X_n = \sum_{i=0}^n \xi_i$, $\mathcal{F}_n = \sigma\{\xi_i, 0 \leq i \leq n\} = \sigma\{X_i, 0 \leq i \leq n\}$. Then*

$$\mathrm{E}(X_{n+1}|\mathcal{F}_n) = \mathrm{E}\left(\sum_{i=0}^n \xi_i + \xi_{n+1}|\mathcal{F}_n\right) = \sum_{i=0}^n \xi_i + \mathrm{E}\xi_{n+1} = X_n + \mathrm{E}\xi_{n+1}.$$

Therefore, in the case where $\mathrm{E}\xi_i = 0$*,* $i \geq 0$*,* $\{X_n, \mathcal{F}_n, n \geq 0\}$ *is a martingale, if* $\mathrm{E}\xi_i \geq 0 \ \ (\leq 0)$*,* $i \geq 0$*, then we have a sub- (super-) martingale with discrete time.*

EXAMPLE 5.2.– **(Process with independent increments).** *Let* $\{X_t, \mathcal{F}_t, t \geq 0\}$ *be an integrable process with independent increments,* $\mathrm{E}X_t = a_t$*. Then*

$$\mathrm{E}(X_t - X_s|\mathcal{F}_s) = \mathrm{E}X_t - \mathrm{E}X_s = a_t - a_s.$$

Therefore, X *is a martingale if* $a_t = a$*, i.e. the same for any* $t \geq 0$*, and* X *is a sub- (super-) martingale if* a_t *is increasing (decreasing) in* t*. In particular, Wiener process* W *is a martingale w.r.t. a natural filtration. Further, let* $N = \{N_t, t \geq 0\}$ *be a homogeneous Poisson process with parameter* λ *(recall that* $\lambda > 0$*). Then* $\mathrm{E}N_t = \lambda t$*, therefore* N *is a submartingale, and a compensated Poisson process* $N_t - \lambda t$ *is a martingale w.r.t. a natural filtration. In general, we see that the process* Y*, where* $Y_t = X_t - a_t$*, is a martingale w.r.t. a natural filtration.*

EXAMPLE 5.3.– **(Multiplicative martingale).** *Let* $\{\xi_i, i \geq 1\}$ *be a sequence of bounded random variables. Consider the process* $X_n = X_0 \prod_{i=1}^n (1 + \xi_i)$*, where* $X_0 \neq 0$ *is a constant. Let* $\mathcal{F}_n = \mathcal{F}_n^X$*. Then*

$$\mathrm{E}(X_{n+1}|\mathcal{F}_n^X) = X_n \mathrm{E}(1 + \xi_{n+1}|\mathcal{F}_n^X) = X_n(1 + \mathrm{E}(\xi_{n+1}|\mathcal{F}_n^X)).$$

We see that $\left\{X_n, \mathcal{F}_n^X, n \geq 0\right\}$ *is a martingale (so-called* multiplicative martingale*) if and only if* $\mathrm{E}\left\{\xi_{n+1}|\mathcal{F}_n^X\right\} = 0, n \geq 0$*, so that* $\xi_i, i \geq 1$ *create a martingale difference, see remark 5.5.*

EXAMPLE 5.4.– **(Likelihood ratio as a martingale).** *Let interval* $[0, T]$ *be fixed, and* $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in [0,T]}, \mathrm{P}\right)$ *be a stochastic basis with filtration. Let* $\mathrm{Q} \ll \mathrm{P}$ *be another probability measure on* $(\Omega, \mathcal{F})$*. Consider the restriction of measures* P *and* Q *on* $\mathcal{F}_t$ *and denote them by* P_t *and* Q_t*, respectively. Evidently,* $\mathrm{Q}_t \ll \mathrm{P}_t$*. Denote by*

$$X_t = \mathrm{E}\left(\frac{d\mathrm{Q}}{d\mathrm{P}}\bigg|\, \mathcal{F}_t\right) = \frac{d\mathrm{Q}_t}{d\mathrm{P}_t}$$

the corresponding Radon–Nikodym derivative that is also called likelihood ratio *or* density process*. Then, for any event* $A \in \mathcal{F}_s$*,* $s \leq t$*, we have that*

$$\int_A \frac{d\mathrm{Q}_t}{d\mathrm{P}_t} d\mathrm{P} = \int_A \frac{d\mathrm{Q}_t}{d\mathrm{P}_t} d\mathrm{P}_t = \mathrm{Q}_t(A) = \mathrm{Q}_s(A) = \int_A \frac{d\mathrm{Q}_s}{d\mathrm{P}_s} d\mathrm{P}_s = \int_A \frac{d\mathrm{Q}_s}{d\mathrm{P}_s} d\mathrm{P}. \quad [5.1]$$

Taking into account that, for any $t \geq 0$*,* $\frac{d\mathrm{Q}_t}{d\mathrm{P}_t}$ *is* $\mathcal{F}_t$*-measurable, we get from* [5.1] *that* $\mathrm{E}\left(\frac{d\mathrm{Q}_t}{d\mathrm{P}_t}\Big|\, \mathcal{F}_s\right) = \frac{d\mathrm{Q}_s}{d\mathrm{P}_s}$*. As a by-product, we get that* $\frac{d\mathrm{Q}_t}{d\mathrm{P}_t} = \mathrm{E}\left(\frac{d\mathrm{Q}}{d\mathrm{P}}\Big|\, \mathcal{F}_t\right)$*. Indeed, similarly to* [5.1]*, for any* $A \in \mathcal{F}_t$*,*

$$\int_A \frac{d\mathrm{Q}_t}{d\mathrm{P}_t} d\mathrm{P} = \mathrm{Q}_t(A) = \mathrm{Q}(A) = \int_A \frac{d\mathrm{Q}}{d\mathrm{P}} d\mathrm{P}.$$

Of course, we can consider the discrete time as well and conclude that for $\mathrm{Q} \ll \mathrm{P}$ $\mathrm{E}\left(\frac{d\mathrm{Q}}{d\mathrm{P}}\middle|\mathcal{F}_n\right) = \frac{d\mathrm{Q}_n}{d\mathrm{P}_n}$ *is a martingale w.r.t. the corresponding discrete-time filtration.*

EXAMPLE 5.5.– **(Geometric Brownian motion)**. *Let* $W = \{W_t, \mathcal{F}_t, t \geq 0\}$ *be a Wiener process,* $X_t = \exp\{at + \sigma W_t\}, a \in \mathbb{R}, \sigma > 0$. *Process* X *is called a* geometric Brownian motion. *Taking into account the independence of the increments of Wiener process on non-overlapping intervals, we get that*

$$\begin{aligned}\mathrm{E}(X_t|\mathcal{F}_s) &= \mathrm{E}\left(\exp\{as + \sigma W_s\}\exp\{a(t-s) + \sigma(W_t - W_s)\}|\mathcal{F}_s\right)\\ &= X_s \exp\{a(t-s)\}\,\mathrm{E}\exp\{\sigma(W_t - W_s)\}\\ &= X_s \exp\{a(t-s)\}\exp\left\{\frac{1}{2}\sigma^2(t-s)\right\}.\end{aligned}$$

Therefore,

$$\text{the process } X \text{ is a} \begin{cases} \text{martingale if } a + \frac{1}{2}\sigma^2 = 0;\\ \text{submartingale if } a + \frac{1}{2}\sigma^2 \geq 0;\\ \text{supermartingale if } a + \frac{1}{2}\sigma^2 \leq 0.\end{cases}$$

EXAMPLE 5.6.– **(Martingale transformation)**. *Consider an arbitrary martingale* $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ *and a bounded process* $\varphi = \{\varphi_n, n \geq 1\}$, *predictable w.r.t. the same filtration, with zero initial value,* $\varphi_0 = 0$. *Create an integral sum of the form* $S_n = \sum_{k=0}^{n-1} \varphi_k(X_{k+1} - X_k), S_0 = 0$. *Then* S *is an integrable adapted process, and*

$$\mathrm{E}(S_n - S_{n-1}|\mathcal{F}_{n-1}) = \varphi_n \mathrm{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = 0.$$

Therefore, S *is a martingale. It is called a* martingale transformation *of martingale* X.

EXAMPLE 5.7.– **(Lévy martingale)**. *Let* ξ *be an integrable random variable on some probability space and* $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ *be any filtration with arbitrary linearly ordered parameter set* $\mathbb{T}$ *on this probability space. Then* $X = \{X_t = \mathrm{E}(\xi|\mathcal{F}_t)\}$ *creates a martingale that is called the Lévy martingale.*

THEOREM 5.1.–

1) Let $\{X_t, \mathcal{F}_t, t \geq 0\}$ *be a martingale,* $f = f(x) : \mathbb{R} \to \mathbb{R}$ *be a convex function, and* $\mathrm{E}|f(X_t)| < \infty$ *for any* $t \geq 0$. *Then* $\{f(X_t), \mathcal{F}_t, t \geq 0\}$ *is a submartingale.*

2) Let $\{X_t, \mathcal{F}_t, t \geq 0\}$ *be a submartingale,* $f = f(x) : \mathbb{R} \to \mathbb{R}$ *be a convex increasing function, and* $\mathrm{E}|f(X_t)| < \infty$ *for any* $t \geq 0$. *Then* $\{f(X_t), \mathcal{F}_t, t \geq 0\}$ *is a submartingale.*

PROOF.– 1) Evidently, $f(X_t)$ is a $\mathcal{F}_t$-adapted and integrable process. Furthermore, it follows from Jensen's inequality for convex functions that, for any $0 \le s \le t$,

$$\mathrm{E}(f(X_t)|\mathcal{F}_s) \ge f(\mathrm{E}(X_t|\mathcal{F}_s)) = f(X_s).$$

2) Similar to the previous statement, $f(X_t)$ is a $\mathcal{F}_t$-adapted and integrable process. Furthermore, for any $0 \le s \le t$,

$$\mathrm{E}(f(X_t)|\mathcal{F}_s) \ge f(\mathrm{E}(X_t|\mathcal{F}_s)) = f(X_s),$$

since $\mathrm{E}(X_t|\mathcal{F}_s) \ge X_s$, and function f is increasing. □

EXAMPLE 5.8.–

i) Obviously, functions $f(x) = x^2$, $f(x) = |x|$, $f(x) = e^{-x}$ *are convex. Therefore, if* $\{X_t, \mathcal{F}_t, t \ge 0\}$ *is a martingale, then* $|X| = \{|X_t|, \mathcal{F}_t, t \ge 0\}$ *is a submartingale. If* $\mathrm{E}\exp\{-X_t\} < \infty$ *for any* $t \ge 0$*, then* $\exp\{-X\} = \{\exp\{-X_t\}, \mathcal{F}_t, t \ge 0\}$ *is a submartingale. If* X *is a square-integrable martingale, then* $X^2 = \left\{X_t^2, \mathcal{F}_t, t \ge 0\right\}$ *is a submartingale. Further, functions* $f(x) = (x-K)^+$*, where* $K \ge 0$ *is a constant, and* $f(x) = e^x$*, are convex and increasing. Therefore, if* X *is a submartingale, then* $(X-K)^+ = \{(X_t-K)^+, \mathcal{F}_t, t \ge 0\}$ *is a submartingale. If* X *is a submartingale and* $\mathrm{E}e^{X_t} < \infty$ *for any* $t > 0$*, then* $e^X = \left\{e^{X_t}, \mathcal{F}_t, t \ge 0\right\}$ *is a submartingale.*

ii) Let X *be a non-negative martingale,* $a > 0$*,* $f(x) = x \wedge a$*. Note that* f *is a concave function, and* $(-X)$ *is a non-positive martingale. Additionally,* $f(x) = -(x \wedge a)$ *is a bounded convex function in the range of* X*; therefore,* $-(X \wedge a)$ *is a submartingale whence* $X \wedge a$ *is a supermartingale. Obviously, we can consider any non-negative bounded concave function and conclude as above.*

5.4. Markov moments and stopping times

Let $\mathbb{T} = \mathbb{R}^+$ or $\mathbb{Z}^+$, $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \in \mathbb{T}}, \mathrm{P}\right)$ be a stochastic basis with filtration.

DEFINITION 5.5.–

1) Random variable $\tau = \tau(\omega) : \Omega \to \mathbb{T} \cup \{+\infty\}$ *is called* Markov moment *if, for any* $t \in \mathbb{T}$*, the event* $\{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$*.*

2) Markov moment $\tau = \tau(\omega)$ *is called a* stopping time *if* $\tau < \infty$ *a.s.*

3) The σ*-algebra generated by the Markov moment* τ *is the class of events*

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \le t\} \in \mathcal{F}_t, t \in \mathbb{T}\}.$$

DEFINITION 5.6.– *The Markov moment* $\tau = \tau(\omega)$ *is called* predictable *if there exists a sequence* $\{\tau_n, n \ge 1\}$ *of Markov moments, such that* *i)* $\tau_n(\omega)$ *is an increasing sequence a.s. and* $\lim_{n\to\infty} \tau_n(\omega) = \tau(\omega)$ *a.s.*

ii) For any $n \geq 1$*, it follows that* $\tau_n(\omega) < \tau(\omega)$ *a.s. on the set* $\{\tau(\omega) > 0\}$.

Sometimes, it is said that the sequence $\tau_n(\omega)$ described above *predicts* the Markov moment τ.

Now we can consider the general definition of predictable σ-algebra and predictable stochastic process.

DEFINITION 5.7.– *A* σ*-algebra is called predictable on* $\mathbb{T} \times \Omega$ *if it is generated by random intervals* $[\tau, \sigma) := \{(t, \omega) : \tau(\omega) \leq t < \sigma(\omega)\}$*, where* τ *and* σ *are predictable Markov moments.*

DEFINITION 5.8.– *A real-valued stochastic process* X *on a stochastic basis* $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{T}}, \mathrm{P})$ *with filtration is said to be predictable if the mapping* $X : \mathbb{T} \times \Omega \to \mathbb{R}$ *is measurable with respect to the predictable* σ*-algebra on* $\mathbb{T} \times \Omega$*.*

REMARK 5.6.– Note that, in discrete time, we cannot define predictable Markov moments because, in this case, conditions $\{\tau_n < \tau\}$ on $\{\tau > 0\}$ and $\tau_n \to \tau$ contradict each other.

Stochastic process $X = \{X_t, t \geq 0\}$, whose trajectories are a.s. left-continuous at any point $t > 0$ and continuous from the right at zero, is predictable w.r.t. the natural filtration. In particular, the process with a.s. continuous trajectories is predictable. For the proof, see e.g. theorem 7.2.4 and corollary 7.2.6 from [COH 15].

Introduce the notion of the process with càdlàg trajectories. It means "continue à droite avec des limites à gauche" in French, and English abbreviation is "corlol", "continuous on the right and with the limits on the left". See also definition A1.4, part (2).

DEFINITION 5.9.– *Stochastic process has a.s.* càdlàg *trajectories, or simply is càdlàg on some interval* $[0, T]$*, if with probability 1 its trajectories are continuous from the right and have the left limit at any interior point, and the trajectory is continuous from the right at the origin and is continuous from the left at* T*.*

EXAMPLE 5.9.– *Consider the examples of Markov moments for discrete and continuous time.*

i) Let $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ *be a real-valued process. Then, for any set* $A \in \mathcal{B}(\mathbb{R})$*, a random variable* $\tau = \inf\{n \geq 0 : X_n \in A\}$ *is a Markov moment because, for any* $k \geq 0$*, the event* $\{\omega \in \Omega : \tau(\omega) \leq k\} = \{X_0 \in A\} \cup \left(\bigcup_{i=1}^{k} \{X_0 \notin A, \ldots, X_{i-1} \notin A, X_i \in A\}\right)$ *and any event* $\{X_0 \notin A, \ldots, X_{i-1} \notin A, X_i \in A\} \in \mathcal{F}_i \subset \mathcal{F}_k$*.*

ii) Let $X = \{X_t, \mathcal{F}_t, t \geq 0\}$ be a real-valued right-continuous process with continuous time. In addition, let $A \in \mathcal{B}(\mathbb{R})$ be an open set. Then $\tau = \inf\{t \geq 0 : X_t \in A\}$ is a Markov moment. Indeed, in this case, the set $A^c = \mathbb{R} \setminus A$ is closed and

$$\{\tau > t\} = \{X_s \notin A, s \in [0,t]\} = \{X_s \in A^c, s \in [0,t]\}$$
$$= \bigcap_{n=1}^{\infty} \left\{X_{\frac{kt}{2^n}} \in A^c, 0 \leq k \leq 2^n\right\},$$

and this event belongs to $\mathcal{F}_t$, whence $\{\tau \leq t\} = \Omega \setminus \{\tau > t\} \in \mathcal{F}_t$.

iii) Let $A \in \mathcal{B}(\mathbb{R})$ be a closed set, and process X be continuous. Then $\tau = \inf\{t \geq 0 : X_t \in A\}$ is a Markov moment. Indeed, the event

$$\{\tau \leq t\} = \left\{\inf_{n\geq 1} \inf_{0\leq k\leq 2^n} \inf_{y\in A} |X_{\frac{kt}{2^n}} - y| = 0\right\} \in \mathcal{F}_t.$$

REMARK 5.7.– The last two examples can be significantly generalized. Namely, if X is a right-continuous adapted process, and $A \in \mathcal{B}(\mathbb{R})$, then the *hitting time* of A, $\tau_A = \inf\{t \geq 0 : X_t \in A\}$, is a stopping time, and this is the so-called *Debut theorem*, see [DEL 78, theorem 50].

THEOREM 5.2.–

1) In the case $\mathbb{T} = \mathbb{R}_+$, a random variable $\tau \colon \Omega \to [0, +\infty]$ is a Markov moment if and only if, for any $t \in \mathbb{R}_+$, $\{\tau < t\} \in \mathcal{F}_t$.

2) If τ is a Markov moment, then for any non-decreasing $f \colon \mathbb{T} \cup \{+\infty\} \to \mathbb{T} \cup \{+\infty\}$ such that $f(t) \geq t$ for any $t \in \mathbb{T}$, $f(\tau)$ is a Markov moment.

3) Let σ and τ be Markov moments. Then $\sigma + \tau$, $\sigma \wedge \tau$, $\sigma \vee \tau$ are Markov moments.

4) Let $\{\tau_k, k \geq 1\}$ be Markov moments. Then $\sum_{k=1}^{\infty} \tau_k$, $\sup_{k\geq 1} \tau_k$, $\inf_{k\geq 1} \tau_k$, $\limsup_{k\to\infty} \tau_k$, $\liminf_{k\to\infty} \tau_k$ are Markov moments.

PROOF.– 1) For necessity, note that $\{\{\}\tau < t\} = \bigcup n \geq 1 \{\tau_n \leq t - 1/n\} \in \mathcal{F}_t$. For sufficiency, note that, for any $n \geq 1$, $\{\tau \leq t\} = \bigcap_{k\geq n} \{\tau < t + 1/k\} \in \mathcal{F}_{t+1/n}$. Therefore, $\{\tau \leq t\} \in \bigcap_{n\geq 1} \mathcal{F}_{t+1/n} = \mathcal{F}_t$ by the right-continuity of filtration.

2) Clearly, for any $t \in \mathbb{T}$, there exists some $g(t) \leq t$ such that either

$$\{\omega : f(\tau) \leq t\} = \{\omega : \tau \leq g(t)\} \in \mathcal{F}_t$$

or

$$\{\omega : f(\tau) \le t\} = \{\omega : \tau < g(t)\} \in \mathcal{F}_t.$$

3) Consider the case $\mathbb{T} = \mathbb{R}^+$ (the discrete time case is similar, but simpler) and random variable $\sigma + \tau$. For any $t > 0$, define a sequence of random variables $\sigma_n = \sum_{k=0}^{\infty} \frac{t(k+1)}{2^n} \mathbb{1}_{\sigma \in \left[\frac{tk}{2^n}, \frac{t(k+1)}{2^n}\right)}$. Clearly, $\sigma_n \to \sigma$, $n \to \infty$, and by (2), these are Markov moments. Then

$$\{\omega : \sigma + \tau \ge t\} = \bigcap_{n=1}^{\infty} \{\omega : \sigma_n + \tau \ge t\}$$
$$= \bigcap_{n=1}^{\infty} \left(\left(\bigcup_{k=0}^{2^n-1} \left\{ \omega : \sigma_n = \frac{t(k+1)}{2^n}, \tau \ge t - \frac{t(k+1)}{2^n} \right\} \right) \cup \{\sigma_n > t\} \right) \in \mathcal{F}_t.$$

Therefore, $\{\omega : \sigma + \tau < t\} = \Omega \setminus \{\sigma + \tau \ge t\} \in \mathcal{F}_t$, so by (1), $\sigma + \tau$ is a stopping time. Further,

$$\{\omega : \sigma \wedge \tau \le t\} = \{\omega : \sigma \le t\} \cup \{\omega : \tau \le t\} \in \mathcal{F}_t,$$
$$\{\omega : \sigma \vee \tau \le t\} = \{\omega : \sigma \le t\} \cap \{\omega : \tau \le t\} \in \mathcal{F}_t,$$

so $\sigma \wedge \tau$, $\sigma \vee \tau$ are stopping times.

4) For any $t \ge 0$,

$$\left\{\omega : \sup_{n \ge 1} \tau_n \le t\right\} = \bigcap_{n \ge 1} \{\tau_n \le t\} \text{ and } \left\{\omega : \inf_{n \ge 1} \tau_n < t\right\} = \bigcup_{n \ge 1} \{\tau_n < t\},$$

whence, with the help of (1), $\sup_{n \ge 1} \tau_n$ and $\inf_{n \ge 1} \tau_n$ are Markov moments. Since $\limsup_{n \to \infty} \tau_n = \inf_{n \ge 1} \sup_{k \ge n} \tau_k$ and $\liminf_{n \to \infty} \tau_n = \sup_{n \ge 1} \inf_{k \ge n} \tau_k$, these are Markov moments too. □

THEOREM 5.3.–

1) Let τ be a Markov moment and let the collection of sets $\mathcal{F}_\tau$ be defined according to definition 5.5, (3). Then $\mathcal{F}_\tau$ is indeed a σ-algebra and τ is a $\mathcal{F}_\tau$-measurable random variable.

2) If $\mathbb{T} = \mathbb{R}_+$, then $A \in \mathcal{F}_\tau$ if and only if for any $t \in \mathbb{T}$, $A \cap \{\tau < t\} \in \mathcal{F}_t$.

3) Let $\sigma \le \tau$ be two Markov moments. Then $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

4) For any two Markov moments σ and τ, $\mathcal{F}_{\sigma \wedge \tau} = \mathcal{F}_\sigma \cap \mathcal{F}_\tau$.

5) For any sequence of stopping times $\{\tau_n, n \ge 1\}$, $\mathcal{F}_{\inf_{n \ge 1} \tau_n} = \bigcap_{n \ge 1} \mathcal{F}_{\tau_n}$.

6) Let σ and τ be two Markov moments. Then the events $\{\sigma = \tau\}$, $\{\sigma \le \tau\}$, and $\{\sigma < \tau\}$ belong to $\mathcal{F}_{\sigma\wedge\tau}$.

7) Let $\{X_t, \mathcal{F}_t, t \in \mathbb{T}\}$ be an adapted stochastic process and τ be a Markov moment. In the case where $\mathbb{T} = \mathbb{R}_+$, let also X be right-continuous. Then X_τ is $\mathcal{F}_\tau$-measurable.

PROOF.– 1) Let $\{A_n, n \ge 1\}$ be the events from $\mathcal{F}_\tau$. This means that for any $t > 0$ and any $n \ge 1$, $\{\omega : \tau(\omega) \le t\} \cap A_n \in \mathcal{F}_t$. Then $\{\omega : \tau(\omega) \le t\} \cap (\bigcup_{n=1}^\infty A_n) = \bigcup_{n=1}^\infty (\{\omega : \tau(\omega) \le t\} \cap A_n) \in \mathcal{F}_t$. Therefore, $\bigcup_{n=1}^\infty A_n \in \mathcal{F}_\tau$. Similarly, we can prove that, for any $A, B \in \mathcal{F}_\tau$, $A\backslash B \in \mathcal{F}_\tau$ as well. Evidently, $\Omega \in \mathcal{F}_\tau$ because, for any $t > 0$, $\Omega \cap \{\tau \le t\} = \{\tau \le t\} \in \mathcal{F}_t$. Therefore, $\mathcal{F}_\tau$ is a σ-algebra.

Further, for any $t \ge 0$, the event $\{\tau \le t\} \in \mathcal{F}_\tau$ because, for any other $s \ge 0$, $\{\tau \le t\} \cap \{\tau \le s\} = \{\tau \le t \wedge s\} \in \mathcal{F}_{t\wedge s} \subset \mathcal{F}_s$. Then, for any interval $(u, t]$, the event $\{\tau \in (u, t]\} = \{\tau \le t\} \setminus \{\tau \le u\} \in \mathcal{F}_\tau$. As the Borel σ-field $\mathcal{B}(\mathbb{R}^+)$ is generated by the intervals $(u, t], 0 \le u \le t$, we get from lemma 1.1 that for any Borel set $A \in \mathcal{B}(\mathbb{R}^+)$ $\tau^{-1}(A) = \{\omega : \tau(\omega) \in A\} \in \mathcal{F}_t$. It means that τ is $\mathcal{F}_\tau$-measurable.

2) For any $t \in \mathbb{T}$, $A \in \mathcal{F}_\tau$, we have

$$\begin{aligned} A \cap \{\tau < t\} &= A \cap \left(\bigcup_{n\ge1} \{\tau \le t - 1/n\} \right) \\ &= \bigcup_{n\ge1} \big(A \cap \{\tau \le t - 1/n\}\big) \in \mathcal{F}_t, \end{aligned}$$

which implies the necessity. Concerning the sufficiency, let A be such that $A \cap \{\tau < s\}$ for any $s \in \mathbb{T}$. Then, for any $t \in \mathbb{T}$ and $n \ge 1$,

$$\begin{aligned} A \cap \{\tau \le t\} &= A \cap \left(\bigcap_{k\ge n} \{\tau < t + 1/k\} \right) \\ &= \bigcap_{k\ge n} \big(A \cap \{\tau < t + 1/k\}\big) \in \mathcal{F}_{t+1/n}. \end{aligned}$$

Therefore, $A \cap \{\tau \le t\} \in \bigcap_{n\ge1} \mathcal{F}_{t+1/n}$. In view of the right-continuity of the filtration, this means $A \cap \{\tau \le t\} \in \mathcal{F}_t$, so $A \in \mathcal{F}_\tau$.

3) Consider the Markov moments σ and τ such that $\sigma \le \tau$ a.s. Let $A \in \mathcal{F}_\sigma$. This means that, for any $t \ge 0$, $A \cap \{\sigma \le t\} \in \mathcal{F}_t$. Now, $A \cap \{\tau \le t\} = A \cap \{\sigma \le t\} \cap \{\tau \le t\}$ is an intersection of $A \cap \{\sigma \le t\}$ and $\{\tau \le t\}$ and both events belong to $\mathcal{F}_t$, so $A \cap \{\tau \le t\} \in \mathcal{F}_t$ and $A \in \mathcal{F}_t$. This means that $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$.

4) It follows from (2) that $\mathcal{F}_{\sigma\wedge\tau} \subset \mathcal{F}_\sigma$ and $\mathcal{F}_{\sigma\wedge\tau} \subset \mathcal{F}_\tau$, so that $\mathcal{F}_{\sigma\wedge\tau} \subset \mathcal{F}_\tau \cap \mathcal{F}_\sigma$. Further, let event $A \in \mathcal{F}_\tau \cap \mathcal{F}_\sigma$. Then $A \in \mathcal{F}_\tau$ and $A \in \mathcal{F}_\sigma$. Therefore, $A \cap \{\sigma \le t\} \in \mathcal{F}_t$ and $A \cap \{\tau \le t\} \in \mathcal{F}_t$. Consider $A \cap \{\sigma \wedge \tau \le t\} = A \cap (\{\sigma \le t\} \cup \{\tau \le t\}) = (A \cap \{\sigma \le t\}) \cup (A \cap \{\tau \le t\}) \in \mathcal{F}_t$. This means that $A \in \mathcal{F}_{\sigma\wedge\tau}$. Therefore, $\mathcal{F}_\sigma \cap \mathcal{F}_\tau \subset \mathcal{F}_{\sigma\wedge\tau}$.

5) From (3), by induction we have that, for any $m \ge 1$, $\mathcal{F}_{\min_{1\le n\le m}\tau_n} = \bigcap_{n=1}^m \mathcal{F}_{\tau_n}$. Using (3), we get

$$\mathcal{F}_{\inf_{n\ge 1}\tau_n} \subset \mathcal{F}_{\min_{1\le n\le m}\tau_n} = \bigcap_{n=1}^{m} \mathcal{F}_{\tau_n}$$

for every $m \ge 1$, whence $\mathcal{F}_{\inf_{n\ge 1}\tau_n} \subset \bigcap_{n\ge 1} \mathcal{F}_{\tau_n}$.

Vice versa, take any $A \in \bigcap_{n\ge 1} \mathcal{F}_{\tau_n}$ and $t \in \mathbb{T}$. Noting that $\tau := \inf_{n\ge 1}\tau_n = \lim_{m\to\infty} \min_{1\le n\le m}\tau_n$, we can write

$$\begin{aligned} A \cap \{\tau < t\} &= A \cap \left(\bigcup_{k\ge 1} \bigcap_{m\ge k} \left\{ \min_{1\le n\le m} \tau_n < t \right\} \right) \\ &= \bigcup_{k\ge 1} \bigcap_{m\ge k} A \cap \left\{ \min_{1\le n\le m} \tau_n < t \right\}. \end{aligned}$$

From (2) and the above argument, we have that $A \cap \{\min_{1\le n\le m}\tau_n < t\} \in \mathcal{F}_t$, so $A \cap \{\tau < t\} \in \mathcal{F}_t$. Using (2) again, we get $A \in \mathcal{F}_\tau$.

6) We prove this for $\mathbb{T} = \mathbb{R}_+$; in the case of discrete time, the proofs are similar, but simpler. For any $t \in \mathbb{T}$, let $\mathbb{T}_t$ be a countable subset of $\mathbb{T} \cap (0, t]$ containing t. Then

$$\{\sigma < \tau\} \cap \{\tau \le t\} = \bigcup_{s\in\mathbb{T}_t} \{\sigma < s\} \cap \{\tau \in (s, t]\} \in \mathcal{F}_t,$$

so $\{\sigma < \tau\} \in \mathcal{F}_\tau$. Further,

$$\{\sigma < \tau\} \cap \{\sigma \le t\} = \bigcup_{s\in\mathbb{T}_t} \{\sigma \le s\} \cap \{\tau > s\} \in \mathcal{F}_t,$$

so $\{\sigma < \tau\} \in \mathcal{F}_\sigma$. As a result, $\{\sigma < \tau\} \in \mathcal{F}_\tau \cap \mathcal{F}_\sigma = \mathcal{F}_{\tau\wedge\sigma}$. Therefore, $\{\tau \le \sigma\} = \Omega \setminus \{\sigma < \tau\} \in \mathcal{F}_{\tau\wedge\sigma}$. By symmetry, $\{\sigma \le \tau\} \in \mathcal{F}_{\tau\wedge\sigma}$. Therefore, $\{\tau = \sigma\} = \{\sigma \le \tau\} \cap \{\tau \le \sigma\} \in \mathcal{F}_{\tau\wedge\sigma}$.

7) Similar to (6), we show this only in the case of continuous time; the discrete time case is simpler. It is enough to prove that for any open set U, $\{X_\tau \in U\} \in \mathcal{F}_\tau$. For any $t \in \mathbb{T}$, due to the right-continuity of X,

$$\{X_\tau \in U\} \cap \{\tau < t\} = \bigcup_{\substack{s\in\mathbb{Q} \\ \tau<s<t}} \bigcap_{\substack{u\in\mathbb{Q} \\ \tau<u<s}} \{X_u \in U\} \cap \{\tau < t\}$$

$$= \bigcup_{s\in\mathbb{Q}, s<t} \bigcap_{u\in\mathbb{Q}, u<s} \big(\{X_u \in U\} \cap \{\tau < u\}\big) \cup \{\tau \geq u\}.$$

As X is adapted, for any $u < t$, $\big(\{X_u \in U\} \cap \{\tau < u\}\big) \cup \{\tau \geq u\} \in \mathcal{F}_u \subset \mathcal{F}_t$. Thus, we get $\{X_\tau \in U\} \cap \{\tau < t\} \in \mathcal{F}_t$, so by (2), $\{X_\tau \in U\} \in \mathcal{F}_\tau$, as required. □

5.5. Martingales and related processes with discrete time

In this section, we concentrate on the discrete-time martingale processes. This specific field is much simpler than the corresponding field for continuous-time processes. However, it allows us to clarify and understand the main properties of martingales.

5.5.1. *Upcrossings of the interval and existence of the limit of submartingale*

Let $X = \{X_n, n \geq 0\}$ be a stochastic process and $a < b$, $[a, b] \subset \mathbb{R}$ be a fixed interval. Define the following Markov moments

$$\tau_1 = \inf\{n \geq 0 : X_n \leq a\}, \ \ \tau_2 = \inf\{n > \tau_1 : X_n \geq b\},$$

$$\tau_3 = \inf\{n > \tau_2 : X_n \leq a\}, \ \ \tau_4 = \inf\{n > \tau_3 : X_n \geq b\},$$

$$\ldots$$

$$\tau_{2k-1} = \inf\{n > \tau_{2k-2} : X_n \leq a\}, \ \ \tau_{2k} = \inf\{n > \tau_{2k-1} : X_n \geq b\}.$$

If the corresponding jth event does not hold on some $\omega \in \Omega$, we put $\tau_j(\omega) = \infty$.

DEFINITION 5.10.– *We say that the process X has $k \geq 1$* upcrossings *of the interval $[a, b]$ on the time interval $[0, N]$ if $\tau_{2k} \leq N < \tau_{2k+2}$. In the case where $\tau_2 > N$, we say that the number of upcrossings equals zero (see Figure 5.1).*

REMARK 5.8.–

1) The number of downcrossings of the interval can be defined in a similar way, and all subsequent theories can be based on the number of downcrossings.

2) Evidently, for stochastic process X, the number of up- (down-) crossings is a random variable.

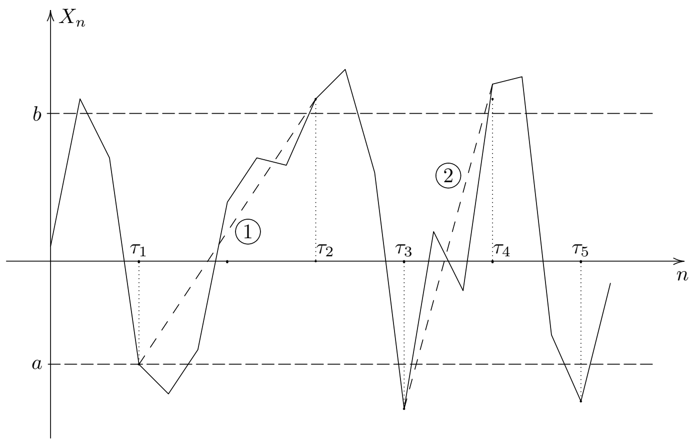


Figure 5.1. *Upcrossings*

Denote by $k_{N,X}([a,b]) = k_{N,X}([a,b])(\omega)$ the number of upcrossings of the interval $[a,b]$ on the time interval $[0,N]$ defined for the process X.

THEOREM 5.4.– *Let* $X - \{X_n, \mathcal{F}_n, n \geq 0\}$ *be a submartingale. Then*

$$\mathrm{E}k_{N,X}([a,b]) \leq \frac{\mathrm{E}(X_N - a)^+}{b-a}.$$

PROOF.– First, note that the process $Y = \{Y_n, \mathcal{F}_n, n \geq 0\}$, where $Y_n = (X_n - a)^+$ is a non-negative submartingale. Further, note that $k_{N,X}([a,b]) = k_{N,Y}([0,b-a])$ (see Figure 5.2).

Now, denote $\chi_i = \mathbb{1}_{\tau_{2k+1} \leq i < \tau_{2k+2} \text{ for some } k \geq 0}$. Then the event

$$\{\chi_i = 1\} = \bigcup_{k=0}^{\infty} (\{\tau_{2k+1} \leq i\} \setminus \{\tau_{2k+2} \leq i\}) \in \mathcal{F}_i.$$

The next inequality is the key point of the proof:

$$(b-a)k_{N,Y}([0,b-a]) \leq \sum_{i=0}^{N-1} (Y_{i+1} - Y_i)\chi_i.$$

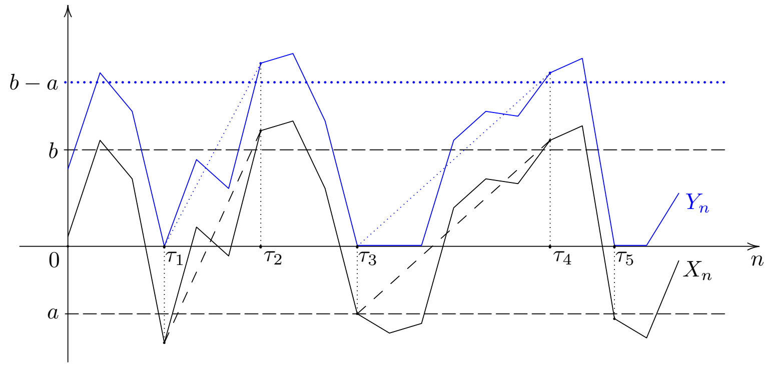


Figure 5.2. $k_{N,X}([a,b]) = k_{N,Y}([0,b-a])$. *For a color version of the figure, see www.iste.co.uk/mishura/stochasticprocesses.zip*

Note, for better understanding, that the latter inequality holds for two reasons: first, any subsequent non-zero series of summands in the right-hand side is equal or overcomes $b-a$, since $Y_{\tau_{2k+1}} = 0$ and $Y_{\tau_{2k+2}} \geq b-a$, and also the right-hand side can contain an "additional" non-zero group of summands, in the case where, for all $k \geq 0$, we have that $N-1 \neq \tau_{2k+2}$, but this group equals $Y_{N-1} - Y_{\tau_{2k+1}} = Y_{N-1}$ and is non-negative due to our replacement of X with non-negative Y. Therefore,

$$(b-a)\mathrm{E}k_{N,Y}([0,b-a]) \leq \sum_{i=0}^{N-1} \mathrm{E}(Y_{i+1} - Y_i)\chi_i$$

$$= \sum_{i=0}^{N-1} \mathrm{E}(\mathrm{E}(Y_{i+1} - Y_i|\mathcal{F}_i)\chi_i) \leq \sum_{i=0}^{N-1} \mathrm{E}(Y_{i+1} - Y_i) \leq \mathrm{E}Y_N = \mathrm{E}(X_N - a)^+,$$

whence the proof follows. □

Consider a submartingale $X = \{X_n, \mathcal{F}_n, n \geq 0\}$. When does there exist a limit $\lim_{n\to\infty} X_n$, at least in some sense? Theorems 5.5 and 5.6 give a partial answer to this question.

THEOREM 5.5.– *Let* $\sup_{n\geq 0} \mathrm{E}|X_n| < \infty$. *Then there exists* $X_\infty = \lim_{n\to\infty} X_n$ *a.s. and* $\mathrm{E}|X_\infty| \leq \sup_{n\geq 0} \mathrm{E}|X_n|$, *so that* $X_\infty \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$.

PROOF.– Let

$$A = \left\{\omega \in \Omega : \lim_{n\to\infty} X_n \text{ does not exist}\right\} = \left\{\omega \in \Omega : \limsup_{n\to\infty} X_n > \liminf_{n\to\infty} X_n\right\}.$$

As both $\limsup_{n\to\infty} X_n$ and $\liminf_{n\to\infty} X_n$ are random variables, we have that $A \in \mathcal{F}$. Moreover, $A = \bigcup_{q_2>q_1, q_1,q_2\in\mathbb{Q}} A_{q_1,q_2}$, where

$$A_{q_1,q_2} = \left\{ \limsup_{n\to\infty} X_n > q_2 > q_1 > \liminf_{n\to\infty} X_n \right\},$$

and $\mathbb{Q}$ is the set of rational numbers. Consider any event A_{q_1,q_2} and introduce the limit of non-decreasing sequence:

$$k_{\infty,X}([a,b])(\omega) = \lim_{N\to\infty} k_{N,X}([a,b])(\omega).$$

If $\omega \in A_{q_1,q_2}$, then $k_{\infty,X}([q_1,q_2])(\omega) = +\infty$. However, it follows from the Lebesgue monotone convergence theorem that $\mathrm{E}k_{\infty,X}([q_1,q_2]) = \lim_{N\to\infty} \mathrm{E}k_{N,X}([q_1,q_2])$, and it follows from theorem 5.4 that

$$\mathrm{E}k_{N,X}([q_1,q_2]) \le \frac{\mathrm{E}(X_N - q_1)^+}{q_2 - q_1} \le \frac{\mathrm{E}|X_N| + |q_1|}{q_2 - q_1},$$

whence

$$\mathrm{E}k_{\infty,X}([q_1,q_2]) \le \limsup_{N\to\infty} \frac{\mathrm{E}|X_N| + |q_1|}{q_2 - q_1} \le \frac{\sup_{n\ge 0} \mathrm{E}|X_N| + |q_1|}{q_2 - q_1} < \infty.$$

It means that $k_{\infty,X}([q_1,q_2]) < \infty$ a.s., and consequently $\mathrm{P}\{A_{q_1,q_2}\} = 0$. Finally, $\mathrm{P}\{A\} = 0$ and $\lim_{n\to\infty} X_n$ exists a.s. The second statement is the direct consequence of Fatou's lemma:

$$\mathrm{E}|X_\infty| = \mathrm{E} \lim_{n\to\infty} |X_n| = \mathrm{E} \liminf_{n\to\infty} |X_n| \le \liminf_{n\to\infty} \mathrm{E}|X_n| \le \sup_{n\ge 0} \mathrm{E}|X_n|. \quad \square$$

THEOREM 5.6.– *Let* $\{X_n, \mathcal{F}_n, n \ge 0\}$ *be a martingale and let, for some* $p > 1$,

$$\sup_{n\ge 0} \mathrm{E}|X_n|^p < \infty.$$

Then there exists a limit $\lim_{n\to\infty} X_n =: X_\infty$ *a.s. and in* $\mathcal{L}_p(\Omega, \mathcal{F}, \mathrm{P})$.

PROOF.– Existence of a.s. limit $X_\infty = \lim_{n\to\infty} X_n$ follows from theorem 5.5. Therefore,

$$|X_n - X_\infty|^p \to 0 \qquad [5.2]$$

a.s. as $n \to \infty$. Moreover, by Fatou's lemma, $\mathrm{E}|X_\infty|^p \le \liminf_{n\to\infty} \mathrm{E}|X_n|^p < \infty$. Let us establish that $\mathrm{E}|X_n - X_\infty|^p \to 0$, $n \to \infty$. First, by theorem A2.5, the sequence $\{X_n, n \geq 0\}$ is uniformly integrable; therefore, by theorem A2.4

$$\mathrm{E}|X_n - X_\infty| \to 0, \; n \to \infty.$$

Consequently, for any $k \geq 0$ and $m > k$,

$$\begin{aligned}\mathrm{E}|X_k - \mathrm{E}(X_\infty|\mathcal{F}_k)| &= \mathrm{E}\,|\mathrm{E}(X_m|\mathcal{F}_k) - \mathrm{E}(X_\infty|\mathcal{F}_k)| \\ &\leq \mathrm{E}|X_m - X_\infty| \to 0, \;\; m \to \infty.\end{aligned}$$

Therefore, $\mathrm{E}(X_\infty|\mathcal{F}_k) = X_k$ a.s. for any $k \geq 0$. By Jensen's inequality, for any $a, C > 0$,

$$\begin{aligned}&\sup_{n\geq 0} \mathrm{E}|X_n|^p \mathbb{1}_{|X_n|\geq C} = \sup_{n\geq 0} \mathrm{E}|\mathrm{E}(X_\infty|\mathcal{F}_n)|^p \mathbb{1}_{|X_n|\geq C} \\ &\leq \sup_{n\geq 0} \mathrm{E}\left(\mathrm{E}(|X_\infty|^p|\mathcal{F}_n)\mathbb{1}_{|X_n|\geq C}\right) \leq \sup_{n\geq 0} \mathrm{E}\left(|X_\infty|^p \mathbb{1}_{|X_n|\geq C}\right) \\ &\leq \sup_{n\geq 0} a^p \mathrm{P}\left\{|X_n| \geq C\right\} + \mathrm{E}|X_\infty|^p \mathbb{1}_{|X_\infty|\geq a} \\ &\leq \sup_{n\geq 0} \frac{a^p}{C^p} \mathrm{E}|X_n|^p + \mathrm{E}|X_\infty|^p \mathbb{1}_{|X_\infty|\geq a},\end{aligned}$$

whence

$$\lim_{C\to\infty} \sup_{n\geq 0} \mathrm{E}|X_n|^p \mathbb{1}_{|X_n|\geq C} \leq \mathrm{E}|X_\infty|^p \mathbb{1}_{|X_\infty|\geq a},$$

and letting $a \to \infty$, we get that

$$\lim_{C\to\infty} \sup_{n\geq 0} \mathrm{E}|X_n|^p \mathbb{1}_{|X_n|\geq C} = 0.$$

It means that the sequence $\{|X_n|^p, n \geq 0\}$ is uniformly integrable and now the proof follows from the relation [5.2] and theorem A2.4. □

How to formulate a similar result for $p = 1$? We see from the proof of theorem 5.6 that such results are closely connected to the uniform integrability property. This connection is demonstrated by the following result as well.

THEOREM 5.7.– *Let* $X = \{X_n, \mathcal{F}_n, n \geq 1\}$ *be a martingale. Then the following statements are equivalent.*

i) Martingale X *is uniformly integrable.*

ii) $\sup_{n\geq 1} \mathrm{E}|X_n| < \infty$ *(consequently, there exists* $X_\infty = \lim_{n\to\infty} X_n$ *a.s.) and*

$$X_n = \mathrm{E}(X_\infty|\mathcal{F}_n) \text{ a.s.}$$

iii) There exists $\lim_{n\to\infty} X_n$ *in* $\mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$ *(then there exists* X_∞ *defined at point (ii) and the limit in* $\mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$ *coincides with* X_∞ *a.s.).*

iv) $X_n = \mathrm{E}(X|\mathcal{F}_n)$ *a.s. for some integrable random variable* X *(then there exists* X_∞ *defined at point (ii) and* $X = X_\infty$ *a.s.).*

PROOF.– (i) ⇒ (ii). Indeed, according to theorem A2.4, for any uniformly integrable sequence $\{X_n, n \geq 0\}$, we have that $\sup_{n\geq 1} \mathrm{E}|X_n| < \infty$. Then the existence of the random variable X_∞ that is, a limit with probability 1 of X_n, $X_\infty = \lim_{n\to\infty} X_n$, follows from theorem 5.5. Moreover, uniform integrability implies that

$$\mathrm{E}|X_n - X_\infty| \to 0$$

as $n \to \infty$, and, for any $m > k$,

$$\mathrm{E}|X_k - \mathrm{E}(X_\infty|\mathcal{F}_k)| = \mathrm{E}|\mathrm{E}(X_m|\mathcal{F}_k) - \mathrm{E}(X_\infty|\mathcal{F}_k)| \leq \mathrm{E}|X_m - X_\infty| \to 0 \quad [5.3]$$

as $m \to \infty$. Therefore, we get that $X_k = \mathrm{E}(X_\infty|\mathcal{F}_k)$ a.s.

(ii) ⇒ (iii). Consider

$$\mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} = \mathrm{E}|\mathrm{E}(X_\infty|\mathcal{F}_n)|\mathbb{1}_{|X_n|\geq C} \leq \mathrm{E}(\mathrm{E}(|X_\infty||\mathcal{F}_n))\mathbb{1}_{|X_n|\geq C} \quad [5.4]$$

$$= \mathrm{E}|X_\infty|\mathbb{1}_{|X_n|\geq C} \leq a\mathrm{P}\{|X_n| \geq C\} + \mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|\geq a} \quad [5.5]$$

$$\leq \frac{a}{C}\mathrm{E}|X_n| + \mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|\geq a}, \quad [5.6]$$

whence

$$\lim_{C\to\infty} \sup_{n\geq 1} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} \leq \mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|\geq a} \to 0 \text{ as } a \to \infty.$$

This means that $\lim_{C\to\infty} \sup_{n\geq 1} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} = 0$, and $\{X_n, n \geq 1\}$ is a uniformly integrable sequence. Then $\mathrm{E}|X_n - X_\infty| \to 0$, $n \to \infty$ according to theorem A2.4.

(iii) ⇒ (iv). Let $\mathrm{E}|X_n - X| \to 0$ as $n \to \infty$. Then, similarly to [5.3],

$$\mathrm{E}|X_k - \mathrm{E}(X|\mathcal{F}_k)| = \mathrm{E}|\mathrm{E}(X_m|\mathcal{F}_k) - \mathrm{E}(X|\mathcal{F}_k)| \leq \mathrm{E}|X_m - X| \to 0$$

as $m \to \infty$, and $X_k = \mathrm{E}(X|\mathcal{F}_k)$ a.s. Moreover, it follows from convergence in $\mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$ that the sequence is bounded in this space, whence $\sup_{n\geq 1} \mathrm{E}|X_n| < \infty$; therefore, there exists X_∞ defined at point (ii) and then obviously the limit X in $\mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$ coincides with X_∞ a.s.

(iv) ⇒ (i). Similar to [5.4], consider

$$\begin{aligned}\mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} &= \mathrm{E}|\mathrm{E}(X|\mathcal{F}_n)|\mathbb{1}_{|X_n|\geq C} \leq \mathrm{E}(\mathrm{E}(|X||\mathcal{F}_n)\mathbb{1}_{|X_n|\geq C}) \\ &= \mathrm{E}|X|\mathbb{1}_{|X_n|\geq C} \leq a\mathrm{P}\left\{|X_n| \geq C\right\} + \mathrm{E}|X|\mathbb{1}_{|X|\geq a} \\ &\leq \frac{a}{C}\mathrm{E}|X_n| + \mathrm{E}|X|\mathbb{1}_{|X|\geq a},\end{aligned}$$

whence

$$\lim_{C\to\infty} \sup_{n\geq 1} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} \leq \mathrm{E}|X|\mathbb{1}_{|X|\geq a} \to 0 \text{ as } a \to \infty.$$

This means that $\lim_{C\to\infty} \sup_{n\geq 1} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} = 0$, and $\{X_n, n \geq 1\}$ is a uniformly integrable sequence. From these reasons, the existence of $X_\infty = X$ a.s. follows immediately. □

5.5.2. *Examples of martingales having a limit and of uniformly and non-uniformly integrable martingales*

EXAMPLE 5.10.– *Consider a sequence of independent random variables* $\{\xi_n, n \geq 1\}$ *such that* $|\xi_n| < 1$ *and* $\mathrm{E}\xi_n = 0$. *Define a stochastic process and the corresponding* σ*-fields:*

$$X_0 = x_0 > 0,\ X_n = x_0 \prod_{k=1}^{n} (1 + \xi_k),\ n \geq 1,$$

$$\mathcal{F}_0 = \{\emptyset, \Omega\},\ \mathcal{F}_n = \sigma\{\xi_1, \ldots, \xi_n\} = \sigma\{X_1, \ldots, X_n\},\ n \geq 1.$$

Then $\{X_n, \mathcal{F}_n, n \geq 1\}$ *is a martingale,* $X_n > 0$ *a.s., and* $\mathrm{E}X_n = x_0$, $n \geq 1$. *Therefore,* $\sup_{n\geq 0} \mathrm{E}|X_n| = \sup_{n\geq 0} \mathrm{E}X_n = x_0$, *which means that there exists a limit* $X_\infty = \lim_{n\to\infty} X_n = x_0 \prod_{k=1}^{\infty}(1 + \xi_k)$ *a.s. Further,*

$$\mathrm{E}|X_n - X_\infty| = x_0\mathrm{E}\left|\prod_{k=1}^{\infty}(1 + \xi_k) - \prod_{k=1}^{n}(1 + \xi_k)\right|$$

$$= x_0\mathrm{E}\prod_{k=1}^{n}(1 + \xi_k)\mathrm{E}\left|\prod_{k=n+1}^{\infty}(1 + \xi_k) - 1\right| = x_0\mathrm{E}\left|\prod_{k=n+1}^{\infty}(1 + \xi_k) - 1\right|.$$

For any $m > n$,

$$\prod_{k=n+1}^{m}(1+\xi_k) - 1 = \prod_{k=n+1}^{m}(1+\xi_k) - \prod_{k=n+1}^{m-1}(1+\xi_k) + \prod_{k=n+1}^{m-1}(1+\xi_k) - \prod_{k=n+1}^{m-2}(1+\xi_k) + \ldots + (1+\xi_k) - 1$$

$$= \sum_{j=n+1}^{m}\left(\prod_{k=n+1}^{j}(1+\xi_k) - \prod_{k=n+1}^{j-1}(1+\xi_k)\right) = \sum_{j=n+1}^{m}\prod_{k=n+1}^{j-1}(1+\xi_k)\xi_j,$$

where $\prod_{k=n+1}^{n} = 1$. *Therefore,*

$$\mathrm{E}\left|\prod_{k=n+1}^{m}(1+\xi_k) - 1\right| \leq \sum_{j=n+1}^{m}\mathrm{E}\prod_{k=n+1}^{j-1}(1+\xi_k)\mathrm{E}|\xi_j| = \sum_{j=n+1}^{m}\mathrm{E}|\xi_j|,$$

whence

$$\mathrm{E}\left|\sum_{k=n+1}^{m}(1+\xi_k) - 1\right| \leq \sum_{k=n+1}^{\infty}\mathrm{E}|\xi_k|.$$

This means that if the series $\sum_{k=1}^{\infty}\mathrm{E}|\xi_k|$ *converges, then* $\mathrm{E}|X_n - X_\infty| \to 0$, $n \to \infty$ *and according to theorem A2.4,* $\{X_n, n \geq 0\}$ *is a uniformly integrable martingale. For example, if we put* $\xi_k = \eta_k a_k$, *where* η_k *are iid random variables, series* $\sum_{k=1}^{\infty}|a_k|$ *converges, and* $|\eta_k a_k| < 1$, *then* $\sum_{k=1}^{\infty}\mathrm{E}|\xi_k| = \mathrm{E}|\eta_1| \sum_{k=1}^{\infty}|a_k| < \infty$. *As an example of non-uniformly integrable martingale, consider the simplest case where* $\xi_k = \pm\frac{1}{2}$ *with* $\mathrm{P}\left\{\xi_k = \pm\frac{1}{2}\right\} = \frac{1}{2}$, $\{\xi_k, k \geq 0\}$ *are independent. Then* $\{X_n, n \geq 0\}$ *is a martingale,* $X_n > 0$ *a.s. and* $\mathrm{E}X_n = x_0$, *whence the limit* $X_\infty = x_0\prod_{k=1}^{\infty}(1+\xi_k)$ *exists. We can identify* X_∞ *via the strong law of large numbers. Indeed,* $\{\log(1+\xi_k), k \geq 1\}$ *is a sequence of bounded iid random variables with*

$$\mathrm{E}\log(1+\xi_k) = \frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{3}{2} = \frac{1}{2}\log\frac{3}{4} < 0.$$

According to SLLN,

$$\frac{\sum_{k=1}^{n}\log(1+\xi_k)}{n} \to \mathrm{E}\log(1+\xi_1) = \frac{1}{2}\log\frac{3}{4} < 0$$

a.s. This means that $\sum_{k=1}^{n}\log(1+\xi_k) \to -\infty$ *a.s. and consequently* $\prod_{k=1}^{n}(1+\xi_k) \to 0$ *a.s. Therefore,* $X_\infty = 0$. *However,* $\mathrm{E}X_n = x_0 \nrightarrow \mathrm{E}X_\infty = 0$, *which means that* $\{X_n, n \geq 0\}$ *is not a uniformly integrable martingale.*

EXAMPLE 5.11.– *Consider a model of population dynamics with discrete time, called a* Halton–Watson process. *Let some population develop in such a way: at the* *initial moment $n = 0$, we have an integer number $\xi_0 \geq 0$ of individuals. Each initial individual generates a random integer number $\xi_i^{(1)} \in \mathbb{N} \cup \{0\}$ of individuals in the next generation, $1 \leq i \leq \xi_0$ and so on. In nth generation, under the condition that it is not degenerated, we have ξ_n of individuals, and $\xi_n = \sum_{i=1}^{\xi_{n-1}} \xi_i^{(n)}$. Denote $\sum_{i=1}^{0} = 0$. Assume that all random variables $\left\{\xi_i^{(n)}, n \geq 1, i \geq 1\right\}$ are mutually independent and $\mathrm{E}\xi_i^{(n)} = \mu_n$. Denote $\mathcal{F}_n = \sigma\left\{\xi_1, \ldots, \xi_n, \xi_i^{(k)}, i \geq 1, 1 \leq k \leq n\right\}$. Then it follows from the Fubini theorem that*

$$\mathrm{E}(\xi_n|\mathcal{F}_{n-1}) = \mathrm{E}\left(\sum_{i=1}^{\infty} \xi_i^{(n)} \mathbb{1}_{i \leq \xi_{n-1}} |\mathcal{F}_{n-1}\right)$$

$$= \sum_{i=1}^{\infty} \mathrm{E}\left(\xi_i^{(n)} \mathbb{1}_{i \leq \xi_{n-1}} |\mathcal{F}_{n-1}\right) = \sum_{i=1}^{\infty} \mathbb{1}_{i \leq \xi_{n-1}} \mathrm{E}\xi_i^{(n)} = \mu_n \xi_{n-1}.$$

Denote $X_n = \frac{\xi_n}{\mu_1 \mu_2 \ldots \mu_n}$. Then $\{X_n, \mathcal{F}_n, n \geq 1\}$ is a non-negative martingale, $\mathrm{E}X_n = \xi_0$. Therefore, there exists a limit

$$X_\infty := \lim_{n\to\infty} X_n = \lim_{n\to\infty} \xi_n \prod_{k=1}^{n} \mu_k^{-1}, \quad \textit{a.s.}$$

The dynamics of population depends on $\{\mu_k, k \geq 1\}$. If $\mu_k = \mu < 1$, then $\mu^{-n}\xi_n \to X_\infty$ whence $\xi_n \to 0$ a.s. and the population asymptotically degenerates. If, e.g. $\mu_k = 1 + \frac{1}{k}$, then $\prod_{k=1}^{n}\left(1 + \frac{1}{k}\right) \to \infty$, $n \to \infty$. However, we cannot conclude that $\xi_n \to \infty$ a.s. because it can be $X_\infty = 0$. The same doubtful situation is in the case where $\mu \geq 1$. To study the asymptotic behavior of population in these cases, more advanced methods from the theory of branching processes should be involved, see e.g. [HAC 07]. Obviously, in the case $\mu_k = \mu < 1$, we conclude that $\{X_n, n \geq 1\}$ is not a uniformly integrable martingale.

5.5.3. *Lévy convergence theorem*

Now we apply the martingale methods to establish a useful result concerning the convergence of conditional expectations.

THEOREM 5.8.– *Let X be an integrable random variable and let the sequence of σ-fields $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \ldots \subset \mathcal{G}_n \ldots \subset \mathcal{F}$ create a filtration. Denote*

$$\mathcal{G}_\infty = \sigma\left(\bigcup_{n=1}^{\infty} \mathcal{G}_n\right).$$

Then $\mathrm{E}(X|\mathcal{G}_n) \to \mathrm{E}(X|\mathcal{G}_\infty)$ a.s. as $n \to \infty$.

PROOF.– Introduce the sequence $X_n = \mathrm{E}(X|\mathcal{G}_n)$. Then $\mathrm{E}(X|\mathcal{G}_{n-1}) = X_{n-1}$, i.e. $\{X_n, \mathcal{G}_n, n \geq 1\}$ is a martingale. Moreover, according to theorem 5.7, $\{X_n, \mathcal{G}_n, n \geq 1\}$ is a uniformly integrable martingale. Therefore, there exists a limit

$$X_\infty = \lim_{n\to\infty} X_n = \lim_{n\to\infty} \mathrm{E}(X|\mathcal{G}_n) \text{ a.s.}$$

Further, it follows from uniform integrability and theorem A2.4 that, for any set $A \in \mathcal{G}_n$,

$$\int_A X_\infty d\mathrm{P} = \lim_{m\to\infty} \int_A X_m d\mathrm{P} = \lim_{m\to\infty} \int_A \mathrm{E}(X|\mathcal{G}_m) d\mathrm{P} = \int_A X d\mathrm{P},$$

because, for $m > n$, $\int_A \mathrm{E}(X|\mathcal{G}_m) d\mathrm{P} = \int_A X d\mathrm{P}$. Therefore,

$$\int_A X_\infty d\mathrm{P} = \int_A X d\mathrm{P} \qquad [5.7]$$

for any $A \in \mathcal{F}_n$, $n \geq 1$. The left-hand and right-hand sides of [5.7] are the finite measures coinciding on the algebra $\bigcup_{n=1}^\infty \mathcal{F}_n$. They can be uniquely extended to the measure on $\mathcal{F}_\infty$; therefore, [5.7] is valid for any $A \in \mathcal{F}_\infty$. This means that for any $A \in \mathcal{G}_\infty$

$$\int_A X_\infty d\mathrm{P} = \int_A X d\mathrm{P} = \int_A \mathrm{E}(X|\mathcal{G}_\infty) d\mathrm{P}. \qquad [5.8]$$

Note that X_∞ as the limit of X_n is $\mathcal{F}_\infty$-measurable. Then it follows immediately from [5.8] that $X_\infty = \mathrm{E}(X|\mathcal{G}_\infty)$ a.s., and the proof follows. □

5.5.4. *Optional stopping*

Consider a process $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ with discrete time. The next result shows that (sub-, super-) martingale property preserves under random stopping, if you stop in a reasonable way. This result is called "Doob's theorem on optional stopping", or "Doob's optional stopping theorem". We formulate it in the following way.

THEOREM 5.9.– *Let $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ be an integrable stochastic process with discrete time. Then the following statements are equivalent:*

1) X is a (sub-,super-) $\mathcal{F}_n$-martingale.

2) For any bounded stopping time τ and any stopping time σ

$$\mathrm{E}(X_\tau|\mathcal{F}_\sigma)(\geq, \leq) = X_{\tau\wedge\sigma}. \qquad [5.9]$$

3) For any bounded stopping times $\sigma \le \tau$

$$\mathrm{E}X_\tau (\ge, \le) = \mathrm{E}X_\sigma.$$

PROOF.– Consider a submartingale; (super-) martingales are considered similarly. 1) $\Rightarrow$ 2). Let τ be a bounded stopping time, $\tau \le N$. Then X_τ is an integrable random variable, because

$$\mathrm{E}|X_\tau| = \sum_{k=1}^{N} \mathrm{E}|X_\tau| \mathbb{1}_{\tau=k} \le \sum_{k=1}^{N} \mathrm{E}|X_k| < \infty.$$

Additionally, $X_{\tau\wedge\sigma} = X_{\tau\wedge\sigma\wedge N}$ and according to theorem 5.8, which can be applied with $\mathcal{G}_N = \mathcal{F}_{\sigma\wedge N}$ and $\mathcal{G}_\infty = \mathcal{F}_\sigma$, we have that a.s.

$$\mathrm{E}(X_\tau|\mathcal{F}_{\sigma\wedge N}) \to \mathrm{E}(X_\tau|\mathcal{F}_\sigma), \quad N \to \infty.$$

Furthermore, it follows from theorem 5.3 that for $\sigma_N = \sigma \wedge N$

$$\begin{aligned} \mathrm{E}(X_\tau|\mathcal{F}_{\sigma_N}) &= \mathrm{E}(X_{\tau\vee\sigma_N} \mathbb{1}_{\tau\ge\sigma_N}|\mathcal{F}_{\sigma_N}) + \mathrm{E}(X_{\tau\wedge\sigma_N} \mathbb{1}_{\tau<\sigma_N}|\mathcal{F}_{\sigma_N}) \\ &= \mathrm{E}(X_{\tau\vee\sigma_N}|\mathcal{F}_{\sigma_N}) \mathbb{1}_{\tau\ge\sigma_N} + X_{\tau\wedge\sigma_N} \mathbb{1}_{\tau<\sigma_N}. \end{aligned} \qquad [5.10]$$

It follows immediately from [5.10], that in order to prove [5.9], it is enough to prove that, for two bounded stopping times ν and ϱ, such that $\varrho \le \nu \le N$, we have that $\mathrm{E}(X_\nu|\mathcal{F}_\varrho) \ge X_\varrho$. Consider any event $A \in \mathcal{F}_\varrho$. Then

$$\int_A \mathrm{E}(X_\nu|\mathcal{F}_\varrho) d\mathrm{P} = \int_A X_\nu d\mathrm{P} = \sum_{k=1}^{N} \int_{A\cap\{\varrho=k\}} X_\nu d\mathrm{P} = \sum_{k=1}^{N} \int_{A\cap\{\varrho=k,\nu\ge k\}} X_\varrho d\mathrm{P}.$$

Therefore, it is sufficient to prove that

$$\int_{A\cap\{\varrho=k,\nu\ge k\}} X_\nu d\mathrm{P} \ge \int_{A\cap\{\varrho=k,\nu\ge k\}} X_k d\mathrm{P}. \qquad [5.11]$$

However,

$$\begin{aligned} \int_{A\cap\{\varrho=k,\nu\ge k\}} X_k d\mathrm{P} &= \int_{A\cap\{\varrho=k,\nu=k\}} X_\nu d\mathrm{P} + \int_{A\cap\{\varrho=k\}\cap\{\nu>k\}} X_k d\mathrm{P} \\ &\le \int_{A\cap\{\varrho=k,\nu=k\}} X_\nu d\mathrm{P} + \int_{A\cap\{\varrho=k\}\cap\{\nu>k\}} \mathrm{E}(X_{k+1}|\mathcal{F}_k) d\mathrm{P}. \end{aligned} \qquad [5.12]$$

Since the event $A \cap \{\varrho = k\} \cap \{\nu > k\} \in \mathcal{F}_k$, we can continue as follows:

$$\begin{aligned}
&\int_{A\cap\{\varrho=k\}\cap\{\nu>k\}} \mathrm{E}(X_{k+1}|\mathcal{F}_k)d\mathrm{P} = \int_{A\cap\{\varrho=k\}\cap\{\nu>k\}} X_{k+1}d\mathrm{P} \\
&= \int_{A\cap\{\varrho=k\}\cap\{\nu=k+1\}} X_\nu d\mathrm{P} + \int_{A\cap\{\varrho=k\}\cap\{\nu>k+1\}} X_{k+1}d\mathrm{P} \\
&\leq \int_{A\cap\{\varrho=k\}\cap\{\nu=k+1\}} X_\nu d\mathrm{P} + \int_{A\cap\{\varrho=k\}\cap\{\nu>k+1\}} \mathrm{E}(X_{k+2}|\mathcal{F}_{k+1})d\mathrm{P} \\
&= \int_{A\cap\{\varrho=k\}\cap\{\nu=k+1\}} X_\nu d\mathrm{P} + \int_{A\cap\{\varrho=k\}\cap\{\nu>k+1\}} X_{k+2}d\mathrm{P} \\
&\leq \ldots \leq \int_{A\cap\{\varrho=k\}\cap\{\nu=k+1\}} X_\nu d\mathrm{P} + \int_{A\cap\{\varrho=k\}\cap\{\nu=k+2\}} X_\nu d\mathrm{P} \\
&+ \ldots + \int_{A\cap\{\varrho=k\}\cap\{\nu=N\}} X_N d\mathrm{P} = \int_{A\cap\{\varrho=k\}\cap\{\nu>k\}} X_\nu d\mathrm{P}.
\end{aligned} \tag{5.13}$$

Combining [5.12] and [5.13], we get [5.11].

2) $\Rightarrow$ 3). For $\sigma \leq \tau$, $\sigma \wedge \tau = \sigma$, so $\mathrm{E}(X_\tau|\mathcal{F}_\sigma) = X_\sigma$. Taking expectation, we get 3).

3) $\Rightarrow$ 1). Let $0 \leq n < N$, the event $A \in \mathcal{F}_n$, and put $\tau = N$ a.s. while $\sigma = n\mathbb{1}_A + N\mathbb{1}_{A^c}$. Then, for any $0 \leq l \leq N$, we have that the event

$$\{\sigma \leq l\} = \begin{cases} \emptyset, & l < n, \\ A & n \leq l < N \in \mathcal{F}_l, \\ \Omega, & l = N. \end{cases}$$

so that σ is a bounded stopping time. Therefore,

$$\mathrm{E}X_\sigma \leq \mathrm{E}X_N,$$

or, that is equivalent,

$$\mathrm{E}X_n\mathbb{1}_A + \mathrm{E}X_N\mathbb{1}_{A^c} \leq \mathrm{E}X_N,$$

$$\mathrm{E}X_n\mathbb{1}_A \leq \mathrm{E}X_N\mathbb{1}_A. \tag{5.14}$$

Inequality [5.14] means that $\mathrm{E}(X_N|\mathcal{F}_n) \geq X_n$. □

5.5.5. *Maximal inequalities for (sub-, super-) martingales*

An explicit calculation of maximum probabilities, i.e. probabilities of the form $\mathrm{P}\left\{\max_{0\le i\le N} X_i \ge a\right\}$, $\mathrm{P}\left\{\max_{0\le i\le N} |X_i| \ge a\right\}$ or $\mathrm{P}\left\{\sup_{i\ge 0} |X_i| \ge a\right\}$ are, as a rule, impossible even for processes with discrete time. For processes with continuous time, the situation can be characterized as even more involved. Even the reasonable upper and lower bounds for such probabilities are often not easy to find. However, for (sub-, super-) martingales, we can get the reasonable and applicable upper bounds for such probabilities.

THEOREM 5.10.– *1) Let* $\{X_n, \mathcal{F}_n, n \ge 0\}$ *be a submartingale. Then*

i) For any $a > 0$ *and* $N \in \mathbb{N}$,

$$\mathrm{P}\left\{\max_{0\le n\le N} X_n \ge a\right\} \le \frac{\mathrm{E}\left(X_N \mathbb{1}_{\max_{0\le n\le N} X_n \ge a}\right)}{a} \le \frac{\mathrm{E}X_N^+}{a};$$

ii) For any $a > 0$,

$$\mathrm{P}\left\{\min_{0\le n\le N} X_n \le -a\right\} \le \frac{\mathrm{E}\left(X_N \mathbb{1}_{\min_{0\le n\le N} X_n > -a}\right) - \mathrm{E}X_0}{a} \le \frac{\mathrm{E}X_N^+ - \mathrm{E}X_0}{a}.$$

2) Let $\{X_n, \mathcal{F}_n, n \ge 0\}$ *be a supermartingale. Then*

$$\mathrm{P}\left\{\max_{0\le n\le N} X_n \ge a\right\} \le \frac{\mathrm{E}X_0 + \mathrm{E}X_N^-}{a} \le \frac{2\mathrm{E}|X_N|}{a}.$$

PROOF.– 1), $i)$ Let $\tau = \inf\{n \ge 0 : X_n \ge a\} \wedge N$. Then $\tau \le N$, τ is a bounded stopping time; therefore, $\mathrm{E}X_\tau \le \mathrm{E}X_N$. Furthermore, let $A = \left\{\sup_{0\le n\le N} X_n \ge a\right\}$. Then $X_\tau \mathbb{1}_{A^c} = X_N \mathbb{1}_{A^c}$, and

$$\mathrm{E}X_\tau = \mathrm{E}X_\tau \mathbb{1}_A + \mathrm{E}X_\tau \mathbb{1}_{A^c} \ge a\mathrm{P}\{A\} + \mathrm{E}X_N \mathbb{1}_{A^c},$$

whence

$$a\mathrm{P}\{A\} \le \mathrm{E}X_N - \mathrm{E}X_N \mathbb{1}_{A^c} = EX_N \mathbb{1}_A = \mathrm{E}\left(X_N \mathbb{1}_{\max_{0\le n\le N} X_n \ge a}\right) \le \mathrm{E}X_N^+;$$

ii) Similarly, let $\sigma = \inf\{n \ge 0 : X_n \le -a\} \wedge N$. Let $B = \{\min_{0\le n\le N} X_n \le -a\}$. Then

$$\mathrm{E}X_0 \le \mathrm{E}X_\sigma \mathbb{1}_B + \mathrm{E}X_\sigma \mathbb{1}_{B^c} \le -a\mathrm{P}\{B\} + \mathrm{E}X_N \mathbb{1}_{B^c},$$

whence

$$\mathrm{P}\{B\} \le \frac{\mathrm{E}X_N \mathbb{1}_{B^c} - \mathrm{E}X_0}{a} \le \frac{\mathrm{E}X_N^+ - \mathrm{E}X_0}{a}.$$

Note that $\mathrm{E}X_N^+ \ge \mathrm{E}X_N \ge \mathrm{E}X_0$.

2) Let τ and A be the same as in 1), (i). Then the following relations hold: $X_N \mathbb{1}_{A^c} = (X_N^+ - X_N^-)\mathbb{1}_{A^c} \ge -X_N^- \mathbb{1}_{A^c}$, and

$$\mathrm{E}X_0 \ge \mathrm{E}X_\tau = \mathrm{E}X_\tau \mathbb{1}_A + \mathrm{E}X_N \mathbb{1}_{A^c} \ge a\mathrm{P}\{A\} - \mathrm{E}X_N^-.$$

Therefore,

$$\mathrm{P}\{A\} \le \frac{\mathrm{E}X_0 + \mathrm{E}X_N^-}{a} \le \frac{2\mathrm{E}|X_N|}{a}. \qquad \square$$

From this point for a random process $\{X_n, n \ge 0\}$, we denote

$$X_N^* = \max_{0 \le n \le N} |X_n|$$

the running maximum of the absolute value of X. Observe that for a non-negative process X, $X_n^* = \max_{0 \le k \le n} X_k$.

REMARK 5.9.– Let $\{X_n, \mathcal{F}_n, n \ge 0\}$ be a non-negative supermartingale. Then $\mathrm{E}X_N^- = 0$; therefore, it follows from theorem 5.10 2) that for any $a > 0$ and any $N \ge 0$,

$$\mathrm{P}\{X_N^* \ge a\} \le \frac{\mathrm{E}X_0}{a}.$$

THEOREM 5.11.– *Let $\{X_n, \mathcal{F}_n, n \ge 0\}$ be a martingale. Then, for any $p \ge 1$, any $a > 0$ and any $N \ge 0$,*

$$\mathrm{P}\{X_N^* \ge a\} \le \frac{\mathrm{E}|X_N|^p}{a^p}.$$

PROOF.– The statement is evident in the case where $\mathrm{E}|X_N|^p = \infty$. Now, let $\mathrm{E}|X_N|^p < \infty$. Then $\{|X_n|^p, \mathcal{F}_n, 0 \le n \le N\}$ is a submartingale, since for $p \ge 1$ $f(x) = |x|^p$ is a convex function. Therefore, according to theorem 5.10, 1), (i)

$$\mathrm{P}\{X_N^* \ge a\} = \mathrm{P}\left\{\left(X_N^*\right)^p \ge a^p\right\} \le \frac{\mathrm{E}|X_N|^p}{a^p}. \qquad \square$$

THEOREM 5.12.– *Let $\{X_n, \mathcal{F}_n, n \ge 0\}$ be a martingale or non-negative submartingale. Then*

1) For any $p > 1$ and any $N \geq 0$,

$$\mathrm{E}\big(X_N^*\big)^p \leq \left(\frac{p}{p-1}\right)^p \mathrm{E}|X_N|^p, \qquad [5.15]$$

or, in other words,

$$\|X_N^*\|_{\mathcal{L}_p(\Omega,\mathcal{F},\mathrm{P})} \leq \frac{p}{p-1} \|X_N\|_{\mathcal{L}_p(\Omega,\mathcal{F},\mathrm{P})}.$$

2) For any $N \geq 0$,

$$\mathrm{E}X_N^* \leq 2(1 + \mathrm{E}(|X_N| \log^+ |X_N|)),$$

where, for any $a > 0$, $\log^+ a = (\log a)\mathbb{1}_{a>1} = \log(a \vee 1)$.

PROOF.– 1) If X is a martingale, then $\{|X_n|, n \geq 0\}$ is a non-negative submartingale. Therefore, it is sufficient to establish [5.15] only for a non-negative submartingale X_n. Evidently, it is sufficient to consider the case where $\mathrm{E}X_N^p < \infty$. In this case, $\mathrm{E}X_n^p < \infty$ for any $0 \leq n \leq N$, because X_n^p is a submartingale, and $\max_{0\leq n\leq N} \mathrm{E}X_n^p < \mathrm{E}X_N^p < \infty$. Therefore,

$$\mathrm{E}(X_N^*)^p \leq \sum_{n=1}^{N} \mathrm{E}X_n^p < \infty.$$

Now, let $F_N(x)$ be the cumulative distribution function of X_N^*. Then, applying theorem 5.10, $1), (i)$, integrating by parts and applying the Fubini theorem and the Hölder inequality, we get the following relations

$$\begin{aligned} E(X_N^*)^p &= \int_0^\infty z^p dF_N(z) = p\int_0^\infty z^{p-1}(1 - F_N(z))dz \\ &= p\int_0^\infty z^{p-1}\mathrm{P}\left\{X_N^* \geq z\right\} dz \leq p\int_0^\infty z^{p-1}\frac{\mathrm{E}X_N}{z}\mathbb{1}_{X_N^*\geq z}dz \\ &= p\mathrm{E}\left(X_N\int_0^\infty z^{p-2}\mathbb{1}_{X_N^*\geq z}dz\right) = p\mathrm{E}\left(X_N\int_0^{X_N^*} z^{p-2}dz\right) \\ &= \frac{p}{p-1}\mathrm{E}\left(X_N(X_N^*)^{p-1}\right) \leq \frac{p}{p-1}\left(\mathrm{E}X_N^p\right)^{\frac{1}{p}}\left(\mathrm{E}(X_N^*)^p\right)^{\frac{p-1}{p}}. \end{aligned} \qquad [5.16]$$

Dividing left- and right-hand sides of [5.16] by $(\mathrm{E}(X_N^*)^p)^{\frac{p-1}{p}} < \infty$, we get the desired inequality.

2) Let $p = 1$. Then it follows from theorem 5.10 that for any $a > 0$ and $N \geq 0$

$$\mathrm{P}\{X_N^* \geq 2a\} \leq \frac{\mathrm{E}\left(X_N \mathbb{1}_{X_N^* \geq 2a}\right)}{2a} \leq \frac{\mathrm{E}\left(X_N \mathbb{1}_{X_N \geq a}\right)}{2a} + \frac{a\mathrm{P}\{X_N^* \geq 2a\}}{2a}$$
$$= \frac{\mathrm{E}\left(X_N \mathbb{1}_{X_N \geq a}\right)}{2a} + \frac{1}{2}\mathrm{P}\{X_N^* \geq 2a\},$$

whence

$$\mathrm{P}\{X_N^* \geq 2a\} \leq \frac{1}{a}\mathrm{E}\left(X_N \mathbb{1}_{X_N \geq a}\right).$$

Now, let $G_N(x) = \mathrm{P}\{X_N^* \geq 2x\}$. Then, again integrating by parts, we get that

$$\mathrm{E}\left(\frac{X_N^*}{2} - 1\right)^+ = \int_0^\infty (x-1)^+ dG_N(x) = \int_1^\infty (x-1) dG_N(x)$$
$$= \int_1^\infty (1 - G_N(x))dx \leq \int_1^\infty \frac{1}{x}\mathrm{E}\left(X_N \mathbb{1}_{X_N \geq x}\right)dx = \mathrm{E}\left(X_N \int_1^{X_N \vee 1} \frac{1}{x}dx\right)$$
$$= \mathrm{E}\left(X_N \left(\log(X_N \vee 1)\right)\right) = \mathrm{E}\left(X_N \log^+ X_N\right).$$

Therefore,

$$\mathrm{E}X_N^* \leq 2\frac{\mathrm{E}\left(X_N^* - 2\right)}{2} + 2$$
$$\leq 2\mathrm{E}\left(\frac{X_N^*}{2} - 1\right)^+ + 2 \leq 2\mathrm{E}\left(X_N \log^+ X_N\right) + 2.$$

□

5.5.6. *Doob decomposition for the integrable processes with discrete time*

THEOREM 5.13.– *Let $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ be an integrable process with discrete time. Then there exists a unique decomposition*

$$X_n = M_n + A_n, \quad n \geq 0, \qquad [5.17]$$

where $M = \{M_n, \mathcal{F}_n, n \geq 0\}$ is a martingale, $A_0 = 0$, $A = \{A_n, n \geq 1\}$ is a predictable process, i.e. A_n is a $\mathcal{F}_{n-1}$-measurable random variable for any $n \geq 1$.

PROOF.– Put

$$A_0 = 0, \quad A_1 - A_0 = \mathrm{E}(X_1 - X_0|\mathcal{F}_0),$$
$$A_2 - A_1 = \mathrm{E}(X_2 - X_1|\mathcal{F}_1), \ \ldots, \ A_n - A_{n-1} = \mathrm{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}), \ \ldots$$

Then it is very easy to see that A_n is a $\mathcal{F}_{n-1}$-measurable random variable. Moreover, A_n are integrable by definition of conditional expectation. Now, put $M_n = X_n - A_n$. Then M_n is a $\mathcal{F}_n$-measurable integrable random variable, and

$$\begin{aligned}&\mathrm{E}(M_n - M_{n-1}|\mathcal{F}_{n-1}) = \mathrm{E}(X_n - X_{n-1} - \mathrm{E}(X_n - X_{n-1}|\mathcal{F}_{n-1})|\mathcal{F}_{n-1})\\ &= \mathrm{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) - \mathrm{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) = 0.\end{aligned}$$

Therefore, $\{M_n, \mathcal{F}_n, n \geq 1\}$ is a martingale, and the existence of decomposition [5.17] is established. To prove the uniqueness, assume that $X_n = M_n^{'} + A_n^{'}$ is another decomposition, where $M^{'}$ is a martingale, $A^{'}$ is predictable and $A_0^{'} = 0$. Then, for any $n \geq 1$, $M_n - M_n^{'} = A_n^{'} - A_n$, so the difference $M_n - M_n^{'}$ is $\mathcal{F}_{n-1}$-measurable, $n \geq 1$. Consider conditional expectation $\mathrm{E}(M_n - M_n^{'}|\mathcal{F}_{n-1})$. It equals $M_{n-1} - M_{n-1}^{'}$ because $M_n - M_n^{'}$ is a martingale and simultaneously it equals $M_n - M_n^{'}$ due to the $\mathcal{F}_{n-1}$-measurability of $M_n - M_n^{'}$. Therefore,

$$\begin{aligned}M_n - M_n^{'} &= M_{n-1} - M_{n-1}^{'} = \ldots = M_0 - M_0^{'}\\ &= X_0 - A_0 - (X_0 - A_0^{'}) = 0,\end{aligned}$$

whence $M_n - M_n^{'}$ a.s. and $A_n - A_n^{'} = 0$ a.s. Uniqueness is proved. □

EXAMPLE 5.12.–

1) Let $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ be a submartingale. Then, in its Doob decomposition, $A_n - A_{n-1} = \mathrm{E}(X_n - X_{n-1}|\mathcal{F}_{n-1}) \geq 0$. This means that A is a non-decreasing and, consequently, non-negative process. Obviously, in the case where X is a supermartingale, A is a non-positive and non-increasing process.

2) Let $M = \{M_n, \mathcal{F}_n, n \geq 0\}$ be a square-integrable martingale. Then $M^2 = \left\{M_n^2, \mathcal{F}_n, n \geq 0\right\}$ is a submartingale. Consider its Doob decomposition. Put $A_0 = 0$,

$$\begin{aligned}A_n - A_{n-1} &= \mathrm{E}(M_n^2 - M_{n-1}^2|\mathcal{F}_{n-1}) = \mathrm{E}((M_{n-1} + \Delta M_n)^2 - M_{n-1}^2|\mathcal{F}_{n-1})\\ &= \mathrm{E}(2M_{n-1}\Delta M_n + (\Delta M_n)^2|\mathcal{F}_{n-1}) = \mathrm{E}((\Delta M_n)^2|\mathcal{F}_{n-1}),\end{aligned}$$

where $\Delta M_n = M_n - M_{n-1}$. Therefore,

$$A_n = \sum_{k=1}^{n} \mathrm{E}((\Delta M_k)^2|\mathcal{F}_{k-1}). \qquad [5.18]$$

We shall consider this process in the next section.

5.5.7. *Quadratic variation and quadratic characteristics: Burkholder–Davis–Gundy inequalities*

Let $X = \{X_n, n \geq 0\}$ be a stochastic process. Denote $\Delta X_n = X_n - X_{n-1}$, $n \geq 1$.

DEFINITION 5.11.– Quadratic variation *of the process X is a stochastic process of the form*

$$[X]_0 = 0, \;\; [X]_n = \sum_{i=1}^{n} (\Delta X_i)^2, n \geq 1.$$

Obviously, $[X]$ is a non-negative and non-decreasing process, adapted to the same filtration as X.

The next result is called the Burkholder–Davis–Gundy inequality. More precisely, the result for $p > 1$ belongs to Burkholder and Gundy, and for $p = 1$ it was proved by Davis.

THEOREM 5.14.– *Let $M = \{M_n, \mathcal{F}_n, n \geq 0\}$ be a martingale with $M_0 = 0$. Then, for any $p \geq 1$, there exist the constants $c_p > 0$, $C_p > 0$ such that, for any $n \geq 1$,*

$$c_p \mathrm{E}[M]_n^{\frac{p}{2}} \leq \mathrm{E} \max_{0 \leq k \leq n} |M_k|^p \leq C_p \mathrm{E}[M]_n^{\frac{p}{2}}. \qquad [5.19]$$

PROOF.– During this proof, we denote by c_p and C_p different constants depending only on p.

i) Let $p > 1$. We shall use the Khinchin inequality of the following form: let $\{a_n, n \geq 0\}$ be a sequence of real numbers such that $\sum_{n=0}^{\infty} a_n^2 < \infty$ and $\{\xi_n, n \geq 0\}$ are a sequence of iid symmetric Bernoulli random variables with $\mathrm{P}\{\xi_n = \pm 1\} = \frac{1}{2}$. Then, for any $p > 0$, there exist such constants c_p and C_p, such that

$$c_p \left(\sum_{n=0}^{\infty} a_n^2 \right)^{\frac{p}{2}} \leq \mathrm{E} \left| \sum_{n=0}^{\infty} a_n \xi_n \right|^p \leq C_p \left(\sum_{n=0}^{\infty} a_n^2 \right)^{\frac{p}{2}}. \qquad [5.20]$$

Now, consider the sequence $\{r_n(t), n \geq 0, t \in [0,1]\}$ of Rademacher functions, i.e. $r_n(t) = \pm 1$ for any $n \geq 0$, $t \in [0,1]$, $\int_0^1 r_n(t)dt = 0$ and

$$\int_0^1 r_n(t) r_m(t) dt = \delta_{nm} := \mathbb{1}_{n=m}.$$

The Rademacher function can be defined as $r_n(t) = \mathrm{sign}(\sin(2^n \pi \mathrm{t}))$, $n \geq 0$. They can be considered as independent r.v., if we put $(\Omega, \mathcal{F}, \mathrm{P}) = ([0,1], \mathcal{B}[0,1], \lambda_1)$. Now, for any $t \in [0,1]$, consider the following sequence of martingale transformations: $S_0(t) = 0$,

$$S_n(t) = \sum_{k=1}^{n} r_k(t)(M_k - M_{k-1}),\ n \geq 1.$$

Then $\{S_n(t), \mathcal{F}_n, n \geq 0\}$ is a martingale for any $t \in [0,1]$. Moreover, if we similarly transform $S_n(t)$, we get M_n, because

$$\sum_{k=1}^{n} r_k(t)(S_k(t) - S_{k-1}(t)) = \sum_{k=1}^{n} r_k^2(t)(M_k - M_{k-1}) = M_n,\ n \geq 1.$$

It follows immediately from theorem A2.17 that, for any $p > 1$,

$$\mathrm{E}|M_n|^p \leq C_p \mathrm{E}|S_n(t)|^p \leq C_p^2 \mathrm{E}|M_n|^p,\ \ n \geq 1,\ \ t \in [0,1],$$

where C_p depends only on p but not on n, t and M.

According to Khinchin's inequality [5.20],

$$c_p[M]_n^{p/2} \leq \int_0^1 |S_n(s)|^p ds \leq C_p[M]_n^{p/2} \text{ for any } n \geq 1. \qquad [5.21]$$

Taking expectation of all sides of [5.21], we get that

$$c_p \mathrm{E}[M]_n^{p/2} \leq \mathrm{E}|M_n|^p \leq C_p \mathrm{E}[M]_n^{p/2}. \qquad [5.22]$$

Since $p > 1$, it follows from inequality [5.15] that

$$\mathrm{E}(M_n^*)^p \leq \left(\frac{p}{p-1}\right)^p \mathrm{E}|M_n|^p, \qquad [5.23]$$

and the proof follows from [5.22] and [5.23].

ii) Now, let $p = 1$. Consider the Davis decomposition of martingale M, i.e. $M_n = M_n' + M_n''$. Its components satisfy evident inequalities:

$$M_n^* \leq (M')_n^* + (M'')_n^* \leq (M')_n^* + \sum_{k=1}^{n} |\Delta M_k''|, \qquad [5.24]$$

and

$$[M']_n^{1/2} \leq [M]_n^{1/2} + [M'']_n^{1/2} \leq [M]_n^{1/2} + \sum_{k=1}^{n} |\Delta M_k''|. \qquad [5.25]$$

According to lemma A2.7, [A2.34] and [5.25],

$$E(M')^*_n \le 3E[M']^{1/2}_n + E(\Delta M)^*_n \le 3E[M]^{1/2}_n + E\sum_{k=1}^n |\Delta M''_k| + E[M]^{1/2}_n, \quad [5.26]$$

and it follows from [A2.36] that

$$E\sum_{k=1}^n |\Delta M''_k| \le 4E(\Delta M)^*_n \le 4E[M]^{1/2}_n. \quad [5.27]$$

Summarizing, we get that $E(M')^*_n \le 8E[M]^{1/2}_n$ and $E(M'')^*_n \le 4E[M]^{1/2}_n$. Therefore, it follows from [5.24]–[5.26] and the last relations that

$$EM^*_n \le E(M')^*_n + E(M'')^*_n \le 8E[M]^{1/2}_n + 4E[M]^{1/2}_n \le 12E[M]^{1/2}_n.$$

This means that the right-hand side of [5.19] holds for $p = 1$ with the constant $C_1 = 12$. To get the left-hand side, we again use the Davis decomposition. This implies that

$$[M]^{1/2}_n \le [M']^{1/2}_n + [M'']^{1/2}_n \le [M']^{1/2}_n + \sum_{k=1}^n |\Delta M''_k|, \quad [5.28]$$

$$(M')^*_n \le M^*_n + (M'')^*_n \le M^*_n + \sum_{k=1}^n |\Delta M''_k|.$$

According to lemma A2.7, [A2.34], [A2.36] and the last inequality,

$$\begin{aligned} E[M']^{1/2}_n &\le 3E(M')^*_n + 4E(\Delta M)^*_n \le 3E(M')^*_n + 8EM^*_n \\ &\le 11EM^*_n + 3E\sum_{k=1}^n |\Delta M''_k| \le 11EM^*_n + 12E(\Delta M)^*_n \le 35EM^*_n. \end{aligned} \quad [5.29]$$

Substituting [5.29] and [5.27] into [5.28] and taking expectation, we get the left-hand side of [5.22] for $p = 1$ with $c_1 = 1/39$. □

REMARK 5.10.– Denote $[M] = \lim_{n\to\infty}[M]_n,\ M^* = \sup_{k\ge 0} |M_k|$. Then it follows immediately from [5.19] that for the same martingale as in theorem 5.14 and any $p \ge 1$, there exists constants $c_p > 0$, $C_p > 0$, such that

$$c_p E[M]^{p/2} \le E(M^*)^p \le C_p E[M]^{p/2}.$$

Now let $M = \{M_n, \mathcal{F}_n, n \ge 0\}$ be a square-integrable martingale. Consider the process of the form [5.18].

DEFINITION 5.12.– *A process* $\langle M \rangle_n = \sum_{k=1}^{n} \mathrm{E}((\Delta M_k)^2 | \mathcal{F}_{k-1})$ *is called a* quadratic characteristic *of martingale* M *it follows.*

It follows from example 5.12, 2) that $\langle M \rangle$ is a predictable process in the Doob decomposition of submartingale M^2. However, $\langle M \rangle$ is also a predictable process in the Doob decomposition of the quadratic variation, which is established in the following lemma.

LEMMA 5.2.– *The Doob decomposition of* $[M]$ *has a form*

$$[M]_n = N_n + \langle M \rangle_n\,, \ \ n \geq 0,$$

where $N = \{N_n, n \geq 0\}$ *is a martingale.*

PROOF.– The proof immediately follows from the formula $A_n - A_{n-1} = \mathrm{E}(X_n - X_{n-1} | \mathcal{F}_{n-1})$ in the Doob decomposition. Indeed, now $X_n = [M]_n$, $X_n - X_{n-1} = [M]_n - [M]_{n-1} = (\Delta M_n)^2$, and $A_n - A_{n-1} = \mathrm{E}((\Delta M_n)^2 | \mathcal{F}_{n-1})$, so that $A_n = \langle M \rangle_n$, $n \geq 0$. □

REMARK 5.11.– Quadratic characteristic $\langle M \rangle$ is also called a *dual predictable projection* of the quadratic variation $[M]$.

The next result is, to some extent, similar to the Burkholder–Davis–Gundy inequality; however, the reader should pay attention that its different parts hold for different values of p and take place simultaneously only for $p = 2$. We omit the proof of this result.

THEOREM 5.15.– *Let* $M = \{M_n, \mathcal{F}_n, n \geq 0\}$ *be a square integrable martingale,* $M_0 = 0$.

1) For any $p \in (0, 2]$, *there exists a constant* C_p, *such that, for any* $N \geq 0$,

$$\mathrm{E}(M_N^*)^p \leq C_p \mathrm{E} \langle M \rangle_N^{p/2}\,;$$

2) For any $p \geq 2$, *there exists a constant* c_p, *such that, for any* $N \geq 0$,

$$c_p \mathrm{E} \langle M \rangle_N^{p/2} \leq \mathrm{E}(M_N^*)^p.$$

5.5.8. *Change of probability measure and Girsanov theorem for discrete-time processes*

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space. Recall the notion of the equivalent probability measure (this notion is stronger than the notion of absolutely continuous probability measures, but we do not go into the details of absolutely continuous probability measures now).

DEFINITION 5.13.– *Probability measure* $\widetilde{\mathrm{P}}$ *on* $(\Omega, \mathcal{F})$ *is equivalent to measure* P $(\widetilde{\mathrm{P}} \sim \mathrm{P})$ *if for any set* $A \in \mathcal{F}$ $\mathrm{P}\{A\} = 0$ *if and only if* $\widetilde{\mathrm{P}}\{A\} = 0$.

According to the Radon–Nikodym theorem, for $\widetilde{\mathrm{P}} \sim \mathrm{P}$, there exists a non-negative integrable random variable $\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}$, such that, for any $B \in \mathcal{F}$

$$\widetilde{\mathrm{P}}\{B\} = \int_B \frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}(\omega)\mathrm{P}\{d\omega\}.$$

Moreover, $\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}} > 0$ a.s., $\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}} = \left(\frac{d\mathrm{P}}{d\widetilde{\mathrm{P}}}\right)^{-1}$ and $\mathrm{E}\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}} = 1$. We can state also that, for any positive a.s. random variable η, such that $\mathrm{E}\eta = 1$, the measure $\widetilde{\mathrm{P}}\{B\}$ defined as $\widetilde{\mathrm{P}}\{B\} = \int_B \eta d\mathrm{P}$ is equivalent to P: $\widetilde{\mathrm{P}} \sim \mathrm{P}$. Therefore, any random variable $\eta > 0$ with $\mathrm{E}\eta = 1$ defines the new equivalent probability measure.

Now, let us have a stochastic basis with filtration $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 0}, \mathrm{P})$ and let $\widetilde{\mathrm{P}} \sim \mathrm{P}$. Then, according to example 5.4, $\left\{\mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\middle| \mathcal{F}_n\right), n \geq 0\right\}$ is a martingale. Now our goal is to find the representation of this martingale and to study the transformation of a martingale by changing a probability measure to an equivalent one.

THEOREM 5.16.– *Let the probability measures be equivalent,* $\widetilde{\mathrm{P}} \sim \mathrm{P}$. *Then there exists a* P*-martingale* $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ *with initial value* $X_0 = 1$ *and increments* $\Delta X_{n+1} = X_{n+1} - X_n > -1$ *a.s., for any* $n \geq 0$, *such that the martingale* $Y_n := \mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\middle| \mathcal{F}_n\right)$ *admits the representation*

$$Y_0 = 1,\ Y_n = \prod_{k=1}^{n}(1 + \Delta X_k),\ n \geq 1. \qquad [5.30]$$

PROOF.– Put $X_0 = 1$ and let

$$X_{k+1} = X_k + \frac{Y_{k+1} - Y_k}{Y_k},\ k \geq 0. \qquad [5.31]$$

Then $\Delta X_{k+1} = \frac{Y_{k+1}}{Y_k} - 1 > -1$ a.s. because $Y_k > 0$ for any $k \geq 0$. Let us show that X is a martingale. Obviously, it is adapted to the filtration $\{\mathcal{F}_n, n \geq 0\}$. Integrability can be established by induction: $X_0 = 1$ is integrable, and if X_k is integrable, then

$$\mathrm{E}|X_{k+1}| \leq \mathrm{E}|X_k| + 1 + \mathrm{E}\left|\frac{Y_{k+1}}{Y_k}\right| = \mathrm{E}|X_k| + 1 + \mathrm{E}\frac{Y_{k+1}}{Y_k}.$$

Since Y is positive with probability 1 process, $\mathrm{E}\left(\frac{Y_{k+1}}{Y_k}\Big|\mathcal{F}_k\right)$ is well defined and, moreover, it equals $\frac{1}{Y_k}\mathrm{E}\left(Y_{k+1}|\mathcal{F}_k\right) = 1$. Therefore, $\mathrm{E}\frac{Y_{k+1}}{Y_k} = 1$, and $\mathrm{E}|X_{k+1}| \le \mathrm{E}|X_k| + 2$, i.e. X_{k+1} is integrable. Now, martingale property of X follows from the relation

$$\begin{aligned}\mathrm{E}(X_{k+1}|\mathcal{F}_k) &= X_k + \mathrm{E}\left(\frac{Y_{k+1}-Y_k}{Y_k}\Big|\mathcal{F}_k\right)\\ &= X_k + \frac{1}{Y_k}\mathrm{E}\left(Y_{k+1}-Y_k|\mathcal{F}_k\right) = X_k.\end{aligned}$$

Finally, [5.31] implies that $Y_{k+1} = Y_k(1+\Delta X_{k+1})$, and we get [5.30]. □

THEOREM 5.17.– *Let* $\widetilde{\mathrm{P}} \sim \mathrm{P}$. *An adapted process* $\widetilde{M}$ *is a* $\widetilde{\mathrm{P}}$*-martingale if and only if the process* $\widetilde{M}\cdot Y$ *is a* P*-martingale, where* $Y = \left\{Y_n = \mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}|\widetilde{\mathcal{F}}_n\right), n \ge 0\right\}$.

PROOF.– Denote $\mathrm{E}_{\widetilde{\mathrm{P}}}$ expectation w.r.t. a measure $\widetilde{\mathrm{P}}$. First, note that

$$\mathrm{E}_{\widetilde{\mathrm{P}}}|\widetilde{M}_n| = \mathrm{E}\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}|\widetilde{M}_n| = \mathrm{E}\left(\mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\Big|\widetilde{\mathcal{F}}_n\right)|\widetilde{M}_n|\right) = \mathrm{E}Y_n|\widetilde{M}_n|,$$

therefore, $\widetilde{M}$ is $\widetilde{\mathrm{P}}$-integrable if and only if $\widetilde{M}\cdot Y$ is P-integrable. Now, we use the following relation for conditional expectations w.r.t. different probability measures: for any non-negative r.v. ξ and $\mathcal{G} \subset \mathcal{F}$,

$$\mathrm{E}_{\widetilde{\mathrm{P}}}(\xi|\mathcal{G}) = \frac{\mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\xi\Big|\mathcal{G}\right)}{\mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\Big|\mathcal{G}\right)}.$$

According to this relation,

$$\begin{aligned}\mathrm{E}_{\widetilde{\mathrm{P}}}(\widetilde{M}_{n+1}|\mathcal{F}_n) &= \frac{\mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\widetilde{M}_{n+1}\Big|\mathcal{F}_n\right)}{\mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\Big|\mathcal{F}_n\right)}\\ &= \frac{\mathrm{E}\left(\mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\widetilde{M}_{n+1}\Big|\mathcal{F}_{n+1}\right)\Big|\mathcal{F}_n\right)}{\mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\Big|\mathcal{F}_n\right)} = \frac{\mathrm{E}(Y_{n+1}\widetilde{M}_{n+1}|\mathcal{F}_n)}{Y_n}.\end{aligned}$$

Therefore, $\mathrm{E}_{\widetilde{\mathrm{P}}}(\widetilde{M}_{n+1}|\mathcal{F}_n) = \widetilde{M}_n$ if and only if $\mathrm{E}(Y_{n+1}\widetilde{M}_{n+1}|\mathcal{F}_n) = Y_n\widetilde{M}_n$, whence the proof follows. □

The next result is called the Girsanov theorem for discrete-time processes (the Girsanov theorem for continuous-time processes is considered in section 8.9).

THEOREM 5.18.– *(Transformation of martingale under transformation of probability measure) Let $Z = \{Z_n, \mathcal{F}_n, 1 \le n \le N\}$ be an integrable process with Doob decomposition*

$$Z_n = M_n + A_n,\ n \ge 0,\ 0 \le n \le N,$$

where $M = \{M_n, \mathcal{F}_n, 0 \le n \le N\}$ is a square-integrable martingale, $A_0 = 0$ and $A = \{A_n, \mathcal{F}_n, 1 \le n \le N\}$ is a predictable process. Let the random variables

$$\Delta X_n := -\frac{\Delta A_n \Delta M_n}{\mathrm{E}\left((\Delta M_n)^2 | \mathcal{F}_{n-1}\right)},\ 1 \le n \le N$$

be bounded, and $\Delta X_n > -1$ a.s. Then the process Z is a martingale w.r.t. an equivalent probability measure $\widetilde{\mathrm{P}}$, such that

$$\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}} = \prod_{k=1}^{N}(1 + \Delta X_k). \qquad [5.32]$$

PROOF.– Define $\widetilde{\mathrm{P}}$ by relation [5.32]. Then $\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}} > 0$ a.s. and $\mathrm{E}\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}} = \mathrm{E}\sum_{k=1}^{N}(1 + \Delta X_k) = 1$, because

$$\mathrm{E}(1 + \Delta X_k | \mathcal{F}_{k-1}) = 1 + \frac{\Delta A_k \mathrm{E}(\Delta M_k | \mathcal{F}_{k-1})}{\mathrm{E}((\Delta M_k)^2 | \mathcal{F}_{k-1})} = 1.$$

Therefore, $\widetilde{\mathrm{P}} \sim \mathrm{P}$. Further, according to theorem 5.17, we need to check that $Z_n \cdot Y_n := Z_n \cdot \mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\middle| \mathcal{F}_n\right)$ is a P-martingale. To this end, we evaluate

$$\begin{aligned}
\mathrm{E}(Z_{n+1}Y_{n+1}|\mathcal{F}_n) &= \mathrm{E}\left(Z_{n+1} \cdot \mathrm{E}\left(\frac{d\widetilde{\mathrm{P}}}{d\mathrm{P}}\middle| \mathcal{F}_{n+1}\right)\middle| \mathcal{F}_n\right) \\
&= \mathrm{E}\left((M_{n+1} + A_{n+1}) \prod_{k=1}^{n+1}(1 + \Delta X_k) | \mathcal{F}_n\right) \\
&= \prod_{k=1}^{n}(1 + \Delta X_k)\left(\mathrm{E}(M_{n+1}(1 + \Delta X_{n+1}) | \mathcal{F}_n)\right. \\
&\quad \left. + A_{n+1}\mathrm{E}(1 + \Delta X_{n+1} | \mathcal{F}_n)\right) \qquad [5.33] \\
&= Y_n\left(M_n + \mathrm{E}(M_{n+1}\Delta X_{n+1} | \mathcal{F}_n) + A_{n+1}\right) \\
&= Y_n\left(M_n - \Delta A_{n+1}\frac{\mathrm{E}(M_{n+1}\Delta M_{n+1} | \mathcal{F}_n)}{\mathrm{E}((\Delta M_{n+1})^2 | \mathcal{F}_n)} + A_{n+1}\right) \\
&= Y_n\left(M_n - \Delta A_{n+1}\frac{\mathrm{E}((\Delta M_{n+1})^2 | \mathcal{F}_n)}{\mathrm{E}((\Delta M_{n+1})^2 | \mathcal{F}_n)} + A_{n+1}\right) \\
&= Y_n(M_n + A_n) = Y_n \cdot Z_n.
\end{aligned}$$

Here, we use the fact that, for the square integrable martingale, the following expectation vanishes: $\mathrm{E}(M_n \Delta M_{n+1} | \mathcal{F}_n) = 0$. The proof follows from [5.33]. □

5.5.9. *Strong law of large numbers for martingales with discrete time*

We assume that a stochastic basis $(\Omega, \mathcal{F}, \{\mathcal{F}_n\}_{n\geq 0}, \mathrm{P})$ is fixed and all processes are adapted to this basis. Let $X = \{X_n, n \geq 0\}$ be a stochastic process. Denote by $X_\infty = \lim_{n\to\infty} X_n$ and $\{X \to\}$ the set where this limit exists. Write $A \subset B$ a.s. if $\mathrm{P}\{A \backslash B\} = 0$. Introduce the notation for stopped process: for a Markov moment τ and random process $X = \{X_n, n \geq 0\}$, we denote $X^\tau = \{X_{n\wedge\tau}, n \geq 0\}$.

LEMMA 5.3.– *Let $Y = \{Y_n, n \geq 0\}$ and $Z = \{Z_n, n \geq 0\}$ be two processes with discrete time. Let, for any $a > 0$, $\mathrm{P}\{Z^{\tau_a} \to\} = 1$, where $\tau_a = \inf\{n > 0 : |Y_n| \geq a\}$. Then*

$$\left\{\sup_{n\geq 0} |Y_n| < \infty\right\} \subset \{Z \to\} \ \textit{a.s.}$$

PROOF.– For any $n \in \mathbb{N}$, we have that

$$\mathrm{P}\{Z \not\to, \tau_n = \infty\} = \mathrm{P}\{Z^{\tau_n} \not\to, \tau_n = \infty\} = 0.$$

Therefore,

$$\mathrm{P}\left\{Z \to, \bigcup_{n\geq 0} \{\tau_n = \infty\}\right\} = 0.$$

Now,

$$\left\{\sup_{k\geq 0} |Y_k| < \infty\right\} = \bigcup_{n\geq 0} \{\tau_n = \infty\}$$

$$= \left(\bigcup_{n\geq 0} \{\tau_n = \infty, Z \to\}\right) \cup \left(\bigcup_{n\geq 0} \{\tau_n = \infty, Z \not\to\}\right) \subset \{Z \to\} \ a.s.,$$

whence the proof follows. □

Now, let $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ be a non-negative and non-decreasing integrable process, $X_0 = 0$, $X = M + A$ its Doob decomposition.

THEOREM 5.19.– *We have that $\{A_\infty < \infty\} \subset \{X_\infty < \infty\}$ a.s.*

PROOF.– Recall that A is a non-negative and non-decreasing process as well.

i) Assume that $\mathrm{E}\left(A_{\tau}-A_{\tau-1}\right)\mathbb{1}_{\tau<\infty}<\infty$ for any Markov moment τ. Define $\tau_a=\inf\{n\geq 1: A_n\geq a\}$, $a>0$. Then

$$\mathrm{E}A_{\tau_a}=\mathrm{E}\left(A_{\tau_a-1}+\Delta A_{\tau_a}\mathbb{1}_{\tau_a<\infty}\right)\leq a+\mathrm{E}\Delta A_{\tau_a}\mathbb{1}_{\tau_a<\infty}<\infty.$$

Further, since X and A are non-integrable and non-decreasing, it follows from the Lebesgue monotone convergence theorem that

$$\mathrm{E}X_{\tau_a}=\lim_{n\to\infty}\mathrm{E}X_{\tau_a\wedge n},\ \mathrm{E}A_{\tau_a}=\lim_{n\to\infty}\mathrm{E}A_{\tau_a\wedge n}<\infty.$$

Moreover, it follows from Doob's optional stopping theorem that $\mathrm{E}M_{\tau_a\wedge n}=0$. Therefore,

$$\begin{aligned}0&=\lim_{n\to\infty}\mathrm{E}M_{\tau_a\wedge n}=\lim_{n\to\infty}\mathrm{E}\left(X_{\tau_a\wedge n}-A_{\tau_a\wedge n}\right)\\&=\lim_{n\to\infty}\left(\mathrm{E}X_{\tau_a\wedge n}-\mathrm{E}A_{\tau_a\wedge n}\right)=\mathrm{E}X_{\tau_a}-\mathrm{E}A_{\tau_a},\end{aligned}$$

whence $\mathrm{E}X_{\tau_a}=\mathrm{E}A_{\tau_a}<\infty$. In turn, it means that $\mathrm{P}\{X_{\tau_a}<\infty\}=1$, and $\mathrm{P}\{X^{\tau_a}\to\}=1$. Now we can apply lemma 5.3 with $Z=X$ and $Y=A$, getting that

$$\left\{\sup_{n\geq 0}A_n<\infty\right\}=\{A_\infty<\infty\}\subset\{X\to\}=\{X_\infty<\infty\}\ \text{a.s.}$$

ii) Consider the general case. Define two increasing processes with zero initial values, $A_0^{(1)}=A_0^{(2)}=0$, having the form

$$A_n^{(1)}=\sum_{k=1}^{n}\mathbb{1}_{\Delta A_k>1}\Delta A_k,\ \text{and}\ A_n^{(2)}=\sum_{k=1}^{n}\mathbb{1}_{\Delta A_k\leq 1}\Delta A_k.$$

Evidently, $A_n^{(1)}$ and $A_n^{(2)}$ are predictable components in the Doob decomposition of the process

$$X_n^{(1)}=\sum_{k=1}^{n}\mathbb{1}_{\Delta A_k>1}\Delta X_k,\ \text{and}\ X_n^{(2)}=\sum_{k=1}^{n}\mathbb{1}_{\Delta A_k\leq 1}\Delta X_k.$$

Moreover, $\Delta A_k^2\leq 1$ for any $k\geq 1$. Applying arguments from (i), we get that

$$\left\{A_\infty^{(2)}<\infty\right\}\subset\left\{X_\infty^{(2)}<\infty\right\}\ \text{a.s.}$$

Further,

$$A_\infty^{(1)} = \sum_{k=1}^{\infty} \mathbb{1}_{\Delta A_k > 1} \Delta A_k, \text{ and } X_\infty^{(1)} = \sum_{k=1}^{\infty} \mathbb{1}_{\Delta A_k > 1} \Delta X_k.$$

Therefore,

$$\left\{A_\infty^{(1)} < \infty\right\} \subset \left\{\sum_{k=1}^{\infty} \mathbb{1}_{\Delta A_k > 1} < \infty\right\}$$
$$\subset \left\{\sum_{k=1}^{\infty} \mathbb{1}_{\Delta A_k > 1} \Delta X_k < \infty\right\} = \left\{X_\infty^{(1)} < \infty\right\}.$$

Finally,

$$\{A_\infty < \infty\} = \left\{A_\infty^{(1)} < \infty\right\} \cap \left\{A_\infty^{(2)} < \infty\right\}$$
$$\subset \left\{X_\infty^{(1)} < \infty\right\} \cap \left\{X_\infty^{(2)} < \infty\right\} = \{X_\infty < \infty\}. \qquad \square$$

Let $M = \{M_n, n \geq 0\}$ be a square-integrable martingale and $L = \{L_n, n \geq 0\}$ be a predictable integrable non-decreasing and non-negative process, both of them starting from zero, $M_0 = 0$ and $L_0 = 0$. We say that the pair (M, L) satisfies the *strong law of large numbers* (SLLN) if $\frac{M_n}{L_n} \to 0$ a.s. as $n \to \infty$. In general, the set, on which $\lim_{n\to\infty} \frac{M_n}{L_n} = 0$, will be denoted as $\{\frac{M}{L} \to 0\}$. Denote $U_n = \sum_{k=1}^{n} \frac{\Delta M_k}{1+L_k}$, $n \geq 1$, $U_0 = 0$.

THEOREM 5.20.– *The following relation holds:*

$$\{L_\infty = \infty\} \cap \{U \to\} \subset \left\{\frac{M}{L} \to 0\right\} \textit{ a.s.}$$

PROOF.– For any $n \geq 1$,

$$\sum_{k=1}^{n} (1 + L_k)(U_k - U_{k-1}) = \sum_{k=1}^{n} (1 + L_k) \frac{\Delta M_k}{1 + L_k} = M_n, \qquad [5.34]$$

and evidently, the sets $\{L_\infty = \infty\} \cap \left\{\frac{M}{L} \to 0\right\}$ and $\{L_\infty = \infty\} \cap \left\{\frac{M}{1+L} \to 0\right\}$ coincide. Therefore, it is enough to prove that

$$\{L_\infty = \infty\} \cap \{U \to\} \subset \left\{\frac{M}{1+L} \to 0\right\} \text{ a.s.}$$

To this end, note that, for any $n \geq 1$, according to [5.34],

$$(1+L_n)U_n = \sum_{k=1}^{n}((1+L_k)U_k - (1+L_{k-1})U_{k-1})$$
$$= \sum_{k=1}^{n}(1+L_k)(U_k - U_{k-1}) + \sum_{k=1}^{n} U_{k-1}(L_k - L_{k-1})$$
$$= M_n + \sum_{k=1}^{n} U_{k-1}(L_k - L_{k-1}),$$

and

$$L_n U_n = \sum_{k=1}^{n} L_k(U_k - U_{k-1}) + \sum_{k=1}^{n} U_{k-1}(L_k - L_{k-1}).$$

This means that

$$\frac{M_n}{1+L_n} = U_n - \frac{\sum_{k=1}^{n} U_{k-1}(L_k - L_{k-1})}{1+L_n}$$
$$= \frac{U_n + L_n U_n - \sum_{k=1}^{n} U_{k-1}(L_k - L_{k-1})}{1+L_n}.$$

Evidently, $\{L_\infty = \infty\} \cap \{U \to\} \subset \left\{\frac{U}{1+L} \to 0\right\}$. Consider

$$\left| L_n U_n - \sum_{k=1}^{n} U_{k-1}(L_k - L_{k-1}) \right| = \left| \sum_{k=1}^{n} (U_n - U_{k-1})(L_k - L_{k-1}) \right|.$$

For any $\varepsilon > 0$ on the set $\{U \to\}$, there exists a number $n(\varepsilon, \omega)$, such that

$|U_\infty - U_n| < \varepsilon$ for $n > n(\varepsilon, \omega)$.

Therefore, for $n > n(\varepsilon, \omega)$,

$$\left| \sum_{k=1}^{n} (U_n - U_{k-1})(L_k - L_{k-1}) \right| \leq \sum_{k=1}^{n(\varepsilon,\omega)} |U_n - U_{k-1}|(L_k - L_{k-1})$$
$$+ \sum_{k=n(\varepsilon,\omega)+1}^{n} (|U_\infty - U_n| + |U_\infty - U_{k-1}|)(L_k - L_{k-1})$$
$$\leq 2 \max_{1 \leq k \leq n} |U_k| L_{n(\varepsilon,\omega)} + 2\varepsilon \cdot L_n,$$

whence on the set $\{U \to\} \cap \{L_\infty = \infty\}$,

$$\lim_{n\to\infty} \frac{|L_n U_n - \sum_{k=1}^n U_{k-1}(L_k - L_{k-1})|}{1+L_n}$$
$$\leq \lim_{n\to\infty} \left(\frac{2\max_{1\leq k\leq n} |U_k| L_{n(\varepsilon,\omega)}}{1+L_n} + 2\varepsilon \frac{L_n}{1+L_n} \right) \leq 2\varepsilon.$$

Since $\varepsilon > 0$ was chosen arbitrary, we get the proof. □

THEOREM 5.21.– *Let $M = \{M_n, n \geq 0\}$ be a square-integrable martingale, and $\langle M \rangle_n = \sum_{k=1}^n \mathrm{E}((\Delta M_k)^2 | \mathcal{F}_{k-1})$ be its quadratic characteristic. Let $\langle M \rangle_\infty = \infty$ a.s. Then the pair $(M, \langle M \rangle)$ satisfies SLLN, i.e.*

$$\frac{M_n}{\langle M \rangle_n} \to 0 \text{ a.s. as } n \to \infty.$$

PROOF.– Note that, for any $n \geq 1$,

$$\sum_{k=1}^n \frac{\mathrm{E}((\Delta M_k)^2 | \mathcal{F}_{k-1})}{(1+\langle M \rangle_k)^2} \leq \sum_{k=1}^n \frac{\mathrm{E}((\Delta M_k)^2 | \mathcal{F}_{k-1})}{(1+\langle M \rangle_k)(1+\langle M \rangle_{k-1})}$$
$$= \sum_{k=1}^n \frac{\langle M \rangle_k - \langle M \rangle_{k-1}}{(1+\langle M \rangle_k)(1+\langle M \rangle_{k-1})} = 1 - \frac{1}{1+\langle M \rangle_n} \leq 1.$$

According to theorem 5.20, it is sufficient to prove that the process $U_n = \sum_{k=1}^n \frac{\Delta M_k}{1+\langle M \rangle_k}$ converges a.s. Process U is a square-integrable martingale with quadratic characteristics

$$\langle U \rangle_n = \sum_{k=1}^n \frac{\mathrm{E}((\Delta M_k)^2 | \mathcal{F}_{k-1})}{(1+\langle M \rangle_k)^2} \leq 1,$$

as we established above. According to Burkholder–Davis–Gundy inequality (theorem 5.14),

$$\mathrm{E} \sup_{n\geq 0} |U_n| \leq C \mathrm{E} \langle U \rangle_\infty^{1/2} \leq C,$$

whence the proof follows. □

THEOREM 5.22.– *Let $M = \{M_n, n \geq 0\}$ be a square-integrable martingale, $M_0 = 0$. Then*

$$\{\langle M \rangle_\infty < \infty\} \subset \{M \to\} \text{ a.s.}$$

PROOF.– i) Let $\mathrm{E}\Delta \langle M\rangle_\tau \mathbb{1}_{\tau<\infty} < \infty$ for any Markov moment τ. Then

$$\mathrm{E}\langle M\rangle_{\tau_a} \le a + \mathrm{E}\Delta \langle M\rangle_{\tau_a} \mathbb{1}_{\tau_a<\infty} < \infty \text{ for } \tau_a = \inf\{n \ge 1 : \langle M\rangle_n \ge a\}.$$

According to Burkholder–Davis–Gundy inequality which can be applied to Markov moments as well,

$$\mathrm{E} \sup_{0\le n\le \tau_a} |M_n| \le C\mathrm{E}\langle M\rangle_{\tau_a}^{1/2} < \infty,$$

i.e. $\sup_{0\le n\le\tau_a} |M_n| < \infty$ a.s. Now we apply lemma 5.3 and get that

$$\{\langle M\rangle_\infty < \infty\} = \left\{\sup_{n\ge 0} \langle M\rangle_n < \infty\right\} \subset \{M \rightarrow\}.$$

ii) In general, consider the expansion $M = M^{(1)} + M^{(2)}$, where

$$M_n^{(1)} = \sum_{k=1}^n \mathbb{1}_{\Delta\langle M\rangle_k > 1} \Delta M_k,\ M_n^{(2)} = \sum_{k=1}^n \mathbb{1}_{\Delta\langle M\rangle_k \le 1} \Delta M_k.$$

Both $M^{(1)}$ and $M^{(2)}$ are square-integrable martingales,

$$\left\langle M^{(1)}\right\rangle_n = \sum_{k=1}^n \mathbb{1}_{\Delta\langle M\rangle_k > 1} \mathrm{E}((\Delta M_k)^2 | \mathcal{F}_{k-1}) = \sum_{k=1}^n \mathbb{1}_{\Delta\langle M\rangle_k > 1} \Delta \langle M\rangle_k,$$

$$\left\langle M^{(2)}\right\rangle_n = \sum_{k=1}^n \mathbb{1}_{\Delta\langle M\rangle_k \le 1} \Delta \langle M\rangle_k.$$

We can apply (i) to $\langle M^{(2)}\rangle$ and get that $\{\langle M^{(2)}\rangle_\infty < \infty\} \subset \{M^{(2)} \rightarrow\}$. Further,

$$\left\{\left\langle M^{(1)}\right\rangle_\infty < \infty\right\} \subset \left\{\sum_{k=1}^\infty \mathbb{1}_{\Delta\langle M\rangle_k > 1} < \infty\right\}$$

$$\subset \left\{\sum_{k=1}^\infty \mathbb{1}_{\Delta\langle M\rangle_k > 1} \Delta \langle M\rangle_k < \infty\right\} = \left\{M_\infty^{(1)} = \infty\right\}.$$

Now we conclude as in theorem 5.19. □

REMARK 5.12.– Concerning the strong law of large numbers for the martingale-type processes with continuous time, see section 8.7.1.

5.6. Lévy martingale stopped

Now we formulate two results that are true both for discrete- and continuous-time uniformly integrable martingales.

THEOREM 5.23.– *Let X be an integrable random variable, let $\{\mathcal{F}_t\}_{t\geq 0}$ be a flow of σ-fields satisfying assumptions (A) and let martingale $X_t = \mathrm{E}(X|\mathcal{F}_t)$. Then, for any stopping time τ, $X_\tau = \mathrm{E}(X|\mathcal{F}_\tau)$ a.s.*

PROOF.– First, consider discrete stopping times $\tau_n = \sum_{k=0}^{\infty} \frac{k+1}{2^n} \mathbb{1}_{\{\tau \in [\frac{k}{2^n}, \frac{k+1}{2^n})\}}$. Then, according to theorem 5.3, 4), for any number $k \geq 0$ and any event $A \in \mathcal{F}_{\tau_n}$, the event $A \cap \{\tau_n = \frac{k}{2^n}\} \in \mathcal{F}_{\tau_n \wedge \frac{k}{2^n}}$. Therefore,

$$\begin{aligned}\int_{A\cap\{\tau_n=\frac{k}{2^n}\}} \mathrm{E}(X|\mathcal{F}_{\tau_n}) d\mathrm{P} &= \int_{A\cap\{\tau_n=\frac{k}{2^n}\}} X d\mathrm{P} \\ &= \int_{A\cap\{\tau_n=\frac{k}{2^n}\}} \mathrm{E}(X|\mathcal{F}_{\frac{k}{2^n}}) d\mathrm{P} \\ &= \int_{A\cap\{\tau_n=\frac{k}{2^n}\}} X_{\frac{k}{2^n}} d\mathrm{P} = \int_{A\cap\{\tau_n=\frac{k}{2^n}\}} X_{\tau_n} d\mathrm{P}.\end{aligned} \qquad [5.35]$$

Since X_{τ_n} is $\mathcal{F}_{\tau_n}$-measurable, we get from the definition of conditional expectation and [5.35] that $\mathrm{E}(X|\mathcal{F}_{\tau_n}) = X_{\tau_n}$. Furthermore, note that according to corollary 5.1.9 from [COH 15], martingale X_t has a càdlàg modification. Now, let $n \to \infty$. Then $\tau_n \to \tau+$, therefore, $X_{\tau_n} \to X_\tau$. Apply theorem 5.8 which states that $\mathrm{E}(\xi|G_n) \to \mathrm{E}(\xi|G)$ when $\mathrm{E}|\xi| < \infty$ and σ-algebras G_n increase and $G = \sigma\left(\bigcup_{k=1}^{\infty} G_k\right)$. Then $\mathrm{E}(X|\mathcal{F}_{\tau_n}) \to \mathrm{E}(X|\mathcal{F}_\tau)$, and the proof follows. □

THEOREM 5.24.– *Let σ and τ be two stopping times and Y be an integrable random variable. Then*

$$\mathrm{E}(\mathrm{E}(Y|\mathcal{F}_\tau)|\mathcal{F}_\sigma) = \mathrm{E}(\mathrm{E}(Y|\mathcal{F}_\sigma)|\mathcal{F}_\tau) = \mathrm{E}(Y|\mathcal{F}_{\tau\wedge\sigma}).$$

PROOF.– Without loss of generality, we can assume that $Y \geq 0$. For any $n, m > 0$, introduce bounded stopping times $\tau_n = \tau \wedge n$ and $\sigma_m = \sigma \wedge m$. Consider the martingale $X_t = \mathrm{E}(Y|\mathcal{F}_t)$. Then, according to Doob's optional stopping theorem, we have that $\mathrm{E}(X_{\tau_n}|\mathcal{F}_{\sigma_m}) = X_{\tau_n\wedge\sigma_m}$. Furthermore, applying theorem 5.23, we can rewrite the last equality as

$$\mathrm{E}(\mathrm{E}(Y|\mathcal{F}_{\tau_n})|\mathcal{F}_{\sigma_m}) = \mathrm{E}(Y|\mathcal{F}_{\tau_n\wedge\sigma_m}).$$

Let n be fixed, $m \to \infty$. Apply theorem 5.8 which supplies that $\mathrm{E}(Y|\mathcal{F}_{\tau_n\wedge\sigma_m}) \to \mathrm{E}(Y|\mathcal{F}_{\tau_n\wedge\sigma})$ a.s. Further, $\mathrm{E}(Y|\mathcal{F}_{\tau_n})$ is an integrable random variable; therefore, by the same theorem, we have that $\mathrm{E}(\mathrm{E}(Y|\mathcal{F}_{\tau_n})|\mathcal{F}_{\sigma_m}) \to \mathrm{E}(\mathrm{E}(Y|\mathcal{F}_{\tau_n})|\mathcal{F}_{\sigma})$, $m \to \infty$. Therefore, we get that

$$\mathrm{E}(\mathrm{E}(Y|\mathcal{F}_{\tau_n})|\mathcal{F}_{\sigma}) = \mathrm{E}(Y|\mathcal{F}_{\tau_n \wedge \sigma}).$$

Similarly, $\mathrm{E}(Y|\mathcal{F}_{\tau_n \wedge \sigma}) \to \mathrm{E}(Y|\mathcal{F}_{\tau \wedge \sigma})$, and $\mathrm{E}(Y|\mathcal{F}_{\tau_n}) \to \mathrm{E}(Y|\mathcal{F}_{\tau})$ as $n \to \infty$. Therefore, in order to finalize the proof, it is sufficient to establish that the random variables $\{\xi_n = \mathrm{E}(Y|\mathcal{F}_{\tau_n}), n \geq 1\}$ are uniformly integrable. Now, for any $b > 0$ and $C > 0$,

$$\int_{\{\xi_n \geq C\}} \xi_n d\mathrm{P} = \mathrm{E}(Y \mathbb{1}_{\xi_n \geq C}) \leq b\mathrm{P}\{\xi_n \geq C\} + \mathrm{E}(Y \mathbb{1}_{Y \geq b}) \leq \frac{b}{C}\mathrm{E}\xi_n$$
$$+\mathrm{E}(Y \mathbb{1}_{Y \geq b}) \leq \frac{b}{C}\mathrm{E}Y + \mathrm{E}(Y \mathbb{1}_{Y \geq b}).$$

For any fixed $\varepsilon > 0$, choose $b > 0$, such that $\mathrm{E}(Y \mathbb{1}_{Y \geq b}) \leq \frac{\varepsilon}{2}$. Then choose $C > 0$, such that $\frac{b}{C}\mathrm{E}Y < \frac{\varepsilon}{2}$, and get that $\lim_{C \to \infty} \sup_{n \geq 1} \int_{\{\xi_n \geq C\}} \xi_n d\mathrm{P} \leq \varepsilon$. Since $\varepsilon > 0$ is arbitrary, we get that

$$\lim_{C \to \infty} \sup_{n \geq 1} \int_{\{\xi_n \geq C\}} \xi_n d\mathrm{P} = 0$$

which means that $\{\xi_n, n \geq 1\}$ are uniformly integrable. Therefore,

$$\lim_{n \to \infty} \mathrm{E}(\mathrm{E}(Y|\mathcal{F}_{\tau \wedge n})|\mathcal{F}_{\sigma}) = \mathrm{E}(\mathrm{E}(Y|\mathcal{F}_{\tau})|\mathcal{F}_{\sigma}) = \mathrm{E}(Y|\mathcal{F}_{\tau \wedge \sigma}). \qquad \square$$

5.7. Martingales with continuous time

Let us consider the case of continuous-time parameter, i.e. the parameter set $\mathbb{T}$ is either $\mathbb{R}_+$ or $[0, T]$ with some $T > 0$. Many of the results for martingales with discrete time are also valid for the continuous-time case. We will prove only those results which will be important in the proceeding. We start with optional stopping theorem.

THEOREM 5.25.– *Let $X = \{X_t, t \in\}$ be an integrable right-continuous stochastic process. Then the following statements are equivalent:*

1) X is a $\mathcal{F}_t$-martingale.

2) For any bounded stopping time τ and any stopping time σ,

$$\mathrm{E}(X_\tau \mid \mathcal{F}_\sigma) = X_{\tau \wedge \sigma}.$$

3) For any bounded stopping times $\sigma \le \tau$,

$$\mathrm{E}X_\tau = \mathrm{E}X_\sigma.$$

PROOF.– The implication 2) $\Rightarrow$ 1) is obvious: for arbitrary $0 \le s < t$, we can set $\sigma = s$ and $\tau = t$, getting $\mathrm{E}(X_t|\mathcal{F}_s) = X_s$, as needed.

To show 3) $\Rightarrow$ 2), first assume that $\sigma \le \tau \le T$ with some non-random T. For any $A \in \mathcal{F}_\sigma$, define $\sigma_A = \sigma \mathbb{1}_A + T\mathbb{1}_{A^c}$ and $\tau_A = \tau \mathbb{1}_A + T\mathbb{1}_{A^c}$. For any $t \in [0, T)$, we have $\{\sigma_A \le t\} = A \cap \{\sigma \le t\} \in \mathcal{F}_t$ by the definition of stopping time; for $t \ge T$, $\{\sigma_A \le t\} = \Omega \in \mathcal{F}_t$, so σ_A is a stopping time. Similarly, τ_A is a stopping time; moreover, $\sigma_A \le \tau_A$, so $\mathrm{E}X_{\sigma_A} = \mathrm{E}X_{\tau_A}$, equivalently, $\mathrm{E}X_\sigma \mathbb{1}_A = \mathrm{E}X_\tau \mathbb{1}_A$. Since by theorem 5.2, X_σ is $\mathcal{F}_\sigma$-measurable, we have from the last equality $\mathrm{E}(X_\tau \mid \mathcal{F}_\sigma) = X_\sigma$ by the definition of conditional expectation.

In general, we have

$$\mathrm{E}(X_\tau \mid \mathcal{F}_\sigma) = \mathrm{E}(X_\tau \mathbb{1}_{\tau<\sigma} \mid \mathcal{F}_\sigma) + \mathrm{E}(X_\tau \mathbb{1}_{\tau\ge\sigma} \mid \mathcal{F}_\sigma).$$

Since the process $X_t \mathbb{1}_{t<\sigma}$ is adapted and right-continuous, we have

$$\mathrm{E}(X_\tau \mathbb{1}_{\tau<\sigma} \mid \mathcal{F}_\sigma) = X_\tau \mathbb{1}_{\tau<\sigma} \qquad [5.36]$$

by virtue of theorem 5.2. Moreover, from the previous section, $\mathrm{E}(X_{\tau\vee\sigma} \mid \mathcal{F}_\sigma) = X_\sigma$. Since $\{\tau \ge \sigma\} \in \mathcal{F}_\sigma$ by theorem 5.2, we have $\mathrm{E}(X_{\tau\vee\sigma}\mathbb{1}_{\tau\ge\sigma} \mid \mathcal{F}_\sigma) = X_\sigma \mathbb{1}_{\tau\ge\sigma}$. Adding this to [5.36], we get $\mathrm{E}(X_\tau \mid \mathcal{F}_\sigma) = X_{\tau\wedge\sigma}$.

It remains to prove 1) $\Rightarrow$ 3). For any bounded stopping times $\sigma \le \tau$, consider their discrete approximations $\tau_n = f_n(\tau)$, $\sigma_n = f_n(\sigma)$ with $f_n(t) = \sum_{k=1}^{\infty} \frac{k}{n} \mathbb{1}_{(\frac{k-1}{n}, \frac{k}{n}]}(t)$. Evidently, these are bounded stopping times with $\sigma_n \le \tau_n$. Further, $\sigma_n \le \tau_n$, $\sigma_n \ge \sigma$, $\tau_n \to \tau$, and $\sigma_n \to \sigma$, $\tau_n \to \tau$, $n \to \infty$. Since $\{X_{k/n}, k \ge 1\}$ is a martingale with a discrete-time parameter, it follows from theorem 5.9 that $\mathrm{E}X_{\tau_n} = \mathrm{E}X_{\sigma_n}$. Thanks to the right-continuity, $X_{\sigma_n} \to X_\sigma$ and $X_{\tau_n} \to X_\tau$, $n \to \infty$. Therefore, in order to prove that $\mathrm{E}X_\tau = \mathrm{E}X_\sigma$, it suffices to show that the sequences $\{X_{\tau_n}, n \ge 1\}$ and $\{X_{\sigma_n}, n \ge 1\}$ are uniformly integrable. We have $\tau_n \le T$, and $\{X_{k/n}, k \ge 1\}$ is a martingale with discrete time, so by theorem 5.9, $X_{\tau_n} = \mathrm{E}(X_T \mid \mathcal{F}_{\tau_n})$. Therefore, for any $C > 0$,

$$\mathrm{E}(X_{\tau_n}\mathbb{1}_{X_{\tau_n}\ge C}) = \mathrm{E}(X_T \mathbb{1}_{X_{\tau_n}\ge C}).$$

Moreover, by the Jensen inequality,

$$\sup_{n\geq 1} \mathrm{P}\left\{X_{\tau_n} \geq C\right\} \leq \sup_{n\geq 1} \mathrm{P}\left\{|X_{\tau_n}| \geq C\right\}$$
$$\leq \frac{1}{C}\sup_{n\geq 1} \mathrm{E}\,|X_{\tau_n}| \leq \frac{1}{C}\mathrm{E}\,|X_T| \to 0, C \to \infty.$$

Therefore,

$$\sup_{n\geq 1} \mathrm{E}(X_{\tau_n} \mathbb{1}_{X_{\tau_n}\geq C}) = \sup_{n\geq 1} \mathrm{E}(X_T \mathbb{1}_{X_{\tau_n}\geq C}) \to 0,\ C \to \infty.$$

Similarly,

$$\sup_{n\geq 1} \mathrm{E}(-X_{\tau_n} \mathbb{1}_{X_{\tau_n}\leq -C}) \to 0,\ C \to \infty,$$

yielding the required uniform integrability. The one of $\{X_{\sigma_n}, n \geq 1\}$ is shown similarly, concluding the proof. □

Let us now address the maximal inequalities. For $T \geq 0$, denote

$$X_T^* = \sup_{t\in[0,T]} |X_t|$$

the running maximum of the absolute value of X.

THEOREM 5.26.– *Let $\{X_t, t \geq 0\}$ be a right-continuous martingale.*

1) For any $p \geq 1$, $a > 0$ and $T \geq 0$,

$$\mathrm{P}\left\{X_T^* \geq a\right\} \leq \frac{\mathrm{E}|X_T|^p}{a^p}.$$

2) For any $p > 1$ and any $T \geq 0$,

$$\mathrm{E}\big(X_T^*\big)^p \leq \left(\frac{p}{p-1}\right)^p \mathrm{E}|X_T|^p.$$

3) For any $T \geq 0$,

$$\mathrm{E}X_T^* \leq 2\big(1 + \mathrm{E}(|X_T| \log^+ |X_T|)\big),$$

where, for any $a > 0$, $\log^+ a = (\log a)\mathbb{1}_{a>1} = \log(a \vee 1)$.

PROOF.– For $n \geq 1$, consider a uniform partition $\Pi_n = \{t_k^n = kT/2^n, 0 \leq k \leq 2^n\}$ of $[0, T]$. Due to right continuity, $X_T^* = \lim_{n\to\infty} \max_{t\in\Pi_n} |X_t|$.

1) The sequence $\max_{t\in\Pi_n} |X_t|$ increases, so $\{X_T^* > b\}$ is the union of increasing events $\{\max_{t\in\Pi_n} |X_t| > b\}$. Due to continuity of probability,

$$\mathrm{P}\{X_T^* > b\} = \lim_{n\to\infty} \mathrm{P}\left\{\max_{t\in\Pi_n} |X_t| > b\right\}.$$

The process $\{X_t, t \in \Pi_n\}$ is a martingale with discrete time, so by theorem 5.11,

$$\mathrm{P}\left\{\max_{t\in\Pi_n} |X_t| > b\right\} \leq \frac{\mathrm{E}\,|X_T|}{b^p}.$$

Consequently,

$$\mathrm{P}\{X_T^* > b\} \leq \frac{\mathrm{E}\,|X_T|}{b^p}.$$

Setting $b = a - 1/k$ and letting $k \to \infty$, we get in view of continuity of probability

$$\mathrm{P}\{X_T^* \geq a\} \leq \frac{\mathrm{E}\,|X_T|}{a^p}.$$

2) Similar to 1), using theorem 5.12, we have for any $n \geq 1$

$$\mathrm{E}\max_{t\in\Pi_n} |X_t|^p \leq \left(\frac{p}{p-1}\right)^p \mathrm{E}|X_T|^p.$$

By the Fatou lemma,

$$\mathrm{E}\big(X_T^*\big)^p = \mathrm{E}\liminf_{n\to\infty} \max_{t\in\Pi_n} |X_t|^p \leq \liminf_{n\to\infty} \mathrm{E}\max_{t\in\Pi_n} |X_t|^p \leq \left(\frac{p}{p-1}\right)^p \mathrm{E}|X_T|^p.$$

The last statement is proved similarly. □

6

Regularity of Trajectories of Stochastic Processes

6.1. Continuity in probability and in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$

Let $\mathbb{T} = \mathbb{R}_+$, $X = \{X_t, t \in \mathbb{R}_+\}$ be a real-valued stochastic process. We can consider its restriction on some interval $[0, T]$, or some other subset $\mathbb{T}' \subset \mathbb{R}_+$, if necessary. If we restrict X on some subset $\mathbb{T}' \subset \mathbb{R}_+$, we suppose that $\mathbb{T}'$ consists of limit points.

DEFINITION 6.1.–

1) Stochastic process X is continuous in probability *(stochastically continuous) at a point $t_0 \in \mathbb{R}_+$ if $X_t \to X_{t_0}$ in probability as $t \to t_0$.*

2) Stochastic process X is continuous in probability (stochastically continuous) on the subset $\mathbb{T} \subset \mathbb{R}_+$ if it is continuous in probability at any point $t \in \mathbb{T}'$.

LEMMA 6.1.– *Let $X = \{X_t, t \in [0, T]\}$ be continuous in probability on $[0, T]$. Then, for any $\varepsilon > 0$,*

$$\lim_{\delta \to 0+} \sup_{\substack{|t-s| \le \delta \\ t,s \in [0,T]}} \mathrm{P}\{|X_s - X_t| \ge \varepsilon\} = 0.$$

PROOF.– Let $\varepsilon > 0$ be fixed. Denote

$$P_\delta := \sup_{\substack{|t-s| \le \delta \\ t,s \in [0,T]}} \mathrm{P}\{|X_s - X_t| \ge \varepsilon\}.$$

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

As P_δ is non-decreasing in δ, $\lim_{\delta\to 0+} P_\delta$ exists. Assume that $\lim_{\delta\to 0+} P_\delta = \alpha > 0$. According to the definition of supremum, there exist two sequences, $\{t_n, s_n, n \geq 1\}$, such that $|t_n - s_n| \leq \frac{1}{n}$, $t_n, s_n \in [0,T]$, and

$$\mathrm{P}\left\{|X_{s_n} - X_{t_n}| \geq \varepsilon\right\} \geq \frac{\alpha}{2}.$$

Consider a convergent subsequence $t_{n_k} \to t_0 \in [0,T]$, which exists since the closed bounded set $[0,T] \subset \mathbb{R}$ is compact. Then $s_{n_k} \to t_0$, $k \to \infty$, and $\mathrm{P}\left\{|X_{t_0} - X_{t_{n_k}}| \geq \frac{\varepsilon}{2}\right\} \geq \frac{\alpha}{4}$ or $\mathrm{P}\left\{|X_{t_0} - X_{s_{n_k}}| \geq \frac{\varepsilon}{2}\right\} \geq \frac{\alpha}{4}$ for any $k \geq 1$, which contradicts to continuity in probability of X at point t_0. □

REMARK 6.1.–

i) Lemma 6.1 is valid for any compact subset $\mathbb{T}' \subset \mathbb{R}_+$ consisting of limit points.

ii) Let X be continuous on $[0,T]$, so that its trajectories are a.s. continuous. Then, for any $t_0 \in [0,T]$, we have that $X_t \to X_{t_0}$ a.s. as $t \to t_0$; therefore, X is continuous in probability. However, the inverse statement fails. As an example, consider $\Omega = [0,1]$, $T = 1$, $\mathcal{F} = \mathcal{B}([0,1])$ a σ-algebra of the Borel sets, $\mathrm{P} = \lambda_1$, the Lebesgue measure on $\mathcal{F}$. Let $X_t(\omega) = \mathbb{1}_{t=\omega}$. Then all trajectories of X are discontinuous. More precisely, they have a discontinuity at the point $t = \omega$. However, for any $0 < \varepsilon < 1$

$$\begin{aligned}\mathrm{P}\left\{|X_t - X_s| \geq \varepsilon\right\} &= \mathrm{P}\left\{X_t = 1, X_s = 0\right\} + \mathrm{P}\left\{X_t = 0, X_s = 1\right\} \\ &\leq \lambda_1(\{t\}) + \lambda_1(\{s\}) = 0.\end{aligned}$$

LEMMA 6.2.– *Let the stochastic process X be continuous in probability on $[0,T]$. Then*

$$\lim_{C\to\infty} \sup_{t\in[0,T]} \mathrm{P}\left\{|X_t| \geq C\right\} = 0.$$

PROOF.– As $\rho_C := \sup_{t\in[0,T]} \mathrm{P}\left\{|X_t| \geq C\right\}$ is non-increasing in $C > 0$, the limit $\lim_{C\to\infty} \rho_C$ exists. Suppose that $\lim_{C\to\infty} \rho_C = \alpha > 0$. Then, according to the definition of supremum, there exists a sequence $\{t_n, n \geq 1\} \subset [0,T]$, such that

$$\mathrm{P}\left\{|X_{t_n}| \geq n\right\} \geq \frac{\alpha}{2}.$$

Consider a convergent subsequence $t_{n_k} \to t_0 \in [0,T]$. Then

$$\frac{\alpha}{2} \leq \mathrm{P}\left\{|X_{t_{n_k}}| \geq n_k\right\} \leq \mathrm{P}\left\{|X_{t_{n_k}} - X_{t_0}| \geq \frac{n_k}{2}\right\} + \mathrm{P}\left\{|X_{t_0}| \geq \frac{n_k}{2}\right\} \to 0,$$

as $k \to \infty$, and we get a contradiction. □

REMARK 6.2.– Lemma 6.2 holds if we replace interval $[0, T]$ with any compact subset $\mathbb{T}' \subset \mathbb{R}_+$. For completeness, we give an example of the process, which is not stochastically continuous at any point.

EXAMPLE 6.1.– *Let $X = \{X_t, t \in \mathbb{R}_+\}$ be a stochastic process with independent values, i.e. the random variables $\{X_t, t \in \mathbb{R}_+\}$ are mutually independent. Assume that X_t are identically distributed and their distribution is non-degenerate. Then the process X is not stochastically continuous at any point $t \in \mathbb{R}_+$. Indeed, for any $t > 0$ $(t = 0$ can be treated similarly$)$ and any $s \geq 0$*

$$\begin{aligned} \mathrm{P}\{|X_t - X_s| \geq \varepsilon\} &= \mathrm{P}\{X_s \geq X_t + \varepsilon\} + \mathrm{P}\{X_s \leq X_t - \varepsilon\} \qquad [6.1] \\ &= \int_{\mathbb{R}} \big(1 - F(x + \varepsilon -)\big) dF(x) + \int_{\mathbb{R}} F(x - \varepsilon) dF(x), \end{aligned}$$

where F is the cumulative distribution function of X_t. We see that the right-hand side of [6.1] *is strictly positive if F is not degenerate, and it does not depend on t and s; therefore, it does not converge to 0 as $s \to t$.*

DEFINITION 6.2.–

1) Stochastic process $X = \{X_t, t \in \mathbb{R}_+\}$, such that $\mathrm{E}X_t^2 < \infty$ for any $t \in \mathbb{R}_+$, is continuous in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ *(mean-square continuous) at the point $t_0 \in \mathbb{R}_+$ if $\mathrm{E}|X_t - X_{t_0}|^2 \to 0$ as $t \to t_0$.*

2) Stochastic process X is continuous in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ on some $\mathbb{T}' \subset \mathbb{R}_+$ if it is continuous in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ at any point $t \in \mathbb{T}'$.

Obviously, a stochastic process X continuous on $\mathbb{T}'$ in $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ is continuous on this set in probability.

6.2. Modification of stochastic processes: stochastically equivalent and indistinguishable processes

Let $X = \{X_t, t \in \mathbb{T}\}$ and $Y = \{Y_t, t \in \mathbb{T}\}$ be two stochastic processes defined on the same parametric set and on the same probability space.

DEFINITION 6.3.– *Processes X and Y are* stochastically equivalent *if for any $t \in \mathbb{T}$*

$$\mathrm{P}\{X_t = Y_t\} = 1.$$

DEFINITION 6.4.– *If X and Y are stochastically equivalent, we say that Y is a modification of X (and vice versa, X is a modification of Y).*

DEFINITION 6.5.– *Processes X and Y are* indistinguishable *if*

$$\mathrm{P}\{X_t = Y_t \ \forall t \in \mathbb{T}\} = 1.$$

Obviously,

$$\mathrm{P}\{X_t = Y_t, \forall t \in \mathbb{T}\} = \mathrm{P}\Big\{\bigcap_{t\in\mathbb{T}} \{X_t = Y_t\}\Big\}$$
$$\leq \mathrm{P}\{X_t = Y_t\} \quad \text{for any} \quad t \in \mathbb{T}.$$

Therefore, any indistinguishable processes are stochastically equivalent. Indistinguishability means that X and Y have the same trajectories $X_\cdot(\omega) = Y_\cdot(\omega)$, $\omega \in \Omega'$ where P$\{\Omega'\}$=1. It means that X and Y coincide, up to a negligible set. Vice versa, stochastic equivalence, generally speaking, does not imply indistinguishability, as shown in the following example.

EXAMPLE 6.2.– *Let* $\Omega = \mathbb{T} = [0,1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $\mathrm{P} = \lambda_1$ *(the Lebesgue measure on* $\mathcal{B}([0,1])$*),* $X_t \equiv 0$, $t \in [0,1]$, $Y_t = \mathbb{1}_{t=\omega}$. *Then* $\mathrm{P}\{X_t = Y_t, t \in [0,1]\} = 0$ *since the set* $\{\omega \in \Omega : X_t = Y_t, t \in [0,1]\}$ *is empty. However*

$$\mathrm{P}\{X_t = Y_t\} = \mathrm{P}\{t \neq \omega\} = 1 - \mathrm{P}\{t = \omega\} = 1 - \lambda_1(\{t\}) = 1,$$

from which X *and* Y *are stochastically equivalent.*

Under some additional assumptions, we may deduce indistinguishability from stochastic equivalence, as formulated in the following theorem.

THEOREM 6.1.– *Let* X *and* Y *be stochastically equivalent processes.*

1) If $\mathbb{T}$ *is at most countable, then* X *and* Y *are indistinguishable.*

2) If X *and* Y *are right-continuous, then they are indistinguishable.*

REMARK 6.3.– The second conclusion also holds for left-continuous processes.

PROOF.– 1) Since

$$\{X_t = Y_t \ \forall t \in \mathbb{T}\} = \bigcap_{t\in\mathbb{T}} \{X_t = Y_t\}$$

is at most a countable intersection of sets of probability 1, it has probability 1 as well.

2) Let $\mathbb{T}' \subset \mathbb{T}$ be a countable set, which is dense everywhere in $\mathbb{T}$. Also let $\mathbb{T}''$ be the set of right limit points of $\mathbb{T}$, i.e. points $t \in \mathbb{T}$ for which there exists a sequence $\{t_n, n \geq 1\} \subset \mathbb{T} \cap (t, +\infty)$ with $t_n \to t$, $n \to \infty$. The set $\mathbb{T} \setminus \mathbb{T}''$ is at most countable, since each of its points has a right neighborhood containing no other points from $\mathbb{T} \setminus \mathbb{T}''$. Thanks to right-continuity,

$$\{X_t = Y_t \ \forall t \in \mathbb{T}\} = \{X_t = Y_t \ \forall t \in \mathbb{T}'\} \cap \{X_t = Y_t \ \forall t \in \mathbb{T} \setminus \mathbb{T}''\}.$$

By 1),

$$\mathrm{P}\{X_t = Y_t \; \forall t \in \mathbb{T}'\} = \mathrm{P}\{X_t = Y_t \; \forall t \in \mathbb{T} \setminus \mathbb{T}''\} = 1,$$

whence the statement follows. □

6.3. Separable stochastic processes: existence of separable modification

Separability of stochastic processes is an important property that helps to establish regularity properties of its trajectories, like continuity or absence of second-kind discontinuities.

Assume that the set $\mathbb{T}$ consists of its limit points, and that $(\Omega, \mathcal{F}, \mathrm{P})$ is a complete probability space.

DEFINITION 6.6.– *A real-valued stochastic process* $X = \{X_t, t \in \mathbb{T}\}$ *is* separable *on* $\mathbb{T}$ *if there exists a set* $\Phi \subset \Omega$, $\Phi \in \mathcal{F}$, *such that* $\mathrm{P}\{\Phi\} = 0$, *and a countable subset* $M \subset \mathbb{T}$, *dense in* $\mathbb{T}$, *such that for any* $\omega \in \Omega \backslash \Phi$ *and any* $t \in \mathbb{T}$,

$$X_t(\omega) \in \left[\liminf_{s \to t, s \in M} X_s(\omega), \limsup_{s \to t, s \in M} X_s(\omega) \right].$$

The countable dense set $M \in \mathbb{T}$ is called a *separant* of $\mathbb{T}$. Separability is a rather weak property; in a sense, any "reasonable" real-valued stochastic process has a separable modification.

THEOREM 6.2.– *Let* $X = \{X_t, \; t \in \mathbb{T}\}$ *be a real-valued stochastically continuous process and* $\mathbb{T}$ *be a separable set. Then there exists a separable process* $Y = \{Y_t, t \in \mathbb{T}\}$ *taking values in the extended phase space* $\overline{\mathbb{R}} = [-\infty, \infty]$ *and stochastically equivalent to* X, *wherein any countable dense set* $M \subset \mathbb{T}$ *can serve as a separant.*

PROOF.– Define the process Y as follows. Let M be any separant; then for any $t \in M$ put $Y_t := X_t$. For $t \in \mathbb{T} \backslash M$ and $\omega \in \Omega$, such that

$$X_t(\omega) \in \left[\liminf_{s \to t, s \in M} X_s(\omega), \limsup_{s \to t, s \in M} X_s(\omega) \right],$$

we put $Y_t(\omega) := X_t(\omega)$. For any $t \in \mathbb{T} \backslash M$ and for any $\omega \in \Omega$, such that

$$X_t(\omega) \notin \left[\liminf_{s \to t, s \in M} X_s(\omega), \limsup_{s \to t, s \in M} X_s(\omega) \right],$$

we put $Y_t(\omega) := \limsup_{s\to t, s\in M} X_s(\omega)$. Alternatively, it is possible to put $Y_t(\omega) := \liminf_{s\to t, s\in M} X_s(\omega)$. Note that both conventions can lead to the expansion of phase space, since the limits can be infinite.

Now the rest of the proof is divided into several steps.

i) Let us prove that $Y = \{Y_t, t \in \mathbb{T}\}$ is a stochastic process. In this connection, it is necessary to establish that, for any $t \in \mathbb{T}$, Y_t is a random variable. It is evident for $t \in M$. Further, for $t \in \mathbb{T}\backslash M$

$$\overline{X}_t(\omega) := \limsup_{s\to t, s\in M} X_s(\omega) = \inf_{m\geq 1} \sup_{\substack{|s-t|\leq \frac{1}{m} \\ s\in M}} X_s(\omega)$$

is a random variable, and similarly

$$\underline{X}_t(\omega) := \liminf_{s\to t, s\in M} X_s(\omega)$$

is a random variable. Therefore, if we denote for any $t \in \mathbb{T}$

$$A_t := \left\{\omega \in \Omega : X_t(\omega) \in \left[\liminf_{s\to t, s\in M} X_s(\omega), \limsup_{s\to t, s\in M} X_s(\omega)\right]\right\},$$

then for $t \in \mathbb{T}\backslash M$

$$A_t = \left\{\omega \in \Omega : \underline{X}_t(\omega) \leq X_t(\omega) \leq \overline{X}_t(\omega)\right\} \in \mathcal{F}.$$

Therefore, $Y_t(\omega) = X_t(\omega)\mathbb{1}_{A_t} + \overline{X}_t(\omega)\mathbb{1}_{A_t^c}$ is a random variable for any $t \in \mathbb{T}\backslash M$.

ii) The process Y is a separable process. Indeed, for $t \in \mathbb{T}\backslash M$ $Y_t(\omega) \in \left[\underline{X}_t(\omega), \overline{X}_t(\omega)\right]$, but $\underline{X}_t(\omega)$ and $\overline{X}_t(\omega)$ are defined by the values of X and M, and on M, X coincides with Y. Therefore, for $t \in \mathbb{T}\backslash M$

$$Y_t(\omega) \in \left[\liminf_{s\to t, s\in M} Y_s(\omega), \limsup_{s\to t, s\in M} Y_s(\omega)\right]$$

for any $\omega \in \Omega$. Now let $t \in M$. Then we have

$$\mathrm{P}\left\{Y_t(\omega) \notin \left[\liminf_{s\to t, s\in M} Y_s(\omega), \limsup_{s\to t, s\in M} Y_s(\omega)\right]\right\} = \mathrm{P}\{A_t^c\}$$

$$= \mathrm{P}\left\{X_t(\omega) \notin \left[\liminf_{s\to t, s\in M} X_s(\omega), \limsup_{s\to t, s\in M} X_s(\omega)\right]\right\}$$

$$= \mathrm{P}\left\{\liminf_{s\to t, s\in M} |X_s - X_t| > 0\right\}$$

$$= \mathrm{P}\left\{\bigcup_{m\geq 1}\bigcup_{l\geq 1}\bigcap_{|s-t|\leq \frac{1}{l}, s\in M}\left\{|X_s - X_t| \geq \frac{1}{m}\right\}\right\} \qquad [6.2]$$

$$\leq \sum_{m=1}^{\infty}\lim_{l\to\infty}\mathrm{P}\left\{\bigcap_{|s-t|\leq \frac{1}{l}, s\in M}\left\{|X_s - X_t| \geq \frac{1}{m}\right\}\right\}$$

$$\leq \sum_{m=1}^{\infty}\lim_{l\to\infty}\inf_{\substack{|s-t|\leq \frac{1}{l}\\ s\in M}}\mathrm{P}\left\{|X_s - X_t| \geq \frac{1}{m}\right\}$$

$$\leq \sum_{m=1}^{\infty}\lim_{\substack{s\to t\\ s\in M}}\mathrm{P}\left\{|X_s - X_t| \geq \frac{1}{m}\right\} = 0,$$

since X is a stochastically continuous process. Denote $\Phi = \bigcup_{t\in M} A_t^c$. Then $\mathrm{P}\{\Phi\} = 0$ and for any $\omega \in \Omega\backslash\Phi$ and any $t \in \mathbb{T}$

$$Y_t(\omega) \in \left[\liminf_{s\to t, s\in M} Y_s(\omega), \limsup_{s\to t, s\in M} Y_s(\omega)\right],$$

from which Y is a separable process.

iii) The process Y is stochastically equivalent to X. Indeed, for any $t \in M$ and any $\omega \in \Omega$, $X_t = Y_t$. For $t \in \mathbb{T}\backslash M$, we have that

$$\mathrm{P}\{X_t \neq Y_t\} = \mathrm{P}\left\{\omega \in \Omega : X_t(\omega) \notin \left[\underline{X}(\omega), \overline{X}_t(\omega)\right]\right\} = 0$$

because this equality was proved in [6.2] for $t \in M$, but the proof is based on the fact that X is stochastically continuous at point t and irrespective of whether t belongs to M or $\mathbb{T}\backslash M$. □

6.4. Conditions of D-regularity and absence of the discontinuities of the second kind for stochastic processes

The notion of D-regularity of the function is introduced in section A1.4, where the criteria of D-regularity are formulated in terms of the modulus of continuity Δ_d. The notion of function without discontinuities of the second kind is well known, and in section A1.4, the criteria for a function to have no discontinuities of the second kind is formulated in terms of ε-oscillations. Note that obviously the property to be D-regular is stronger than to have no discontinuities of the second kind. In this section, we consider first the Skorokhod conditions of D-regularity of the trajectories of the stochastic process in terms of three-dimensional distributions and the conditions for the process to have no discontinuities of the second kind in terms of conditional probabilities of the big increments.

6.4.1. *Skorokhod conditions of D-regularity in terms of three-dimensional distributions*

Now our goal is to consider the sufficient conditions for the D-regularity of a stochastic process, which means that its trajectories have no discontinuities of the second kind and, at any point, have at least one of the one-sided limits. First, we prove an auxiliary result. Let the interval $[0,T]$ be fixed.

THEOREM 6.3.– *Let $X=\{X_t, t\in[0,T]\}$ be a real-valued separable stochastically continuous process, satisfying the condition: there exists a strictly positive non-decreasing function $g(h)$ and a function $q(C,h)$, $h\geq 0$, such that, for any $0\leq h\leq t\leq T-h$ and any $C>0$*

$$\mathrm{P}\left\{\min(|X_{t+h}-X_t|,|X_t-X_{t-h}|)>Cg(h)\right\}\leq q(C,h),$$

and $G(0)<\infty,\ Q(1,C)<\infty$. Define

$$\delta_n=\frac{T}{2^n},\ G(n)=\sum_{k=n}^{\infty} g\left(\delta_k\right),\ Q(n,C):=\sum_{k=n}^{\infty} 2^k q\left(C,\delta_k\right).$$

Then, for any $\varepsilon>0$

i) $$\mathrm{P}\left\{\sup_{0\leq s\leq t\leq T}|X_t-X_s|>\varepsilon\right\}\leq \mathrm{P}\left\{|X_T-X_0|>\frac{\varepsilon g(T)}{2G(0)}\right\}+Q\left(\frac{\varepsilon}{2G(0)}\right).$$

ii) $$\mathrm{P}\left\{\Delta_d^1(X,[0,T],\varepsilon)>2CG\left(\left[\log_2\frac{T}{2\varepsilon}\right]\right)\right\}\leq Q\left(\left[\log_2\frac{T}{2\varepsilon}\right],C\right),$$

where

$$\Delta_d^1(X,[0,T],\varepsilon)=\sup_{\substack{0\leq t_1<t_2<t_3\leq T,\\ t_3-t_1<\varepsilon}}\min(|X_{t_2}-X_{t_1}|,|X_{t_3}-X_{t_2}|).$$

REMARK 6.4.– Compare the modulus $\Delta_d^1(X,[0,T],\varepsilon)$, which is just introduced, with the modulus $\Delta_d(X,[0,T],\varepsilon)$ introduced in [A1.4]. We see that the latter also contains the values of the increments in the left and right ends of the interval $[0,T]$.

PROOF.– Define $t_k^n = k\delta_n$, $k = 0,1,\ldots,2^n$.

i) Introduce the following events. Let

$$A_{n,k} = \left\{\omega \in \Omega : \left|X_{t_{k+1}^n} - X_{t_k^n}\right| \le Cg\left(\delta_n\right)\right\}, \quad n \ge 0,\ 0 \le k \le 2^n - 1,$$

$$B_{n,k} = A_{n,k-1} \cup A_{n,k}, \quad n \ge 1,\ 1 \le k \le 2^{n-1},$$

$$D_n = \bigcap_{m=n}^{\infty} \bigcap_{k=1}^{2^m-1} B_{mk}, \;\; D_0 = D_1 \cap A_{0,0}. \qquad [6.3]$$

Since X is a separable stochastically continuous process, we can consider any dense countable set, e.g. the set $M = \{t_k^n, n \ge 0, 0 \le k \le 2^n\}$, as a separant. Recall that for the set A we denote its complement in the whole space by A^c, which is Ω now, so that $A^c = \Omega \setminus A$. Note that $\mathrm{P}\left\{B_{nk}^c\right\} \le q\left(C,\delta_n\right)$, and therefore,

$$\mathrm{P}\left\{D_n^c\right\} \le \sum_{m=n}^{\infty} \sum_{k=1}^{2^m-1} \mathrm{P}\left\{B_{mk}^c\right\} \le \sum_{m=n}^{\infty} 2^m q\left(C,\delta_m\right) = Q(n,C),$$

and

$$\mathrm{P}\left\{D_0^c\right\} \le \mathrm{P}\left\{A_{0,0}^c\right\} + \mathrm{P}\left\{D_1^c\right\} \le \mathrm{P}\left\{|X_T - X_0| > Cg(T)\right\} + Q(1,C). \quad [6.4]$$

Let the event D_0 hold. Then $A_{0,0}$ holds, so that $|X_T - X_0| \le Cg(T)$, and $B_{1,1}$ holds so that at least one of the two events holds:

$$|X_T - X_{\delta_1}| \le Cg\left(\delta_1\right) \;\text{ or }\; |X_{\delta_1} - X_0| \le Cg\left(\delta_1\right).$$

Then it follows from triangle inequality that in any case

$$|X_{\delta_1} - X_0| \le Cg\left(T\right) + Cg\left(\delta_1\right), \text{ and } |X_T - X_{\delta_1}| \le Cg\left(T\right) + Cg\left(\delta_1\right).$$

Now we shall apply induction. Assume that D_0 holds and suppose that the inequality

$$\left|X_{t_k^n} - X_{t_j^n}\right| \le Cg\left(T\right) + 2C\sum_{k=1}^{n} g\left(\delta_k\right) \qquad [6.5]$$

is true for some $n \geq 1$ any $k, j = 0, 1, \ldots, 2^n$. Let us establish the same inequality for $n+1$. Let, e.g. k and j be even numbers: $k = 2k_1$, $j = 2j_1$. Then it follows from [6.5] that

$$\left|X_{t_{2k_1}^{n+1}} - X_{t_{2j_1}^{n+1}}\right| = \left|X_{t_{k_1}^n} - X_{t_{j_1}^n}\right| \leq Cg(T) + 2C\sum_{k=1}^{n} g\left(\delta_k\right). \qquad [6.6]$$

Now, let k and j be odd numbers: $k = 2k_1 + 1$, $j = 2j_1 + 1$ (the cases where one of the numbers is even and another is odd can be considered similarly). Then k and j are situated between two even numbers: $2k_1 < k < 2k_1 + 2$ and $2j_1 < j < 2j_1 + 2$, and it follows that $B_{n+1,2k_1+2}$ and B_{n,j_1+1} hold, i.e.

$$\left|X_{t_{2k_1+2}^{n+1}} - X_{t_{2k_1+1}^{n+1}}\right| \leq Cg\left(\delta_{n+1}\right) \text{ or } \left|X_{t_{2k_1+1}^{n+1}} - X_{t_{2k_1}^{n+1}}\right| \leq Cg\left(\delta_{n+1}\right)$$

holds, and

$$\left|X_{t_{2j_1+2}^{n+1}} - X_{t_{2j_1+1}^{n+1}}\right| \leq Cg\left(\delta_{n+1}\right) \text{ or } \left|X_{t_{2j_1+1}^{n+1}} - X_{t_{2j_1}^{n+1}}\right| \leq Cg\left(\delta_{n+1}\right)$$

holds. Let, e.g.

$$\left|X_{t_{2k_1+1}^{n+1}} - X_{t_{2k_1}^{n+1}}\right| \leq Cg\left(\delta_{n+1}\right)$$

and

$$\left|X_{t_{2j_1+2}^{n+1}} - X_{t_{2j_1+1}^{n+1}}\right| \leq Cg\left(\delta_{n+1}\right).$$

Then

$$\left|X_{t_{2k_1+1}^{n+1}} - X_{t_{2j_1+1}^{n+1}}\right| \leq \left|X_{t_{2k_1+1}^{n+1}} - X_{t_{2k_1}^{n+1}}\right| + \left|X_{t_{2j_1+1}^{n+1}} - X_{t_{2j_1+2}^{n+1}}\right|$$
$$+ \left|X_{t_{k_1}^n} - X_{t_{j_1+1}^n}\right| \leq Cg(T) + 2C\sum_{k=1}^{n} g\left(\delta_k\right) + 2Cg\left(\delta_{n+1}\right).$$

Therefore, we prove [6.5] for all $n \geq 0$. It means that, for $\omega \in D_0$,

$$\left|X_{t_{2k_1}^{n+1}} - X_{t_{2j_1}^{n+1}}\right| \leq 2CG(0), \qquad [6.7]$$

and it follows from separability of X that for any $0 \leq s < t \leq T$ $|X_t - X_s| \leq 2CG(0)$ for $\omega \in D_0$. Now, for any $\varepsilon > 0$, put $C = \frac{\varepsilon}{2G(0)}$. Then it follows from [6.4] that

$$\mathrm{P}\left\{\sup_{0\leq s<t\leq T}|X_t - X_s| > \varepsilon\right\} \leq \mathrm{P}\{D_0^c\} \leq \mathrm{P}\left\{|X_T - X_0| > \frac{\varepsilon g(T)}{2G(0)}\right\} + Q\left(1, \frac{\varepsilon}{2G(0)}\right).$$

ii) Set $n = \left[\log_2 \frac{T}{\varepsilon}\right]$ so that $\varepsilon \leq \delta_n$ and fix some $i \in \{1, \ldots, 2^n - 1\}$ temporarily. Denote $M_m = \{t_k^m, k = 0, \ldots, 2^m\}$, $m \geq 1$. Let us prove by induction that, for any $\omega \in D_n$ and any $m \geq n$, there exists $\sigma_m \in [t_{i-1}^n, t_{i+1}^n) \cap M_m$, such that

$$\max_{t_{i-1}^n \leq t \leq \sigma_m,\ t \in M_m} \left|X_{t_{i-1}^n} - X_t\right| \leq C\sum_{k=n}^{m} g\left(\delta_k\right), \qquad [6.8]$$

$$\max_{\sigma_m < t \leq t_{i+1}^n,\ t \in M_m} \left|X_t - X_{t_{i+1}^n}\right| \leq C\sum_{k=n}^{m} g\left(\delta_k\right), \qquad [6.9]$$

and σ_m is non-decreasing in m.

Let $m = n$. For $\omega \in D_n$

$$\left|X_{t_{i+1}^n} - X_{t_i^n}\right| \leq Cg\left(\delta_n\right)$$

or

$$\left|X_{t_i^n} - X_{t_{i-1}^n}\right| \leq Cg\left(\delta_n\right).$$

In the first case, we can choose $\sigma_n = t_{i-1}^n$, and, in the second case, $\sigma_n = t_i^n$; if both inequalities hold, the choice can be t_{i-1}^n or t_i^n.

Let σ_m be already chosen. For $\omega \in D_n$,

$$\left|X_{\sigma_m+\delta_{m+1}} - X_{\sigma_m}\right| \leq Cg\left(\delta_{m+1}\right)$$

or

$$\left|X_{\sigma_m+\delta_{m+1}} - X_{\sigma_m+\delta_m}\right| \leq Cg\left(\delta_{m+1}\right).$$

In the first case, we can choose $\sigma_{m+1} = \sigma_m + \delta_{m+1}$, and, in the second case, $\sigma_{m+1} = \sigma_m$; if both inequalities hold, the choice of these two values is arbitrary.

Denote $\sigma = \lim_{m\to\infty} \sigma_m$ and recall that $M = \bigcup_m M_m$. Then we have from [6.8] and [6.9]

$$\sup_{t_{i-1}^n \le t < \sigma, t\in M} \left| X_{t_{i-1}^n} - X_t \right| \le CG(n)$$

and

$$\sup_{\sigma < t \le t_{i+1}^n, t\in M} \left| X_t - X_{t_{i+1}^n} \right| \le CG(n).$$

However, if $\sigma \in M$ so that $\sigma = t_j^{m_0}$ for some m_0, we can say more than that. Namely, two situations are possible.

1) There exists m_1 such that $\sigma_m = \sigma$ for all $m \ge m_1$. In this case, equation [6.8] yields

$$\sup_{t_{i-1}^n \le t \le \sigma, t\in M} \left| X_{t_{i-1}^n} - X_t \right| \le CG(n). \qquad [6.10]$$

2) For any $m \ge 0$, $\sigma_m < \sigma$. Then, for $m > m_0$, $\sigma_m + \delta_m \le \sigma$. Therefore, for $m > m_0$, we get from equation [6.9] that

$$\max_{\sigma \le t \le t_{i+1}^n, t\in M_m} \left| X_t - X_{t_{i+1}^n} \right| \le CG(n).$$

By letting $m \to \infty$, we get

$$\sup_{\sigma \le t \le t_{i+1}^n, t\in M} \left| X_t - X_{t_{i+1}^n} \right| \le CG(n). \qquad [6.11]$$

Now let $t_1, t_2, t_3 \in M$, $t_{i-1}^n \le t_1 < t_2 < t_3 \le t_{i+1}^n$. If $t_2 > \sigma$, then from [6.11] $\left| X_{t_2} - X_{t_3} \right| \le 2CG(n)$, and for $t_2 < \sigma$ $\left| X_{t_1} - X_{t_2} \right| \le 2CG(n)$. Finally, for $t_2 = \sigma$, which means that $\sigma \in M$, as explained above, either [6.10] or [6.11] holds. Hence, we get either $\left| X_{t_1} - X_{t_2} \right| \le 2CG(n)$ or $\left| X_{t_2} - X_{t_3} \right| \le 2CG(n)$ respectively.

Since $\varepsilon \le \delta_n$, for any $t_1, t_2, t_3 \in M$ satisfying $t_1 < t_2 < t_3 \le t_1 + \varepsilon$, there exists $i \in \{1, \ldots, 2^n - 1\}$ such that $t_{i-1}^n \le t_1 < t_2 < t_3 \le t_{i+1}^n$. From the above paragraph, we have

$$\min \left(|X_{t_2} - X_{t_1}|, |X_{t_3} - X_{t_2}| \right) \le 2CG(n)$$

for any $\omega \in D_n$. Since X is a separable process, there then exists $\Phi \in \mathcal{F}$ such that $\mathrm{P}\{\Phi\} = 0$ and for any $\omega \in D_n \backslash \Phi$

$$\Delta_\varepsilon^1 := \sup_{0 \le t_1 < t_2 < t_3 \le t_1 + \varepsilon} \min \left(|X_{t_2} - X_{t_1}|, |X_{t_3} - X_{t_2}| \right) \le 2CG(n).$$

Therefore,

$$\mathrm{P}\left\{\Delta_\varepsilon^1 > 2CG(n)\right\} \le \mathrm{P}\{D_n^c\} \le Q(n, C),$$

as required. □

THEOREM 6.4.– *Suppose that conditions of theorem 6.3 hold. Then with probability 1 the process X is D-regular, i.e. it has no discontinuities of the second kind, and moreover, for each $t \in (0, T)$, it is left- or right-continuous at t, i.e. $X(t-) = X(t)$ or $X(t-) = X(t)$, and $X(0+) = X(0)$, $X(T-) = X(T)$.*

PROOF.– Using the notation of the proof of theorem 6.3, we will show that X is D-regular for $\omega \in D := \bigcup_{n=1}^{\infty} D_n$, where D_n were introduced in [6.3]. The assertion will then follow since $\mathrm{P}\{D^c\} = \lim_{n\to\infty} \mathrm{P}\{D_n^c\} \le \limsup_{n\to\infty} Q(n, C) = 0$.

It follows from the proof of theorem 6.3 (ii) that if $\varepsilon \in (\delta_{n+1}, \delta_n]$, then $\Delta_d^1(X, [0, T], \varepsilon) \le 2CG(n)$ on D_n. Since $G(n) \to 0$ as $n \to \infty$, we get that $\Delta_d^1(X, [0, T], \varepsilon) \to 0$ as $\varepsilon \to 0$ on the event D.

To apply theorem A1.9, we need to show that

$$\sup_{0\le t\le\varepsilon} |X_t - X_0| \to 0 \quad \text{and} \quad \sup_{T-\varepsilon\le t\le T} |X_t - X_T| \to 0,\ \varepsilon \to 0,$$

on the event D. We will show only the first convergence, the second one being similar.

From the proof of theorem 6.3 (ii) it follows that, for any $n \ge 1$ and $\omega \in D_n$, there exists $\sigma(n) \in [0, 2\delta_n]$ such that

$$\sup_{0\le t<\sigma(n)} |X_t - X_0| \le CG(n)$$

and

$$\sup_{\sigma(n)<t\le 2\delta_n} |X_t - X_{2\delta_n}| \le CG(n). \qquad [6.12]$$

Define $D'_n = \{\omega \in D_n : \sigma(n) > 0\}$,

$$D' = \overline{\lim_{n\to\infty}}\, D'_n = \{\omega \in D : \sigma(n) > 0 \text{ for infinitely many } n\}\,.$$

Clearly, $\sup_{0\le t\le\varepsilon} |X_t - X_0| \le CG(n)$ on D'_n for any $\varepsilon < \sigma(n)$, hence

$$\overline{\lim_{\varepsilon\to 0}} \sup_{0\le t\le\varepsilon} |X_t - X_0| \le CG(n)$$

on D'_n. Using that $G(n) \to 0$, $n \to \infty$, we get that

$$\lim_{\varepsilon \to 0} \sup_{0 \le t \le \varepsilon} |X_t - X_0| = 0 \qquad [6.13]$$

on D'. On the other hand, if $\omega \in D \setminus D'$, then there exists n_0, such that $\sigma_n = 0$ for all $n \ge n_0$. Then from [6.12]

$$\sup_{0 < t_1 < t_2 \le 2\delta_n} |X_{t_1} - X_{t_2}| \le 2CG(n)$$

for $n \ge n_0$, which obviously implies that there exists the limit $\lim_{t \to 0+} X_t$. Thanks to stochastic continuity, this limit is equal to X_0, but then [6.12] yields

$$\sup_{0 \le t \le 2\delta_n} |X_t - X_{2\delta_n}| \le CG(n),$$

whence

$$\sup_{0 \le t_1 < t_2 \le 2\delta_n} |X_{t_1} - X_{t_2}| \le 2CG(n)$$

for all $n \ge n_0$ on $D \setminus D'$. Setting $t_1 = 0$ and combining this with [6.13], we get

$$\lim_{\varepsilon \to 0} \sup_{0 \le t \le \varepsilon} |X_t - X_0| = 0$$

on D, as claimed. As a result, $\lim_{\varepsilon \to 0} \Delta_d(X, [0,T], \varepsilon) = 0$ on D, so the statement follows from theorem A1.9. □

6.4.2. *Conditions of absence of the discontinuities of the second kind formulated in terms of conditional probabilities of large increments*

In this section, consider a complete probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Recall that the completeness means the following: for any $A \in \mathcal{F}$ with $\mathrm{P}\{A\} = 0$ and any $B \subset A$, we have that $B \in \mathcal{F}$, and consequently, $\mathrm{P}\{B\} = 0$. Furthermore, consider stochastic basis with filtration $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \ge 0}, \mathrm{P}\right)$ constructed on $(\Omega, \mathcal{F}, \mathrm{P})$ and assume that this basis is complete, i.e. $\mathcal{F}_0$ contains all P-zero sets of $\mathcal{F}$. Consider a stochastic process $X = \{X_t, \mathcal{F}_t, t \ge 0\}$ adapted to the filtration mentioned above. For any $\varepsilon > 0$ and $0 \le s < t$, consider a conditional probability $\mathrm{P}\{|X_t - X_s| \ge \varepsilon | \mathcal{F}_s\}$ (we call it the conditional probability of a big increment). It is a bounded random variable; therefore, for any interval $[a,b] \subset \mathbb{R}_+$, and any $\Omega' \in \mathcal{F}$ with $\mathrm{P}\{\Omega'\} = 1$, we can consider

$$\alpha(\varepsilon, \delta, \Omega', [a,b]) = \sup_{\omega \in \Omega'} \sup_{a \le s < t \le (s+\delta) \wedge b} \mathrm{P}\{|X_t - X_s| \ge \epsilon \mid \mathcal{F}_s\}.$$

Let

$$\alpha(\varepsilon, \delta, [a, b]) = \inf_{\Omega' : \mathrm{P}\{\Omega'\}=1} \alpha(\varepsilon, \delta, \Omega', [a, b]).$$

Note that both $\alpha(\varepsilon, \delta, \Omega', [a, b])$ and $\alpha(\varepsilon, \delta, [a, b])$ are real numbers between 0 and 1. According to the definition of infimum, there exists a sequence $\{\Omega'_n, n \geq 1\}$, such that $\mathrm{P}\{\Omega'_n\} = 1$ and $\alpha(\varepsilon, \delta, \Omega'_n, [a, b]) \to \alpha(\varepsilon, \delta, [a, b])$. If we put $\Omega'_0 = \bigcup_{n=1}^{\infty} \Omega'_n$, then

$$\alpha(\varepsilon, \delta, \Omega'_0, [a, b]) \leq \lim_{n\to\infty} \alpha(\varepsilon, \delta, \Omega'_n, [a, b]),$$

therefore, $\alpha(\varepsilon, \delta, [a, b]) = \alpha(\varepsilon, \delta, \Omega'_0, [a, b])$.

Now establish the auxiliary result. Let $a \leq t_1 < t_2 < \ldots < t_n \leq b$ be any finite number of points. Denote $\pi = \{t_1, \ldots, t_n\}$ and introduce the following events: $A^k(\varepsilon, \pi) = \{\omega \in \Omega : X \text{ has at least } k \ \varepsilon\text{-oscillations on the set } \pi\}$.

LEMMA 6.3.– *The following upper bound holds:*

$$\mathrm{P}\left\{A^k(\varepsilon, \pi) \mid \mathcal{F}_a\right\} \leq \left(2\alpha\left(\frac{\varepsilon}{4}, b-a, [a, b]\right)\right)^k \ \textit{a.s.} \qquad [6.14]$$

REMARK 6.5.– Since the conditional probability $\mathrm{P}\{|X_t - X_s| \geq \epsilon \mid \mathcal{F}_s\}$ for each s and t is defined up to a set of zero probability, then the expression $\alpha(\varepsilon, \delta, \Omega', [a, b])$ is defined non-uniquely. Moreover, since the supremum in its definition is taken over an uncountable collection, the non-uniqueness is essential: the union of exceptional sets of zero probability can even be equal to whole Ω. Consequently, $\alpha(\varepsilon, \delta, [a, b])$ is also defined non-uniquely. Nevertheless, all results that follow are valid for any choice of $\alpha(\varepsilon, \delta, [a, b])$.

PROOF.– 1) Consider the case $k = 1$ and introduce the events

$$A_l = \left\{|X_{t_i} - X_a| < \frac{\varepsilon}{2}, 1 \leq i \leq l-1, |X_{t_l} - X_a| > \frac{\varepsilon}{2}\right\},$$
$$B_l = \left\{|X_{t_l} - X_b| \geq \frac{\varepsilon}{4}\right\}, C_l = A_l \cap B_l, 1 \leq l \leq n, B_0 = \left\{|X_b - X_a| \geq \frac{\varepsilon}{4}\right\}.$$

The events A_l, consequently C_l, are disjoint, and $A^1(\varepsilon, \pi) \subset (\bigcup_{l=1}^{n} C_l) \cup B_0$. Indeed, if the trajectory of X has at least one ε-oscillation on π, then one of A_l

happens, and if, at the same time, B_l does not happen, then $|X_b - X_a| \geq |X_{t_l} - X_a| - |X_b - X_{t_l}| \geq \frac{\varepsilon}{4}$, i.e. B_0 happens. Therefore,

$$\mathrm{P}\left\{A^1(\varepsilon, \pi)|\mathcal{F}_a\right\} \leq \mathrm{P}\left\{B_0|\mathcal{F}_a\right\} + \sum_{l=1}^{n} \mathrm{P}\left\{C_l|\mathcal{F}_a\right\}$$

$$= \mathrm{P}\left\{|X_b - X_a| \geq \frac{\varepsilon}{4}|\mathcal{F}_a\right\} + \sum_{l=1}^{n} \mathrm{E}\left(\mathbb{1}_{A_l}\mathrm{E}\left(\mathbb{1}_{B_l}|\mathcal{F}_{t_l}\right)|\mathcal{F}_a\right)$$

$$\leq \alpha\left(\frac{\varepsilon}{4}, b-a, [a,b]\right) + \sum_{l=1}^{n} \mathrm{E}\left(\mathbb{1}_{A_l}\alpha\left(\frac{\varepsilon}{4}, b-a, [a,b]\right)|\mathcal{F}_a\right)$$

$$\leq \alpha\left(\frac{\varepsilon}{4}, b-a, [a,b]\right)\left(1 + \sum_{l=1}^{n} \mathrm{P}\{A_l|\mathcal{F}_a\}\right) \leq 2\alpha\left(\frac{\varepsilon}{4}, b-a, [a,b]\right).$$

2) Now we can apply induction. Assume that [6.14] is checked for some $k-1$. Introduce the event

$D_l = \{$on the set $\{t_1, \ldots, t_l\}$ the trajectory of X has at least $k-1$ ε-oscillations, while on the set $\{t_1, \ldots, t_{l-1}\}$ the number of ε-oscillations is less than $k-1\}$.

The events $D_l,\ 1 \leq l \leq n$ are disjoint, $\bigcup_{l=1}^n D_l = A^{k-1}(\varepsilon, \pi) \supset A^k(\varepsilon, \pi)$.

Now, let $A^k(\varepsilon, \pi) \cap D_l$ hold. Then on the set $\{t_l, t_{l+1}, \ldots, t_n\}$, we have at least one ε-oscillation. Therefore, $A^k(\varepsilon, \pi) = \bigcup_{l=1}^n \left(D_l \cap A^k(\varepsilon, \pi)\right) \subset \bigcup_{l=1}^n (D_l \cap E_l)$, where

$E_l = \{$on the set $\{t_l, t_{l+1}, \ldots, t_n\}$ the trajectory of X has at least one ε-oscillation$\}$.

Note that $\mathrm{P}\left\{\mathrm{E}_l|\mathcal{F}_{t_l}\right\} \leq 2\alpha\left(\frac{\varepsilon}{4}, b-t_l, [t_l, b]\right) \leq 2\alpha\left(\frac{\varepsilon}{4}, b-a, [a,b]\right)$. Therefore, applying induction, we get

$$\mathrm{P}\left\{A^k(\varepsilon, \pi)|\mathcal{F}_a\right\} \leq \sum_{l=1}^{n} \mathrm{P}\left\{D_l \cap E_l|\mathcal{F}_a\right\}$$

$$= \sum_{l=1}^{n} \mathrm{E}\left(\mathbb{1}_{D_l}\mathrm{E}\left(\mathbb{1}_{\mathrm{E}_l}|\mathcal{F}_{t_l}\right)|\mathcal{F}_a\right) = \sum_{l=1}^{n} \mathrm{E}\left(\mathbb{1}_{D_l}\mathrm{P}\left\{\mathrm{E}_l|\mathcal{F}_{t_l}\right\}|\mathcal{F}_a\right)$$

$$\le 2\alpha\left(\frac{\varepsilon}{4}, b-a, [a,b]\right) \mathrm{P}\left\{\bigcup_{l=1}^{n} D_l | \mathcal{F}_a\right\}$$

$$= 2\alpha\left(\frac{\varepsilon}{4}, b-a, [a,b]\right) \mathrm{P}\left\{A^{k-1}(\varepsilon, \pi) | \mathcal{F}_a\right\}$$

$$\le \left(2\alpha\left(\frac{\varepsilon}{4}, b-a, [a,b]\right)\right)^k. \qquad \square$$

THEOREM 6.5.– *Let the stochastic process $X = \{X_t, \mathcal{F}_t, t \ge 0\}$ be separable and for some $T > 0$ and any $\varepsilon > 0$ $\lim_{\delta \to 0+} \alpha(\varepsilon, \delta, [0,T]) = 0$. Then X has no discontinuities of the second kind on the interval $[0,T]$.*

PROOF.– Denote by $M \subset [0,T]$ the separant of the process X. Since M is countable, it can be presented as $M = \bigcup_{n=1}^{\infty} M_n$, where M_n are increasing finite sets. Let $\varepsilon > 0$ be fixed. We can choose $m \in \mathbb{N}$ in such a way that $2\alpha\left(\frac{\varepsilon}{4}, \frac{T}{m}, [0,T]\right) = \alpha < 1$. Since $\alpha(\varepsilon, \delta, \cdot)$ increases when the interval increase, $2\alpha\left(\frac{\varepsilon}{4}, \frac{T}{m}, \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right) \le \alpha$ for any $1 \le k \le m$. Therefore, according to lemma 6.3,

$$\mathrm{P}\left\{A^k\left(\varepsilon, M_n \cap \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right) \middle| \mathcal{F}_{\frac{(k-1)T}{m}}\right\} \le \alpha^k. \qquad [6.15]$$

Denote

$$A^{\infty}(\varepsilon, \widetilde{\mathbb{T}}) = \left\{\text{the trajectory of } X \text{ has an infinite number of } \varepsilon\text{-oscillations on } \widetilde{\mathbb{T}}\right\}.$$

Then for a separable process

$$A^{\infty}(\varepsilon, [a,b]) = A^{\infty}(\varepsilon, M \cap [a,b]) = \bigcap_{r=1}^{\infty} A^r(\varepsilon, M \cap [a,b]),$$

and the events $A^r(\varepsilon, M \cap [a,b])$ are decreasing in r; therefore,

$$\mathrm{P}\left\{A^{\infty}\left(\varepsilon, \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right) \middle| \mathcal{F}_{\frac{(k-1)T}{m}}\right\} \qquad [6.16]$$
$$\le \lim_{r\to\infty} \mathrm{P}\left\{A^r\left(\varepsilon, M \cap \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right) \middle| \mathcal{F}_{\frac{(k-1)T}{m}}\right\}.$$

Further,

$$A^r\left(\varepsilon, M \cap \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right) = \bigcup_{n=1}^{\infty} A^r\left(\varepsilon, M_n \cap \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right),$$

and the events $A^r\left(\varepsilon, M_n \cap \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right)$ are increasing in n, hence, using [6.15],

$$\mathrm{P}\left\{A^r\left(\varepsilon, M \cap \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right)\middle| \mathcal{F}_{\frac{(k-1)T}{m}}\right\} \qquad [6.17]$$

$$= \lim_{n\to\infty} \mathrm{P}\left\{A^r\left(\varepsilon, M_n \cap \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right)\middle| \mathcal{F}_{\frac{(k-1)T}{m}}\right\} \le \alpha^r.$$

It follows from [6.16] and [6.17] that $\mathrm{P}\left\{A^\infty\left(\varepsilon, \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right)\middle| \mathcal{F}_{\frac{(k-1)T}{m}}\right\} = 0$, from which $\mathrm{P}\left\{A^\infty\left(\varepsilon, \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right)\right\} = 0$ for any $1 \le k \le m$, and

$$\mathrm{P}\left\{A^\infty\left(\varepsilon, [0,T]\right)\right\} \le \sum_{k=1}^m \mathrm{P}\left\{A^\infty\left(\varepsilon, \left[\frac{(k-1)T}{m}, \frac{kT}{m}\right]\right)\right\} = 0 \text{ for any } \varepsilon > 0.$$

The proof now follows from theorem A1.8. □

COROLLARY 6.1.– *Let $X = \{X_t, t \in [0,T]\}$ be a separable continuous in probability stochastic process with independent increments. Then it has no discontinuities of the second kind.*

PROOF.– Calculate $\alpha(\varepsilon, \delta)$ for any $\varepsilon > 0$. Owing to independent increments, it equals

$$\alpha(\varepsilon, \delta) = \sup_{\substack{\omega\in\Omega_0,\\ 0\le s<t\le(s+\delta)\wedge T}} \mathrm{P}\left\{|X_t - X_s| \ge \varepsilon \mid \mathcal{F}_s\right\}$$

$$= \sup_{0\le s<t\le(s+\delta)\wedge T} \mathrm{P}\left\{|X_t - X_s| \ge \varepsilon\right\}$$

and this value tends to 0 as $\delta \to 0+$ according to lemma 6.1. □

6.5. Conditions of continuity of trajectories of stochastic processes

6.5.1. *Kolmogorov conditions of continuity in terms of two-dimensional distributions*

Similarly to theorem 6.4 and in the same notations, we can formulate and prove the following result.

THEOREM 6.6.– *Let $X = \{X_t, t \in [0,T]\}$ be a real-valued separable stochastic process satisfying the condition: there exists a strictly positive non-decreasing function $g(h)$ and a function $q(C,h)$, $h \ge 0$, such that for any $0 \le h \le t \le T-h$ and $C > 0$,*

$$\mathrm{P}\left\{|X_{t+h} - X_t| > Cg(h)\right\} \le q(C,h),$$

and

$$G(0)=\sum_{n=0}^{\infty} g(\delta_n)<\infty,\ \ Q(C)=\sum_{n=1}^{\infty} 2^n q(C,\delta_n)<\infty.$$

Then

i) for any $\varepsilon>0$

$$\mathrm{P}\left\{\sup_{0\le s\le t\le T}|X_t-X_s|>\varepsilon\right\}\le Q\left(\frac{\varepsilon}{2G(0)}\right);$$

ii) $$\mathrm{P}\left\{\Delta_c(X,[0,T],\varepsilon)>2CG\left(\left[\log_2\frac{T}{2\varepsilon}\right]\right)\right\}\le Q\left(\left[\log_2\frac{T}{2\varepsilon}\right],C\right),$$

where

$$G(n)=\sum_{k=n}^{\infty} g(\delta_k),\ \ Q(n,C)=\sum_{k=n}^{\infty} 2^k q(C,\delta_k).$$

PROOF.– As before, denote $t_k^n=k\delta_n$, $k=0,1,\dots,2^n$ and consider the events

$$A_{n,k}=\left\{\omega\in\Omega:\left|X_{t_{k+1}^n}-X_{t_k^n}\right|\le Cg\left(\frac{T}{2^n}\right)\right\},\ \ n\ge 0,0\le k\le 2^n-1$$

and let

$$D_n=\bigcap_{m=n}^{\infty}\bigcap_{k=0}^{2^m-1}A_{m,k}.$$

Then, for any $\omega\in D_0$ and any $n\ge 0$, $0\le j<k\le 2^n$,

$$\left|X_{t_k^n}-X_{t_j^n}\right|\le 2CG(0)$$

which can be proved similarly to [6.7]. It follows from separability of the process X that, for $\omega\in D_0$, $|X_t-X_s|\le 2CG(0)$, and

$$\mathrm{P}\left\{\sup_{0\le s\le t\le T}|X_t-X_s|>2CG(0)\right\}\le \mathrm{P}\left\{D_0^c\right\},$$

whence

$$\mathrm{P}\left\{\sup_{0\le s\le t\le T}|X_t-X_s|\ge\varepsilon\right\}\le\sum_{n=0}^{\infty}\sum_{k=0}^{2^n-1} q\left(\frac{\varepsilon}{2G(0)},\delta_n\right)=Q\left(\frac{\varepsilon}{2G(0)}\right).$$

Let $\varepsilon > 0$, $n = \left[\log_2 \frac{T}{\varepsilon}\right]$, so that $\varepsilon \le \frac{T}{2^n}$, and let $\omega \in D_n$. Then $\left|X_{t^n_{k+1}} - X_{t^n_k}\right| < Cg\left(\delta_n\right)$. Assume that for some $m > n$, $|k - j| \le 2^{m-n}$

$$\left|X_{t^m_k} - X_{t^m_j}\right| \le 2C \sum_{k=n}^{m} g\left(\delta_k\right).$$

Then for $m + 1$ and $|k - j| \le 2^{m+1-n}$ such that $k > j$, $k = 2k_1 + 1$, $j = 2j_1 + 1$ we have that $|k_1 - 1 - j_1| \le 2^{m-n}$, therefore

$$\left|X_{t^{m+1}_k} - X_{t^{m+1}_j}\right| \le \left|X_{t^{m+1}_k} - X_{t^{m+1}_{k_1+1}}\right| + \left|X_{t^{m+1}_{k_1-1}} - X_{t^m_j}\right|$$
$$+ \left|X_{t^{m+1}_{2j_1}} - X_{t^{m+1}_{2j_1+1}}\right| \le 2g\left(\delta_m\right) + 2C \sum_{k=n}^{m} g\left(\delta_k\right) \le 2CG(n).$$

Other points t^{m+1}_k and t^{m+1}_j are considered similarly. Finally, we get that, for any t^m_k, t^m_j, $m \ge n$, such that $|k - j|\delta_m \le \delta_n$

$$\left|X_{t^m_k} - X_{t^m_j}\right| \le 2C \sum_{k=n}^{\infty} g\left(\delta_k\right) = 2CG(n).$$

Since X is separable, we get that, for $0 < \varepsilon \le \delta_n$ and $|s - t| \le \varepsilon$,

$$|X_t - X_s| \le 2CG(n) \text{ on } D_n. \qquad [6.18]$$

It means that

$$\mathrm{P}\left\{\sup_{\substack{|t-s|\le\varepsilon \\ 0\le s<t\le T}} |X_t - X_s| \ge 2CG(n)\right\} \le \mathrm{P}\left\{D^c_n\right\} \le Q(n, C),$$

as required. □

The next result is an obvious corollary of theorem 6.6. It can be proved similarly to theorem 6.4 but much simpler.

THEOREM 6.7.– *Under the conditions of theorem 6.6, the process X is continuous on $[0, T]$.*

THEOREM 6.8.– (*Kolmogorov–Chentsov*) *Let $X = \{X_t, t \in [0, T]\}$ be a separable stochastic process satisfying the assumption: there exist constants $K > 0$, $\alpha > 0$ and $\beta > 0$, such that*

$$\mathrm{E}|X_t - X_s|^\alpha \le K|t - s|^{1+\beta}, \; 0 \le s < t \le T. \qquad [6.19]$$

Then X is a continuous process.

PROOF.– Let $g(h) = h^{\beta_1/\alpha}$, where $0 < \beta_1 < \beta$. Then

$$G(m) = \sum_{k=m}^{\infty} (\delta_k)^{\beta_1/\alpha} = \frac{T^{\beta_1/\alpha} \left(\frac{1}{2}\right)^{\beta_1 m/\alpha}}{1 - \left(\frac{1}{2}\right)^{\beta_1/\alpha}} = C_1 (\delta_m)^{\frac{\beta_1}{\alpha}},$$

$$\begin{aligned} &\mathrm{P}\left\{|X_t - X_{t+h}| \geq Cg(h)\right\} \leq C^{-\alpha} \left[g(h)\right]^{-\alpha} \mathrm{E}\left|X_{t+h} - X_t\right|^{\alpha} \\ &\leq C^{-\alpha} K h^{-\beta_1} h^{1+\beta} = C^{-\alpha} K h^{1+\beta-\beta_1}. \end{aligned}$$

Therefore, we can put $q(C, h) = KC^{-\alpha} h^{1+\beta-\beta_1}$. In this case,

$$\sum_{n=0}^{\infty} g(\delta_n) = \sum_{n=0}^{\infty} \left(\frac{T}{2^n}\right)^{\beta_1/\alpha} < \infty,$$

and

$$\begin{aligned} \sum_{n=0}^{\infty} 2^n g(C, \delta_n) &= \sum_{n=0}^{\infty} KC^{-\alpha} 2^n \left(\delta_n\right)^{1+\beta-\beta_1} \\ &= KC^{-\alpha} T^{1+\beta-\beta_1} \sum_{n=0}^{\infty} 2^{-n(\beta-\beta_1)} < \infty, \end{aligned}$$

and the proof follows from theorem 6.7. □

REMARK 6.6.– Theorems 6.7 and 6.8 can be reformulated in such a way that a stochastic process, satisfying their assumptions except the assumption of separability, has a continuous modification. From now on assume that we consider a separable modification of any process X.

REMARK 6.7.– Condition $\mathrm{E}|X_t - X_s|^{\beta} \leq C|t - s|$ does not supply the continuity of X. Indeed, if X is a homogeneous Poisson process with parameter λ, then, for any $k \in \mathbb{N}$,

$$\begin{aligned} \mathrm{E}|X_t - X_s|^k &= \sum_{l=0}^{\infty} l^k e^{-\lambda|t-s|} \frac{(\lambda|t-s|)^l}{l!} \leq \lambda|t-s| \sum_{l=1}^{\infty} l^{k-1} \frac{(\lambda|t-s|)^{l-1}}{(l-1)!} \\ &= \lambda|t-s| \sum_{l=0}^{\infty} (l+1)^{k-1} \frac{(\lambda|t-s|)^l}{l!} \leq C|t-s|, \end{aligned}$$

however, almost all trajectories of the Poisson process have jumps on $\mathbb{R}^+$.

REMARK 6.8.– Let $X = \{X_t, t \in [0,T]\}$ be a Gaussian process with zero mean and covariance function $R(s,t)$. Then

$$\mathrm{E}(X_t - X_s)^2 = R(t,t) - 2R(s,t) + R(s,s).$$

According to the formula for higher moments of Gaussian distribution, if $\xi \sim \mathcal{N}(0,\sigma^2)$, then $\mathrm{E}\xi^{2k} = C_k\sigma^k$, where $k \in \mathbb{N}$, C_k are some constants depending only on k. Therefore,

$$\mathrm{E}(X_t - X_s)^2 = C_k\left(R(t,t) - 2R(s,t) + R(s,s)\right)^k, \; k \in \mathbb{N}.$$

Assume that there exist $C > 0$ and $\gamma > 0$ such that, for any $s,t \in [0,T]$,

$$R(t,t) - 2R(s,t) + R(s,s) \le C|t-s|^\gamma. \qquad [6.20]$$

Then, for $k > j^{-1}$, we have that

$$\mathrm{E}(X_t - X_s)^{2k} \le C_k C^k |t-s|^{\gamma k}.$$

It means that X is a continuous process.

In particular, let X be a Gaussian process with zero mean and $\mathrm{E}(X_t - X_s)^2 = R(|t-s|)$, where $R : \mathbb{R}_+ \to \mathbb{R}_+$ is some function (such processes are called *processes with stationary increments*). If $R(x) \le Cx^\gamma$, $x \in [0,T]$, then X is a continuous process.

Consider some examples.

EXAMPLE 6.3.– *Let $X = W$ be a Wiener process. Then* $\mathrm{E}|W_t - W_s|^2 = |t-s|$, *so W is a process with stationary increments with $R(x) = x$ and W is a continuous process on any* $[0,T]$.

EXAMPLE 6.4.– *Let $X = B^H$ be a fractional Brownian motion with Hurst index* $H \in (0,1)$. *Then*

$$\mathrm{E}|B_t^H - B_s^H|^2 = t^{2H} - 2\cdot\frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right) + s^{2H} = |t-s|^{2H}.$$

Therefore, B^H is a process with stationary increments, $R(x) = x^{2H}$, and B^H is a continuous process.

6.5.2. *Hölder continuity of stochastic processes: a sufficient condition*

Recall that function $f : [0,T] \to \mathbb{R}$ is said to be *Hölder continuous* of order $0 < \alpha \le 1$ if there exists $C > 0$, such that, for any $s,t \in [0,T]$,

$$|f(t) - f(s)| \le C|t-s|^\alpha.$$

Obviously, if f is Hölder continuous of some order $\alpha \in (0,1)$ on $[0,T]$, then it is Hölder continuous of any order $\beta \in (0,\alpha)$ on this interval.

DEFINITION 6.7.– *Function* $f : [0,T] \to \mathbb{R}$ *is* Hölder continuous up to order $\alpha \in (0,1)$ *on* $[0,T]$ *if it is Hölder continuous of any order* $\beta \in (0,\alpha)$.

THEOREM 6.9.– *Let stochastic process* $X = \{X_t, t \in [0,T]\}$ *be separable and satisfy condition* [6.19]. *Then a.s. its trajectories are Hölder continuous up to order* β/α.

PROOF.– Let us attentively check the proof of theorem 6.8. It follows from [6.18] that for $0 < \varepsilon \le \frac{T}{2^n}$, and $|s-t| \le \varepsilon$

$$|X_t - X_s| \le 2CG(n) = 2C\sum_{k=n}^{\infty}(\delta_k)^{\beta_1/\alpha} \le 2C\frac{(\delta_n)^{\beta_1/\alpha}}{1-(\delta_n)^{\beta_1/\alpha}}$$

$$\le 4C\,(\delta_n)^{\beta_1/\alpha} \ \text{ for } n > \log_2\left(T\cdot 2^{\frac{\alpha}{\beta_1}}\right), \ \text{ and for } \omega \in D_n.$$

Furthermore,

$$\sum_{n=1}^{\infty} \mathrm{P}\{D_n^c\} \le \sum_{n=1}^{\infty} 2^n Q(n,C) < \infty.$$

Therefore, it follows from the Borel–Cantelli lemma that, for any $\omega \in \Omega'$ with $\mathrm{P}\{\Omega'\} = 1$, there exists $n_0 = n_0(\omega) > \log_2(T\cdot 2^{\beta_1/\alpha})$, such that, for $n \ge n_0(\omega)$,

$$|X_t - X_s| \le 4C\,(\delta_n)^{\beta_1/\alpha}, \ \text{ for } |t-s| \le \delta_n.$$

Then, for any $t, s \in [0,T]$, such that $|t-s| \ge \delta_{n_0(\omega)}$, the distance $|t-s|$ does not exceed $T = 2^{n_0(\omega)}\cdot\delta_{n_0(\omega)}$, whence

$$|X_t - X_s| \le 2^{n_0(\omega)\cdot\left(\delta_{n_0(\omega)}\right)^{\beta_1/\alpha}},$$

and for $|t-s| \in [\delta_{n+1}, \delta_n]$ for $n \ge n_0(\omega)$

$$|X_t - X_s| \le 4C\,(\delta_n)^{\beta_1/\alpha} \le 4C\,(\delta_{n+1})^{\beta_1/\alpha}\cdot 2^{\beta_1/\alpha} \le C_2|t-s|^{\beta_1/\alpha},$$

and the proof follows. □

In particular, let X be a Gaussian process with zero mean, stationary increments, and let $\mathrm{E}(X_t - X_s)^2 = R(|t-s|)$. If $R(x) \le Cx^\gamma$, $x \in [0,T]$, then, for any $p \in \mathbb{N}$, $\mathrm{E}|X_t - X_s|^{2p} \le C_p|t-s|^{\gamma p}$; therefore, X is Hölder continuous up to order $\frac{\gamma}{2} - \frac{1}{2p}$ for any $p \ge 1$, i.e. it is Hölder continuous up to order $\frac{\gamma}{2}$. Consider some examples.

EXAMPLE 6.5.– *Let $X = W$ be a Wiener process. Then* $\mathrm{E}|W_t - W_s|^2 = |t-s|$, *so* $\gamma = 1$ *and W is Hölder continuous up to order* $\frac{1}{2}$. *It means that there exists* $\Omega' \subset \Omega$ *such that* $\mathrm{P}\{\Omega'\} = 1$ *and for any* $\omega \in \Omega'$, *any* $\delta > 0$ *and any* $T > 0$, *there exists a constant* $C = C(T, \omega, \delta)$, *such that, for any* $s, t \in [0, T]$, $|W_t - W_s| \le C(T, \omega, \delta)|t-s|^{\frac{1}{2}-\delta}$.

EXAMPLE 6.6.– *Let* $X = B^H$ *be a fractional Brownian motion with Hurst index* $H \in (0, 1)$. *Then*

$$\mathrm{E}|B_t^H - B_s^H|^2 = |t-s|^{2H}.$$

Therefore, $\gamma = 2H$ *and* B^H *are Hölder continuous up to order* H. *It means that there exists* $\Omega' \subset \Omega$, *such that* $\mathrm{P}\{\Omega'\} = 1$ *and for any* $\omega \in \Omega'$, *any* $\delta > 0$ *and any* $T > 0$, *there exists a constant* $C = C(T, \omega, \delta)$, *such that, for any* $s, t \in [0, T]$,

$$|B_t^H - B_s^H| \le C(T, \omega, \delta)|t-s|^{H-\delta}.$$

Of course, these reasons do not deny that W *and* B^H *have smoother trajectories; however, it can be established that the statements above concerning their Hölder properties are sharp.*

6.5.3. *Conditions of continuity in terms of conditional probabilities*

THEOREM 6.10.– *Consider the interval* $[0, T]$ *and sequence of partitions*

$$\pi_n = \left\{0 = t_0^n < \ldots < t_{k_n}^n = T\right\}.$$

Denote $|\pi_n| = \max_{1 \le k \le k_n}(t_k^n - t_{k-1}^n)$. *Let* $X = \{X_t, t \in [0, T]\}$ *be a separable stochastic process without discontinuities of second kind, such that, for any sequence* $\{\pi_n, n \ge 1\}$ *of partitions with* $|\pi_n| \to 0$ *as* $n \to \infty$ *and any* $\varepsilon > 0$,

$$\sum_{k=1}^{k_n} \mathrm{P}\left\{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \ge \varepsilon\right\} \to 0 \text{ as } n \to \infty. \qquad [6.21]$$

Then the process X *is continuous, i.e. it has a.s. continuous trajectories on* $[0, T]$.

PROOF.– For any $\varepsilon > 0$, denote by ν_ε the number of points $t \in (0, T)$ for which $X_{t+} \ne X_{t-}$ and let ν_ε^n be the number of points t_n^k, for which $\left|X_{t_k^n} - X_{t_{k-1}^n}\right| > \frac{\varepsilon}{2}$. Then $\nu_\varepsilon \le \lim_{n\to\infty} \inf \nu_\varepsilon^n$, while

$$\mathrm{E}\nu_\varepsilon^n = \mathrm{E}\sum_{k=1}^{k_n} \mathbb{1}_{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \ge \frac{\varepsilon}{2}} = \sum_{k=1}^{k_n} \mathrm{P}\left\{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \ge \frac{\varepsilon}{2}\right\}.$$

It follows from the Fatou lemma and the theorem assumptions that

$$\mathrm{E}\nu_\varepsilon \le \mathrm{E}\liminf_{n\to\infty} \nu_\varepsilon^n \le \liminf_{n\to\infty} \mathrm{E}\nu_\varepsilon^n = 0.$$

Therefore, $\nu_\varepsilon = 0$ with probability 1 for any $\varepsilon > 0$. It means that $X_{t+} = X_{t-}$, $t \in (0,T)$ with probability 1. Then it follows from separability of X that $X_t = X_{t+} = X_{t-}$, $t \in (0,T)$ with probability 1. Moreover, it follows from separability of X on $[0,T]$ that $X_0 = X_{0+}$ and $X_T = X_{T-}$ with probability 1. It means that X is continuous on $[0,T]$. □

THEOREM 6.11.– *Let a process* $X = \{X_t, t \in [0,T]\}$ *be separable and, for any* $\varepsilon > 0$, $\lim_{\delta\to 0+} \frac{\alpha(\varepsilon,\delta,[0,T])}{\delta} = 0$. *Then the process* X *is continuous on* $[0,T]$.

PROOF.– Under the theorem's condition $\lim_{\delta\to 0+}\alpha(\varepsilon,\delta,[0,T]) = 0$. Therefore, X has no discontinuities of the second kind according to theorem 6.5. Consequently, it is sufficient to check condition [6.21]. Consider a sequence of partitions π_n with $|\pi_n| \to 0$ as $n \to \infty$. Note that

$$\begin{aligned}
&\mathrm{P}\left\{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \geq \varepsilon\right\} \leq \mathrm{E}\left(\mathbb{1}_{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \geq \varepsilon}\right) \\
&= \mathrm{E}\left(\mathrm{E}\left(\mathbb{1}_{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \geq \varepsilon} | \mathcal{F}_{t_{k-1}^n}\right)\right) = \mathrm{E}\left(\mathrm{P}\left\{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \geq \varepsilon | \mathcal{F}_{t_{k-1}^n}\right\}\right) \\
&\leq \alpha(\varepsilon, t_k^n - t_{k-1}^n, [t_k^n, t_{k-1}^n]) \leq \alpha(\varepsilon, t_k^n - t_{k-1}^n, [0,T]),
\end{aligned}$$

we conclude that

$$\begin{aligned}
&\sum_{k=1}^{k_n} \mathrm{P}\left\{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \geq \varepsilon\right\} \leq \sum_{k=1}^{k_n} (t_k^n - t_{k-1}^n) \frac{\alpha(\varepsilon, t_k^n - t_{k-1}^n, [0,T])}{t_k^n - t_{k-1}^n} \\
&\leq \max_{1\leq k\leq n} \frac{\alpha(\varepsilon, t_k^n - t_{k-1}^n, [0,T])}{t_k^n - t_{k-1}^n} \cdot T \leq \sup_{0<\gamma\leq|\pi_n|} \frac{\alpha(\varepsilon,\gamma,[0,T])}{\gamma} \cdot T.
\end{aligned} \qquad [6.22]$$

Therefore, it follows from the definition of $\limsup$ and [6.22] that

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \mathrm{P}\left\{\left|X_{t_k^n} - X_{t_{k-1}^n}\right| \geq \varepsilon\right\} \leq \limsup_{\delta\to 0+} \frac{\alpha(\varepsilon,\delta,[0,T])}{\delta} \cdot T = 0.$$ □

7

Markov and Diffusion Processes

7.1. Markov property

Consider a complete probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and a parameter set $\mathbb{T} \subset \mathbb{R}$, playing the role of time.

DEFINITION 7.1.– *A stochastic process* $\{X_t, t \in \mathbb{T}\}$ *in* $(\mathcal{S}, \Sigma)$ *is called a* Markov process *if for any* $s, t \in \mathbb{T}$*,* $s < t$ *and any* $A \in \Sigma$

$$\mathrm{P}\{X_t \in A \mid \mathcal{F}_s\} = \mathrm{P}\{X_t \in A \mid X_s\} \qquad [7.1]$$

almost surely, where $\mathcal{F}_t = \mathcal{F}_t^X = \sigma\{X_s, s \le t, s \in \mathbb{T}\}$ *is the natural filtration of* X*.*

Equation [7.1], called the *Markov property*, is the absence of memory: it means that the probability distribution of future values of the process depends only on its current value but not on the path which led to this value. In other words, conditionally on the current state of the process, its future values are independent of its past values.

PROPOSITION 7.1.– *The stochastic process* X *is Markov if and only if for any* $n \ge 1$*, any* $s_1, s_2, \ldots, s_n, t \in \mathbb{T}$ *with* $s_1 < s_2 < \cdots < s_n < t$ *and any* $A \in \Sigma$

$$\mathrm{P}\{X_t \in A \mid X_{s_1}, X_{s_2}, \ldots, X_{s_n}\} = \mathrm{P}\{X_t \in A \mid X_{s_n}\} \qquad [7.2]$$

almost surely.

PROOF.– Let X be a Markov process. Then, for any $n \ge 1$, any $s_1, s_2, \ldots, s_n, t \in \mathbb{T}$ with $s_1 < s_2 < \cdots < s_n < t$ and any $A \in \Sigma$, it follows from the properties of conditional expectation that

$$\mathrm{P}\{X_t \in A \mid X_{s_1}, X_{s_2}, \ldots, X_{s_n}\} = \mathrm{E}\big(\mathrm{P}\{X_t \in A \mid \mathcal{F}_{s_n}\} \mid X_{s_1}, X_{s_2}, \ldots, X_{s_n}\big)$$
$$= \mathrm{E}\big(\mathrm{P}\{X_t \in A \mid X_{s_n}\} \mid X_{s_1}, X_{s_2}, \ldots, X_{s_n}\big) = \mathrm{P}\{X_t \in A \mid X_{s_n}\}.$$

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

Vice versa, let [7.2] hold. Fix some $s, t \in \mathbb{T}$ with $s < t$. When $\mathrm{P}\{X_t \in A\} = 0$, equation [7.1] is obvious, since both its sides are zero almost surely. So assume that $A \in \Sigma$ is such that $K := \mathrm{P}\{X_t \in A\} > 0$. Denote

$$\eta = \mathrm{P}\{X_t \in A \mid X_s\}$$

and consider set functions

$$\mathrm{Q}_1(B) = \frac{1}{K}\mathrm{E}(\mathbb{1}_{X_t \in A}\mathbb{1}_B),\ \mathrm{Q}_2(B) = \frac{1}{K}\mathrm{E}(\eta\mathbb{1}_B), \quad B \in \mathcal{F}_s.$$

It is obvious that Q_1 is a probability measure and it follows from the tower property of conditional expectation that $\mathrm{E}\eta = \mathrm{P}\{X_t \in A\} = K$, so Q_2 is a probability measure too. Now for B of the form

$$B = \{(X_{s_1}, \ldots, X_{s_n}) \in D\},\ n \geq 1, s_1 < \cdots < s_n \leq s, D \in \Sigma^{(n)}, \qquad [7.3]$$

we have, thanks to equation [7.2], that $\mathrm{E}(\mathbb{1}_{X_t \in A}\mathbb{1}_B) = \mathrm{E}(\eta\mathbb{1}_B)$, whence $\mathrm{Q}_1(B) = \mathrm{Q}_2(B)$. The sets of the form [7.3] form a π-system and generate $\mathcal{F}_s$, so we get from theorem A2.2 that $\mathrm{Q}_1(B) = \mathrm{Q}_2(B)$; hence, $\mathrm{E}(\mathbb{1}_{X_t \in A}\mathbb{1}_B) = \mathrm{E}(\eta\mathbb{1}_B)$ for all $B \in \mathcal{F}_s$. Since η is $\mathcal{F}_s$-measurable, the latter equality means that

$$\eta = \mathrm{E}(\mathbb{1}_{X_t \in A} \mid \mathcal{F}_s) = \mathrm{P}\{X_t \in A \mid \mathcal{F}_s\},$$

as required. □

Assume further that the state space $(\mathcal{S}, \Sigma)$ is a Polish space, i.e. a separable complete metric space with Borel σ-algebra. By theorem A2.6 and remark A2.2, we can write

$$\mathrm{P}\{X_t \in A \mid X_s\} = P(s, X_s, t, A), \qquad [7.4]$$

where the function $P(s, x, t, A)$ is measurable in x for any fixed s, t, A. Moreover, by theorem A2.12, a regular conditional distribution exists, so we can assume without any loss of generality that $P(s, x, t, A)$ is a measure as a function of A for any fixed s, t, x. Further, using theorem A2.3, we find that for any bounded measurable function $g\colon \mathcal{S} \to \mathbb{R}$,

$$\mathrm{E}(g(X_t) \mid \mathcal{F}_s) = \int_{\mathcal{S}} g(y)P(s, X_s, t, dy) \qquad [7.5]$$

almost surely. Therefore, using the tower property of conditional expectation, we can write for any $s, u, t \in \mathbb{T}$ with $s < u < t$ and any $A \in \Sigma$

$$P(s, X_s, t, A) = \mathrm{P}\left\{X(t) \in A \mid \mathcal{F}_s\right\} = \mathrm{E}\left(\mathrm{P}\left\{X_t \in A \mid \mathcal{F}_u\right\} \mid \mathcal{F}_s\right)$$
$$= \mathrm{E}\left(P(u, X_u, t, A) \mid \mathcal{F}_s\right) = \int_{\mathcal{S}} P(u, y, t, A) P(s, X_s, u, dy)$$

almost surely. This is equivalent to saying that

$$P(s, x, t, A) = \int_{\mathcal{S}} P(u, y, t, A) P(s, x, u, dy) \qquad [7.6]$$

for all $x \in \mathcal{S}$ except some set N such that $\mathrm{P}\{X_s \in N\} = 0$. This equation, called the *Chapman–Kolmogorov equation*, motivates the following definition, where we denote $\mathbb{T}^{2<} = \big\{(s, t) \in \mathbb{T}^2 : s < t\big\}$.

DEFINITION 7.2.– *A* transition probability function *is a function* $P\colon \mathbb{T}^{2<} \times \mathcal{S} \times \Sigma \to \mathbb{R}$ *such that:*

1) $P(s, x, t, A)$ *is measurable in* x*;*

2) as a function of A*,* $P(s, x, t, A)$ *is a probability measure;*

3) equation [7.6] *holds for all* $(s, t) \in \mathbb{T}^{2<}$*,* $x \in \mathcal{S}$*,* $A \in \Sigma$.

A function $Q\colon \mathcal{S} \times \Sigma$, such that $Q(x, A)$ is measurable in x and a probability measure in A, is called a *transition kernel*, *stochastic kernel* or *Markov kernel*. Alternative names for transition probability function are *transition probability* and *Markov transition probability* (function).

It is natural and convenient to define the “zero-time” transition probabilities as

$$P(t, x, t, A) = \delta_x(A) := \mathbb{1}_A(x).$$

With this extension, equations [7.4] and [7.6] obviously hold when some of the parameters s, u, t coincide.

DEFINITION 7.3.– *A Markov process* X *is said to have the transition probability function* P*, or the transition probability function* P *is said to correspond to* X*, if equation* [7.4] *is satisfied for any* $(s, t) \in \mathbb{T}^{2<}$ *and* $A \in \Sigma$ *almost surely.*

If a Markov process X has transition probability function P, then, appealing as before to theorem A2.3, [7.5] holds for any bounded measurable function g. Vice versa, substituting an indicator function into [7.5] leads to [7.4], so [7.5] can be regarded as an alternative definition of the fact that P corresponds to X.

The discussion preceding Definition 7.2 shows that to each Markov process corresponds a function, which satisfies [7.6] for *almost* all $x \in \mathcal{S}$ with respect to the distribution of X_s, so it is in this sense *almost* a transition probability function. It turns out that for a Markov process taking values in a complete separable metric space, there exists a genuine transition probability function, i.e. one that satisfies [7.6] for any $(s,t) \in \mathbb{T}^{2<}, A \in \Sigma$ and all $x \in \mathcal{S}$; this was proved by Kuznetsov [KUZ 84]. The converse is not always the case: not all transition probability functions have Markov processes corresponding to them.

EXAMPLE 7.1.– *Let* $\mathbb{T} = \mathbb{Z}$, $\mathcal{S} = \mathbb{N}$, $\Sigma = 2^{\mathbb{N}}$. *Define for integer* $s \le t$, $x \in \mathbb{N}$, $A \subset \mathbb{N}$

$$P(s,x,t,A) = \mathbb{1}_A(x+t-s).$$

This function is easily seen to be a transition probability. Assume that a corresponding Markov process $\{X_t, t \in \mathbb{Z}\}$ *exists. Note that*

$$\mathrm{P}\{X_{s+1} \in A \mid X_s = x\} = P(s,x,s+1,A) = \mathbb{1}_A(x+1),$$

so the underlying evolution is deterministic: the process simply increases its value by 1 *on each step. In particular, thanks to the law of total probability, for every* $n \ge 2$,

$$\mathrm{P}\{X_{t+1} = n\} = \sum_{k=1}^{\infty} \mathrm{P}\{X_{t+1} = n \mid X_t = k\}\mathrm{P}\{X_t = k\} \qquad [7.7]$$

$$= \mathrm{P}\{X_{t+1} = n \mid X_t = n-1\}\mathrm{P}\{X_t = n-1\} = \mathrm{P}\{X_t = n-1\},$$

and $\mathrm{P}\{X_{t+1} = 1\} = 0$. *Obviously,* $\mathrm{P}\{X_0 = m\} > 0$ *for some* $m \in \mathbb{N}$. *From* [7.8], *we get*

$$\mathrm{P}\{X_0 = m) = \mathrm{P}\{X_{-1} = m-1\} = \mathrm{P}\{X_{-2} = m-2\} = \cdots = \mathrm{P}\{X_{1-m} = 1\} = 0,$$

a contradiction.

Nevertheless, if there is a “starting point”, then the corresponding process exists.

THEOREM 7.1.– *Assume that* $\mathbb{T}$ *has a minimal element* t_0 *and* P *is a transition probability function. Then, for any probability distribution* μ *on* (S, Σ), *there exists a Markov process* X *such that* P *is its transition probability function and* X_{t_0} *has distribution* μ. *Moreover, finite-dimensional distributions of the Markov process* X *are uniquely determined by* μ *and* P.

PROOF.– Let us start with the second statement, at the same time determining the finite-dimensional distributions of X. For arbitrary integer $n \ge 2$, let $A_1, \ldots, A_n \in \Sigma$

and $t_1, \dots, t_n \in \mathbb{T}$ be such that $t_1 < \dots < t_n$. Denote $\mathbf{I}_k = \mathbb{1}_{X_{t_k} \in A_k}$, $k = 1, \dots, n$, and write, using the Markov property and [7.5],

$$\begin{aligned}
&\mathrm{P}\left\{X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\right\} = \mathrm{E}\left(\mathrm{E}\left(\mathbf{I}_1 \cdots \mathbf{I}_n \mid \mathcal{F}_{t_{n-1}}\right)\right) \\
&= \mathrm{E}\left(\mathbf{I}_1 \cdots \mathbf{I}_{n-1}\, \mathrm{E}\left(\mathbf{I}_n \mid \mathcal{F}_{t_{n-1}}\right)\right) = \mathrm{E}\left(\mathbf{I}_1 \cdots \mathbf{I}_{n-1}\, \mathrm{E}\left(\mathbf{I}_n \mid X_{t_{n-1}}\right)\right) \\
&= \mathrm{E}\left(\mathbf{I}_1 \cdots \mathbf{I}_{n-1} P\big(t_{n-1}, X_{t_{n-1}}, t_n, A_n\big)\right) \\
&= \mathrm{E}\left(\mathbf{I}_1 \cdots \mathbf{I}_{n-2}\, \mathrm{E}\left(\mathbf{I}_{n-1} P\big(t_{n-1}, X_{t_{n-1}}, t_n, A_n\big) \mid \mathcal{F}_{t_{n-2}}\right)\right) \\
&= \mathrm{E}\left(\mathbf{I}_1 \cdots \mathbf{I}_{n-2}\, \mathrm{E}\left(\mathbf{I}_{n-1} P\big(t_{n-1}, X_{t_{n-1}}, t_n, A_n\big) \mid X(t_{n-2})\right)\right) \\
&= \mathrm{E}\bigg(\mathbf{I}_1 \cdots \mathbf{I}_{n-2} \int_{A_{n-1}} P\big(t_{n-1}, y_{n-1}, t_n, A_n\big) P\big(t_{n-2}, X_{t_{n-2}}, t_{n-1}, dy_{n-1}\big)\bigg).
\end{aligned}$$

Repeating this chain of reasoning and noting for better appearance that

$$P\big(t_{n-1}, y_{n-1}, t_n, A_n\big) = \int_{A_n} P\big(t_{n-1}, y_{n-1}, t_n, dy_n\big),$$

we get

$$\begin{aligned}
&\mathrm{P}\Big(X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\Big) \\
&= \mathrm{E}\bigg(\int_{A_1}\int_{A_2} \cdots \int_{A_n} \bigg(\prod_{k=2}^{n} P\big(t_{k-1}, y_{k-1}, t_k, dy_k\big)\bigg) P\big(t_0, X_{t_0}, t_1, dy_1\big)\bigg) \\
&= \int_{\mathcal{S}}\int_{A_1}\int_{A_2} \cdots \int_{A_n} \bigg(\prod_{k=1}^{n} P\big(t_{k-1}, y_{k-1}, t_k, dy_k\big)\bigg) \mu(dy_0), \qquad [7.8]
\end{aligned}$$

which implies the uniqueness.

On the other hand, given a transition probability function P and an initial distribution μ, we can define finite-dimensional distributions through [7.8]. This is easily seen to be a consistent family, so thanks to theorem 1.2, there exists a stochastic process $\{X_t, t \in \mathbb{T}\}$ such that its finite-dimensional distributions are given by [7.8]. Moreover, it follows immediately from [7.8] that the distribution of X_{t_0} is μ. Therefore, it remains to check that X is a Markov process with transition probability function P.

To this end, first observe from [7.8] that for any $s \in \mathbb{T}$, $B \in \Sigma$,

$$\mathrm{P}\big\{X_s \in B\big\} = \int_{\mathcal{S}} P(t_0, y_0, s, B)\mu(dy_0).$$

By using theorem A2.3, we get

$$\mathrm{E}f(X_s) = \int_{\mathcal{S}} \int_{\mathcal{S}} f(y) P(t_0, y_0, s, dy) \mu(dy_0)$$

for any bounded measurable function f. Therefore, using [7.8] again, for any $t \in \mathbb{T}$, $t > \mathbb{T}$ and $A, B \in \Sigma$, we have

$$\begin{aligned} \mathrm{P}\big\{X_s \in B, X_t \in A\big\} &= \int_{\mathcal{S}} \int_B P(s, y_1, t, A) P(t_0, y_0, s, dy_1) \mu(dy_0) \\ &= \mathrm{E}\left(\mathbb{1}_{X_s \in B} P(s, X_s, t, A)\right). \end{aligned}$$

whence [7.4] follows. Similarly, from [7.8], for any $n \geq 2$, $t_1 < \cdots < t_n$ and $A_1, \ldots, A_n \in \Sigma$,

$$\begin{aligned} &\mathrm{P}\big\{X_{t_1} \in A_1, \ldots, X_{t_n} \in A_n\big\} \\ &= \int_{\mathcal{S}} \int_{A_1} \cdots \int_{A_{n-1}} P\big(t_{n-1}, y_{n-1}, t_n, A_n\big) \left(\prod_{k=1}^{n-1} P\big(t_{k-1}, y_{k-1}, t_k, dy_k\big) \right) \mu(dy_0) \\ &= \mathrm{E}\left(\prod_{k=1}^{n-1} \mathbb{1}_{X_{t_k} \in A_k} P\big(t_{n-1}, X_{t_{n-1}}, t_n, A_n\big) \right), \end{aligned}$$

whence

$$\mathrm{P}\big\{X_{t_n} \in A_n \mid X_{t_1}, \ldots, X_{t_{n-1}}\big\} = P\big(t_{n-1}, X_{t_{n-1}}, t_n, A_n\big)$$

almost surely. In particular, $\mathrm{P}\big\{X_{t_n} \in A_n \mid X_{t_1}, \ldots, X_{t_{n-1}}\big\}$ is $\sigma(X_{t_{n-1}})$-measurable, whence by the properties of conditional expectation,

$$\begin{aligned} \mathrm{P}\big\{X_{t_n} \in A_n \mid X_{t_{n-1}}\big\} &= \mathrm{E}\left(\mathrm{P}\big\{X_{t_n} \in A_n \mid X_{t_1}, \ldots, X_{t_{n-1}}\big\} \mid X_{t_{n-1}}\right) \\ &= \mathrm{P}\big\{X_{t_n} \in A_n \mid X_{t_1}, \ldots, X_{t_{n-1}}\big\}, \end{aligned}$$

which yields the Markov property, since $t_1, \ldots, t_n$ and A_n are arbitrary. □

In the case where $(\mathcal{S}, \Sigma) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, if the transition probability P, as a function of A, has a density p with respect to the Lebesgue measure, i.e.

$$P(s, x, t, A) = \int_A p(s, x, t, y) dy,$$

then p is called a *transition probability density*. The Chapman–Kolmogorov equation [7.6] can be rewritten with the help of the Fubini theorem as

$$\begin{aligned} \int_A p(s, x, t, y) dy &= \int_{\mathbb{R}^d} \int_A p(u, z, t, y) dy \, p(s, x, u, z) dz \\ &= \int_A \int_{\mathbb{R}^d} p(s, x, u, z) p(u, z, t, y) dz \, dy. \end{aligned}$$

Since this is true for any $A \in \Sigma$, we get

$$p(s, x, t, y) = \int_{\mathbb{R}^d} p(s, x, u, z) p(u, z, t, y) dz \qquad [7.9]$$

for almost all $y \in \mathbb{R}^d$.

An important particular case is where the transition probability depends only on the distance between time instances.

DEFINITION 7.4.– *A transition probability function* P *is* homogeneous *if for all* $s_1, t_1, s_2, t_2 \in \mathbb{T}$ *with* $t_2 - s_2 = t_1 - s_1 > 0$ *and any* $x \in \mathcal{S}$*,* $A \in \Sigma$*,*

$$P(s_1, x, t_1, A) = P(s_2, x, t_2, A).$$

A Markov process is homogeneous if it has a homogeneous transition probability function.

A homogeneous transition probability function may be regarded as a function of three arguments, i.e. $P(s, x, t, A) = P(t - s, x, A)$, equivalently, $P(s, x, s + t, A) = P(t, x, A)$. Then, the Chapman–Kolmogorov equation may be rewritten as

$$P(t + s, x, A) = \int_{\mathcal{S}} P(t, y, A) P(s, x, dy). \qquad [7.10]$$

If $\mathcal{S} = \mathbb{R}^d$ and the transition probability density exists, then we can write the following homogeneous version of the Chapman–Kolmogorov equation for densities:

$$p(t + s, x, y) = \int_{\mathbb{R}^d} p(s, x, z) p(t, y, z) dz \qquad [7.11]$$

for almost all $y \in \mathbb{R}^d$.

7.2. Examples of Markov processes

7.2.1. *Discrete-time Markov chain*

Let the state space $\mathcal{S}$ be finite or countable, $\Sigma = 2^{\mathcal{S}}$ and the parameter set be the set of non-negative integers: $\mathbb{T} = \{0, 1, 2 \dots\}$. In this case, a Markov process $\{X_t, t \in \mathbb{T}\} = \{X_n, n \geq 0\}$ is called a (discrete-time) *Markov chain*. Without loss of generality, we can assume that $\mathcal{S} = \{1, 2, \dots, N\}$ or $\mathcal{S} = \mathbb{N}$.

It is not hard to see that in this case the Markov property can be reformulated as follows: for any $n > m \geq 0$ and $x_0, x_1, \ldots, x_m, x \in \mathcal{S}$ such that $\mathrm{P}\{X_m = x_m\} > 0$,

$$\mathrm{P}\big\{X_n = x \mid X_0 = x_0, X_1 = x_1, \ldots, X_m = x_m\big\} = \mathrm{P}\big\{X_n = x \mid X_m = x_m\big\}.$$

For any $i, j \in \mathcal{S}$ such that $\mathrm{P}\left\{X_m = j\right\} > 0$, define the *transition probabilities*

$$p_{ij}(m, n) = \mathrm{P}\big\{X_n = j \mid X_m = i\big\};$$

for definiteness set $p_{ij}(m, n) = \mathbb{1}_{i=j}$ if $\mathrm{P}\left\{X_m = j\right\} = 0$. These probabilities form a matrix (of infinite size if $\mathcal{S} = \mathbb{N}$), called the *transition probability matrix* (or simply the transition matrix) of X:

$$P(m, n) = \big(p_{ij}(m, n)\big)_{i,j\in\mathcal{S}}.$$

The transition probability function is now easily seen to be

$$P(m, i, n, A) = \sum_{j\in A} p_{ij}(m, n),$$

so, setting in [7.6] $x = i$, $s = m$, $t = n$, $u = l$, $A = \{j\}$, we get

$$p_{ij}(m, n) = \sum_{k\in\mathbb{S}} p_{ik}(m, l)p_{kj}(l, n).$$

Thus, in the Markov chain setting, the Chapman–Kolmogorov equation turns into matrix multiplication

$$P(m, n) = P(m, l)P(l, n).$$

In particular, transition probabilities can be expressed in terms of one-step transition probabilities

$$P(m, n) = \prod_{k=m}^{n-1} P(k, k + 1). \qquad [7.12]$$

If $P(m, n)$ is a function of $n - m$, then the corresponding Markov chain is homogeneous; this is equivalent to saying that the one-step transition probabilities are independent of time: $p_{ij}(m, m + 1) = p_{ij}, i, j \in \mathcal{S}$, so

$$P(m, m + 1) = P = \left\{p_{ij}\right\}_{i,j\in\mathcal{S}},\ m \geq 0.$$

Then, the n-step transition matrix, thanks to [7.12], is a power of the one-step transition matrix $P(m, m+n) = P^n$.

Finally, it is worth mentioning that the term "Markov chain" is often used for a homogeneous Markov chain, while a general one is called a *time-inhomogeneous Markov chain*.

7.2.2. *Continuous-time Markov chain*

Consider a generalization of the previous situation: the state space is again finite or countable, but the parameter set $\mathbb{T} = \mathbb{R}_+$. Let X be a Markov process. As above, the transition probability function can be written as a matrix

$$P(s,t) = (p_{ij}(t,s))_{i,j \in \mathcal{S}},$$

where $p_{ij}(s,t) = \mathrm{P}\{X_t = j \mid X_s = i\}$ if $\mathrm{P}\{X_s = i\} > 0$ and $p_{ij}(s,t) = \mathbb{1}_{i=j}$ otherwise. Then, the Chapman–Kolmogorov equation reads

$$P(s,t) = P(s,u)P(u,t)$$

for all $t \geq u \geq s \geq 0$. In the present case of continuous time argument, this is the so-called *cocycle property*.

Let us now turn to the homogeneous case where

$$P(s,t) = P_{t-s} = \big(p_{ij}(t-s)\big)_{i,j \in \mathcal{S}},$$

so

$$P_{s+t} = P_s P_t. \qquad [7.13]$$

Together with the "initial condition" $P_0 = I$, the identity matrix, this equation resembles properties of an exponential function and gives the idea that

$$P_t = e^{tA} := \sum_{n=0}^{\infty} \frac{t^n}{n!} A^n,$$

the matrix exponential. We will show this under the additional assumptions that $\mathcal{S}$ is finite and that the transition probabilities are continuous. The matrix A in this representation is called the *generator* matrix.

PROPOSITION 7.2.– *Let $\mathcal{S}$ be finite and the transition probability matrix P_t be continuous at $t = 0$, i.e. $P_t \to I$, $t \to 0+$. Then, for each $i, j \in \mathcal{S}$, the limit $a_{ij} := \lim_{t\to 0+} t^{-1}(p_{ij}(t) - \delta_{ij})$ exists, and $P_t = e^{tA}$, $t \geq 0$, where $A = \big(a_{ij}\big)_{ij\in\mathcal{S}} = \lim_{t\to 0+} t^{-1}(P_t - I)$.*

PROOF.– It follows from the continuity at 0 and [7.13] that P_t is continuous on $\mathbb{R}_+$. Then, we can integrate it elementwise, moreover,

$$S_t := \frac{1}{t}\int_0^t P_s ds \to I,\ t \to 0+\,.$$

In particular, $\det S_t \to 1$, $t \to 0+$, so there exists $\varepsilon > 0$ such that S_t is invertible for any $t \in (0, \varepsilon)$. Taking some $a \in (0, \varepsilon)$ and using [7.13], we get for any $h > 0$

$$V_h = P_h S_a = P_h \int_0^a P_t dt = \int_0^a P_{t+h} dt = \int_h^{h+a} P_t dt.$$

Thanks to the continuity of P_t if the right-hand side of the last equation is continuously differentiable in h, then the left-hand side is differentiable as well; in particular, it is differentiable at zero. Therefore, $P_h = V_h S_a^{-1}$ is differentiable at zero as well, which implies the existence of the matrix A.

From [7.13], we have

$$\frac{1}{s}\left(P_{t+s} - P_t\right) = \frac{1}{s}\left(P_s - I\right) P_t \to AP_t,\ s \to 0+,$$

so we get the differential equation

$$\frac{d}{dt} P_t = AP_t,\ t \geq 0,$$

or, in integral form,

$$P_t = I + \int_0^t AP_s ds,\ t \geq 0.$$

Then, we can write

$$\begin{aligned} P_t &= I + \int_0^t A\Big(I + \int_0^s AP_u du\Big) ds = I + tA + \int_0^t \int_0^s A^2 P_u du\, ds \\ &= I + tA + \int_0^t \int_0^s A^2\Big(I + \int_0^v AP_v dv\Big) du\, ds \\ &= I + tA + \frac{t^2}{2} A^2 + \int_0^t \int_0^s \int_0^u A^3 P_v dv\, du\, ds. \end{aligned}$$

Continuing this line of reasoning, we get

$$P_t = I + I + tA + \frac{t^2}{2}A^2 + \cdots + \frac{t^n}{n!}A^n + R_{n,t},$$

where

$$R_{n,t} = \int_0^t \int_0^{t_1} \cdots \int_0^{t_n} A^{n+1} P_{t_{n+1}} dt_{n+1} dt_n \ldots dt_1.$$

It is easy to see that the elements of $R_{n,t}$ do not exceed $(ta^* |\mathcal{S}|)^{n+1}/(n+1)!$, where $a^* = \max_{i,j\in\mathcal{S}} a_{i,j}$. Therefore, $R_{n,t}$ vanishes as $n \to \infty$, so letting $n \to \infty$, we get $P_t = e^{tA}$. □

Assume further that the conditions of proposition 7.2 are fulfilled.

Let us now describe the evolution of X. It follows from our assumptions that $p_{ii}(t) \to 1$ for each $t \in \mathcal{S}$, which means that X is stochastically continuous. Therefore, by theorem 6.2, it has a separable modification, so we will assume that X is separable. Denote $\lambda_i = -a_{ii}$, $i \in \mathcal{S}$. Let us identify the distribution of the exit time $\tau = \inf\{t \geq 0 : X_t \neq i\}$ given that $X_0 = i$. Note that $\{\tau > t\} = \{X_s = i \;\forall s \in [0,t]\}$. Let $t_k^n = tk/n$, $k = 0, 1, \ldots, n$. Then, thanks to the separability of X,

$$\begin{aligned} \mathrm{P}\{\tau > t \mid X_0 = i\} &= \lim_{n\to\infty} \mathrm{P}\big\{X_{t_k^n} = i \;\forall k = 0, 1, \ldots, n \mid X_0 = i\big\} \\ &= \lim_{n\to\infty} \prod_{k=1}^{n} \mathrm{P}\big\{X_{t_k^n} = i \mid X_{t_{k-1}^n} = i\big\} = \lim_{n\to\infty} p_{ii}(t/n)^n \\ &= \exp\Big\{\lim_{n\to\infty} n \log p_{ii}(t/n)\Big\} = \exp\Big\{\lim_{n\to\infty} n\big(p_{ii}(t/n) - 1\big)\Big\} = e^{-\lambda_i t}, \end{aligned}$$

so τ has an exponential distribution with parameter λ_i. Further, to identify the distribution of the value of X_t after jump, we observe that for $j \neq i$, the probability $\mathrm{P}\big\{X_{t+s} = j \mid X_{t+s} \neq i, X_t = i\big\}$ is independent of t and is equal to

$$\frac{p_{ij}(s)}{1 - p_{ii}(s)},$$

which converges to $q_{ij} := a_{ij}/\lambda_i$ as $s \to 0+$. Consequently, independently of where the jump occurs, the distribution of the value $X_{\tau+}$ after the jump is

$$\mathrm{P}\big\{X_{\tau+} = j \mid X_0 = i\big\} = q_{ij},\; j \in \mathcal{S} \setminus \{i\}.$$

Summing up, the behavior of the continuous-time Markov chain X is as follows. It spends an exponential time at state i and then switches to another state according to probabilities q_{ij}. The sequence of values of X is a discrete-time homogeneous Markov chain with one-step transition probability matrix $Q = \left(q_{ij}\right)_{i,j\in\mathcal{S}}$, where $q_{ii} = 0$ for all $i \in \mathcal{S}$; this is the so-called *embedded Markov chain*. Such behavior is close to that of the Poisson process (see section 2.3.3).

7.2.3. *Process with independent increments*

Assume that $\mathcal{S} = \mathbb{R}^d$ and that the process $\{X_t, t \in \mathbb{T}\}$ in $\mathcal{S}$ has independent increments. Then, for any $s, t \in \mathbb{T}$ with $s < t$ and any $A \in \mathcal{B}(\mathbb{R}^d)$,

$$\begin{aligned} \mathrm{P}\big\{X(t) \in A \mid \mathcal{F}_s\big\} &= \mathrm{P}\big\{X_t - X_s \in A - X_s \mid \mathcal{F}_s\big\} \\ &= \mathrm{P}\big\{X_t - X_s \in A - x\big\}_{X_s = x} \end{aligned} \quad [7.14]$$

almost surely, since X_s is $\mathcal{F}_s$-measurable and $X_t - X_s$ is independent of $\mathcal{F}_s$; here, $A - x := \{y - x, y \in A\}$. Noting that the last term in [7.14] is a function of X_s, we establish the Markov property of X.

In particular, any Lévy process is a homogeneous Markov process.

7.3. Semigroup resolvent operator and generator related to the homogeneous Markov process

Let $\mathcal{S}$ be a complete separable metric space, $\Sigma = \mathcal{B}(\mathcal{S})$ be the Borel σ-field and $X = \{X_t, t \geq 0\}$ be a homogeneous Markov process with transition probability function $P(t, x, A)$, $t \geq 0$, $x \in \mathcal{S}$, $A \in \Sigma$. Since $P(t, x, \cdot)$ is a probability measure in $A \in \Sigma$, for any bounded measurable function $f \colon \mathcal{S} \to \mathbb{R}$, we can define the integral

$$T_t f(x) = \int_{\mathcal{S}} f(y) P(t, x, dy), \;\; t > 0. \quad [7.15]$$

7.3.1. *Semigroup related to Markov process*

Denote $\mathbb{B}(\mathcal{S})$ the space of bounded measurable functions $f \colon \mathcal{S} \to \mathbb{R}$ with the norm $\|f\| = \sup_{x\in\mathcal{S}} |f(x)|$. Formula [7.15] defines the operator $T_t : \mathbb{B}(\mathcal{S}) \to \mathbb{B}(S)$. Indeed,

$$\|T_t f\| \leq \sup_{x\in\mathcal{S}} |T_t f(x)| \leq \|f\| \int_{\mathcal{S}} P(t, x, dy) = \|f\|. \quad [7.16]$$

The operator T_t is obviously linear on $\mathbb{B}(\mathcal{S})$ and the relation [7.16] means that $\|T_t\| = \sup_{f\in\mathbb{B}(\mathcal{S})} \|T_t f\| \leq 1$ (for the definition of the operator norm, see section A1.7).

Moreover, if we put $f \equiv 1$, then $\|T_t f\| = 1$. It means that $\|T_t\| = 1$ for any $t > 0$. Defining

$$T_0 f(x) = f(x), \tag{7.17}$$

we get a family of linear isometric operators $\{T_t, t \geq 0\} : \mathbb{B}(\mathcal{S}) \to \mathbb{B}(\mathcal{S})$. With the general definition of semigroup from section A1.7, we can prove the following result.

THEOREM 7.2.– *The family of operators* $\{T_t, t \geq 0\}$ *defined by the relations* [7.15], [7.17] *is a semigroup.*

PROOF.– We should check only semigroup equation [A1.6], but it follows immediately from the Chapman–Kolmogorov equation, because for any $f \in \mathbb{B}(\mathcal{S})$ and $t, s \geq 0$, the Fubini theorem and [7.16] imply that

$$T_{t+s} f(x) = \int_{\mathcal{S}} f(y) P(t+s, x, dy) = \int_{\mathcal{S}} f(y) \int_{\mathcal{S}} P(t, z, dy) P(s, x, dz)$$

$$= \int_{\mathcal{S}} P(s, x, dz) \int_{\mathcal{S}} f(y) P(t, z, dy) = \int_{\mathcal{S}} T_t f(z) P(s, x, dz) = T_t(T_s f)(x). \qquad \square$$

The semigroup T_t is called a *Markov semigroup*.

7.3.2. *Resolvent operator and resolvent equation*

Let $\{T_t, t \geq 0\}$ be a semigroup defined by relations [7.15] and [7.17]. Consider the Laplace transform of the following form: for any $\lambda > 0$ and $f \in \mathbb{B}(\mathcal{S})$, let

$$R_\lambda f(x) = \int_0^\infty e^{-\lambda t} T_t f(x) dt. \tag{7.18}$$

Operators $\{R_\lambda, \lambda > 0\}$ form a family of *resolvent operators* of the semigroup T_t.

LEMMA 7.1.– *The family* $\{R_\lambda, \lambda > 0\}$ *defined by relation* [7.18] *consists of bounded linear operators from* $\mathbb{B}(\mathcal{S})$ *to* $\mathbb{B}(\mathcal{S})$*, and* $\|R_\lambda\| = \lambda^{-1}$.

PROOF.– Let us estimate $\|R_\lambda\|$. We have that

$$|R_\lambda f(x)| \leq \int_0^\infty e^{-\lambda t} |T_t f(x)| dt \leq \int_0^\infty e^{-\lambda t} \sup_{x\in S} |T_t f(x)| dt$$

$$\leq \|f\| \int_0^\infty e^{-\lambda t} dt = \lambda^{-1} \|f\|,$$

whence $R_\lambda \colon \mathbb{B}(\mathcal{S}) \to \mathbb{B}(\mathcal{S})$ and $\|R_\lambda\| \le \lambda^{-1}$. Let $f = 1$. Then, $R_\lambda \mathbb{1}(x) = \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda}$, whence $\|R_\lambda\| = \frac{1}{\lambda}$. Linear property is evident. □

THEOREM 7.3.– (*Resolvent equation*). *For any* $\lambda > 0$, $\mu > 0$, *we have the following operator equation:*

$$R_\lambda - R_\mu = (\mu - \lambda) R_\lambda R_\mu = (\lambda - \mu) R_\mu R_\lambda. \qquad [7.19]$$

PROOF.– For any $f \in \mathbb{B}(\mathcal{S})$, $x \in \mathcal{S}$ and $\lambda, \mu > 0$, $\lambda \neq \mu$, we can apply the Fubini theorem and theorem A1.14 to get the following equality:

$$\begin{aligned}
R_\lambda f(x) - R_\mu f(x) &= \int_0^\infty (e^{-\lambda t} - e^{-\mu t}) T_t f(x) dt \\
&= \int_0^\infty e^{-\lambda t}(1 - e^{-(\mu-\lambda)t}) T_t f(x) dt \\
&= \int_0^\infty e^{-\lambda t} \int_0^t e^{-(\mu-\lambda)s}\, ds\, T_t f(x)\, dt\, (\mu - \lambda) \\
&= (\mu - \lambda) \int_0^\infty e^{-(\mu-\lambda)s} \left(\int_s^\infty e^{-\lambda t} T_t f(x)\, dt \right) ds \\
&= (\mu - \lambda) \int_0^\infty e^{-(\mu-\lambda)s} \left(\int_0^\infty e^{-\lambda(s+u)} T_{s+u} f(x)\, du \right) ds \\
&= (\mu - \lambda) \int_0^\infty e^{-\mu s} T_s \left(\int_0^\infty e^{-\lambda u} T_u f(x)\, du \right) ds \\
&= (\mu - \lambda) \int_0^\infty e^{-\mu s} T_s (R_\lambda f(x)) ds = (\mu - \lambda) R_\mu R_\lambda f(x).
\end{aligned}$$

□

REMARK 7.1.– In the theory of linear operators, the resolvent operator traditionally is defined in the following way. Let $\mathbb{Z}$ be some linear normed space and $A : \mathbb{Z} \to \mathbb{Z}$ be a bounded linear operator. Denote

$$\mathcal{R}_A = \left\{ \lambda \in R : (A - \lambda I)^{-1} \text{ exists as a linear bounded operator from } \mathbb{Z} \text{ to } \mathbb{Z} \right\}$$

the resolvent set of A. Define $\widetilde{R}_\lambda = (A - \lambda I)^{-1}$, $\lambda \in \mathcal{R}_A$ the resolvent operator of A. Now, let $\lambda, \mu \in \mathcal{R}_A$. Then

$$\begin{aligned}
\widetilde{R}_\lambda - \widetilde{R}_\mu &= (A - \lambda I)^{-1} - (A - \mu I)^{-1} \\
&= (A - \lambda I)^{-1}(I - (A - \lambda I)(A - \mu I)^{-1}) \qquad [7.20] \\
&= (A - \lambda I)^{-1}(I - (A - \mu I)(A - \mu I)^{-1} - (\mu - \lambda)(A - \mu I)^{-1}) \\
&= (\lambda - \mu)(A - \lambda I)^{-1}(A - \mu I) = (\lambda - \mu)\widetilde{R}_\lambda \widetilde{R}_\mu.
\end{aligned}$$

Equation [7.20] differs from [7.19] only in sign. Therefore, for $R_\lambda = -\widetilde{R}_\lambda = (\lambda I - A)^{-1}$, we get equation [7.19]. This leads to the idea of looking for an operator for R_λ from equation [7.18] as a resolvent operator. This idea will be realized with the notion of generator, which generalizes that from section 7.2.2. However, at first, we establish some auxiliary results.

LEMMA 7.2.– *For any $t > 0$, $\lambda > 0$ and $f \in \mathbb{B}(S)$,*

$$T_t R_\lambda f - R_\lambda f = (e^{\lambda t} - 1) R_\lambda f - e^{\lambda t} \int_0^t T_s f e^{-\lambda s} ds. \qquad [7.21]$$

PROOF.– Because of theorem A1.14, we can write

$$\begin{aligned} T_t R_\lambda f &= T_t \int_0^\infty e^{-\lambda s} T_s f ds = \int_0^\infty e^{-\lambda s} T_{t+s} f ds \\ &= \int_t^\infty e^{-\lambda(u-t)} T_u f du = e^{\lambda t} \int_t^\infty e^{-\lambda u} T_u f du \\ &= e^{\lambda t} \left(R_\lambda f - \int_0^t e^{-\lambda u} T_u f d_u \right) \\ &= R_\lambda f + (e^{\lambda t} - 1) R_\lambda f - e^{\lambda t} \int_0^t e^{-\lambda u} T_u f du, \end{aligned}$$

as required. □

7.3.3. *Generator of a semigroup*

The general definition of the generator of a semigroup is given in section A1.7. We can consider a particular case of a Markov semigroup T_t to transition probability via [7.15].

DEFINITION 7.5.– *Let $\{T_t, t \geq 0\}$ be defined by* [7.15] *and* [7.17]. *The* generator A *of the semigroup T_t is the operator*

$$Af(x) = \lim_{t \downarrow 0} \frac{T_t f(x) - f(x)}{t}$$

whenever the limit exists in the norm $\|\cdot\|$ on the space $\mathbb{B}(\mathcal{S}) : \|g\| = \sup_{x \in \mathcal{S}} |g(x)|$.

Denote D_A the domain of the operator A. Sometimes A is called the *infinitesimal operator* of the semigroup T_t.

By theorem A1.13, for any $t \geq 0$ and $f \in D_A$,

$$\frac{d}{dt} T_t f = A T_t f$$

and

$$\frac{d}{dt} T_t f = T_t A f. \qquad [7.22]$$

These are so-called *forward and backward Kolmogorov equations* for the Markov semigroup $\{T_t, t \geq 0\}$.

THEOREM 7.4.–

i) $\mathbb{B}_0(\mathcal{S})$ *is the subspace in* $\mathbb{B}(\mathcal{S})$ *and for any* $s > 0$ $T_s : \mathbb{B}_0(\mathcal{S}) \to \mathbb{B}_0(\mathcal{S})$.

ii) For any $f \in \mathbb{B}_0(\mathcal{S})$,

$$\lim_{\lambda \uparrow \infty} \|\lambda R_\lambda f - f\| = 0.$$

iii) For any $f \in \mathbb{B}_0(\mathcal{S})$,

$$A R_\lambda f = \lambda R_\lambda f - f.$$

iv) For any $f \in D_A$,

$$R_\lambda A f = \lambda R_\lambda f - f.$$

PROOF.– i) Obviously, $\mathbb{B}_0(\mathcal{S})$ is a linear set. Further, let $\{f_n, n \geq 1\} \in \mathbb{B}_0(\mathcal{S})$ and $\|f_n - f\| \to 0$ as $n \to \infty$, $f \in \mathbb{B}(\mathcal{S})$. Then

$$\begin{aligned}\|T_t f - f\| &\leq \|T_t f_n - f_n\| + \|T_t f_n - T_t f\| + \|f_n - f\| \\ &\leq \|T_t f_n - f_n\| + 2\|f_n - f\|.\end{aligned}$$

Therefore, for any $n \geq 1$,

$$\limsup_{t \downarrow 0} \|T_t f - f\| \leq \limsup_{t \downarrow 0} \|T_t f_n - f_n\| + 2\|f_n - f\| = 2\|f_n - f\|. \qquad [7.23]$$

Taking limits as $n \to \infty$ in the left- and right-hand sides of [7.23], we get that

$$\lim_{t \downarrow 0} \|T_t f - f\| = 0.$$

Now, let $f \in \mathbb{B}_0(\mathcal{S})$. Then, for any $s \leq 0$,

$$\lim_{t \downarrow 0} \|T_t T_s f - T_s f\| = \lim_{t \downarrow 0} \|T_s T_t f - T_s f\| \leq \lim_{t \downarrow 0} \|T_t f - f\| = 0.$$

Therefore, $T_s f \in \mathbb{B}_0(\mathcal{S})$.

ii) Let $f \in \mathbb{B}_0(\mathcal{S})$. Then, for any $\lambda > 0$,

$$\lambda R_\lambda f - f = \lambda \int_0^\infty e^{-\lambda t}(T_t f - f)dt = \int_0^\infty e^{-u} \left(T_{u/\lambda} f - f\right) du.$$

Therefore,

$$\|\lambda R_\lambda f - f\| \leq \int_0^\infty e^{-u} \left\|T_{u/\lambda} f - f\right\| du.$$

For any $u \geq 0$, $\left\|T_{u/\lambda} f - f\right\| \to 0$ as $\lambda \to \infty$. Additionally, there exists an integrable dominant $2e^{-u} \|f\|$. Therefore, the Lebesgue dominated convergence theorem supplies that $\int_0^\infty e^{-u} \left\|T_{u/\lambda} f - f\right\| du \to 0$, and obviously $\|\lambda R_\lambda f - f\| \to 0$, $\lambda \to +\infty$.

iii) Let $f \in \mathbb{B}_0(\mathcal{S})$. Then, it follows from [7.21] that

$$\frac{T_t R_\lambda f - R_\lambda f}{t} = \frac{e^{\lambda t} - 1}{t} R_\lambda - e^{\lambda t} \frac{1}{t} \int_0^t T_s f e^{-\lambda s} ds.$$

Now, let $t \downarrow 0$. Then, the real-number multiplier $\frac{e^{\lambda t}-1}{t} \to \lambda$. Furthermore,

$$\begin{aligned}
&\left\| \frac{1}{t} \int_0^t T_s f e^{-\lambda s} ds - f \right\| \\
&= \left\| \frac{1}{t} \int_0^t T_s f e^{-\lambda s} ds - \frac{1}{t} \int_0^t e^{-\lambda s} ds \cdot f + \left(\frac{1 - e^{-\lambda t}}{\lambda t} - 1 \right) f \right\| \\
&= \left\| \int_0^t (T_s f - f) \frac{1 - e^{-\lambda s}}{t} ds \right\| + \frac{|1 - e^{-\lambda t} - \lambda t|}{\lambda t} \|f\| \\
&\leq \lambda \int_0^t \|T_s f - f\| \frac{s}{t} ds + \frac{|1 - e^{-y} - y|}{y} \|f\|,
\end{aligned}$$

where $y = \lambda t \downarrow 0$ as $t \downarrow 0$. Further,

$$\int_0^t \|T_s f - f\| \frac{s}{t} ds \leq \int_0^t \|T_s f - f\| \, ds \leq 2 \|f\| \, t \to 0$$

as $t \downarrow 0$, and by the L'Hôpital rule,

$$\lim_{y \downarrow 0} \frac{1 - e^{-y} - y}{y} = \lim_{y \downarrow 0} (e^{-y} - 1) = 0.$$

Therefore,

$$\left\| \frac{1}{t} \int_0^t e^{-\lambda s} T_s f ds - f \right\| \to 0 \text{ as } t \downarrow 0,$$

and

$$AR_\lambda f = \lim_{t \downarrow 0} \frac{T_t R_\lambda f - R_\lambda f}{t} = \lambda R_\lambda f - f,$$

as required.

iv) Let $f \in D(A)$. Then

$$R_\lambda A f = \int_0^\infty e^{-\lambda t} T_t A f dt = \int_0^\infty e^{-\lambda t} T_t \left(\lim_{s \downarrow 0} \frac{T_s f - f}{s} \right) dt$$
$$= \int_0^\infty e^{-\lambda t} \lim_{s \downarrow 0} \frac{T_{t+s} f - T_t f}{s} dt.$$

To swap the integral and the limit sign in the last expression, we only need to check that $e^{-\lambda t} \frac{1}{s} (T_{t+s} f - T_t f)$ admits an integrable dominant independent of s. Thanks to equation [7.22], $T_{s+t} f - T_t f = \int_t^{t+s} T_u A f du$, and

$$\frac{\|T_{s+t} f - T_t f\|}{s} \leq \|A f\| \, .$$

Therefore, the integrable dominant is $e^{-\lambda t} \|A f\|$, whence we get that

$$R_\lambda A f = \lim_{s \downarrow 0} \int_0^\infty e^{-\lambda t} \frac{T_{s+t} f - T_t f}{s} dt$$
$$= \lim_{s \downarrow 0} \frac{1}{s} \left(e^{\lambda s} \int_s^\infty e^{-\lambda u} T_u f d\lambda \, du - \int_0^\infty e^{-\lambda t} T_t f dt \right)$$
$$= \lim_{s \downarrow 0} \frac{1}{s} \left((e^{\lambda s} - 1) \int_0^\infty T_u f du - e^{\lambda s} \int_0^s e^{-\lambda u} T_u f du \right)$$
$$= \lambda R_\lambda f - f.$$

□

REMARK 7.2.– It follows from (iii) that $R_\lambda : \mathbb{B}_0(\mathcal{S}) \to D_A$. The operator R_λ is a bijection between $\mathbb{B}_0(\mathcal{S})$ and D_A. Indeed, if $R_\lambda f = 0$, then it follows from (iii) that $f = 0$. Also, any $g \in D_A$ can be represented as $g = R_\lambda f$ with $f \in \mathbb{B}_0(\mathcal{S})$ if we put $f = \lambda g - Ag$. The only question is why $Ag \in \mathbb{B}_0(\mathcal{S})$. It is true because $g \in D_A \subset \mathbb{B}_0(\mathcal{S})$, $T_t g \in D_A \subset \mathbb{B}_0(\mathcal{S})$, $T_t g - g \in \mathbb{B}_0(\mathcal{S})$ and $Ag = \lim_{t\downarrow 0} \frac{T_t g - g}{t} \in \mathbb{B}_0(\mathcal{S})$ because this limit exists and $\mathbb{B}_0(\mathcal{S})$ is a closed set. Then, it follows from (iii) and (iv) of theorem 7.4 that $R_\lambda = (\lambda I - A)^{-1}$.

7.4. Definition and basic properties of diffusion process

As the name suggests, a diffusion process is a mathematical model for the physical phenomenon of diffusion. In physics, diffusion can be understood either at a macroscopic level as a movement of substance from a region with high concentration to a region with low concentration or at a microscopic level as a chaotic movement of an individual particle, say, of a gas, which results from its interaction with other particles. We are interested in this second notion, which in its simplest form is the celebrated *Brownian motion* (mathematically modeled by the Wiener process).

Let the state space be a finite-dimensional Euclidean space: $\mathcal{S} = \mathbb{R}^d$, and the parameter set be non-negative half-line: $\mathbb{T} = \mathbb{R}_+$.

Denote by $B(x,r) = \{y \in \mathbb{R}^d : |y - x| \le r\}$ the ball of radius r centered at x, with $B(x,r)^c = \mathbb{R}^d \backslash B(x,r)$ its complement. Also, let (x,y) denote the inner product in $\mathbb{R}^d$and M_d the set of symmetric non-negative matrices of size d.

DEFINITION 7.6.– *A continuous Markov process X with the transition probability P is called a* diffusion process *if there exist measurable functions $a\colon \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}_+ \times \mathbb{R}^d \to M_d$ such that for all $\varepsilon > 0$, $t \in \mathbb{R}_+$, $x, z \in \mathbb{R}^d$,*

$$\lim_{h\to 0+} \frac{1}{h} P\big(t, x, t+h, B(x,\varepsilon)^c\big) = 0, \tag{7.24}$$

$$\lim_{h\to 0+} \frac{1}{h} \int_{B(x,\varepsilon)} (y-x) P(t,x,t+h,dy) = a(t,x), \tag{7.25}$$

$$\lim_{h\to 0+} \frac{1}{h} \int_{B(x,\varepsilon)} \big(y-x,z\big)^2 P(t,x,t+h,dy) = \big(\sigma(t,x)z,z\big). \tag{7.26}$$

The functions a and b are called the drift coefficient *and the* diffusion matrix*, respectively.*

The drift coefficient plays the role of local average speed of a particle, the deterministic part of evolution. The diffusion matrix, which corresponds to the stochastic part of evolution, measures the amplitude of noise, namely, for $z \in \mathbb{R}^d$

with $\|z\| = 1$, the expression $\big(\sigma(t,x)z, z\big)\,dt$ is the variance of the infinitesimal displacement projected to z.

REMARK 7.3.– It follows from [7.26] that for any $z_1, z_2 \in \mathbb{R}^d$,

$$\frac{1}{h}\int_{B(x,\varepsilon)} \big(y-x, z_1\big)\big(y-x, z_2\big)P(t,x,t+h,dy) = \big(\sigma(t,x)z_1, z_2\big). \qquad [7.27]$$

Indeed, both sides of this equality are symmetric bilinear forms as functions of z_1, z_2, so [7.26] implies [7.27] through the polarization identity: for any symmetric bilinear function f,

$$f(z_1,z_2) = \frac{1}{4}\big(f(z_1+z_2, z_1+z_2) - f(z_1-z_2, z_1-z_2)\big).$$

Let us give simpler sufficient conditions for a diffusion process, which are often easier to check.

PROPOSITION 7.3.– *Assume that for some $\delta > 0$ and any $z \in \mathbb{R}^d$, the following conditions hold:*

$$\lim_{h\to 0+} \frac{1}{h}\int_{\mathbb{R}^d} |y-x|^{2+\delta}P(t,x,t+h,dy) = 0,$$

$$\lim_{h\to 0+} \frac{1}{h}\int_{\mathbb{R}^d} (y-x)P(t,x,t+h,dy) = a(t,x),$$

$$\lim_{h\to 0+} \frac{1}{h}\int_{\mathbb{R}^d} \big(y-x, z\big)^2 P(t,x,t+h,dy) = \big(\sigma(t,x)z, z\big),$$

where $a\colon \mathbb{R}_+ \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}_+ \to M_d$ are measurable functions. Then, X is a diffusion process with drift $a(t,x)$ and diffusion matrix $\sigma(t,x)$.

PROOF.– For any $\varepsilon > 0$, by the Markov inequality

$$\frac{1}{h}P\big(t,x,t+h,B(x,\varepsilon)^c\big) \le \frac{1}{h\varepsilon^{2+\delta}}\int_{\mathbb{R}^d} |y-x|^{2+\delta}P(t,x,t+h,dy) \to 0,\ h\to 0+\,.$$

Further,

$$\frac{1}{h}\int_{B(x,\varepsilon)} (y-x)P(t,x,t+h,dy)$$
$$= \frac{1}{h}\int_{\mathbb{R}^d} (y-x)P(t,x,t+h,dy) - \frac{1}{h}\int_{B(x,\varepsilon)^c} (y-x)P(t,x,t+h,dy)$$
$$=: I_1(h) + I_2(h).$$

By our assumption, $I_1(h) \to a(t,x)$, $h \to 0+$. Also

$$
\begin{aligned}
|I_2(h)| &\le \frac{1}{h}\int_{B(x,\varepsilon)^c} |y-x|\, P(t,x,t+h,dy) \\
&\le \frac{1}{h\varepsilon^{1+\delta}}\int_{B(x,\varepsilon)^c} |y-x|\, P(t,x,t+h,dy) \\
&\le \frac{1}{h\varepsilon^{1+\delta}}\int_{\mathbb{R}^d} |y-x|\, P(t,x,t+h,dy) \to 0,\ h \to 0+\,.
\end{aligned}
$$

Therefore,

$$
\lim_{h\to 0+} \frac{1}{h}\int_{B(x,\varepsilon)} (y-x)P(t,x,t+h,dy) = a(t,x).
$$

Similarly, for any $z \in \mathbb{R}^d$, we have

$$
\lim_{h\to 0+} \frac{1}{h}\int_{B(x,\varepsilon)} \big(y-x,z\big)^2 P(t,x,t+h,dy) = \big(\sigma(t,x)z,z\big),
$$

as required. □

Now we are going to obtain an alternative definition of diffusion process, which is often used as a principal definition. Introduce the second-order differential operator

$$
\begin{aligned}
\mathcal{L}_t f(x) &= \sum_{i=1}^{d} a_i(t,x)\frac{\partial}{\partial x_i} f(x) + \frac{1}{2}\sum_{i,j=1}^{d} \sigma_{ij}(t,x)\frac{\partial^2}{\partial x_i \partial x_j} f(x) \\
&= \big(a(t,x), D_x f(x)\big) + \frac{1}{2}\mathrm{tr}\left(\sigma(t,x)D^2_{xx} f(x)\right),\ f \in C^2_b(\mathbb{R}^d),
\end{aligned}
$$

where $C^2_b(\mathbb{R}^d)$ denotes the set of twice-continuously differentiable bounded functions with bounded derivatives of first and second orders.

THEOREM 7.5.– *A continuous Markov process X is a diffusion process if and only if for any $f \in C^2_b(\mathbb{R}^d)$ and all $t \in \mathbb{R}^+$, $x \in \mathbb{R}^d$*

$$
\frac{1}{h}\int_{\mathbb{R}^d} \big(f(y) - f(x)\big)P(t,x,t+h,dy) \to \mathcal{L}_t f(x),\ h \to 0+\,. \qquad [7.28]
$$

PROOF.– First, assume that X is a diffusion process. Take any $f \in C_b^2(\mathbb{R}^d)$ and write for $\varepsilon > 0$

$$\begin{aligned}
&\frac{1}{h}\int_{\mathbb{R}^d} \big(f(y) - f(x)\big)P(t, x, t+h, dy) \\
&= \left(\int_{B(x,\varepsilon)} + \int_{B(x,\varepsilon)^c}\right)\big(f(y) - f(x)\big)P(t, x, t+h, dy) \\
&=: I_1(h) + I_2(h).
\end{aligned}$$

Thanks to [7.24]:

$$\begin{aligned}
|I_2(\varepsilon)| &\le \frac{1}{h}\int_{B(x,\varepsilon)^c} |f(y) - f(x)|\, P(t, x, y, dy) \\
&\le \frac{2}{h}\sup_{z\in\mathbb{R}^d} |f(z)|\, P\big(t, x, t+h, B(x,\varepsilon)^c\big) \to 0,\ h \to 0+\,.
\end{aligned}$$

Further, write for $y \in B(x,\varepsilon)$, using the Taylor formula,

$$f(y) - f(x) = (D_x f(x), y - x) + \frac{1}{2}\big(\big(D^2_{xx} f(x)\big)(y-x), y-x\big) + R(x,y),$$

where $|R(x,y)| \le c(\varepsilon)\,|y-x|^2$ and $c(\varepsilon) \to 0$, $\varepsilon \to 0+$. Then

$$\begin{aligned}
I_2(h) &= \frac{1}{h} D_x f(x) \cdot \int_{B(x,\varepsilon)} (y-x) P(t, x, t+h, dy) \\
&\quad + \frac{1}{2h}\int_{B(x,\varepsilon)} \big(\big(D^2_{xx} f(x)\big)(y-x), y-x\big) P(t, x, t+h, dy) \\
&\quad + \frac{1}{h}\int_{B(x,\varepsilon)} R(y) P(t, x, t+h, dy) =: I_{21}(h) + I_{22}(h) + I_{23}(h).
\end{aligned}$$

By [7.25],

$$I_{21}(h) \to a(t,x),\ h \to 0+;$$

by [7.27],

$$\begin{aligned}
I_{22}(h) &= \frac{1}{2h}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} f(x) \int_{B(x,\varepsilon)} (y_i - x_i)(y_j - x_j) P(t, x, t+h, dy) \\
&\longrightarrow \frac{1}{2}\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} f(x) \sigma_{ij}(t,x) = \frac{1}{2}\operatorname{tr}\left(\sigma(t,x) D^2_{xx} f(x)\right),\ h \to 0+\,.
\end{aligned}$$

Further,

$$|I_{23}(h)| \le c(\varepsilon)\frac{1}{h}\int_{B(x,\varepsilon)} |y-x|^2 P(t,x,t+h,dy),$$

and from [7.27], we obtain

$$\lim_{h\to 0+} \frac{1}{h}\int_{B(x,\varepsilon)} |y-x|^2 P(t,x,t+h,dy) = \operatorname{tr}\sigma(t,x),$$

whence

$$\limsup_{h\to 0+} |I_2(h) - \mathcal{L}_t f(x)| = \limsup_{h\to 0+} |I_{23}(h)| \le c(\varepsilon)\operatorname{tr}\sigma(t,x) \le c(\varepsilon)\operatorname{tr}\sigma(t,x).$$

By letting $\varepsilon \to 0+$, we arrive at the necessary part of the statement.

Concerning the sufficiency part, to prove [7.24], we consider for fixed $x \in \mathbb{R}^d$ a non-negative function $f \in C_b^2(\mathbb{R}^d)$ such that f vanishes at x together with its first- and second-order derivatives, and $f(y) = 1$ for $y \notin B(x,\varepsilon)$. Then, using [7.28], we have

$$\frac{1}{h}P(t,x,t+h,B(x,\varepsilon)^c) \le \frac{1}{h}\int_{\mathbb{R}^d} (f(y)-f(x))P(t,x,y,dy) \to \mathcal{L}_t f(x), h \to 0+\,.$$

Further, for each $i = 1,\dots,d$, we consider a function $g \in C_b^2(\mathbb{R}^d)$ such that $g(y) = y_i$, $y \in B(x,\varepsilon)$. Then

$$\frac{1}{h}\int_{B(x,\varepsilon)} (x_i - y_i)P(t,x,t+h,dy)$$
$$= \frac{1}{h}\left(\int_{\mathbb{R}^d} - \int_{B(x,\varepsilon)^c}\right) g(x)P(t,x,t+h,dy) \to \mathcal{L}_t g(x) = a_i(t,x),\ h \to 0+,$$

where the second integral vanishes thanks to the already proved [7.24]. This establishes [7.25], and [7.26] is shown similarly. □

7.5. Homogeneous diffusion process. Wiener process as a diffusion process

DEFINITION 7.7.– *A diffusion process X is called* homogeneous *if it is a homogeneous Markov process.*

We can formulate an equivalent definition by adjusting definition 7.6 accordingly.

DEFINITION 7.8.– *A process X with the transition probability P is called a homogeneous diffusion process if there exist measurable functions $a\colon \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma\colon \mathbb{R}^d \to M_d$ such that for all $\varepsilon > 0$, $x, z \in \mathbb{R}^d$,*

$$\lim_{h\to 0+} \frac{1}{h} P\big(h, x, B(x,r)^c\big) = 0,$$
$$\lim_{h\to 0+} \frac{1}{h} \int_{B(x,\varepsilon)} (y-x) P(h,x,dy) = a(x), \qquad [7.29]$$
$$\lim_{h\to 0+} \frac{1}{h} \int_{B(x,\varepsilon)} \big(y-x,z\big)^2 P(h,x,dy) = \big(\sigma(x)z,z\big).$$

We see that the drift and the diffusion matrix of a homogeneous diffusion process are functions of x only, i.e. $a(t,x) = a(x)$, $\sigma(t,x) = \sigma(x)$. The operator $\mathcal{L}$ is also time independent now:

$$\mathcal{L}f(x) = \big(a(x), D_x f(x)\big) + \frac{1}{2}\mathrm{tr}\left(\sigma(x) D^2_{xx} f(x)\right),\ f \in C_b^2(\mathbb{R}^d).$$

Let us formulate a homogeneous counterpart of theorem 7.5.

THEOREM 7.6.– *A homogeneous Markov process X is a diffusion process if and only if for any $f \in C_b^2(\mathbb{R}^d)$ and all $x \in \mathbb{R}^d$,*

$$\frac{1}{h} \int_{\mathbb{R}^d} \big(f(y) - f(x)\big) P(h,x,dy) \to \mathcal{L}f(x),\ h \to 0+ . \qquad [7.30]$$

In the left-hand side of [7.30], we have exactly

$$\frac{1}{h}\big(T_h f(x) - f(x)\big),$$

where

$$T_t f(x) = \int_{\mathbb{R}^d} f(x) P(t,x,dy)$$

is the Markov semigroup corresponding to X; assume that the process X is *Feller*, i.e. the semigroup T_t is strongly continuous on $C_b(\mathbb{R}^d)$. Then, theorem 7.5 states that $\mathcal{L}$ is the generator of this semigroup.

Let us now consider the standard Wiener process $W = \{W_t,\ t \geq 0\}$ in $\mathbb{R}^d$. It has independent increments, so it is a Markov process, as shown above. Since the increment $W_t - W_s$ has the normal distribution $\mathcal{N}(0,(t-s)E_d)$, the transition probability density is the density of that distribution, i.e.

$$p(s,x,t,y) = p(t-s,x,y) = \frac{1}{\big(2\pi(t-s)\big)^{d/2}} \exp\left\{-\frac{(x-y)^2}{2(t-s)}\right\}.$$

Let us check the assumptions of proposition 7.3:

$$\frac{1}{h}\int_{\mathbb{R}^d} |x-y|^{2+\delta} p(h,x,y)dy = \frac{1}{h}\mathrm{E}|W_h|^{2+\delta} = C_\delta h^{\delta/2} \to 0,\ h \to 0+,$$
$$\frac{1}{h}\int_{\mathbb{R}^d} (x-y)p(h,x,y) = \frac{1}{h}\mathrm{E}W_h = 0,$$
$$\frac{1}{h}\int_{\mathbb{R}^d} \big(x-y,z\big)^2 p(h,x,y) = \frac{1}{h}\mathrm{E}\big(W_h,z\big)^2 = |z|^2.$$

As a result, W is a diffusion process with zero drift, which means that the evolution is purely random, and an identity diffusion matrix, which means that the process is isotropic, i.e. its properties are the same in all directions. The generator of the Wiener process is

$$\mathcal{L}^W f(x) = \frac{1}{2}\sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2} f(x) =: \frac{1}{2}\Delta f(x),$$

the Laplace operator times $1/2$. Conversely, if X is a diffusion process with zero drift and a unit diffusion matrix, then X is a Wiener process; this follows from the unique solvability of the Kolmogorov equation, which is the subject of the following section.

7.6. Kolmogorov equations for diffusions

Recall that Kolmogorov equations for a general homogeneous Markov process are

$$\frac{d}{dt}T_t f(x) = AT_t f(x) \quad \text{(backward)},$$
$$\frac{d}{dt}T_t f(x) = T_t Af(x) \quad \text{(forward)}.$$

For diffusion processes, $A = \mathcal{L}$ is a differential operator, so the Kolmogorov equations are partial differential equations. Let us give precise formulations.

For $\mathbb{T} \subset \mathbb{R}_+$, we will say that $u \in C^{0,2}(\mathbb{T} \times \mathbb{R}^d)$ if it is continuous and has continuous derivatives $\frac{\partial}{\partial x_i} u(t,x)$, $\frac{\partial^2}{\partial x_i \partial x_j} u(t,x)$, $i, j = 1, \ldots, d$, on $\mathbb{T} \times \mathbb{R}^d$; we will also denote by $C_b(\mathbb{R}^d)$ the set of bounded continuous functions.

THEOREM 7.7.– *Let a homogeneous diffusion process X have continuous drift $a(x)$ and diffusion matrix $\sigma(x)$. Also, let $g \in C_b(\mathbb{R})$ be such that the function*

$$u(t,x) = \int_{\mathbb{R}^d} g(y) P(T-t, x, dy)$$

belongs to $C^{0,2}([0,T) \times \mathbb{R}^d)$ for some $T > 0$. Then, u satisfies the backward Kolmogorov equation

$$\frac{\partial}{\partial t} u(t,x) + \mathcal{L} u(t,x) = 0, \; t \in [0,T), x \in \mathbb{R}^d. \qquad [7.31]$$

REMARK 7.4.– The formulation is designed to conform with the inhomogeneous case, where we can define

$$u(t,x) = \int_{\mathbb{R}^d} g(y) P(t, x, T, dy),$$

and show, under the same assumptions, that $\frac{\partial}{\partial t} u(t,x) + \mathcal{L}_t u(t,x) = 0$. In addition to those assumptions, if u is continuous at point T, then it is a classical solution to the Cauchy problem

$$\frac{\partial}{\partial t} u(t,x) + \mathcal{L}_t u(t,x) = 0, \; t \in [0,T), x \in \mathbb{R}^d$$
$$u(T,x) = g(x).$$

Now the term "backward" becomes clear: this equation describes an evolution of systems *backwards in time*, starting from time T.

PROOF.– Note from the Chapman–Kolmogorov equation [7.10] that

$$\begin{aligned} u(t,x) &= \int_{\mathbb{R}^d} f(y) P(T-t, x, dy) = \int_{\mathbb{R}^d} f(y) \int_{\mathbb{R}^d} P(h, x, dz) P(T-t-h, z, dy) \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(y) P(T-t-h, z, dy) P(h, x, dz) = \int_{\mathbb{R}^d} u(t+h, z) P(h, x, dz), \end{aligned}$$

so by theorem 7.6, taking into account that $\int_{\mathbb{R}^d} P(h,x,dy) = 1$,

$$\begin{aligned} \frac{1}{h}\big(u(t,x) - u(t+h,x)\big) &= \frac{1}{h} \int_{\mathbb{R}^d} \big(u(t+h,y) - u(t+h,x)\big) P(h,x,dy) \\ &= \mathcal{L} u(t+h, x) + o(1). \end{aligned}$$

It is clear from its proof that the remainder term depends on the moduli of continuity of the second derivatives of $u(t+h,\cdot)$ in a small neighborhood of x. Therefore, in view of the continuity of those derivatives and of $\mathcal{L}u(\cdot,x)$, we arrive at

$$\frac{\partial_+}{\partial t}u(t,x) = -\mathcal{L}u(t,x),$$

where $\frac{\partial_+}{\partial t}$ denotes the right derivative. Since the right-hand side of the last equality is continuous in t, the left-hand side is continuous as well, so u is continuously differentiable in t and

$$\frac{\partial}{\partial t}u(t,x) = -\mathcal{L}u(t,x),$$

as required. □

Now assume that there exists the transition probability density $p(t,x,y)$, i.e.

$$P(t,x,A) = \int_A p(t,x,y)dy.$$

Then, for any $T > t > 0$ and $g \in C_b(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} g(y)P(T-t,x,dy) = \int_{\mathbb{R}^d} p(T-t,x,y)g(y)dy,$$

so in view of theorem 7.5, p, as a function of t,x,y, is a fundamental solution of [7.31]. This gives a good chance that p is itself a solution to this equation. Let us formulate the corresponding result.

THEOREM 7.8.– *We assume that the diffusion process X has a transition probability density p satisfying* [7.11] *for all $s,t \in \mathbb{R}^+$ and $x,y \in \mathbb{R}^d$. Let also for any $y \in \mathbb{R}^d$, $p(\cdot,\cdot,y) \in C^{0,2}((0,\infty)\times\mathbb{R}^d)$ and $p(t,x,y)$ is bounded in x. Then, for any $t \in \mathbb{R}^+$ and $x,y \in \mathbb{R}^d$, p satisfies the backward Kolmogorov equation*

$$\frac{\partial}{\partial t}p(t,x,y) = \mathcal{L}_x p(t,x,y). \qquad [7.32]$$

REMARK 7.5.– The symbol $\mathcal{L}_x$ is used to emphasize the fact that the operator $\mathcal{L}$ acts on $p(t,x,y)$ with respect to the argument x. The reason for the sign change is that we now write the equation for $p(t,x,y)$, not for $p(T-t,x,y)$ as in theorem 7.7.

PROOF.– Using [7.9], we write for some $T > 0$, $s \in (0,T)$, and $u > 0$,

$$u(s,x) := p(T-s+u,x,y) = \int_{\mathbb{R}^d} p(T-s,x,z)p(u,z,y)dz.$$

Denoting $g(z) = p(u, z, y)$, we get from theorem 7.7 that

$$\frac{\partial}{\partial s} u(s, x) + \mathcal{L} u(s, x) = 0,$$

which transforms into [7.32] by the variable change $t = T - s + u$. □

Let us now turn to the forward Kolmogorov equation. Let μ be a σ-finite measure on $\mathbb{R}^d$. If it is a probability measure, then it can be understood as the probabilistic distribution of initial condition $X(0)$ of the underlying diffusion process and we are interested in the evolution of the distribution of $X(t)$. More generally, we can think of μ as initial distribution (in the physical sense) of some substance, which further diffuses according to the drift and the diffusion matrix, and we are interested in the evolution of this mass in time:

$$\mu_t(A) = \int_{\mathbb{R}^d} P(t, x, A) \mu(dx).$$

If μ and P have densities m and p, respectively, we can also look at the evolution of densities

$$v(t, y) = \int_{\mathbb{R}^d} p(t, x, y) m(x) dx. \qquad [7.33]$$

In contrast to the "backward evolution" considered before, in general, this expression can be ill-defined even for good m, say, continuous and bounded. Further, we study this evolution.

Assume that the drift coefficient a is continuously differentiable, the diffusion matrix b is twice continuously differentiable and consider the adjoint to the $\mathcal{L}$ operator:

$$\mathcal{L}^* m(x) = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i} \big(a_i(x) m(x)\big) + \frac{1}{2} \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} \big(\sigma_{ij}(x) m(x)\big).$$

THEOREM 7.9.– *Let $a \in C^1(\mathbb{R}^d)$, $\sigma \in C^2(\mathbb{R}^d)$ and the convergence in* [7.29] *be uniform in x.*

1) If the transition probability density $p(t, x, y)$ has continuous derivatives $\frac{\partial}{\partial t} p(t, x, y)$ and $\frac{\partial^2}{\partial y_i \partial y_j} p(t, x, y)$, $i, j = 1, \ldots, d$, then it satisfies the forward Kolmogorov equation

$$\frac{\partial}{\partial t} p(t, x, y) = \mathcal{L}_y^* p(t, x, y).$$

2) Let the transition probability density $p(t, x, y)$ have continuous derivative $\frac{\partial}{\partial t}p(t, x, y)$, and m be such that the function defined by [7.33] *has continuous derivatives $\frac{\partial}{\partial t}v(t, y)$ and $\frac{\partial^2}{\partial y_i \partial y_j}v(t, y)$, $i, j = 1, \ldots, d$, and $\frac{\partial}{\partial t}v(t, y) = \int_{\mathbb{R}^d} m(x)\frac{\partial}{\partial t}p(t, x, y)dx$. Then, v satisfies the forward Kolmogorov equation*

$$\frac{\partial}{\partial t}v(t, y) = \mathcal{L}^* v(t, y).$$

PROOF.– Take arbitrary $f \in C^2_{fin}(\mathbb{R}^d)$, i.e. a twice continuously differentiable function with compact support. We will first argue that f is in the domain of the infinitesimal generator $\mathcal{L}$. Indeed, inspecting the proof of theorem 7.5, we have

$$\frac{1}{h}\int_{\mathbb{R}^d} \big(f(y) - f(x)\big)P(h, x, dy) = \mathcal{L}f(x) + o(1),\ h \to 0+,$$

where the remainder term depends on $\sup|f|$, the moduli of continuity of second derivatives of f and the speed of convergence in [7.29]. It then follows from the assumption of the theorem that the convergence

$$\frac{1}{h}\int_{\mathbb{R}^d} \big(f(y) - f(x)\big)P(h, x, dy) \to \mathcal{L}f(x),\ h \to 0+,$$

is uniform. Therefore, f is indeed in the domain of $\mathcal{L}$ and

$$\frac{d}{dt}T_t f = T_t \mathcal{L} f.$$

Rewriting in terms of the transition probability density,

$$\frac{\partial}{\partial t}\int_{\mathbb{R}^d} p(t, x, y)f(y)dy = \int_{\mathbb{R}^d} p(t, x, y)\mathcal{L}f(y)dy$$

Since f has bounded support and $\frac{\partial}{\partial t}p(t, x, y)$ is continuous, we have

$$\frac{\partial}{\partial t}\int_{\mathbb{R}^d} p(t, x, y)f(y)dy = \int_{\mathbb{R}^d} f(y)\frac{\partial}{\partial t}p(t, x, y)dy.$$

Recall that

$$\mathcal{L}f(y) = \sum_{i=1}^d a_i(y)\frac{\partial}{\partial y_i}f(y) + \frac{1}{2}\sum_{i,j=1}^d \sigma_{ij}(y)\frac{\partial^2}{\partial y_i \partial y_j}f(y),$$

so

$$\int_{\mathbb{R}^d} p(t,x,y)\mathcal{L}f(y)dy = \sum_{i=1}^{d} \int_{\mathbb{R}^d} p(t,x,y)a_i(y)\frac{\partial}{\partial y_i}f(y)dy$$
$$+ \frac{1}{2}\sum_{i,j=1}^{d} \int_{\mathbb{R}^d} p(t,x,y)\sigma_{ij}(y)\frac{\partial^2}{\partial y_i \partial y_j}f(y)dy.$$

Integrating by parts with respect to y_i and recalling that f has compact support, we get

$$\int_{\mathbb{R}^d} p(t,x,y)a_i(y)\frac{\partial}{\partial y_i}f(y)dy = -\int_{\mathbb{R}^d} f(y)\frac{\partial}{\partial y_i}\big(p(t,x,y)a_i(y)\big)dy.$$

Similarly,

$$\int_{\mathbb{R}^d} p(t,x,y)\sigma_{ij}(y)\frac{\partial^2}{\partial y_i \partial y_j}f(y)dy = \int_{\mathbb{R}^d} f(y)\frac{\partial^2}{\partial y_i \partial y_j}\big(p(t,x,y)\sigma_{ij}(y)\big)dy.$$

Therefore,

$$\int_{\mathbb{R}^d} f(y)\frac{\partial}{\partial t}p(t,x,y)dy = -\sum_{i=1}^{d} \int_{\mathbb{R}^d} f(y)\frac{\partial}{\partial y_i}\big(p(t,x,y)a_i(y)\big)dy$$
$$+\frac{1}{2}\sum_{i,j=1}^{d} \int_{\mathbb{R}^d} f(y)\frac{\partial^2}{\partial y_i \partial y_j}\big(p(t,x,y)\sigma_{ij}(y)\big)dy = \int_{\mathbb{R}^d} f(y)\mathcal{L}_y^* p(t,x,y)dy.$$

From the arbitrariness of f and continuity of p, we get the first statement. The second statement is proved in the same way. □

REMARK 7.6.– A distribution μ (not necessarily probabilistic) is called *invariant* for a diffusion process X if the evolution of X does not alter it, i.e. $\mu_t = \mu$. If an invariant distribution has a density m, then in view of invariance:

$$m(y) = \int_{\mathbb{R}^d} p(t,x,y)m(x)dx,$$

so the function v defined by [7.33] is independent of t, namely $v(t,x) = m(x)$. Consequently, if all assumptions of theorem 7.9 are satisfied, then this density solves a second-order ordinary differential equation

$$\mathcal{L}^* m(x) = 0.$$

For the Wiener process, this transforms to $\Delta m(x) = 0$. The only positive function solving this equation is constant, so all invariant measures are proportional to the Lebesgue measure; in particular, for the Wiener process, there is no invariant probability distribution.

8

Stochastic Integration

8.1. Motivation

Consider a dynamical system with state space $\mathbb{R}^d$, which evolves under the influence of external forces. The dynamics of such a system can be described by a system of differential equations (or simply a differential equation in $\mathbb{R}^d$)

$$\dot{X}(t) = f(X(t)),\ t \geq 0,$$

where X is the current state of the system, $\dot{X}$ denotes the time derivative and $f\colon \mathbb{R}^d \to \mathbb{R}^d$ is the function measuring the external influence at the point x. This is a very popular mathematical model to describe a deterministic evolution. However, as it was mentioned in the Introduction, no evolution can be totally deterministic. The simplest model of perturbation is perhaps the *white noise*, $\eta(t)$, a collection of independent identically distributed random variables indexed by time $t \geq 0$. A slightly more advanced idea, which leads to a much greater versatility of models, is to allow the amplitude of noise to depend on the state of the system, which leads to the following equation modeling a behavior of a *random dynamical system*:

$$\dot{X}(t) = f(X(t)) + g(X(t))\eta(t),\ t \geq 0, \qquad [8.1]$$

where g is a deterministic function measuring the amplitude of noise and η is the white noise. Depending on the particular application, there can be different choices of distribution for η, but the most popular model is Gaussian white noise. Now comes the bad news: if we assume that $\eta(t)$ are independent Gaussian random variables $\mathcal{N}(0, \sigma^2)$, then η, as a function of t, is a very ill-behaved object: the integral $\int_0^T \eta(t)dt$ is not well defined. Indeed, if it were, by lemma A2.5, its value would be a Gaussian

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

random variable; in particular, it would be square integrable. However, by the Fubini theorem,

$$\mathrm{E}\left(\int_0^T \eta(s)ds\right)^2 = \int_0^T \int_0^T \mathrm{E}\eta(t)\eta(s)dt\, ds = 0, \qquad [8.2]$$

whence $\int_0^T \eta(t)dt = 0$ almost surely, which is absurd. So even in the simple case where $f \equiv 0$ and $g \equiv 1$, it is hard to say what X is. The most feasible way is to assume the independence only for variables which are on distance at least ε and to let $\varepsilon \to 0$. If the variance of η is bounded, then through an argument similar to [8.2], we arrive at the boring conclusion $X(t) = X(0)$. This means that in order to get a non-trivial evolution, the variance should be unbounded. Therefore, taking into account the independence, we arrive at the following desired covariance:

$$\mathrm{E}\big(\eta(t)\eta(s)\big) = \delta(t-s),$$

where $\delta = \delta(x)$ is the so-called Dirac delta function, which vanishes for $x \neq 0$ but integrates to 1. Such rather misty discussion may scare off some mathematical purists, but not physicists. Moreover, surprisingly, it has, in a sense, much better properties than the white noise with finite variance. Let us return to the simple case $f \equiv 0$ and $g \equiv 1$, where the solution can be written "explicitly" as

$$X(t) = X(0) + \int_0^t \eta(s)ds =: X(0) + H(t).$$

To identify the integral, note that $\mathrm{E}H(t) = \mathrm{E}\int_0^t \eta(s)ds = \int_0^t \mathrm{E}\eta(s)ds = 0$ and for $s \le t$,

$$\begin{aligned}\mathrm{E}\big(H(s)H(t)\big) &= \mathrm{E}\left(\int_0^s \eta(u)du \int_0^t \eta(v)dv\right) = \int_0^s \int_0^t \mathrm{E}\big(\eta(u)\eta(v)\big)dv\, du \\ &= \int_0^s \int_0^t \delta(u-v)dv\, du = \int_0^s du = s.\end{aligned}$$

Thanks to symmetry,

$$\mathrm{E}\big(H(s)H(t)\big) = t \wedge s,$$

the covariance function of standard Wiener process W. Substituting this into [8.1], we get

$$\dot{X}(t) = f(X(t)) + g(X(t))\dot{W}(t),\ t \ge 0.$$

However, we know from theorem 4.3 that the Wiener process is nowhere differentiable. Nevertheless, we can write this equation in an integral form

$$X(t) = X(0) + \int_0^t f(X(s))ds + \int_0^t g(X(s))dW(s). \quad [8.3]$$

Now it boils down to defining an integral with respect to W. This was already discussed in section 4.3, but for deterministic integrands. The following chapter will be devoted to the construction of such an integral for random integrands and to the investigation of its properties.

8.2. Definition of Itô integral

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \geq 0\}, \mathrm{P})$ be a stochastic basis with the filtration satisfying standard assumptions from section 5.1. We assume that $\{W(t), t \geq 0\}$ is a Wiener process on this basis, which means that W is adapted to the filtration $\{\mathcal{F}_t, t \geq 0\}$ and for any $t > s \geq 0$ the increment $W(t) - W(s)$ is independent of $\mathcal{F}_s$.

REMARK 8.1.– We may restrict ourselves to the case where $\mathcal{F}_t = \mathcal{F}_t^W$, i.e. the filtration is generated by the Wiener process, then the latter assumption is clearly satisfied. However, this filtration might not satisfy the standard assumptions from section 5.1 (e.g. $\mathcal{F}_0$ is trivial, since $W_0 = 0$). However, this is not the only reason to consider integration in a more general setting. Much more important is the necessity of considering multi-dimensional integrals later.

In order to define the class of admissible integrands, we need to recall a notion of progressively measurable process.

DEFINITION 8.1.– *Let* $\mathbb{T} \subset \mathbb{R}^+$ *be a parametric set. A stochastic process* $\{X(t), t \in \mathbb{T}\}$ *with values in a measurable space* $(\mathcal{S}, \Sigma)$ *is called progressively measurable if for any* $t > 0$ *and* $B \in \Sigma$

$$\{(s, \omega) \in (\mathbb{T} \cap [0, t]) \times \Omega : X(s, \omega) \in B\} \in \mathcal{F}_t \otimes \mathcal{B}([0, t]),$$

where $\mathcal{B}([0, t])$ *is the Borel* σ*-algebra on* $[0, t]$.

REMARK 8.2.– In layman's terms, a progressively measurable process is an adapted process jointly measurable in (t, ω). Measurability in t is now important, as otherwise it is impossible to define even the deterministic integral $\int X(t)dt$.

It is worth mentioning, see e.g. [DEL 78, theorem 39], that an adapted *measurable* process (i.e. such that the map $X\colon \mathbb{R}^+ \times \Omega \to \mathcal{S}$ is measurable) has a progressively measurable modification. So, in a sense, the above "layman's" definition turns out to be quite close to reality.

Now, for $a, b \in \mathbb{R}^+$, $a < b$, we introduce the class $\mathcal{H}_2([a,b])$ of real-valued processes $\{\xi(t), t \in [a,b]\}$ such that:

– ξ is progressively measurable;

– $\|\xi\|^2_{\mathcal{H}_2([a,b])} := \int_a^b \mathrm{E}\xi(t)^2 dt < \infty$.

Provided that we identify indistinguishable processes, this space can be regarded as a Banach or even Hilbert space. Because of the progressive measurability requirement, there are some hidden rocks on this path, for example, when proving completeness. We put aside this subtle matter, referring an interested reader to [DEL 78].

It is natural to include to $\mathcal{H}_2([a,b])$ also the processes defined on a larger interval, so that, in particular, $\mathcal{H}_2([a,b]) \subset \mathcal{H}_2([c,d])$ whenever $[a,b] \subset [c,d]$.

The construction of integral will follow the same scheme that was used for the Wiener integral. Let us first consider simple processes of the form

$$\eta(t) = \sum_{k=1}^{n} \alpha_k \mathbb{1}_{[a_k,b_k)}(t), \qquad [8.4]$$

where $n \geq 1$ is an integer, $a \leq a_k < b_k \leq b$ are some real numbers and α_k is an $\mathcal{F}_{a_k}$-measurable square-integrable random variable. Clearly, $\eta \in \mathcal{H}_2([a,b])$. Define *Itô integral*, or stochastic integral, of η with respect to W as

$$\int_a^b \eta(t)dW(t) = \sum_{k=1}^{n} \alpha_k \big(W(b_k) - W(a_k)\big).$$

For notation simplicity, we will also denote

$$I(\eta, W, [a,b]) = I(\eta, [a,b]) = \int_a^b \eta(t)dW_t.$$

It is evident that the value of the integral does not depend on the particular representation [8.4] of a simple process.

Further, we establish several properties of the Itô integral.

THEOREM 8.1.– *Let η, ζ be simple processes in $\mathcal{H}([a,b])$. Then, the following properties are true:*

1) $I(\eta + \zeta, [a,b]) = I(\eta, [a,b]) + I(\zeta, [a,b])$;

2) For any $c \in \mathbb{R}$, $I(c\eta, [a,b]) = cI(\eta, [a,b])$;

3) For any $c \in (a,b)$, $I(\eta,[a,b]) = I(\eta,[a,c]) + I(\eta,[c,b])$*;*

4) $\mathrm{E}I(\eta,[a,b]) = 0$*. Moreover,* $\{I(\eta,[a,t]), t \in [a,b]\}$ *is a martingale;*

5) $\mathrm{E}I(\eta,[a,b])^2 = \|\eta\|^2_{\mathcal{H}_2([a,b])} = \int_a^b \mathrm{E}\eta(t)^2 dt$*;*

6) $\mathrm{E}\big(I(\eta,[a,b])I(\zeta,[a,b]) \mid \mathcal{F}_a\big) = \int_a^b \mathrm{E}(\eta(t)\zeta(t) \mid \mathcal{F}_a)dt$*, in particular:*

$$\langle \eta, \zeta \rangle_{\mathcal{H}_2([a,b])} := \mathrm{E}\big(I(\eta,[a,b])I(\zeta,[a,b])\big) = \int_a^b \mathrm{E}(\eta(t)\zeta(t))dt.$$

REMARK 8.3.– Properties 1 and 2 mean that I is a linear operator on the set of simple functions. Properties 4 and 5 are counterparts of the corresponding properties of Wiener integral. However, note that in contrast to the latter, in general, Itô integral does not have Gaussian distribution.

Property 5 is the so-called *Itô isometry*: it says that I maps the family of simple functions (as a subspace of $\mathcal{H}_2([a,b])$) isometrically into a subspace of square-integrable random variables. This property will be crucial in extending I to the whole $\mathcal{H}_2([a,b])$.

PROOF.– Properties 1–3 are obvious from the definition.

To prove 4, assume that η is given by [8.4] and consider the conditional expectation:

$$\begin{aligned}\mathrm{E}(I(\eta,[a,b]) \mid \mathcal{F}_a) &= \sum_{k=1}^{n} \mathrm{E}\left(\alpha_k\big(W(b_k) - W(a_k)\big) \mid \mathcal{F}_a\right) \\ &= \sum_{k=1}^{n} \mathrm{E}\left(\alpha_k \mathrm{E}\big(\big(W(b_k) - W(a_k)\big) \mid \mathcal{F}_{a_k}\big) \mid \mathcal{F}_a\right) = 0,\end{aligned}$$

where we have used that α_k is $\mathcal{F}_{a_k}$-measurable and $W(b_k) - W(a_k)$ is independent of $\mathcal{F}_{a_k}$. It follows that $\mathrm{E}I(\eta,[a,b]) = 0$. Further, for any $t \in (a,b)$:

$$\mathrm{E}(I(\eta,[a,b]) \mid \mathcal{F}_t) = \mathrm{E}(I(\eta,[a,t]) \mid \mathcal{F}_t) + \mathrm{E}(I(\eta,[t,b]) \mid \mathcal{F}_t) = I(\eta,[a,t]), \quad [8.5]$$

since $I(\eta,[a,t])$ is clearly $\mathcal{F}_t$-measurable and $\mathrm{E}(I(\eta,[t,b]) \mid \mathcal{F}_t) = 0$. This implies the martingale property.

To prove 6 (5 would follow), first note that the both sides of equality are linear in η and ζ, so it is enough to prove it in the case where $\eta(t) = \alpha_1 \mathbb{1}_{[a_1,b_1)}(t)$, $\zeta(t) = \alpha_2 \mathbb{1}_{[a_2,b_2)}(t)$, where α_i is $\mathcal{F}_{a_i}$-measurable, $i = 1, 2$. In turn, when splitting the intervals, if necessary, into smaller parts and using the linearity again, it is sufficient to consider the cases $[a_1, b_1) = [a_2, b_2)$ and $[a_1, b_1) \cap [a_2, b_2) = \varnothing$. In the first case, recalling that $W(b_1) - W(a_1)$ is independent of $\mathcal{F}_{a_1}$, we get

$$\begin{aligned}
&\mathrm{E}\big(I(\eta,[a,b])I(\zeta,[a,b]) \mid \mathcal{F}_a\big) = \mathrm{E}\big(\alpha_1\alpha_2\big(W(b_1) - W(a_1)\big)^2 \mid \mathcal{F}_a\big) \\
&= \mathrm{E}\big(\alpha_1\alpha_2\mathrm{E}\big(\big(W(b_1) - W(a_1)\big)^2 \mid \mathcal{F}_{a_1}\big) \mid \mathcal{F}_a\big) = \mathrm{E}\big(\alpha_1\alpha_2(b_1 - a_1) \mid \mathcal{F}_a\big) \\
&= (b_1 - a_1)\mathrm{E}\big(\alpha_1\alpha_2 \mid \mathcal{F}_a\big) = \int_a^b \mathrm{E}\big(\eta(t)\zeta(t) \mid \mathcal{F}_a\big)dt.
\end{aligned}$$

In the second case, assuming that $a_2 > b_1$, we have that $\alpha_1\big(W(b_1) - W(a_1)\big)\alpha_2$ is $\mathcal{F}_{a_2}$-measurable, while $W(b_2) - W(a_2)$ is independent of $\mathcal{F}_{a_2}$, so

$$\begin{aligned}
&\mathrm{E}\big(I(\eta,[a,b])I(\zeta,[a,b]) \mid \mathcal{F}_a\big) \\
&= \mathrm{E}\big(\alpha_1\big(W(b_1) - W(a_1)\big)\alpha_2\big(W(b_2) - W(a_2)\big) \mid \mathcal{F}_a\big) \\
&= \mathrm{E}\big(\alpha_1\big(W(b_1) - W(a_1)\big)\alpha_2\,\mathrm{E}\big(W(b_2) - W(a_2) \mid \mathcal{F}_{a_2}\big) \mid \mathcal{F}_a\big) = 0 \\
&= \int_a^b \mathrm{E}\big(\eta(t)\zeta(t) \mid \mathcal{F}_a\big)dt.
\end{aligned}$$

This establishes the first equality; the second follows from the tower property of conditional expectation. □

To extend the definition from simple functions to the whole space, we need the following approximation result.

LEMMA 8.1.– *Let* $\xi \in \mathcal{H}_2([a,b])$. *Then, there exists a sequence* $\{\eta_n, n \geq 1\}$ *of simple processes such that*

$$\|\xi - \eta_n\|_{\mathcal{H}_2([a,b])} \to 0, n \to \infty.$$

PROOF.– It is evident that we can approximate an element of $\mathcal{H}_2([a,b])$ by bounded functions (e.g. by truncation), so it is enough to prove the lemma in the case where $\sup_{t\in[a,b],\omega\in\Omega} |\xi(t,\omega)| < \infty$.

Define a smooth approximation of ξ:

$$\xi_\varepsilon(t) = \varepsilon^{-1}\int_{(t-\varepsilon)\vee a}^{t} \xi(s)ds,\ t \in [a,b].$$

Due to boundedness, $\xi_\varepsilon \in \mathcal{H}_2([a,b])$. Moreover, from theorem A1.4, we have that $\int_a^b \big(\xi(t) - \xi_\varepsilon(t)\big)^2 dt \to 0$, $\varepsilon \to 0+$, for any $\omega \in \Omega$. Then, the bounded convergence theorem implies

$$\|\xi - \xi_\varepsilon\|_{\mathcal{H}_2([a,b])} \to 0,\ \varepsilon \to 0+ . \qquad [8.6]$$

Now define for integer $N \geq 1,\ t_k^N = a + (b-a)k/N$, $k = 0, \dots, n$ and

$$\xi_{\varepsilon,N}(t) = \sum_{k=1}^{N} \xi_\varepsilon(t_{k-1}^N) \mathbb{1}_{[t_{k-1}^N, t_k^N)}(t),\ t \in [a,b].$$

Due to the continuity of ξ_ε,

$$|\xi_{\varepsilon,N}(t) - \xi_\varepsilon(t)| \leq \sup_{\substack{u,s\in[a,b] \\ |u-s| \leq (b-a)/N}} |\xi_\varepsilon(u) - \xi_\varepsilon(s)| \to 0,\ N \to \infty.$$

Moreover, $|\xi_{\varepsilon,N}(t) - \xi_\varepsilon(t)| \leq 2\sup_{t\in[a,b],\omega\in\Omega} |\xi(t,\omega)|$, so the bounded convergence theorem gives $\|\xi_\varepsilon - \xi_{\varepsilon,N}\|_{\mathcal{H}_2([a,b])} \to 0$, $N \to \infty$. Since $\xi_{\varepsilon,N}$ is a simple process, we get the desired approximation, with η_n equal to $\xi_{\varepsilon,N}$ with appropriate ε and N, through [8.6] and the triangle inequality. □

With this at hand, the extension is done in a standard manner. Namely, if $\{\eta_n, n \geq 1\}$ is a sequence of simple processes converging in $\mathcal{H}_2([a,b])$ to $\xi \in \mathcal{H}_2([a,b])$, then, due to the isometry property, the sequence $\{I(\eta_n, [a,b]), n \geq 1\}$ is a Cauchy sequence in $\mathcal{L}_2(\Omega)$. Then, it has a limit in $\mathcal{L}_2(\Omega)$, which justifies the following definition.

DEFINITION 8.2.– *For $\xi \in \mathcal{H}_2([a,b])$,* Itô integral *of ξ with respect to the Wiener process is the limit*

$$I(\xi, [a,b]) = \int_a^b \xi(t)dW(t) = \lim_{n\to\infty} I(\eta_n, [a,b]) \qquad [8.7]$$

in $\mathcal{L}^2(\Omega)$, where $\{\eta_n, n \geq 1\}$ is a sequence of simple processes in $\mathcal{H}_2([a,b])$ such that $\|\xi - \eta_n\|_{\mathcal{H}_2([a,b])} \to 0$, $n \to \infty$.

It is clear from such a definition that the integral is defined modulo P-null sets. Moreover, the limit does not depend on the approximating sequence. Indeed, if $\{\zeta_n, n \geq 1\}$ is another approximating sequence, then

$$\lim_{n\to\infty} \mathrm{E}\big(I(\eta_n, [a,b]) - I(\zeta_n, [a,b])\big)^2 = \lim_{n\to\infty} \|\eta_n - \zeta_n\|^2_{\mathcal{H}_2([a,b])} = 0,$$

so the limits coincide.

The properties of the Itô integral defined by [8.7] are essentially the same as for simple functions. For completeness, we give them in full.

THEOREM 8.2.– *Let* $\eta, \zeta \in \mathcal{H}_2([a,b])$.

1) $I(\eta+\zeta,[a,b]) = I(\eta,[a,b]) + I(\zeta,[a,b])$ *almost surely;*

2) For any $c \in \mathbb{R}$, $I(c\eta,[a,b]) = cI(\eta,[a,b])$ *almost surely;*

3) For any $c \in (a,b)$, $I(\eta,[a,b]) = I(\eta,[a,c]) + I(\eta,[c,b])$ *almost surely;*

4) $\mathrm{E}I(\eta,[a,b]) = 0$. *Moreover,* $\{I(\eta,[a,t]), t \in [a,b]\}$ *is a martingale;*

5) $\mathrm{E}I(\eta,[a,b])^2 = \|\eta\|^2_{\mathcal{H}_2([a,b])} = \int_a^b \mathrm{E}\eta(t)^2 dt$ (*Itô isometry*). *Moreover, the process*

$$M(t) = I(\eta,[a,t])^2 - \int_a^t \eta(s)^2 ds,\ t \in [a,b],$$

is a martingale;

6) $\mathrm{E}\big(I(\eta,[a,b])I(\zeta,[a,b])\big) = \langle \eta,\zeta\rangle_{\mathcal{H}_2([a,b])}$.

PROOF.– Let $\{\eta_n, n \geq 1\}$ and $\{\zeta_n, n \geq 1\}$ be sequences of simple processes converging in $\mathcal{H}_2([a,b])$ to η and ζ, respectively.

To prove 1 and 2, note that $\eta_n + \zeta_n$ and $c\eta_n$ converge in $\mathcal{H}_2([a,b])$ to $\eta+\zeta$ and $c\eta$, respectively, whence the properties follow from those for simple functions.

Property 3 follows from 1, since clearly $I(\eta,[a,t]) = I(\eta\mathbb{1}_{[a,t)},[a,b])$ and $I(\eta,[t,b]) = I(\eta\mathbb{1}_{[t,b]},[a,b])$.

To prove property 4, first observe that

$$\begin{aligned}&\mathrm{E}\big(\mathrm{E}(I(\eta,[a,b]) \mid \mathcal{F}_t) - \mathrm{E}(I(\eta_n,[a,b]) \mid \mathcal{F}_t)\big)^2\\ &\leq \mathrm{E}\big(I(\eta,[a,b]) - I(\eta_n,[a,b])\big)^2 = \|\eta-\eta_n\|^2_{\mathcal{H}_2([a,b])}\end{aligned}$$

for any $t \in [a,b]$, so $\mathrm{E}(I(\eta_n,[a,b]) \mid \mathcal{F}_t) \to \mathrm{E}(I(\eta,[a,b]) \mid \mathcal{F}_t)$, $n \to \infty$, in $\mathcal{L}_2(\Omega)$. On the other hand, by [8.5], $\mathrm{E}(I(\eta_n,[a,b]) \mid \mathcal{F}_t) = I(\eta_n,[a,t]) \to I(\eta,[a,t])$, $n \to \infty$, since $\|\eta-\eta_n\|_{\mathcal{H}_2([a,t])} \leq \|\eta-\eta_n\|_{\mathcal{H}_2([a,b])} \to 0$, $n \to \infty$, whence

$$\mathrm{E}(I(\eta_n,[a,b]) \mid \mathcal{F}_t) \to I(\eta,[a,t]),\ n \to \infty,$$

in $\mathcal{L}_2(\Omega)$. Therefore, for any $t \in [a,b]$:

$$\mathrm{E}(I(\eta,[a,b]) \mid \mathcal{F}_t) = I(\eta,[a,t])$$

almost surely, implying the martingale property.

The isometric property follows from that for simple functions through the Minkowski inequalities:

$$\Big|\big(\mathrm{E}I(\eta,[a,b])^2\big)^{1/2} - \big(\mathrm{E}I(\eta_n,[a,b])^2\big)^{1/2}\Big|$$

$$\leq \Big(\mathrm{E}\big(I(\eta,[a,b]) - I(\eta_n,[a,b])\big)^2\Big)^{1/2} = \|\eta-\eta_n\|_{\mathcal{H}_2([a,b])};$$

$$\Big|\|\eta\|_{\mathcal{H}_2([a,b])} - \|\eta_n\|_{\mathcal{H}_2([a,b])}\Big| \leq \|\eta-\eta_n\|_{\mathcal{H}_2([a,b])}.$$

Let us now show the martingale property for M. Let $t \in [a,b]$. From theorem 8.1, we have

$$\mathrm{E}\big(I(\eta,[t,b])^2 \mid \mathcal{F}_t\big) = \mathrm{E}\left(\int_t^b \eta(s)^2 ds \,\middle|\, \mathcal{F}_t\right),$$

whence

$$\mathrm{E}(M(b) \mid \mathcal{F}_t) = \mathrm{E}\left(\big(I(\xi,[a,t]) + I(\xi,[t,b])\big)^2 \,\middle|\, \mathcal{F}_t\right)$$

$$-\int_a^t \xi(s)^2 ds - \mathrm{E}\left(\int_t^b \xi(s)^2 ds \,\middle|\, \mathcal{F}_t\right)$$

$$= I(\xi,[a,t])^2 + 2I(\xi,[a,t])\mathrm{E}\left(I(\xi,[t,b]) \,\middle|\, \mathcal{F}_t\right) - \int_a^t \xi(s)^2 ds = M(t),$$

since $\mathrm{E}\left(I(\xi,[t,b]) \,\middle|\, \mathcal{F}_t\right) = 0$.

The last property can be shown similarly to the isometric property or can be deduced from the latter by using the polarization identities. □

8.2.1. *Itô integral of Wiener process*

Let us compute the integral

$$I(W,[0,T]) = \int_0^T W(t)dW(t).$$

To this end, we have to approximate W by simple processes. The simplest way is probably to consider for $n \geq 1$ a uniform partition $t_k^n = kT/n$, $k = 0,\ldots,n$, and to set

$$\eta_n(t) = \sum_{k=1}^n W(t_{k-1}^n)\mathbb{1}_{[t_{k-1}^n,t_k^n)}(t).$$

Then

$$\|\eta_n - W\|^2_{\mathcal{H}_2([0,T])} = \sum_{k=1}^{n} \int_{t_{k-1}^n}^{t_k^n} \mathrm{E}\big(W(t) - W(t_{k-1}^n)\big)^2 dt$$

$$= \sum_{k=1}^{n} \int_{t_{k-1}^n}^{t_k^n} (t - t_{k-1}^n) dt = \frac{1}{2}\sum_{k=1}^{n}(t_k^n - t_{k-1}^n)^2 = \frac{T^2}{2n} \to 0,\ n \to \infty,$$

so $I(W, [0,T]) = \lim_{n\to\infty} I(\eta_n, [0,T])$ in $L^2(\Omega)$. Denote $\Delta_k^n W = W(t_k^n) - W(t_{k-1}^n)$ and write

$$2I(\eta_n, [0,T]) = 2\sum_{k=1}^{n} W(t_{k-1}^n)\Delta_k^n W$$

$$= \sum_{k=1}^{n} W(t_{k-1}^n)\Delta_k^n W + \sum_{k=1}^{n} W(t_k^n)\Delta_k^n W - \sum_{k=1}^{n} \left(\Delta_k^n W\right)^2$$

$$= \sum_{k=1}^{n} \left(W(t_k^n)^2 - W(t_{k-1}^n)^2\right) - \sum_{k=1}^{n} \left(\Delta_k^n W\right)^2$$

$$= W(t_n^n)^2 - \sum_{k=1}^{n} \left(\Delta_k^n W\right)^2 = W(T)^2 - \sum_{k=1}^{n} \left(\Delta_k^n W)\right)^2.$$

Consider

$$\mathrm{E}\left(\sum_{k=1}^{n} \left(\Delta_k^n W\right)^2 - T\right)^2 = \mathrm{E}\left(\sum_{k=1}^{n} \left(\left(\Delta_k^n W\right)^2 - \mathrm{E}\big(\Delta_k^n W\big)^2\right)\right)^2$$

$$= \sum_{k=1}^{n} \mathrm{E}\left(\left(\Delta_k^n W\right)^2 - \mathrm{E}\big(\Delta_k^n W\big)^2\right)^2 = \sum_{k=1}^{n} 2(t_k^n - t_{k-1}^n)^2$$

$$= \frac{2T^2}{n} \to 0,\ n \to \infty.$$

Consequently, $\sum_{k=1}^{n} \left(\Delta_k^n W\right)^2 \to T, n \to \infty$, in $\mathcal{L}_2(\Omega)$, whence

$$\int_0^T W(t)dW(t) = \frac{1}{2}\big(W(T)^2 - T\big),$$

not $W(T)^2/2$, as some might have expected. (Note that the answer $W(T)^2/2$ is impossible, as the Itô integral has zero mean.)

8.3. Continuity of Itô integral

Let $\{\xi(t), t \geq 0\}$ be an adapted process such that $\xi \in \mathcal{H}_2([0,t])$ for any $t > 0$; then for each $0 \leq s \leq t$, the Itô integral

$$I(\xi, [s,t]) = \int_s^t \xi(u) dW(u)$$

is well defined. However, we have seen above that the definition is up to sets of probability zero. These exceptional sets can be different for different values of s and t, so the integral may have some deficiencies as a function of s and t. For example, for any fixed $\omega \in \Omega$, it might happen that $I(\xi, [s,u]) + I(\xi, [u,t]) \neq I(\xi, [s,t])$ for the majority of values $s \leq u \leq t$, which is not a desired behavior.

For this reason, we will establish the existence of a nice modification of the Itô integral.

THEOREM 8.3.– *Let* $\{\xi(t), t \geq 0\}$ *be such that* $\xi \in \mathcal{H}_2([0,t])$ *for any* $t > 0$. *Then, there exists a modification* $\{\mathcal{I}(s,t), 0 \leq s \leq t\}$ *of* $\left\{\int_s^t \xi_u dW_u, 0 \leq s \leq t\right\}$ *such that:*

– $\mathcal{I}$ *is a continuous function of s and t,*

– *for all* $0 \leq s \leq u \leq t$,

$$\mathcal{I}(s,t) = \mathcal{I}(s,u) + \mathcal{I}(u,t).$$

PROOF.– Since the Wiener process has a continuous modification, we will assume that this is continuous.

Let us first construct a continuous modification $\{\mathcal{I}(t),\ t \geq 0\}$ of $\{I(\xi, [0,t]), t \geq 0\}$. Note that it is enough to construct a continuous modification $\mathcal{I}_N$ on $[0,N]$ for each integer $N \geq 1$. Indeed, once this is done, for any $N_2 > N_1$, the equality $\mathcal{I}_{N_1}(q) = \mathcal{I}_{N_2}(q)$ holds almost surely for all $q \in [0, N_1] \cap \mathbb{Q}$, hence

$$\mathrm{P}\{\Omega_{N_1,N_2}\} := \mathrm{P}\{\mathcal{I}_{N_1}(t) = \mathcal{I}_{N_2}(t) \text{ for all } t \in [0,N_1]\} = 1.$$

Then, $\mathrm{P}\left\{\bigcap_{1 \leq N_1 < N_2} \Omega_{N_1,N_2}\right\} = 1$, as the intersection is taken over a countable set of indices. Therefore, setting $\mathcal{I}(t) = \sum_{N=1}^{\infty} \mathcal{I}_N(t) \mathbb{1}_{[N-1,N)}(t)$ for $\omega \in \bigcap_{1 \leq N_1 < N_2} \Omega_{N_1,N_2}$ and $\mathcal{I}(t) = 0$ otherwise yields the desired modification continuous on $\mathbb{R}_+$.

Now let $N \geq 1$ be fixed and $\{\eta_n, n \geq 1\}$ be a sequence of simple processes such that $\|\xi - \eta_n\|_{\mathcal{H}_2([0,N])} \to 0$, $n \to \infty$. Then there is a subsequence $\{\eta_{n_k}, k \geq 1\}$ such that $\left\|\eta_{n_k} - \eta_{n_{k+1}}\right\|_{\mathcal{H}_2([0,N])} \leq 2^{-2k}$ for any $k \geq 1$. The process

$$M_k(t) = I(\eta_{n_k} - \eta_{n_{k+1}}, [0,t])$$

is a martingale; moreover, it is easy to check from the definition of the Itô integral for simple functions that M_k is continuous in t. Hence, using theorem 5.26, we get

$$\mathrm{P}\left\{\sup_{t\in[0,N]}|M_k(t)|\geq 2^{-k}\right\}\leq 2^{2k}\mathrm{E}M_k(N)^2$$
$$=2^{2k}\left\|\eta_{n_k}-\eta_{n_{k+1}}\right\|^2_{\mathcal{H}_2([0,N])}\leq 2^{-2k}.$$

Observe that, by definition, $\mathcal{J}_n(t)=I(\eta_n,[0,t])$ is continuous in t for each $n\geq 1$. Thanks to linearity of Itô integral, the last inequality can be rewritten as

$$\mathrm{P}\left\{\sup_{t\in[0,N]}\left|\mathcal{J}_{n_k}(t)-\mathcal{J}_{n_{k+1}}(t)\right|\geq 2^{-k}\right\}\leq 2^{-2k}.$$

Then, the Borel–Cantelli lemma implies that with probability 1, there exists some $k_0=k_0(\omega)$ such that $\sup_{t\in[0,N]}\left|\mathcal{J}_{n_k}(t)-\mathcal{J}_{n_{k+1}}(t)\right|<2^{-k}$ for all $k\geq k_0$. Therefore, the event

$$A:=\left\{\sum_{k=1}^{\infty}\sup_{t\in[0,N]}\left|\mathcal{J}_{n_k}(t)-\mathcal{J}_{n_{k+1}}(t)\right|<\infty\right\}$$

has probability 1. It is easy to see that the sequence $\{\mathcal{I}_{n_k},k\geq 1\}$ is a Cauchy sequence in $C[0,N]$ with respect to the uniform norm for $\omega\in A$. Therefore, for each $\omega\in A$, there exists a continuous process $\{\mathcal{I}(t),t\in[0,N]\}$ such that $\mathcal{J}_{n_k}\to\mathcal{I}$ in $C[0,N]$. Set $\mathcal{I}(t)=0$ for $\omega\in\Omega\setminus A$. To conclude, we need to show for any $t\in[0,N]$ that $\mathcal{I}(t)=\mathcal{I}(\xi,[0,t])$ almost surely. Since $\|\xi-\eta_{n_k}\|_{\mathcal{H}_2([0,t])}\to 0$, $k\to\infty$, we have that $\mathcal{J}_{n_k}(t)\to I(\xi,[0,t])$ in $L_2(\Omega)$. Therefore, there is a subsequence converging almost surely. However, we also know that $\mathcal{J}_{n_k}(t)\to\mathcal{I}(t)$, $k\to\infty$ almost surely. Hence, $\mathcal{I}(t)=I(\xi,[0,t])$ almost surely, which yields the desired continuous modification of $I(\xi,[0,t])$.

Now set $\mathcal{I}(s,t)=\mathcal{I}(t)-\mathcal{I}(s)$. Then, $\mathcal{I}(s,t)$ is continuous as well, and

$$\mathcal{I}(s,t)=I(\xi,[0,t])-I(\xi,[0,s])=I(\xi,[s,t])$$

almost surely thanks to the additivity property of Itô integral. Therefore, $\mathcal{I}(s,t)$ is a modification of $I(\xi,[s,t])$, as required. □

REMARK 8.4.– In the following, we will always assume that the Itô integral is additive (i.e. $I(\xi,[s,u])+I(\xi,[u,t])=I(\xi,[s,t])$ for all $0\leq s\leq u\leq t$) and continuous. We should be careful here, as the exceptional set may depend on the integrand. However, since we will always be dealing with at most countable families of functions, we are on the safe side with this assumption.

8.4. Extended Itô integral

It turns out that the Itô integral can be naturally extended to a larger class of integrands. In order to proceed, we need to prove the "locality" property of the Itô integral, which will also be useful in the sequel.

THEOREM 8.4.– *Let* $\xi \in \mathcal{H}_2([0,T])$*. Then, for any stopping time* τ*,* $\xi\mathbb{1}_{[0,\tau]} \in \mathcal{H}_2([0,T])$ *and*

$$\int_0^T \xi(t)\mathbb{1}_{[0,\tau]}(t)dW(t) = \int_0^{\tau\wedge T} \xi(t)dW(t)$$

almost surely.

REMARK 8.5.– The right-hand side of this formula should be understood as $I(\xi,[0,t])|_{t=\tau\wedge T}$.

The formula is not true when τ is not a stopping time; moreover, its left-hand side is not well defined in general.

PROOF.– It is clear that $\left\|\xi\mathbb{1}_{[0,\tau]}\right\|_{\mathcal{H}_2([0,T])} \le \|\xi\|_{\mathcal{H}_2([0,T])} < \infty$, so we need to show the progressive measurability. The process ξ is progressively measurable, so it suffices to show that $\mathbb{1}_{[0,\tau]}$ is. For any $x \in (0,1]$,

$$\begin{aligned} A_x &:= \left\{(\omega,s)\in\Omega\times[0,t] : \mathbb{1}_{[0,\tau]}(s) < x\right\} = \{(\omega,s)\in\Omega\times[0,t] : \tau(\omega) > s\} \\ &= \bigcup_{q\in(0,t)\cap\mathbb{Q}} \{\omega\in\Omega : \tau(\omega) > q\}\times(q,t] \in \mathcal{F}_t\otimes\mathcal{B}([0,t]), \end{aligned}$$

since $\{\omega\in\Omega : \tau(\omega) > q\} \in \mathcal{F}_q \subset \mathcal{F}_t$. For $x \le 0$, $A_x = \varnothing$ and for $x > 1$, $A_x = \Omega\times[0,t]$. As a result, $\mathbb{1}_{[0,\tau]}$ is progressively measurable.

To prove the desired equality, start by observing that

$$\xi(t)\mathbb{1}_{[0,\tau]}(t) = \xi(t)\mathbb{1}_{[0,\tau\wedge T]}(t),\ t\in[0,T],$$

so $I(\xi\mathbb{1}_{[0,\tau]},[0,T]) = I(\xi\mathbb{1}_{[0,\tau\wedge T]},[0,T])$. Therefore, we can assume that τ takes values in $[0,T]$.

Let us first consider a situation where the integrand is a simple process of the form

$$\eta(t) = \sum_{k=1}^{n} \alpha_k\mathbb{1}_{[t_{k-1},t_k)}(t),$$

where $t_0 = 0 < t_1 < \cdots < t_{n-1} < t_n = T$, α_k is $\mathcal{F}_{t_{k-1}}$-measurable, and τ takes values from the set $\{t_0, t_1, \ldots, t_n\}$. Then,

$$\eta(t)\mathbb{1}_{[0,\tau)}(t) = \sum_{k=1}^{n} \alpha_k \mathbb{1}_{[t_{k-1},t_k)}(t)\mathbb{1}_{[0,\tau)}(t) = \sum_{k=1}^{n} \alpha'_k \mathbb{1}_{[t_{k-1},t_k)}(t),$$

where $\alpha'_k = \alpha_k \mathbb{1}_{\tau > t_{k-1}}$ is $\mathcal{F}_{t_{k-1}}$-measurable, since τ is a stopping time. By the definition of the Itô integral,

$$\begin{aligned}
\int_0^T \eta(t)\mathbb{1}_{[0,\tau)}(t)dW(t) &= \sum_{k=1}^{n} \alpha'_k \big(W(t_k) - W(t_{k-1})\big) \\
&= \sum_{k=1}^{n} \alpha_k \big(W(t_k) - W(t_{k-1})\big)\mathbb{1}_{\tau > t_{k-1}} \\
&= \sum_{k=1}^{n} \alpha_k \big(W(t_k) - W(t_{k-1})\big) \sum_{j=k}^{n} \mathbb{1}_{\tau = t_j} \\
&= \sum_{j=1}^{n} \left(\sum_{k=1}^{j} \alpha_k \big(W(t_k) - W(t_{k-1})\big) \right) \mathbb{1}_{\tau = t_j} \\
&= \sum_{j=1}^{n} \left(\int_0^{t_j} \eta(t)dW(t) \right) \mathbb{1}_{\tau = t_j} = \int_0^{\tau} \eta(t)dW(t).
\end{aligned}$$

Now consider dyadic partitions $t_k^n = k2^{-n}T$, $k = 0, \ldots, 2^n$ and let $\tau \in \{t_k^m, k = 0, \ldots, 2^m\}$ for fixed $m \geq 1$. Take arbitrary $\xi \in \mathcal{H}_2([0,T])$ and approximate it by processes of the form

$$\eta_n(t) = \sum_{k=1}^{2^n} \alpha_k^n \mathbb{1}_{[t_{k-1}^n, t_k^n)}(t),$$

where α_k^n is $\mathcal{F}_{t_{k-1}^n}$-measurable so that $\|\xi - \eta_n\|_{\mathcal{H}_2([0,T])} \to 0$, $n \to \infty$. For any $n \geq m$, we have

$$\begin{aligned}
\int_0^T \eta_n(t)\mathbb{1}_{[0,\tau)}(t)dW(t) &= \int_0^{\tau} \eta_n(t)dW(t) \\
&= \sum_{j=1}^{m} \left(\int_0^{t_j} \eta_n(t)dW(t) \right) \mathbb{1}_{\tau = t_j}.
\end{aligned}$$

Since for each $j = 1, \dots, m$ we have $I(\eta_n, [0, t_j]) \to I(\xi, [0, t_j])$, $n \to \infty$, in $L_2(\Omega)$, it follows that

$$\int_0^T \eta_n(t) \mathbb{1}_{[0,\tau)}(t) dW(t) \to \sum_{j=1}^m \left(\int_0^{t_j} \xi(t) dW(t) \right) \mathbb{1}_{\tau = t_j}$$

$$= \int_0^\tau \xi(t) dW(t), \ n \to \infty,$$

in $L_2(\Omega)$. On the other hand, the obvious fact $\left\| \eta_n \mathbb{1}_{[0,\tau)} - \xi \mathbb{1}_{[0,\tau]} \right\|_{\mathcal{H}_2([0,T])} \to 0$, $n \to \infty$, implies that $I(\eta_n \mathbb{1}_{[0,\tau)}, [0,T]) \to I(\xi \mathbb{1}_{[0,\tau]}, [0,T])$, $n \to \infty$, in $L_2(\Omega)$, yielding $\int_0^\tau \xi(t) dW(t) = \int_0^T \xi(t) \mathbb{1}_{[0,\tau]}(t) dW(t)$ almost surely. Since $\tau \le T$, this is the desired equality.

Finally, let $\xi \in \mathcal{H}_2([0,T])$ and the stopping time $\tau \le T$ be arbitrary. Define for $n \ge 1$ $\tau_n = \sum_{k=1}^{2^n} t_k^n \mathbb{1}_{(t_{k-1}^n, t_k^n]}(\tau)$. By theorem 5.2, this discretized version of τ is a stopping time, too. From the previous paragraph, for each $n \ge 1$,

$$\int_0^T \xi(t) \mathbb{1}_{[0,\tau_n]}(t) dW(t) = \int_0^{\tau_n} \xi(t) dW(t).$$

Since $\tau_n \to \tau$, $n \to \infty$, the dominated convergence theorem yields

$$\left\| \xi \mathbb{1}_{[0,\tau]} - \xi \mathbb{1}_{[0,\tau_n]} \right\|_{\mathcal{H}_2([0,T])} \to 0, \ n \to \infty,$$

so $I(\xi \mathbb{1}_{[0,\tau_n]}, [0,T]) \to I(\xi \mathbb{1}_{[0,\tau]}, [0,T])$, $n \to \infty$, in $L_2(\Omega)$. On the other hand, $I(\xi \mathbb{1}_{[0,\tau_n]}, [0,T]) \to \int_0^\tau \xi(t) dW(t)$, $n \to \infty$, due to continuity of the Itô integral. The proof is now complete. □

Now let $\xi = \{\xi(t), t \in [0,T]\}$ be a progressively measurable process such that

$$\int_0^T \xi(t)^2 dt < \infty \text{ almost surely.}$$

We will denote the class of such processes by $\mathcal{H}([0,T])$; clearly, $\mathcal{H}_2([0,T]) \subset \mathcal{H}([0,T])$. Define

$$\tau_n = \inf \left\{ t \ge 0 : \int_0^t \xi(s)^2 ds \ge n \right\} \wedge T, \ n \ge 1.$$

Obviously, this is a sequence of stopping times increasing to T almost surely. The processes $\xi_n(t) = \xi(t) \mathbb{1}_{[0,\tau_n]}(t)$, $n \ge 1$, belong to $\mathcal{H}_2([0,T])$. Moreover, for any $m \ge n \ge 1$, $\xi_n(t) = \xi_m(t) \mathbb{1}_{[0,\tau_n]}$, so by theorem 8.4, $I(\xi_n, [0,t]) = I(\xi_m, [0,t])$

almost surely on the set $\{\tau_n \geq t\}$. In particular, $I(\xi_n, [0,T]) = I(\xi_m, [0,T])$ almost surely on $\Omega_n := \left\{\int_0^T \xi(s)^2 dt \leq n\right\}$ for any $m \geq n$. Hence, it is easy to deduce the existence of a random variable $\mathcal{I}$ such that $\mathcal{I} = I(\xi_n, [0,T])$ almost surely on Ω_n for any $n \geq 1$. The value of this random variable is, quite naturally, called the *extended Itô integral* of ξ with respect to W. From theorem 8.4 it follows that $\mathcal{I} = I(\xi, [0,t])$ almost surely if $\xi \in \mathcal{H}_2([0,T])$. Therefore, this is indeed an *extension* of the Itô integral. For this reason, we will use the same notation for extended Itô integral:

$$I(\xi, [0,t]) = \int_0^T \xi(t)dW(t) := \mathcal{I};$$

it will always be clear from the context which definition is used, i.e. whether $\xi \in \mathcal{H}_2([0,T])$ or not. However, the extended Itô integral loses some of the properties of the usual Itô integral. For instance, it does not have zero mean in general; moreover, the mean is not guaranteed to exist. For convenience, we gather all the properties which are preserved; they immediately follow from theorem 8.2. The statement below is for arbitrary interval $[a,b]$; the definition is modified obviously.

THEOREM 8.5.– *Let $\xi, \eta \in \mathcal{H}([a,b])$. Then:*

1) $I(\xi + \eta, [a,b]) = I(\xi, [a,b]) + I(\eta, [a,b])$ almost surely;

2) For any $c \in \mathbb{R}$, $I(c\xi, [a,b]) = cI(\xi, [a,b])$ almost surely;

3) For any $c \in (a,b)$, $I(\xi, [a,b]) = I(\xi, [a,c]) + I(\eta, [c,b])$ almost surely;

4) There exists a continuous modification of $\{I(\xi, [a,t]), t \in [a,b]\}$;

5) For any stopping time τ with values in $[a,b]$, $\int_a^b \xi(t)\mathbb{1}_{[a,\tau]}(t)dW_t = \int_a^\tau \xi(t)dW_t$.

We conclude this section by showing the continuity of the extended Itô integral with respect to the integrand.

THEOREM 8.6.– *Let $\{\xi_n, n \geq 1\}$ be a sequence of processes in $\mathcal{H}([0,T])$ such that $\int_0^T |\xi_n(t) - \xi(t)|^2 \, dt \xrightarrow{\mathrm{P}} 0$, $n \to \infty$. Then,*

$$\sup_{t\in[0,T]} |I(\xi_n, [0,T]) - I(\xi, [0,T])| \xrightarrow{\mathrm{P}} 0, \ n \to \infty.$$

PROOF.– Due to linearity, it is sufficient to prove the statement for $\xi = 0$. Define for fixed $\varepsilon > 0$, $\tau_{n,\varepsilon} = \inf\left\{t \geq 0 : \int_0^t \xi_n(s)^2 ds \geq \varepsilon^3\right\} \wedge T$ and $\xi_{n,\varepsilon}(t) = \xi_n(t)\mathbb{1}_{t\leq\tau_{n,\varepsilon}}$.

We have that $I(\xi_n,[0,t]) = I(\xi_{n,\varepsilon},[0,t])$, $t \in [0,T]$, on $A_{n,\varepsilon} := \left\{\int_0^T \xi_n(t)^2 dt < \varepsilon^3\right\}$; moreover, $\mathrm{P}\{A_{n,\varepsilon}\} \to 1$, $n \to \infty$. By theorem 5.26,

$$\mathrm{P}\Big\{\sup_{t\in[0,T]} |I(\xi_{n,\varepsilon},[0,T])| \ge \varepsilon\Big\} \le \frac{1}{\varepsilon^2}\int_0^T \mathrm{E}\,|\xi_{n,\varepsilon}(t)|^2\,dt \le \varepsilon.$$

Consequently,

$$\begin{aligned}\mathrm{P}\big\{|I(\xi_n,[0,T])| \ge \varepsilon\big\} &\le \mathrm{P}\Big\{\sup_{t\in[0,T]} |I(\xi_{n,\varepsilon},[0,T])| \ge \varepsilon\Big\} + \mathrm{P}\{A^c_{n,\varepsilon}\}\\ &\le \varepsilon + \mathrm{P}\{A^c_{n,\varepsilon}\},\end{aligned}$$

whence:

$$\limsup_{n\to\infty} \mathrm{P}\big\{|I(\xi_n,[0,T])| \ge \varepsilon\big\} \le \varepsilon,$$

and the statement follows due to the arbitrariness of ε. □

8.5. Itô processes and Itô formula

DEFINITION 8.3.– *A process $X = \{X(t), t \in [a,b]\}$ is called an* Itô process *if it admits the representation:*

$$X(t) = X(a) + \int_a^t \alpha(s)ds + \int_a^t \beta(s)dW(s) \qquad [8.8]$$

almost surely for all $t \in [a,b]$, where $X(a)$ is an $\mathcal{F}_a$-measurable random variable and α and β are progressively measurable processes such that $\int_a^b \big(|\alpha(t)| + \beta(t)^2\big)\,dt < \infty$ almost surely. The expression

$$dX(t) := \alpha(t)dt + \beta(t)dW(t)$$

is called the stochastic differential *of X.*

In other words, Itô processes are sums of indefinite Lebesgue and Itô integrals. These processes are adapted and, as we know from the previous section, have a continuous modification; we will assume that they are continuous.

THEOREM 8.7.– *Let X_i and $i = 1,2$ be Itô processes on $[a,b]$. Then, for any $c_1, c_2 \in \mathbb{R}$, $c_1X_1 + c_2X_2$ is an Itô process with*

$$d\big(c_1X_1(t) + c_2X_2(t)\big) = c_1dX_1(t) + c_2dX_2(t),$$

where the linear combination of differentials is defined in an obvious way.

PROOF.– Follows straightforwardly from the linearity of the Itô integral. □

As we have seen in section 8.2.1, $dW(t)^2 \neq 2W(t)dW(t)$, so we should not expect in general that $dF(X(t)) = F'(X(t))dX(t)$, as in the deterministic case. The change of variable formula for Itô integral, called *Itô formula*, involves an extra term with second derivative; this is a subject of the following theorem.

THEOREM 8.8.– *Let $\{X(t), t \in [a,b]\}$ be an Itô process with $dX(t) = \alpha(t)dt + \beta(t)dW(t)$ and $F \in C^{1,2}([a,b] \times \mathbb{R})$. Then, $\{F(t, X(t)), t \in [a,b]\}$ is an Itô process with*

$$\begin{aligned} dF(t,X(t)) &= \frac{\partial}{\partial t}F(t,X(t))dt + \frac{\partial}{\partial x}F(t,X(t))dX(t) \\ &\quad + \frac{1}{2}\frac{\partial^2}{\partial x^2}F(t,X(t))\beta(t)^2 dt \\ &= \left(\frac{\partial}{\partial t}F(t,X(t)) + \frac{\partial}{\partial x}F(t,X(t))\alpha(t) + \frac{1}{2}\frac{\partial^2}{\partial x^2}F(t,X(t))\beta(t)^2\right)dt \\ &\quad + \frac{\partial}{\partial x}F(t,X(t))\beta(t)dW(t). \end{aligned}$$

First, we need a lemma on behavior of sums of weighted squares of increments of Itô integral.

LEMMA 8.2.– *Let $\beta \in \mathcal{H}([a,b])$, $\{Y(t), t \in [a,b]\}$ be a continuous process and $\{a = t_0^n < t_n^1 < \cdots < t_n^n = b, n \geq 1\}$ be a sequence of partitions, with the mesh going to zero:* $\max_{1 \leq k \leq n} |t_k^n - t_{k-1}^n| \to 0$*, $n \to \infty$. Then,*

$$\sum_{k=1}^{n} Y_{t_k^n} \left(\int_{t_{k-1}^n}^{t_k^n} \beta(s)dW(s) \right)^2 \xrightarrow{\mathrm{P}} \int_a^b Y(t)\beta(t)^2 dt, \ n \to \infty.$$

PROOF.– To avoid cumbersome notation, denote $t_k = t_k^n$, $k = 0, \ldots, n$. Let us first consider the case where $\beta \in \mathcal{H}_2([a,b])$, $Y \equiv 1$. Let $\{\beta_n, n \geq 1\}$ be a sequence of simple processes from $\mathcal{H}_2([a,b])$ such that for any $n \geq 1$, β_n is constant on each interval $[t_{k-1}, t_k)$, and $\|\beta - \beta_n\|_{\mathcal{H}_2([a,b])} \to 0$, $n \to \infty$; the existence of such sequence is proved similarly to lemma 8.1. For any $x, y \in \mathbb{R}$ and $\theta > 0$, we have the following simple inequality:

$$\left|x^2 - y^2\right| \leq |2x(x-y)| + |x-y|^2 \leq \theta x^2 + (1+\theta^{-1})|x-y|^2.$$

Hence, setting $\theta_n = \|\beta - \beta_n\|_{\mathcal{H}_2([a,b])}$,

$$\int_a^b \mathrm{E}\left|\beta(s)^2 - \beta_n(s)^2\right| ds \le \theta_n \int_a^b \mathrm{E}\beta(s)^2 ds$$

$$+ (1+\theta_n^{-1}) \int_a^b \mathrm{E}\left(\beta(s) - \beta_n(s)\right)^2 ds \qquad [8.9]$$

$$= \|\beta - \beta_n\|_{\mathcal{H}_2([a,b])} \left(\|\beta\|^2_{\mathcal{H}_2([a,b])} + \|\beta - \beta_n\|_{\mathcal{H}_2([a,b])} + 1\right) \to 0, n \to \infty;$$

in other words, $\beta_n^2 \to \beta^2$ in $L_1([a,b] \times \Omega)$; therefore, the sequence $\left\{\beta_n^2, n \ge 1\right\}$ is uniformly integrable in $L_1([a,b] \times \Omega)$. Similarly, we obtain with the help of Itô isometry

$$\mathrm{E}\left|\sum_{k=1}^n \left(\int_{t_{k-1}}^{t_k} \beta(s) dW(s)\right)^2 - \sum_{k=1}^n \left(\int_{t_{k-1}}^{t_k} \beta_n(s) dW(s)\right)^2\right|$$

$$\le \theta_n \sum_{k=1}^n \mathrm{E}\left(\int_{t_{k-1}}^{t_k} \beta(s) dW(s)\right)^2$$

$$+ (1+\theta_n^{-1}) \sum_{k=1}^n \mathrm{E}\left(\int_{t_{k-1}}^{t_k} \left(\beta(s) - \beta_n(s)\right) dW(s)\right)^2$$

$$= \|\beta - \beta_n\|_{\mathcal{H}_2([a,b])} \left(\|\beta\|^2_{\mathcal{H}_2([a,b])} + \|\beta - \beta_n\|_{\mathcal{H}_2([a,b])} + 1\right) \to 0,\ n \to \infty.$$

As a result,

$$\sum_{k=1}^n \left(\int_{t_{k-1}}^{t_k} \beta(s) dW(s)\right)^2 - \sum_{k=1}^n \left(\int_{t_{k-1}}^{t_k} \beta_n(s) dW(s)\right)^2 \xrightarrow{\mathrm{P}} 0,\ n \to \infty. \qquad [8.10]$$

Take $A > 0$ and write

$$\sum_{k=1}^n \left(\int_{t_{k-1}}^{t_k} \beta_n(s) dW(s)\right)^2 = \sum_{k=1}^n \beta_n(t_{k-1})^2 (\Delta_k W)^2$$

$$= \sum_{k=1}^n \beta_n(t_{k-1})^2 \mathbb{1}_{|\beta_n(t_{k-1})| \le A} (\Delta_k W)^2 \qquad [8.11]$$

$$+ \sum_{k=1}^n \beta_n(t_{k-1})^2 \mathbb{1}_{|\beta_n(t_{k-1})| > A} (\Delta_k W)^2,$$

where $\Delta_k W = W(t_k) - W(t_{k-1})$. Denoting $\gamma_k^n = \beta_n(t_{k-1})^2 \mathbb{1}_{|\beta_n(t_{k-1})|\le A}$, $\Delta_k = t_k - t_{k-1}$,

$$\mathrm{E}\left(\sum_{k=1}^n \gamma_k^n \left((\Delta_k W)^2 - \Delta_k\right)\right)^2$$
$$= \sum_{k=1}^n \sum_{l=1}^n \mathrm{E}\left(\gamma_k^n \left((\Delta_k W)^2 - \Delta_k\right) \gamma_l^n \left((\Delta_l W)^2 - \Delta_l\right)\right).$$

Note that for $k < l$,

$$\mathrm{E}\left(\gamma_k^n \left((\Delta_k W)^2 - \Delta_k\right) \gamma_l^n \left((\Delta_l W)^2 - \Delta_l\right)\right)$$
$$= \mathrm{E}\left(\gamma_k^n \left((\Delta_k W)^2 - \Delta_k\right) \gamma_l^n \mathrm{E}\left(\left((\Delta_l W)^2 - \Delta_l\right) \mid \mathcal{F}_{t_{l-1}}\right)\right) = 0,$$

since $\gamma_k^n \left((\Delta_k W)^2 - \Delta_k\right) \gamma_l^n$ is $\mathcal{F}_{t_{l-1}}$-measurable and $\left((\Delta_l W)^2 - \Delta_l\right)$ is independent of $\mathcal{F}_{t_{l-1}}$ and centered. Therefore,

$$\mathrm{E}\left(\sum_{k=1}^n \gamma_k^n \left((\Delta_k W)^2 - \Delta_k\right)\right)^2 = \sum_{k=1}^n \mathrm{E}\left((\gamma_k^n)^2 \left((\Delta_k W)^2 - \Delta_k\right)^2\right)$$
$$\le A^2 \sum_{k=1}^n \mathrm{E}\left(\left((\Delta_k W)^2 - \Delta_k\right)^2\right) = 2A^2 \sum_{k=1}^n \Delta_k^2$$
$$\le 2A^2(b-a) \max_{1\le k\le n} \Delta_k \to 0,\ n \to \infty.$$

Therefore,

$$\sum_{k=1}^n \gamma_k^n (\Delta_k W)^2 - \sum_{k=1}^n \gamma_k^n \Delta_k \xrightarrow{\mathrm{P}} 0,\ n \to \infty. \qquad [8.12]$$

Further,

$$\sup_{n\ge1} \mathrm{E}\left(\sum_{k=1}^n \beta_n(t_{k-1})^2 \mathbb{1}_{|\beta_n(t_{k-1})|>A} (\Delta_k W)^2\right)$$
$$= \sup_{n\ge1} \sum_{k=1}^n \mathrm{E}\left(\beta_n(t_{k-1})^2 \mathbb{1}_{|\beta_n(t_{k-1})|>A}\right) \Delta_k \qquad [8.13]$$
$$= \sup_{n\ge1} \int_a^b \mathrm{E}\left(\beta_n(s)^2 \mathbb{1}_{|\beta_n(s)|>A}\right) ds \to 0,\ A \to \infty,$$

Due to uniform integrability. Similarly,

$$\sup_{n\ge 1} \mathrm{E}\left|\int_0^t \beta_n(s)^2 ds - \sum_{k=1}^n \gamma_k^n \Delta_k\right| = \sup_{n\ge 1} \mathrm{E}\left|\int_0^t \beta_n(s)^2 ds \right.$$

$$\left. - \int_a^b \beta_n(s)^2 \mathbb{1}_{|\beta_n(s)|\le A} ds\right| \qquad [8.14]$$

$$= \sup_{n\ge 1} \int_a^b \mathrm{E}\left(\beta_n(s)^2 \mathbb{1}_{|\beta_n(s)|>A}\right) ds \to 0,\ A\to\infty.$$

Combining [8.11]–[8.14], we get

$$\sum_{k=1}^n \left(\int_{t_{k-1}}^{t_k} \beta_n(s) dW(s)\right)^2 - \int_0^t \beta_n(s)^2 ds \xrightarrow{\mathrm{P}} 0,\ n\to\infty.$$

Recalling [8.9] and [8.14], we get

$$\sum_{k=1}^n \left(\int_{t_{k-1}}^{t_k} \beta(s) dW(s)\right)^2 - \int_0^t \beta(s)^2 ds \xrightarrow{\mathrm{P}} 0,\ n\to\infty,$$

as required in this case.

By linearity, we get the statement in the case where $\beta \in \mathcal{H}_2([a,b])$, Y is a simple process, which, through a standard approximation argument, gives the statement for a continuous Y.

Finally, for $\beta \in \mathcal{H}([a,b])$, define $\tau_N = \inf\left\{t \ge a : \int_a^t \beta(s)^2 ds \ge N\right\} \wedge b$ and set $\beta_N(t) = \beta(t)\mathbb{1}_{t\le\tau_N}$ so that $\beta_N \in \mathcal{H}_2([a,b])$, $N \ge 1$. Then,

$$\sum_{k=1}^n Y(t_k^n) \left(\int_{t_{k-1}^n}^{t_k^n} \beta_N(s) dW(s)\right)^2 \xrightarrow{\mathrm{P}} \int_a^b Y(t)\beta_N(t)^2 dt,\ n\to\infty.$$

The required statement then follows from the observation that $\beta \equiv \beta_N$ on $\left\{\int_a^b \beta(s)^2 ds < N\right\}$, and these events increase, as $N \to \infty$, to an almost sure event. □

PROOF (Proof of theorem 8.8).– We need to show that

$$F(t,X(t)) = F(a,X(a)) + \int_a^t \left(\frac{\partial}{\partial t}F(s,X(s)) + \frac{\partial}{\partial x}F(s,X(s))\alpha(s) \right.$$
$$\left. + \frac{1}{2}\frac{\partial^2}{\partial x^2}F(s,X(s))\beta(s)^2\right) ds + \int_a^t \frac{\partial}{\partial x}F(s,X(s))\beta(s)dW(s) \qquad [8.15]$$

almost surely for all $t \in [a, b]$. As both sides of the equality are continuous processes, it is enough to show this (almost sure) equality for any fixed $t \in [a, b]$, see theorem 6.1.

Let us first assume that $\beta \in \mathcal{H}_2([a, b])$ and $F \in C^\infty([a, b] \times \mathbb{R}^2)$, compactly supported. We fix some $t \in [a, b]$, take arbitrary $n \geq 1$, set $\delta_n = (t - a)/n$ and consider the uniform partition of $[a, t]$: $t_k = a + k\delta_n$, $k = 0, \ldots, n$. We write

$$F(t, X(t)) - F(a, X(a)) = \sum_{k=1}^{n} \big(F(t_k, X(t_k)) - F(t_{k-1}, X(t_{k-1}))\big).$$

We have

$$\begin{aligned}
&F(t_k, X(t_k)) - F(t_{k-1}, X(t_k)) \\
&\quad = \frac{\partial}{\partial t} F(t_{k-1}, X(t_k))\delta_n + \frac{1}{2}\frac{\partial^2}{\partial t^2} F(\theta_k, X(t_k))\delta_n^2 \\
&\quad = \frac{\partial}{\partial t} F(t_{k-1}, X(t_{k-1}))\delta_n \\
&\qquad + \frac{\partial^2}{\partial t\, \partial x} F(t_{k-1}, \nu_k)\big(X(t_k) - X(t_{k-1})\big)\delta_n + \frac{1}{2}\frac{\partial^2}{\partial t^2} F(\theta, X(t_k))\delta_n^2
\end{aligned}$$

and

$$\begin{aligned}
&F(t_{k-1}, X(t_k)) - F(t_{k-1}, X(t_{k-1})) \\
&\quad = \frac{\partial}{\partial x} F(t_{k-1}, X(t_{k-1}))\big(X(t_k) - X(t_{k-1})\big) \\
&\qquad + \frac{1}{2}\frac{\partial^2}{\partial x^2} F(t_{k-1}, X(t_{k-1}))\big(X(t_k) - X(t_{k-1})\big)^2 \\
&\qquad + \frac{1}{6}\frac{\partial^3}{\partial x^3} F(t_{k-1}, \mu_k)\big(X(t_k) - X(t_{k-1})\big)^3.
\end{aligned}$$

As a result,

$$\begin{aligned}
&F(t_k, X(t_k)) - F(t_{k-1}, X(t_k)) \\
&\quad = \frac{\partial}{\partial t} F(t_{k-1}, X(t_{k-1}))\delta_n + \frac{\partial}{\partial x} F(t_{k-1}, X(t_{k-1}))\big(X(t_k) - X(t_{k-1})\big) \\
&\qquad + \frac{1}{2}\frac{\partial^2}{\partial x^2} F(t_{k-1}, X(t_{k-1}))\big(X(t_k) - X(t_{k-1})\big)^2 + R_{n,k},
\end{aligned}$$

where, thanks to our assumptions, the remaining term admits the estimate

$$|R_{n,k}| \leq C\left(\delta_n^2 + \delta_n\big|X(t_k) - X(t_{k-1})\big| + \big|X(t_k) - X(t_{k-1})\big|^3\right) \qquad [8.16]$$

with constant independent of k, n. Summing up, we have

$$F(t, X(t)) - F(a, X(a)) = S_{1,n} + S_{2,n} + \frac{1}{2} S_{3,n} + R_n,$$

where

$$S_{1,n} = \sum_{k=1}^{n} \frac{\partial}{\partial t} F(t_{k-1}, X(t_{k-1}))\delta_n, \; R_n = \sum_{k=1}^{n} R_{n,k}$$

$$S_{2,n} = \sum_{k=1}^{n} \frac{\partial}{\partial x} F(t_{k-1}, X(t_{k-1}))\big(X(t_k) - X(t_{k-1})\big),$$

$$S_{3,n} = \sum_{k=1}^{n} \frac{\partial^2}{\partial x^2} F(t_{k-1}, X(t_{k-1}))\big(X(t_k) - X(t_{k-1})\big)^2.$$

First,

$$\begin{aligned}
&\left| S_{1,n} - \int_a^t \frac{\partial}{\partial t} F(s, X(s)) ds \right| \\
&\leq \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \left| \frac{\partial}{\partial t} F(t_{k-1}, X(t_{k-1})) - \frac{\partial}{\partial t} F(s, X(s)) \right| ds \qquad [8.17] \\
&\leq (t-a) \sup_{\substack{s,u \in [a,t] \\ |s-u| \leq \delta_n}} \left| \frac{\partial}{\partial t} F(u, X(u)) - \frac{\partial}{\partial t} F(s, X(s)) \right| \to 0, \; n \to \infty,
\end{aligned}$$

due to uniform continuity of F and continuity of X. To study the terms $S_{2,n}$ and $S_{3,n}$, we write

$$X(t_k) - X(t_{k-1}) = I_k + J_k,$$

where $I_k = \int_{t_{k-1}}^{t_k} \alpha(s) ds$, $J_k = \int_{t_{k-1}}^{t_k} \beta(s) dW(s)$.

Similarly to [8.2],

$$\sum_{k=1}^{n} \frac{\partial}{\partial x} F(t_{k-1}, X(t_{k-1})) I_k \to \int_a^t \frac{\partial}{\partial x} F(s, X(s)) \alpha(s) ds, \; n \to \infty.$$

Defining $\eta_n(s) = \beta(s) \sum_{k=1}^{n} \frac{\partial}{\partial x} F(t_{k-1}, X(t_{k-1})) \mathbb{1}_{[t_{k-1}, t_k)}(s)$, we have

$$\sum_{k=1}^{n} \frac{\partial}{\partial x} F(t_{k-1}, X(t_{k-1})) J_k = \int_a^t \eta_n(s) dW(s),$$

and by virtue of Itô's isometry,

$$\begin{aligned}
&\mathrm{E}\left(\int_a^t \eta_n(s)dW(s) - \int_a^t F(s,X(s))\beta(s)dW(s)\right)^2 \\
&= \mathrm{E}\left(\int_a^t \big(\eta_n(s) - F(s,X(s))\beta(s)\big)dW(s)\right)^2 \\
&= \int_a^t \mathrm{E}\big(\eta_n(s) - F(s,X(s))\beta(s)\big)^2 ds \le \int_a^t \mathrm{E}\left(D_n^2\beta(s)^2\right) ds,
\end{aligned}$$

where

$$D_n = \sup_{\substack{s,u\in[a,t]\\ |s-u|\le\delta_n}} \left|\frac{\partial}{\partial x}F(u,X(u)) - \frac{\partial}{\partial x}F(s,X(s))\right|.$$

As above, $D_n \to 0$, $n\to\infty$. Moreover, $D_n \le 2\sup_{s\in[a,b],x\in\mathbb{R}}\left|\frac{\partial}{\partial x}F(s,x)\right|$, so by the dominated convergence theorem

$$\mathrm{E}\left(\int_a^t \eta_n(s)dW(s) - \int_a^t F(s,X(s))\beta(s)dW(s)\right)^2 \to 0,\ n\to\infty,$$

whence

$$\sum_{k=1}^n \frac{\partial}{\partial x}F(t_{k-1},X(t_{k-1}))J_k \xrightarrow{\mathrm{P}} \int_a^t F(s,X(s))\beta(s)dW(s),\ n\to\infty.$$

Consequently,

$$S_{2,n} \xrightarrow{\mathrm{P}} \int_a^t \frac{\partial}{\partial x}F(s,X(s))\alpha(s)ds + \int_a^t F(s,X(s))\beta(s)dW(s),\ n\to\infty.$$

Further, consider

$$S_{3,n} = \frac{1}{2}\sum_{k=1}^n \frac{\partial^2}{\partial x^2}F(t_{k-1},X(t_{k-1}))\left(I_k^2 + 2I_kJ_k + J_k^2\right).$$

Thanks to boundedness of $\frac{\partial^2}{\partial x^2}F$,

$$\begin{aligned}
&\left|\sum_{k=1}^n \frac{\partial^2}{\partial x^2}F(t_{k-1},X(t_{k-1}))\left(I_k^2 + 2I_kJ_k\right)\right| \\
&\le C\left(\max_{1\le k\le n}|I_k| + \max_{1\le k\le n}|J_k|\right)\sum_{k=1}^n I_k \\
&\le C\left(\max_{1\le k\le n}|I_k| + \max_{1\le k\le n}|J_k|\right)\int_a^t |\alpha(s)|\,ds.
\end{aligned}$$

The last expression vanishes, since by continuity of Lebesgue and Itô integrals, $\max_{1\le k\le n}|I_k| + \max_{1\le k\le n}|J_k| \to 0$, $n\to\infty$, almost surely. By lemma 8.2,

$$\sum_{k=1}^{n} \frac{\partial^2}{\partial x^2} F(t_{k-1}, X(t_{k-1}))J_k^2 \xrightarrow{\mathrm{P}} \int_a^t \frac{\partial^2}{\partial x^2} F(s, X(s))\beta(s)^2 ds,\ n\to\infty.$$

Summing up, we have

$$S_{3,n} \xrightarrow{\mathrm{P}} \int_a^t \frac{\partial^2}{\partial x^2} F(s, X(s))\beta(s)^2 ds,\ n\to\infty.$$

Finally, by [8.16],

$$\sum_{k=1}^{n} |R_{n,k}| \le C\left(\frac{1}{n} + \max_{k=1,\dots,n} \left|X(t_k) - X(t_{k-1})\right| \left(1 + \sum_{k=1}^{n} \big(X(t_k) - X(t_{k-1})\big)^2\right)\right).$$

Similarly to $S_{3,n}$,

$$\sum_{k=1}^{n} \big(X(t_k) - X(t_{k-1})\big)^2 \xrightarrow{\mathrm{P}} \int_a^t \beta(s)^2 ds,\ n\to\infty,$$

and thanks to continuity, $\max_{k=1,\dots,n}\left|X(t_k) - X(t_{k-1})\right| \to 0$, $n\to\infty$, almost surely. Therefore, $R_n \xrightarrow{\mathrm{P}} 0$, $n\to\infty$, which finishes the proof for the current case.

Let us now turn to the general case, i.e. where $\beta \in \mathcal{H}([a,b])$ and $F\in C^{1,2}([a,b]\times\mathbb{R})$. Take arbitrary $N\ge 1$ and consider a sequence of compactly supported functions $F_n \in C^{1,2}([a,b]\times\mathbb{R})$ such that $F_n\to F$, $\frac{\partial}{\partial t}F_n \to \frac{\partial}{\partial t}F$, $\frac{\partial}{\partial x}F_n \to \frac{\partial}{\partial x}F$, $\frac{\partial^2}{\partial x^2}F_n \to \frac{\partial^2}{\partial x^2}F$, $n\to\infty$, uniformly on $[a,b]\times[-N,N]$. Define also $\tau_N = \inf\left\{t\ge a : \int_a^t \beta(s)^2ds + |X(t)| \ge N\right\} \wedge b$, $\alpha_N(t) = \alpha(t)\mathbb{1}_{t\le\tau_N}$, $\beta_N(t) = \beta(t)\mathbb{1}_{t\le\tau_N}$ and $X_N(t) = X(a) + \int_a^t \alpha_N(s)ds + \int_a^t \beta_N(s)dW(s)$. We write the Itô formula for $F_n(t, X_N(t))$:

$$\begin{aligned} F_n(t, X_N(t)) = F_n(a, X(a)) &+ \int_a^t \left(\frac{\partial}{\partial t} F_n(s, X_N(s))\right. \\ &+ \frac{\partial}{\partial x} F_n(s, X_N(s))\alpha_N(s) \\ &+ \left.\frac{1}{2}\frac{\partial^2}{\partial x^2} F_n(s, X_N(s))\beta_N(s)^2\right) ds \\ &+ \int_a^t \frac{\partial}{\partial x} F_n(s, X_N(s))\beta_N(s)dW(s) \end{aligned}$$

almost surely. Since $|X_N(t)| \leq N$ for $t \in [a,b]$ and in view of uniform convergence of F_n together with its derivatives to F, we get

$$F(t, X_N(t)) = F(a, X(a)) + \int_a^t \left(\frac{\partial}{\partial t} F(s, X_N(s)) + \frac{\partial}{\partial x} F(s, X_N(s)) \alpha_N(s) + \frac{1}{2} \frac{\partial^2}{\partial x^2} F(s, X_N(s)) \beta_N(s)^2 \right) ds + \int_a^t \frac{\partial}{\partial x} F(s, X_N(s)) \beta_N(s) dW(s)$$

almost surely. This coincides with [8.15] on $\left\{ \int_a^b \beta(t)^2 dt + \sup_{t\in[a,b]} |X(t)| \leq N \right\}$. Since these events increase as $N \to \infty$ to an almost sure event, we arrive at [8.15]. □

8.6. Multivariate stochastic calculus

The definition of Itô integral can be extended straightforwardly to the multi-dimensional case. Specifically, let $\left\{ W(t) = \big(W_1(t), \ldots, W_k(t)\big), t \geq 0 \right\}$ be a standard Wiener process in $\mathbb{R}^k$, i.e. its coordinates are independent standard Wiener processes in $\mathbb{R}$. As before, we assume that W is adapted to the filtration $\{\mathcal{F}_t, t \geq 0\}$ and for any $0 \leq s < t$ the increment $W(t) - W(s)$ is independent of $\mathcal{F}_s$. For a matrix-valued process $\{\xi(t) = (\xi_{ij}(t), i = 1, \ldots, d, j = 1, \ldots, k)\}$ such that $\xi_{ij} \in \mathcal{H}([a,b])$ for any $i = 1, \ldots, d$, $j = 1, \ldots, k$, we will understand $I(\xi, [a,b]) = \int_a^b \xi(s) dW(s)$ as an $\mathbb{R}^d$-valued process with ith coordinate equal to

$$I_i(\xi, [a,b]) = \sum_{j=1}^k \int_a^b \xi_{ij}(t) dW_j(t).$$

In particular, for an $\mathbb{R}^k$-valued process $\left\{ \eta(t) = \big(\eta_1(t), \ldots, \eta_k(t)\big), t \geq 0 \right\}$ with $\eta_i \in \mathcal{H}([a,b])$, $i = 1, \ldots, k$, we define

$$I(\eta, [a,b]) = \int_a^b \big(\eta(t), dW(t)\big) = \sum_{i=1}^k \int_a^b \eta_i(t) dW_i(t).$$

As before, (x, y) denotes the inner product. The meaning of $|x|$ will depend on the context: it is the absolute value of a number, Euclidean norm of a vector or a matrix (square root of the sum of squares of elements).

The multi-dimensional version of Itô integral has the same properties as its scalar counterpart; in particular, if $\xi \in \mathcal{H}_2([a,b])$ (which means that all its elements are in $\mathcal{H}_2([a,b])$), then

$$\mathrm{E}I(\xi(t),[a,b]) = 0,$$
$$\mathrm{E}\left|I(\xi(t),[a,b])\right|^2 = \int_a^b \mathrm{E}\left|\xi(t)\right|^2 dt.$$

The definition of the Itô process is carried over to several dimensions in a straightforward way: it is a process of the form

$$X(t) = X(a) + \int_a^t \alpha(s)ds + \int_a^t \beta(s)dW(s),\ t \in [a,b] \qquad [8.18]$$

where $X(a)$ is an $\mathcal{F}_a$-measurable random vector in $\mathbb{R}^d$, and α and β are progressively measurable processes with values in $\mathbb{R}^d$ and $\mathbb{R}^{d\times k}$, respectively, such that $\int_a^b \big(|\alpha(t)| + |\beta(t)|^2\big)dt < \infty$ almost surely. Equation [8.18] can be written coordinatewise:

$$X_i(t) = X_i(a) + \int_a^t \alpha_i(s)ds + \sum_{j=1}^k \int_a^t \beta_{ij}(s)dW_j(s),\ t \in [a,b],\ i = 1,\dots,d.$$

Similarly to the scalar case, the expression $dX(t) = \alpha(t)dt + \beta(t)dW(t)$ is called the stochastic differential of X.

There are no difficulties in generalizing the Itô formula, except notational ones. Let $F = F(t,x)\colon [a,b]\times\mathbb{R}^d \to \mathbb{R}$ be continuously differentiable with respect to t and twice continuously differentiable with respect to x. Then, for X given by [8.18], $F(t,X(t))$ is an Itô process with

$$\begin{aligned} dF(t,X(t)) = {} & \frac{\partial}{\partial t}F(t,X(t))dt + \sum_{i=1}^d \frac{\partial}{\partial x_i}F(t,X(t))\alpha_i(t)dt \\ & + \sum_{i=1}^d \frac{\partial}{\partial x_i}F(t,X(t))\sum_{j=1}^k \beta_{ij}(t)dW_j(t) \\ & + \frac{1}{2}\sum_{i,l=1}^d \frac{\partial^2}{\partial x_i x_j}F(t,X(t))\sum_{j=1}^k \beta_{ij}(t)\beta_{lj}(t)dt. \end{aligned}$$

To avoid cumbersome expressions, we will use a shorter form:

$$dF(t,X(t)) = \frac{\partial}{\partial t}F(t,X(t))dt + \big(D_x F(t,X(t)),\alpha(t)\big)dt$$
$$+D_x F(t,X(t))\beta(t)dW(t) + \frac{1}{2}\mathrm{tr}\left(\beta\beta^\top(t)D^2_{xx}F(t,X(t))\right)dt,$$

where $D_x F(t,x) = \left(\frac{\partial}{\partial x_1}F(t,x),\dots,\frac{\partial}{\partial x_d}F(t,x)\right)$ is the vector of first derivatives and $D^2_{xx}F(t,x) = \left(\frac{\partial^2}{\partial x_i x_l}F(t,x)\right)_{i,l=1}^d$ is the matrix of second derivatives.

REMARK 8.6.– Informally, the Itô formula can be written as

$$dF(t,X(t)) = \frac{\partial}{\partial t}F(t,X(t))dt + \big(D_x F(t,X(t)),dX(t)\big)$$
$$+\frac{1}{2}\big(D^2_{xx}F(t,X(t))dX(t),dX(t)\big),$$

where we use the following rules of multiplying differentials:

$$dt\,dt = dt\,dW_i(t) = dW_i(t)\,dW_l(t) = 0$$

for $i \neq l$ and

$$dW_i(t)\,dW_i(t) = dt.$$

EXAMPLE 8.1.– *One of the important particular cases of the Itô formula is the formula for differential of product. Let X_i and $i=1,2$ be Itô processes on $[a,b]$ with*

$$dX_i(t) = \alpha_i(t)dt + \sum_{j=1}^k \beta_{ij}(t)dW_j(t),\ i=1,2.$$

Then, the product X_1X_2 is an Itô process, and

$$d\big(X_1(t)X_2(t)\big) = X_1(t)dX_2(t) + X_2(t)dX_1(t) + \sum_{j=1}^k \beta_{1j}(t)\beta_{2j}(t)dt$$
$$= X_1(t)dX_2(t) + X_2(t)dX_1(t) + d[X_1,X_2]_t. \qquad [8.19]$$

The process $[X_1,X_2]_t = \sum_{j=1}^k \int_a^t \beta_{1j}(s)\beta_{2j}(s)ds$ is called covariation *of the processes X_1 and X_2. The above formula simplifies when one of the processes has usual differential. Let, for example, X_2 be absolutely continuous, that is, $dX_2(t) = \alpha_2(t)dt$. Then,*

$$d\big(X_1(t)X_2(t)\big) = X_1(t)dX_2(t) + X_2(t)dX_1(t), \qquad [8.20]$$

which coincides with the usual formula for differential of product.

8.7. Maximal inequalities for Itô martingales

In this section, we consider maximal inequalities for *Itô local martingales*, i.e. Itô processes of the form $M(t) = \int_0^t \xi(s)dW(s)$, $t \in [0,T]$, where $\xi(t) \in \mathcal{H}([0,T])$. Consider a sequence of partitions $\{0 = t_0^n < t_n^1 < \cdots < t_n^n = T, n \geq 1\}$, with the mesh going to zero. Then, it follows from lemma 8.2 that

$$\sum_{k=1}^{n} \left(M(t_k^n) - M(t_{k-1}^n)\right)^2 \xrightarrow{\mathrm{P}} \int_0^T \xi(t)^2 dt,\ n \to \infty.$$

Thus, the process $[M]_t := \int_0^t \xi(s)^2 ds$ is a natural generalization of quadratic variation process to a continuous parameter case.

THEOREM 8.9.– *For each $p > 0$, there exist positive c_p, C_p such that for any progressively measurable $\{\xi_t, t \in [0,T]\}$ with $\int_0^T \xi(t)^2 dt < \infty$ almost surely*

$$c_p \mathrm{E}[M]_T^{p/2} \leq \mathrm{E} \sup_{t\in[0,T]} |M(t)|^p \leq C_p \mathrm{E}[M]_T^{p/2}, \qquad [8.21]$$

where $M(t) = \int_0^t \xi(s)dW(s)$ and $[M]_t = \int_0^t \xi(s)^2 ds$.

REMARK 8.7.– For $p \geq 1$, inequality [8.21] is often referred to as the Burkholder–Davis–Gundy inequality (for $p = 1$, this is the Davis inequality, for $p > 1$, the Burkholder–Gundy inequality). Similar inequalities for martingales with discrete parameter were discussed in section 5.5.7.

REMARK 8.8.– It is worth mentioning that the same assertion holds when T is a stopping time: after setting $\widetilde{M}(t) = M(t)\mathbb{1}_{t\leq T}$ and using theorem 8.4, this boils down to the case of non-random T.

PROOF.– Take any $N \geq 1$ and denote

$$\tau_N = \inf\left\{t \geq 0 : \int_0^t \xi(s)^2 ds + |M(t)| \geq N\right\} \wedge T,$$

$M_N(t) = M(t \wedge \tau_N)$, $M_N^*(T) = \sup_{t\in[0,T]} |M_N(t)|$, $\xi_N(t) = \xi(t)\mathbb{1}_{[0,t]}(\tau_N)$. By theorem 8.4, $M_N(t) = \int_0^t \xi_N(s)dW(s)$ and M_N is a continuous martingale.

We first prove the right inequality in [8.21]. Let $p \geq 2$. By theorem 5.26,

$$\mathrm{E}M_N^*(T)^p \leq C_p \mathrm{E}\left|M_N(T)\right|^p.$$

Using the Itô formula,

$$\begin{aligned}\mathrm{E}\,|M_N(T)|^p &= p\mathrm{E}\left(\int_0^T |M_N(t)|^{p-1}\,\mathrm{sign}\,M_N(t)\,\xi_N(t)\,dW(t)\right)\\ &\quad+\frac{1}{2}p(p-1)\mathrm{E}\left(\int_0^T |M_N(t)|^{p-2}\,\xi_N(t)^2dt\right)\\ &=\frac{1}{2}p(p-1)\mathrm{E}\left(\int_0^T |M_N(t)|^{p-2}\,d[M_N]_t\right)\\ &\le\frac{1}{2}p(p-1)\mathrm{E}\left(M_N^*(T)^{p-2}[M_N]_T\right).\end{aligned}$$

If $p = 2$, we arrive at

$$\mathrm{E}M_N^*(T)^2 \le C_2\mathrm{E}[M_N]_T \le C_2\mathrm{E}[M]_T.$$

If $p > 2$, using the Hölder inequality with $q = p/(p-2)$, we get

$$\mathrm{E}M_N^*(T)^p \le C_p\left(\mathrm{E}M_N^*(T)^p\right)^{1/q}\left(\mathrm{E}[M_N]_T^{p/2}\right)^{2/p},$$

whence

$$\mathrm{E}M_N^*(T)^p \le C_p\mathrm{E}[M_N]_T^{p/2} \le C_p\mathrm{E}[M]_T^{p/2}.$$

Letting $N \to \infty$ and using the Fatou lemma, we obtain the right inequality in [8.21].

Now let $p < 2$. Define $\alpha_N(t) = \xi_N(t)[M_N]_t^{p/4-1/2}\mathbb{1}_{[M_N]_t>0}$, $A_N(t) = \int_0^t \alpha_N(s)dW(s)$, $A_N^*(T) = \sup_{t\in[0,T]}|A_N(t)|$. Then, A_N is a martingale and

$$\begin{aligned}\mathrm{E}A_N(T)^2 &= \mathrm{E}\left(\int_0^T \xi_N(t)^2[M_N]_t^{p/2-1}\mathbb{1}_{[M_N]_t>0}dt\right)\\ &=\mathrm{E}\left(\int_0^T [M_N]_t^{p/2-1}\mathbb{1}_{[M_N]_t>0}d[M_N]_t\right)=\frac{2}{p}\mathrm{E}[M_N]_T^{p/2}.\end{aligned}$$

On the other hand, by the Itô formula,

$$A_N(t)[M_N]_t^{1/2-p/4} = \int_0^t \alpha_N(s)[M_N]_s^{1/2-p/4} dW(s) + \int_0^t A_N(s) d[M_N]_s^{1/2-p/4} = M_N(t) + \int_0^t A_N(s) d[M_N]_s^{1/2-p/4},$$

whence

$$M_N^*(T) \le 2A_N^*(T)[M_N]_T^{1/2-p/4}.$$

Thus, using the Hölder inequality with $q = 2/(2-p)$ and theorem 5.26, we arrive at

$$\begin{aligned} \mathrm{E}M_N^*(T)^p &\le 2^p \mathrm{E}\left(A_N^*(T)^p \, [M_N]_T^{p(1/2-p/4)}\right) \\ &\le 2^p \left(\mathrm{E}A_N^*(T)^2\right)^{2/p} \left(\mathrm{E}[M_N]_T^{pq(1/2-p/4)}\right)^{1/q} \\ &\le C_p \left(\mathrm{E}A_N(T)^2\right)^{2/p} \left(\mathrm{E}[M_N]_T^{p/2}\right)^{1/q} = C_p \mathrm{E}[M_N]_T^{p/2}. \end{aligned}$$

Hence, as above, we derive the required inequality by letting $N \to \infty$.

To prove the left inequality, use the Itô formula again:

$$M_N(T)^2 = 2\int_0^T M_N(t)\xi_N(t) dW(t) + [M_N]_T,$$

getting

$$\begin{aligned} \mathrm{E}[M_N]_T^{p/2} &\le C_p \left(\mathrm{E}\left|M_N(T)\right|^p + \mathrm{E}\left|\int_0^T M_N(t)\xi_N(t) dW(t)\right|^{p/2}\right) \\ &\le C_p \left(\mathrm{E}M_N^*(T)^p + \mathrm{E}\left(\int_0^T M_N(t)^2 d[M_N]_t\right)^{p/4}\right) \\ &\le C_p \left(\mathrm{E}M_N^*(T)^p + \mathrm{E}M_N^*(T)^{p/2}[M_N]_T^{p/4}\right) \\ &\le C_p \left(\mathrm{E}M_N^*(T)^p + \left(\mathrm{E}\left(M_N^*(T)^p\right) \cdot \mathrm{E}\left([M_N]_T^{p/2}\right)\right)^{1/2}\right), \end{aligned}$$

whence it follows that

$$\mathrm{E}[M_N]_T^{p/2} \le C_p \mathrm{E} M_N^*(T)^p \le \mathrm{E} \sup_{t\in[0,T]} |M(t)|^p .$$

As before, letting $N \to \infty$ leads to the desired conclusion. □

8.7.1. *Strong law of large numbers for Itô local martingales*

As an application of the maximal inequalities, we will show a strong law of large numbers for Itô martingales, which is of great importance in statistics of stochastic processes.

Let $\{X(t), t \ge 0\}$ be a one-dimensional Itô martingale, i.e.

$$X(t) = X(0) + \int_0^t \xi(s)dW(s),\ t \ge 0,$$

where $\xi \in \mathcal{H}([0,t])$ for each $t > 0$. Recall that the quadratic variation of X is $[X]_t = \int_0^t \xi(s)^2 ds$. Obviously, this is a non-decreasing process, so there exists the limit $\lim_{T\to\infty}[X]_T \in [0,+\infty]$.

THEOREM 8.10.– *For any Itô local martingale* $\{X(t), t \ge 0\}$*,*

$$\frac{X(T)}{[X]_T} \to 0,\ T \to \infty,$$

for almost all $\omega \in \{\lim_{T\to\infty}[X]_T = +\infty\}$.

PROOF.– Set $T_n = \inf\{t \ge 0 : [X]_t \ge 2^n\}$, $n \ge 1$. We have that $T_n \to +\infty$, $n \to \infty$ and $T_n < +\infty$ on $A := \{\lim_{T\to\infty}[X]_T = +\infty\}$.

For any $T > 0$, we define

$$M_n(T) = \sup_{T_n\wedge T \le t < T_{n+1}\wedge T} \frac{|X(t)|^2}{[X]_t^2},$$

and $M_n(T) = 0$ if $T \le T_n$. We estimate

$$\begin{aligned} M_n(T) &\le 2^{1-2n}\left(|X(T_n\wedge T)|^2 + \sup_{T_n\wedge T \le t < T_{n+1}\wedge T} |X(t) - X(T_n\wedge t)|^2\right) \\ &\le 2^{1-2n}\left(|X(T_n\wedge T)|^2 + \sup_{t\in[0,T]} |X(T_{n+1}\wedge t) - X(T_n\wedge t)|^2\right). \end{aligned}$$

It follows easily from theorem 8.4 that

$$X(T_{n+1} \wedge t) - X(T_n \wedge t) = \int_0^t \xi_n(s)dW(s),$$

where $\xi_n(s) = \xi(s)\mathbb{1}_{T_n \le s \le T_{n+1}}$. Therefore, by theorem 8.9, we have

$$\mathrm{E}\left(\sup_{t\in[0,T]} |X(T_{n+1} \wedge t) - X(T_n \wedge t)|^2\right) \le \mathrm{E}\int_0^T \xi_n(t)^2 dt \le 2^{n+1}.$$

Also, from theorem 8.4 and the Itô isometry,

$$\mathrm{E}\,|X(T_n \wedge T)|^2 = \int_0^T \mathrm{E}\left(\xi(t)^2 \mathbb{1}_{t\le T_n}\right) dt \le 2^n.$$

Collecting the estimates, we get

$$\mathrm{E}M_n(T) \le 2^{1-2n}(2^{n+1} + 2^n) \le 2^{3-n}.$$

Therefore,

$$\mathrm{E}\Big(\sup_{T_n \le t < T_{n+1}} \frac{|X(t)|^2}{[X]_t^2}\mathbb{1}_A\Big) \le \mathrm{E}M_n(T) \le 2^{3-n}.$$

Hence

$$\mathrm{E}\Big(\sum_{n=1}^{\infty} \sup_{T_n \le t < T_{n+1}} \frac{|X(t)|^2}{[X]_t^2}\mathbb{1}_A\Big) = \sum_{n=1}^{\infty} \mathrm{E}\Big(\sup_{T_n \le t < T_{n+1}} \frac{|X(t)|^2}{[X]_t^2}\mathbb{1}_A\Big) < \infty.$$

In particular,

$$\sup_{T_n \le t < T_{n+1}} \frac{|X(t)|^2}{[X]_t^2}\mathbb{1}_A \to 0,\ n \to \infty,$$

almost surely. Consequently, $|X(t)|/[X]_t \to 0$, $t \to \infty$, for almost all $\omega \in A$, as claimed. □

8.8. Lévy martingale characterization of Wiener process

As we know, the standard Wiener process W is a martingale with quadratic characteristics t, i.e. $W_t^2 - t$ is a martingale, too. It turns out that this statement can be reversed. The reverse statement is called the *Lévy martingale characterization* (or simply the Lévy characterization) of the Wiener process; we formulate its multi-dimensional version.

THEOREM 8.11.– *Let $\{X(t), t \geq 0\}$ be a continuous process in $\mathbb{R}^d$ such that:*

– for any $i = 1, \ldots, d$, X_i is a martingale;

– for any $i, j = 1, \ldots, d$, $X_i(t)X_j(t) - \delta_{ij}t$, $t \geq 0$, is a martingale.

Then, X is a standard Wiener process in $\mathbb{R}^d$.

REMARK 8.9.– The continuity assumption cannot be omitted: the compensated Poisson process $N(t) - t$, $t \geq 0$, is easily seen to satisfy both conditions.

We start by proving some analogue of the Itô formula.

LEMMA 8.3.– *Let X satisfy the assumptions of theorem 8.11 and $f \in C^3(\mathbb{R}^d)$ be bounded together with its derivatives up to the third order. Then, for any $0 \leq s \leq t$,*

$$\mathrm{E}(f(X(t)) \mid \mathcal{F}_s) = f(X(s)) + \frac{1}{2}\int_s^t \sum_{i=1}^d \mathrm{E}\left(\frac{\partial^2}{\partial x_i^2} f(X(u)) \,\middle|\, \mathcal{F}_s\right) du.$$

PROOF.– For $n \geq 1$, we denote $\delta_n = (t-s)/n$ and consider a uniform partition $\{t_k^n = s + k\delta_n, k = 0, \ldots, n\}$ of $[s,t]$. We fix arbitrary $\varepsilon > 0$ and define $w_n(u) = \sup_{v \in [u-\delta_n, u]} |X(v) - X(u)|$, $u \in [s + \delta_n, t]$, $\tau_{n,\varepsilon} = \inf\{u \geq s : w_n(u) \geq \varepsilon\} \wedge t$. The continuity of X implies

$$\tau_{n,\varepsilon} \to t, \; n \to \infty. \qquad [8.22]$$

We write

$$\mathrm{E}(f(X(t)) \mid \mathcal{F}_s) - f(X(s)) = \mathrm{E}(f(X(t)) - f(X(\tau_{n,\varepsilon})) \mid \mathcal{F}_s)$$
$$+\mathrm{E}(f(X(\tau_{n,\varepsilon})) \mid \mathcal{F}_s) - f(X(s)).$$

Taking into account [8.22] and the boundedness of f, we get

$$\mathrm{E}(f(X(t)) - f(X(\tau_{n,\varepsilon})) \mid \mathcal{F}_s) \to 0, \; n \to \infty,$$

almost surely. Further, we denote $t^n_{k,\varepsilon} = t^n_k \wedge \tau_{n,\varepsilon}$, $\Delta^n_{k,\varepsilon} X_i = X_i(t^n_k) - X_i(t^n_{k-1})$ and write, using the Taylor formula,

$$\mathrm{E}(f(X(\tau_{n,\varepsilon})) \mid \mathcal{F}_s) - f(X(s)) = \sum_{k=1}^{n} \mathrm{E}(f(X(t^n_{k,\varepsilon})) - f(X(t^n_{k-1,\varepsilon})) \mid \mathcal{F}_s)$$

$$= \sum_{k=1}^{n} \mathrm{E}\left(\sum_{i=1}^{d} \frac{\partial}{\partial x_i} f(X(t^n_{k-1,\varepsilon}))\Delta^n_{k,\varepsilon} X_i \;\Big|\; \mathcal{F}_s \right)$$

$$+\frac{1}{2}\sum_{k=1}^{n} \mathrm{E}\left(\sum_{i,j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} f(X(t^n_{k-1,\varepsilon}))\Delta^n_{k,\varepsilon} X_i \Delta^n_{k,\varepsilon} X_j \;\Big|\; \mathcal{F}_s \right)$$

$$+\frac{1}{6}\sum_{k=1}^{n} \mathrm{E}\left(\sum_{i,j,l=1}^{d} \frac{\partial^3}{\partial x_i \partial x_j \partial x_l} f(\theta_k)\Delta^n_{k,\varepsilon} X_i \Delta^n_{k,\varepsilon} X_j \Delta^n_{k,\varepsilon} X_l \;\Big|\; \mathcal{F}_s \right)$$

$$=: S^n_1 + S^n_2 + S^n_3,$$

where θ_k are some points in $\mathbb{R}^d$ between $X(t^n_{k-1,\varepsilon})$ and $X(t^n_{k,\varepsilon})$, $k = 1, \ldots, n$. Note that for any $k = 0, 1, \ldots, n$, $t^n_{k,\varepsilon}$ is a bounded stopping time. Since X is a continuous martingale, from theorem 5.25, we have

$$S^n_1 = \mathrm{E}\left(\frac{\partial}{\partial x_i} f(X(t^n_{k-1,\varepsilon}))\mathrm{E}\left(\Delta^n_{k,\varepsilon} X_i \mid \mathcal{F}_{t^n_{k-1,\varepsilon}} \right) \;\Big|\; \mathcal{F}_s \right) = 0.$$

Further,

$$S^n_2 = \frac{1}{2}\sum_{k=1}^{n}\sum_{i,j=1}^{d} \mathrm{E}\left(\frac{\partial^2}{\partial x_i x_j} f(X(t^n_{k-1,\varepsilon}))\mathrm{E}\left(\Delta^n_{k,\varepsilon} X_i \Delta^n_{k,\varepsilon} X_j \mid \mathcal{F}_{t^n_{k-1,\varepsilon}} \right) \;\Big|\; \mathcal{F}_s \right). \quad [8.23]$$

Since $X_i(t)X_j(t) - \delta_{ij} t$ is a martingale, we have

$$X_i(t^n_{k-1,\varepsilon})X_j(t^n_{k-1,\varepsilon}) - \delta_{ij} t^n_{k-1,\varepsilon} = \mathrm{E}\left(X_i(t^n_{k,\varepsilon})X_j(t^n_{k,\varepsilon}) - \delta_{ij} t^n_{k,\varepsilon} \mid \mathcal{F}_{t^n_{k-1,\varepsilon}} \right)$$

$$= \mathrm{E}\left(\left(X_i(t^n_{k-1,\varepsilon}) + \Delta^n_{k,\varepsilon} X_i \right) \left(X_j(t^n_{k-1,\varepsilon}) + \Delta^n_{k,\varepsilon} X_j \right) - \delta_{ij} t^n_{k,\varepsilon} \mid \mathcal{F}_{t^n_{k-1,\varepsilon}} \right)$$

$$= X_i(t^n_{k-1,\varepsilon})X_j(t^n_{k-1,\varepsilon}) + X_i(t^n_{k-1,\varepsilon})\mathrm{E}\left(\Delta^n_{k,\varepsilon} X_j \mid \mathcal{F}_{t^n_{k-1,\varepsilon}} \right)$$

$$+X_j(t^n_{k-1,\varepsilon})\mathrm{E}\left(\Delta^n_{k,\varepsilon} X_i \mid \mathcal{F}_{t^n_{k-1,\varepsilon}} \right) + \mathrm{E}\left(\Delta^n_{k,\varepsilon} X_i \Delta^n_{k,\varepsilon} X_j - \delta_{ij} t^n_{k,\varepsilon} \mid \mathcal{F}_{t^n_{k-1,\varepsilon}} \right)$$

$$= X_i(t^n_{k-1,\varepsilon})X_j(t^n_{k-1,\varepsilon}) + \mathrm{E}\left(\Delta^n_{k,\varepsilon} X_i \Delta^n_{k,\varepsilon} X_j - \delta_{ij} t^n_{k,\varepsilon} \mid \mathcal{F}_{t^n_{k-1,\varepsilon}} \right),$$

whence

$$\mathrm{E}\left(\Delta_{k,\varepsilon}^n X_i \Delta_{k,\varepsilon}^n X_j \mid \mathcal{F}_{t_{k-1,\varepsilon}^n}\right) = \delta_{ij} \mathrm{E}\left(t_{k,\varepsilon}^n - t_{k-1,\varepsilon}^n \mid \mathcal{F}_{t_{k-1,\varepsilon}^n}\right). \qquad [8.24]$$

Substituting this into [8.23], we get

$$S_2^n = \frac{1}{2}\mathrm{E}\left(\sum_{k=1}^n \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f(X(t_{k-1,\varepsilon}^n))\mathrm{E}\left(t_{k,\varepsilon}^n - t_{k-1,\varepsilon}^n \mid \mathcal{F}_{t_{k-1,\varepsilon}^n}\right) \;\Big|\; \mathcal{F}_s\right).$$

In view of [8.22], the expression under expectation is equal to

$$\sum_{k=1}^n \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f(X(t_{k-1,\varepsilon}^n))\delta_n$$

for all n large enough, which converges, thanks to the continuity of X, to

$$\int_s^t \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f(X(u))du$$

as $n \to \infty$. Moreover, that expression is bounded by assumption, so the dominated convergence theorem yields

$$\begin{aligned} S_2^n &\to \mathrm{E}\left(\int_s^t \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2} f(X(u))du \;\Big|\; \mathcal{F}_s\right) \\ &= \int_s^t \sum_{i=1}^d \mathrm{E}\left(\frac{\partial^2}{\partial x_i^2} f(X(u)) \;\Big|\; \mathcal{F}_s\right) du,\ n \to \infty. \end{aligned}$$

Recalling that $\left|\Delta_{k,\varepsilon}^n X_l\right| \le \varepsilon$ and the third derivatives are bounded, we estimate

$$\begin{aligned} |S_3^n| &\le C\varepsilon \sum_{k=1}^n \mathrm{E}\left(\sum_{i,j=1}^d \left|\Delta_{k,\varepsilon}^n X_i \Delta_{k,\varepsilon}^n X_j\right| \;\Big|\; \mathcal{F}_s\right) \\ &\le \frac{C\varepsilon}{2} \sum_{k=1}^n \mathrm{E}\left(\sum_{i,j=1}^d \left(\left(\Delta_{k,\varepsilon}^n X_i\right)^2 + \left(\Delta_{k,\varepsilon}^n X_j\right)^2\right) \;\Big|\; \mathcal{F}_s\right) \\ &= Cd\varepsilon \sum_{k=1}^n \mathrm{E}\left(\sum_{i=1}^d \left(\Delta_{k,\varepsilon}^n X_i\right)^2 \;\Big|\; \mathcal{F}_s\right) \\ &= Cd\varepsilon \sum_{k=1}^n \mathrm{E}\left(t_{k,\varepsilon}^n - t_{k-1,\varepsilon}^n \mid \mathcal{F}_s\right) \le Cd(t-s)\varepsilon, \end{aligned}$$

where we have used [8.24].

Collecting our findings and letting $n \to \infty$, we arrive at

$$\left| \mathrm{E}(f(X(t)) \mid \mathcal{F}_s) - f(X(s)) - \frac{1}{2} \int_s^t \sum_{i=1}^{d} \mathrm{E}\left(\frac{\partial^2}{\partial x_i^2} f(X(u)) \,\Big|\, \mathcal{F}_s \right) du \right|$$
$$\leq Cd(t-s)\varepsilon.$$

Since ε is arbitrary, the statement follows. □

PROOF (Proof of theorem 8.11).– Take any $\lambda \in \mathbb{R}^d$, $s \in [0,T]$ and use lemma 8.3 to derive that

$$f(t) := \mathrm{E}\left(e^{i(\lambda, X(t))} \mid \mathcal{F}_s \right) = e^{i(\lambda, X(s))}$$
$$+ \frac{1}{2} \int_s^t \sum_{k=1}^{d} (i\lambda_k)^2 \mathrm{E}\left(e^{i(\lambda, X(u))} \,\Big|\, \mathcal{F}_s \right) du$$
$$= e^{i(\lambda, X(s))} - \frac{|\lambda|^2}{2} \int_s^t f(u) du, \ t \in [s,T].$$

Solving this equation for f, we obtain

$$f(t) = e^{i(\lambda, X(s)) - |\lambda|^2 (t-s)/2}, \ t \in [s,T],$$

whence

$$\mathrm{E}\left(e^{i(\lambda, X(t) - X(s))} \mid \mathcal{F}_s \right) = e^{-|\lambda|^2 (t-s)/2},$$

which shows that for any $t \in [s,T]$, the increment $X(t) - X(s)$ is independent of $\mathcal{F}_s$ and has the normal distribution $\mathcal{N}(0, (t-s)E_d)$, where E_d is the identity matrix. Consequently, X is a standard Wiener process in $\mathbb{R}^d$. □

8.9. Girsanov theorem

In section 5.5.8 we discussed how to turn a stochastic process with discrete time into a martingale. This section studies a similar question for continuous time. However, in contrast to general setting, considered in discrete time situation, here we address the particular case of the Wiener process and related processes.

Let $\{W(t) = (W_1(t), \ldots, W_k(t)), t \geq 0\}$ be a standard Wiener process in $\mathbb{R}^k$ and $\{h(t) = (h_1(t), \ldots, h_k(t))\}$ be an $\mathbb{R}^k$-valued process in $\mathcal{H}([0,T])$, i.e.

$\int_0^T h_i(t)^2 dt < \infty$ for $i = 1, \ldots, k$. We are interested in a measure Q such that the Wiener process with drift

$$W^h(t) = W(t) + \int_0^t h(s)ds, t \in [0, T]$$

is a Wiener process under Q.

We start by studying the martingale property of the so-called *stochastic exponential* (or *Doléans–Dade exponential*)

$$\mathcal{E}^h(t) = \exp\left\{\int_0^t \big(h(s), dW(s)\big) - \frac{1}{2}\int_0^t |h(s)|^2 \, ds\right\}.$$

THEOREM 8.12.– *Let* $h \in \mathcal{H}([0, T])$ *be such that*

$$\mathrm{E}\, \mathcal{E}^h(T) = 1. \qquad [8.25]$$

Then, $\left\{\mathcal{E}^h(t), t \in [0, T]\right\}$ *is a martingale.*

REMARK 8.10.– Since $\mathcal{E}^h(0) = 1$, it is obvious that [8.25] is also necessary for $\mathcal{E}^h$ to be a martingale.

PROOF.– Define $\tau_N = \left\{t \geq 0 : \mathcal{E}^h(t) + \int_0^t |h(s)|^2 ds > N\right\} \wedge T$; due to the continuity of the Itô integral, $\tau_N \to T$, $N \to \infty$, almost surely. Therefore, setting $h^N(t) = h(t)\mathbb{1}_{t \leq \tau_N}$ for $N \geq 1$, we have $\mathcal{E}^{h^N}(t) \to \mathcal{E}^h(t)$, $N \to \infty$, almost surely by virtue of theorem 8.4.

By the Itô formula,

$$\begin{aligned} d\mathcal{E}^{h^N}(t) &= \mathcal{E}^{h^N}(t)\left(\big(h^N(t), dW(t)\big) - \frac{1}{2}\left|h^N(t)\right|^2 dt + \frac{1}{2}\sum_{j=1}^k h_j^N(t)^2 dt\right) \\ &= \mathcal{E}^{h^N}(t)\big(h^N(t), dW(t)\big). \end{aligned}$$

Since $\int_0^T \mathcal{E}^{h^N}(t)^2 \left|h^N(t)\right|^2 dt \leq N^3$, we get that $\mathcal{E}^{h^N}(t)$ is a martingale. Then, the Fatou lemma for conditional expectations yields for any $t \in [0, T]$

$$\begin{aligned} \mathcal{E}^h(t) &= \liminf_{N\to\infty} \mathcal{E}^h(t) = \liminf_{N\to\infty} \mathrm{E}\big(\mathcal{E}^{h^N}(T) \,\big|\, \mathcal{F}_t\big) \\ &\geq \mathrm{E}\Big(\liminf_{N\to\infty} \mathcal{E}^{h^N}(T) \,\Big|\, \mathcal{F}_t\Big) = \mathrm{E}(\mathcal{E}^h(T) \mid \mathcal{F}_t). \end{aligned}$$

By integrating, we get

$$\mathrm{E}\,\mathcal{E}^h(t) \geq \mathrm{E}\,\mathcal{E}^h(T) = 1.$$

Similarly,

$$\mathrm{E}\,\mathcal{E}^h(t) \leq \mathrm{E}\,\mathcal{E}^h(0) = 1.$$

This means that all the above inequalities must be equalities, consequently,

$$\mathcal{E}^h(t) = \liminf_{N\to\infty} \mathcal{E}^h(t) = \mathrm{E}(\mathcal{E}^h(T) \mid \mathcal{F}_t)$$

a.s., whence the martingale property follows. □

There are a number of sufficient conditions supplying [8.25]. The most popular one is the *Novikov condition*, which is the subject of the following theorem.

THEOREM 8.13.– *Let $h \in \mathcal{H}([0,T])$ be such that*

$$\mathrm{E}\exp\left\{\frac{1}{2}\int_0^T |h(t)|^2\,dt\right\} < \infty. \qquad [8.26]$$

Then, [8.25] *holds true.*

PROOF.– Let $\{h^N, N \geq 1\}$ be a sequence of bounded processes from $\mathcal{H}_2([0,T])$ such that $|h^N(t)| \leq |h(t)|$ for all $N \geq 1$, $t \in [0,T]$ and $\|h^N - h\|_{\mathcal{H}_2([0,T])} \to 0$, $N \to \infty$ (e.g. we may take the sequence constructed in the proof of previous theorem).

Let us consider $a \in (0,1)$. Thanks to properties of Itô processes, $\mathcal{E}^{ah^N}(T) \to \mathcal{E}^{ah}(T)$, $N \to \infty$, in probability. Using the same reasoning as in the previous theorem, we have $\mathrm{E}\,\mathcal{E}^{ah^N}(T) = 1$. Moreover, by the Hölder inequality,

$$\begin{aligned}
\mathrm{E}\,\mathcal{E}^{ah^N}(T) &= \mathrm{E}\exp\left\{\int_0^T a\big(h^N(t), dW(t)\big) - \frac{1}{2}\int_0^T a^2\left|h^N(t)\right|^2 dt\right\} \\
&= \mathrm{E}\left(\mathcal{E}^{h^N}(T)^a \cdot \exp\left\{\frac{a-a^2}{2}\int_0^T \left|h^N(t)\right|^2 dt\right\}\right) \\
&\leq \left(\mathrm{E}\,\mathcal{E}^{h^N}(T)\right)^a \cdot \left(\mathrm{E}\exp\left\{\frac{a}{2}\int_0^T \left|h^N(t)\right|^2 dt\right\}\right)^{1-a} \qquad [8.27] \\
&\leq \left(\mathrm{E}\exp\left\{\frac{1}{2}\int_0^T |h(t)|^2\,dt\right\}\right)^{1-a}.
\end{aligned}$$

Therefore, the sequence $\left\{\mathcal{E}^{ah^N}(T), N \geq 1\right\}$ is uniformly integrable, whence

$$\mathrm{E}\,\mathcal{E}^{ah}(T) = \lim_{N\to\infty} \mathrm{E}\,\mathcal{E}^{ah^N}(T) = 1.$$

Further, similarly to [8.13],

$$1 = \mathrm{E}\,\mathcal{E}^{ah}(T) \leq \big(\mathrm{E}\,\mathcal{E}^{h}(T)\big)^{a} \cdot \left(\mathrm{E}\exp\left\{\frac{a}{2}\int_0^T |h(t)|^2\,dt\right\}\right)^{1-a},$$

whence, letting $a \to 1$, we get $\mathrm{E}\,\mathcal{E}^h(T) \geq 1$. Recalling from the proof of the previous theorem that $\mathrm{E}\,\mathcal{E}^h(T) \leq 1$, we arrive at the statement. □

Finally, let us prove the Girsanov theorem in continuous time.

THEOREM 8.14.– *Let* $h \in \mathcal{H}([0,T])$ *be such that* [8.26] *holds. We define the probability measure* Q *by*

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \mathcal{E}^{-h}(T).$$

Then, the process

$$W^h(t) = W(t) + \int_0^t h(s)ds, \quad t \in [0,T],$$

is a Wiener process on $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t\in[0,T]\}, \mathrm{Q})$.

REMARK 8.11.– The stochastic exponential

$$\mathcal{E}^{-h}(T) = \exp\left\{-\int_0^T \big(h(t), dW(t)\big) - \frac{1}{2}\int_0^T |h(t)|^2\,ds\right\}$$

is sometimes called the *Girsanov density* corresponding to the drift term h. Note that the Novikov condition is the same for h and $-h$, so theorem 8.13 supplies that $\mathrm{E}\,\mathcal{E}^{-h}(T) = 1$, i.e. it is indeed a density of probability measure under this assumption. Another important observation is that [8.25] suffices for theorem 8.14 to hold true; the proof of this fact is beyond the scope of this book.

PROOF.– Clearly, W^h is adapted to the filtration $\{\mathcal{F}_t\}$, so it is enough to show that for any $s,t \in [0,T]$ with $s < t$, the increment $W^h(t) - W^h(s)$ is independent of $\mathcal{F}_s$ and has the normal distribution $\mathcal{N}(0,(t-s)E_k)$ with respect to the measure Q. To this end, consider the conditional characteristic function

$$\mathrm{E}^{\mathrm{Q}}\big(e^{(\lambda, W^h(t)-W^h(s))} \mid \mathcal{F}_s\big) = \frac{\mathrm{E}\big(e^{(\lambda, W^h(t)-W^h(s))}\mathcal{E}^{-h}(T) \mid \mathcal{F}_s\big)}{\mathrm{E}\big(\mathcal{E}^{-h}(T) \mid \mathcal{F}_s\big)}, \lambda \in \mathbb{R}^k.$$

By theorem 8.12, $\mathrm{E}\big(\mathcal{E}^{-h}(T) \mid \mathcal{F}_s\big) = \mathcal{E}^{-h}(s)$ and

$$\begin{aligned}
&\mathrm{E}\big(e^{(\lambda, W^h(t) - W^h(s))}\mathcal{E}^{-h}(T) \mid \mathcal{F}_s\big) \\
&= \mathrm{E}\big(e^{(\lambda, W^h(t) - W^h(s))}\mathrm{E}\big(\mathcal{E}^{-h}(T) \mid \mathcal{F}_t\big) \mid \mathcal{F}_s\big) \\
&= \mathrm{E}\big(e^{(\lambda, W^h(t) - W^h(s))}\mathcal{E}^{-h}(t) \mid \mathcal{F}_s\big) \\
&= \mathcal{E}^{-h}(s)\,\mathrm{E}\left(\exp\left\{\int_s^t \big(g(u), dW(u)\big) + \int_s^t f(u)du\right\} \,\Big|\, \mathcal{F}_s\right)
\end{aligned}$$

almost surely, where $g(u) = i\lambda - h(u)$, $f(u) = i\big(\lambda, h(u)\big) - \frac{1}{2}\left|h(u)\right|^2$. It is easy to see that $f(u) = |g(u)|^2 - |\lambda|^2/2$ and that g satisfies the Novikov condition [8.26], so $\mathrm{E}(\mathcal{E}^g(t) \mid \mathcal{F}_s) = \mathcal{E}^g(s)$ almost surely (despite g being complex-valued, the proof needs just minor modification, as the imaginary part of g is constant). Therefore,

$$\begin{aligned}
&\mathrm{E}\left(\exp\left\{\int_s^t \big(g(u), dW(u)\big) + \int_s^t f(u)du\right\} \,\Big|\, \mathcal{F}_s\right) \\
&= \mathrm{E}\left(\frac{\mathcal{E}^g(t)}{\mathcal{E}^g(s)} e^{-|\lambda|^2(t-s)/2} \,\Big|\, \mathcal{F}_s\right) = e^{-|\lambda|^2(t-s)/2}.
\end{aligned}$$

Combining our findings, we get

$$\mathrm{E}^{\mathrm{Q}}\big(e^{(\lambda, W^h(t) - W^h(s))} \mid \mathcal{F}_s\big) = e^{-|\lambda|^2(t-s)/2},$$

which shows that the increment $W^h(t) - W^h(s)$ is independent of $\mathcal{F}_s$ and has the normal distribution $\mathcal{N}(0, (t-s)E_k)$ with respect to the measure Q. □

As an immediate consequence, we obtain a result on the change of measure, turning an Itô process into a martingale. This is of great importance for financial modeling (see section 9.8).

COROLLARY 8.1.– *Let X be an Itô process in $\mathbb{R}^d$ with*

$$dX(t) = \alpha(t)dt + \beta(t)dW(t), t \in [0, T],$$

where α is an $\mathbb{R}^d$-valued progressively measurable process and β is a bounded $\mathbb{R}^{d\times k}$-valued progressively measurable process. Assume that there exist some $\mathbb{R}^k$-valued processes $h \in \mathcal{H}([0,T])$ such that [8.26] *holds and $\alpha(t) = \beta(t)h(t)$ almost surely for any $t \in [0, T]$. Then, $\{X(t), t \in [0, T]\}$ is a martingale with respect to the measure* Q *with density*

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \mathcal{E}^{-h}(T).$$

REMARK 8.12.– In the one-dimensional case, where $k = d = 1$, we have $h(t) = \alpha(t)/\beta(t)$ provided that β is non-zero.

PROOF.– By assumption, we can rewrite the stochastic differential of X in the form

$$dX(t) = \beta(t)\big(h(t)dt + dW(t)\big) = \beta(t)dW^h(t).$$

By theorem 8.14, $W^h(t)$ is a Wiener process with respect to Q. Since β is bounded, it follows that $\{X(t), t \in [0,T]\}$ is a martingale with respect to Q. □

8.10. Itô representation

Let, as in the previous section, W be a standard Wiener process in $\mathbb{R}^k$. We will discuss the representation of random variables in the form

$$X = C + \int_0^T \big(\xi(s), dW(s)\big) = C + \sum_{i=1}^k \int_0^T \xi_i(s)dW_i(s)$$

with $\xi_i \in \mathcal{H}_2([0,T])$, $i = 1, \dots, k$, $C \in \mathbb{R}$. Since the Itô integral is centered, we must have $C = \mathrm{E}\, X$ in this representation.

Such representations play an important role in applications, most notably in mathematical modeling of financial markets, where they are related to replicating portfolios for contingent claims (see section 9.8 for details). The following result, establishing the existence and uniqueness of such a representation, called *Itô representation*, is thus of significant importance. Denote by $\big\{\mathcal{F}_t^W, t \in [0,T]\big\}$ the augmented natural filtration of the Wiener process W.

THEOREM 8.15.– *For any $\mathcal{F}_T^W$-measurable random variable X with $\mathrm{E}\, X^2 < \infty$, there exists a unique (up to modification) $\mathbb{R}^k$-valued process ξ, progressively measurable with respect to $\big\{\mathcal{F}_t^W, t \in [0,T]\big\}$ such that $\int_0^T \mathrm{E}|\xi(t)|^2 dt < \infty$ and the following representation holds:*

$$X = \mathrm{E}\, X + \int_0^T \big(\xi(s), dW(s)\big). \qquad [8.28]$$

PROOF.– The uniqueness follows immediately from the Itô isometry.

Denote by $\mathcal{H}_2^W$ the set of $\mathbb{R}^k$-valued process ξ, progressively measurable with respect to $\big\{\mathcal{F}_t^W, t \in [0,T]\big\}$, with $\|\xi\|^2_{\mathcal{H}_2^W} := \int_0^T \mathrm{E}|\xi(t)|^2 dt < \infty$. Similarly to $\mathcal{H}_2([0,T])$, this normed space is complete. Let also $\mathcal{I}$ be the set of square integrable $\mathcal{F}_T^W$-measurable random variables representable in the form [8.28]. It is evident that $\mathcal{I}$ is a linear subset of $\mathcal{L}_2(\Omega)$. Let us show that it is closed. Take any sequence $\{X_n, n \geq 1\} \subset \mathcal{I}$ such that $\mathrm{E}(X_n - X_0)^2 \to 0$, $n \to \infty$, for some $X_0 \in \mathcal{L}_2(\Omega)$. Then, it follows from Hölder's inequality that $\mathrm{E}\, X_n \to \mathrm{E}\, X_0$, $n \to \infty$. Writing now the Itô representation $X_n = \mathrm{E}\, X_n + I(\xi^n, [0,T])$, we get that the sequence $\{I(\xi^n, [0,T]), n \geq 1\}$ is a Cauchy sequence in $\mathcal{L}^2(\Omega)$. Therefore, thanks to the Itô isometry, the sequence $\{\xi^n, n \geq 1\}$ is a Cauchy sequence in $\mathcal{H}_2^W$. Since this space is complete, there exists some $\xi \in \mathcal{H}_2^W$ such that $\xi^n \to \xi$, $n \to \infty$, in $\mathcal{H}_2^W$. Using the Itô isometry again, we obtain $I(\xi^n, [0,T]) \to I(\xi, [0,T])$, $n \to \infty$, whence $X_0 = \mathrm{E}\, X_0 + I(\xi, [0,T])$. As a result, $\mathcal{I}$ is closed in $\mathcal{L}_2(\Omega)$.

For a deterministic function $h\colon [0,T] \to \mathbb{R}^k$ with $\int_0^T |h(t)|^2\, dt < \infty$ consider the process

$$\mathcal{E}^{ih}(t) = \exp\left\{ i \int_0^t \big(h(s), dW(s)\big) + \frac{1}{2}\int_0^t |h(s)|^2\, ds \right\}.$$

This is exactly the same stochastic exponential studied in section 8.9; despite it involving complex quantities, we can treat it similarly. That is, we can use the Itô formula, which easily generalizes to complex-valued functions, to get

$$\begin{aligned} d\mathcal{E}^{ih}(t) &= \mathcal{E}^{ih}(t) \left(i\big(h(t), dW(t)\big) + \frac{1}{2}|h(t)|^2\, dt + \frac{1}{2}\sum_{j=1}^k (ih_j(t))^2\, dt \right) \\ &= i\mathcal{E}^{ih}(t)\big(h(t), dW(t)\big). \end{aligned}$$

In particular,

$$\mathcal{E}^{ih}(T) = 1 + i\int_0^T \mathcal{E}^{ih}(t)\big(h(t), dW(t)\big). \qquad [8.29]$$

Since $\left\{\mathcal{E}^{ih}(t)h(t), t \in [0,T]\right\}$ is progressively measurable with respect to the natural filtration $\{\mathcal{F}_t^W, t \in [0,T]\}$ of the Wiener process and

$$\int_0^T \mathrm{E}\left|\mathcal{E}^{ih}(t)h(t)\right|^2 dt \leq \int_0^T |h(t)|^2\, dt \cdot \exp\left\{\int_0^T |h(t)|^2\, dt\right\} < \infty,$$

we get $\operatorname{Re}\mathcal{E}^{ih}(T) \in \mathcal{I}$ and $\operatorname{Im}\mathcal{E}^{ih}(T) \in \mathcal{I}$. By linearity, $\mathcal{I}$ contains variables of the form $\sin\left(h(t), W(t)\right)$, $\cos\left(h(t), W(t)\right)$. Taking h piecewise constant, we get that for any $n \geq 1$ and any $t_1, \dots, t_n \in [0,T]$, $\mathcal{I}$ contains all trigonometric polynomials of $W_i(t_1), \dots, W_i(t_n)$, $i = 1, \dots, k$. By theorem A1.5, the set of such polynomials is dense in the space of square-integrable random variables of the form $F(W(t_1), \dots, W(t_n))$, so they belong to $\mathcal{I}$ as well. In particular, $\mathbb{1}_A(W(t_1), \dots, W(t_n)) \in \mathcal{I}$ for any $A \in \mathcal{B}(\mathbb{R}^n)$, $n \geq 1$ and $t_1, \dots, t_n \in [0,T]$. Hence, similarly to theorem A2.3, we get that $\mathcal{I} = \mathcal{L}_2(\Omega, \mathcal{F}_T^W)$, as needed. □

The left-hand side of [8.28] is a martingale, as a function of T. Thus, as a corollary to theorem 8.15, we get the so-called *martingale representation*.

THEOREM 8.16.– *For any $\mathcal{F}^W$-martingale $\{M(t), t \in [0,T]\}$ with $\mathrm{E}\, M(T)^2 < \infty$, there exists a unique (up to modification) $\mathbb{R}^k$-valued process ξ, progressively measurable with respect to $\{\mathcal{F}_t^W, t \in [0,T]\}$, satisfying $\int_0^T \mathrm{E}|\xi(t)|^2 dt < \infty$ and such that*

$$M(t) = M(0) + \int_0^t \big(\xi(s), dW(s)\big)$$

almost surely for each $t \in [0,T]$.

PROOF.– Follow from theorem 8.15 by setting $X = M(T)$ and taking the conditional expectation with respect to $\mathcal{F}_t^W$. □

Theorem 8.15 asserts only existence of some integrand in Itô representation, and gives almost no idea how this integrand can be found. In some cases, the answer can be given in terms of the *stochastic derivative* (or Malliavin derivative). We will give only basic information; the details may be found in [NUA 06]. For technical simplicity, we will treat only the case $k = 1$; the generalization to the multi-dimensional case is straightforward.

We call an $\mathcal{F}_T^W$-measurable random variable X *cylindrical* if it can be represented in the form $X = f\big(W(t_1), \dots, W(t_n)\big)$ for some $t_i \in [0,T]$, $1 \le i \le n$, and infinitely differentiable compactly supported function f. The *stochastic derivative* of X is the stochastic process

$$D_t X = \sum_{i=1}^n \frac{\partial}{\partial x_i} f_i'\big(W(t_1), \dots, W(t_n)\big) \mathbb{1}_{[0,t_i]}(t),\ t \in [0,T];$$

we will consider it as an element of $\mathcal{H}_2([0,T])$. In particular, the indistinguishable processes will be identified. We define the following norm:

$$\|X\|_{\mathbb{D}_{1,2}}^2 = \mathrm{E}X^2 + \int_0^T \mathrm{E}(D_t X)^2\, dt.$$

The space $\mathbb{D}_{1,2} \subset L_2(\Omega)$ is defined as a completion of the set of cylindrical random variables with respect to the norm $\|\cdot\|_{\mathbb{D}_{1,2}}$. It can be checked that the operator D is closable, so it admits a unique extension to $\mathbb{D}_{1,2}$.

The following properties of the stochastic derivatives can be checked easily from this definition:

– linearity: $D_t(aX + bY) = a\, D_t X + b\, D_t Y$, $a, b \in \mathbb{R}$, $X, Y \in \mathbb{D}_{1,2}$;

– product rule: $D_t(XY) = X\,D_tY + Y\,D_tX$ if $X, Y, XY \in \mathbb{D}_{1,2}$;

– chain rule: for any $X_1, \dots, X_m \in \mathbb{D}_{1,2}$ and any function $h \in C^1(\mathbb{R}^m)$ of at most linear growth,

$$D_t h(X_1, \dots, X_m) = \sum_{i=1}^{m} \frac{\partial}{\partial x_i} h(X_1, \dots, X_k) D_t X_k.$$

For elements of $\mathbb{D}_{1,2}$, it is possible to write the Itô representation explicitly in terms of stochastic derivative via the so-called *Clark–Ocone formula* (also called Clark formula or Clark–Ocone–Haussmann formula by some authors).

THEOREM 8.17.– *For any* $X \in \mathbb{D}_{1,2}$,

$$X = \mathrm{E}X + \int_0^T \mathrm{E}(D_t X \mid \mathcal{F}_t^W) dW(t) \qquad [8.30]$$

almost surely.

PROOF.– The set $\mathcal{J}$ of variables from $\mathbb{D}_{1,2}$ satisfying [8.30] is clearly linear. Further, let $\{X_n, n \geq 1\} \subset \mathcal{J}$ and $\|X_n - X_0\|_{\mathbb{D}_{1,2}} \to 0$, $n \to \infty$. Then, by Jensen's inequality, $\mathrm{E}X_n \to \mathrm{E}X_0$ and

$$\int_0^t \mathrm{E}\left(\mathrm{E}(D_t X_n \mid \mathcal{F}_t^W) - \mathrm{E}(D_t X_0 \mid \mathcal{F}_t^W)\right)^2 dt \to 0$$

as $n \to \infty$. Thanks to Itô's isometry,

$$\int_0^T \mathrm{E}(D_t X_n \mid \mathcal{F}_t^W) dW(t) \to \int_0^T \mathrm{E}(D_t X_0 \mid \mathcal{F}_t^W) dW(t),\ n \to \infty.$$

Therefore, X_0 satisfies [8.30] as well, so $\mathcal{J}$ is closed in $\mathbb{D}_{1,2}$.

Equation [8.29] establishes [8.30] for the variables $\mathcal{E}^{ih}(T)$; therefore, by linearity, $\mathcal{J}$ contains all trigonometric polynomials of values of W taken at different points. Taking for granted (a not-so-trivial fact) that the set of such polynomials is dense in $\mathbb{D}_{1,2}$, we get that $\mathcal{J} = \mathbb{D}_{1,2}$, as claimed. □

As a corollary, we have the following *stochastic integration by parts formula*. It may be used to define an extension of the Itô integral to non-adapted (*anticipative*) integrands, which is called *Skorokhod integral*; for more information, see [NUA 06].

COROLLARY 8.2.– *Assume that* $X \in \mathbb{D}^{1,2}$ *and* ξ *is progressively measurable with respect to* $\{\mathcal{F}_t^W, t \in [0,T]\}$ *process, satisfying* $\int_0^T \mathrm{E}|\xi(t)|^2 dt < \infty$. *Then,*

$$\mathrm{E}\left(X \int_0^T \xi(t) dW(t)\right) = \int_0^T \mathrm{E}\big(\xi(t)\, D_t X\big) dt.$$

PROOF.– Using theorem 8.17,

$$\begin{aligned}
&\mathrm{E}\left(X \int_0^T \xi(t) dW(t)\right) \\
&= \mathrm{E}\left(\left(\mathrm{E}X + \int_0^T \mathrm{E}(D_t X \mid \mathcal{F}_t^W) dW(t)\right) \int_0^T \xi(t) dW(t)\right) \\
&= \int_0^T \mathrm{E}\big(\mathrm{E}(D_t X \mid \mathcal{F}_t^W)\xi(t)\big) dt = \int_0^T \mathrm{E}(\xi_t\, D_t X) dt,
\end{aligned}$$

where we have used the $\mathcal{F}_t^W$-measurability of ξ_t and property 6 from theorem 8.2. □

9

Stochastic Differential Equations

9.1. Definition, solvability conditions, examples

As explained in section 8.1, the main reason for stochastic integration lies in the necessity of modeling dynamical systems with randomness. This is done through stochastic differential equations, which are the main object of this chapter.

To keep things simpler, we will consider a finite interval $[0,T]$; in the case of whole half-line $[0,+\infty)$, only marginal changes are needed. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in [0,T]\}, \mathrm{P})$ be a stochastic basis and W be a standard $\mathbb{R}^k$-valued Wiener process on this basis. Assume that we have deterministic functions $a\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $b\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times k}$, which serve as coefficients for the equation, and an $\mathbb{R}^d$-valued $\mathcal{F}_0$-measurable random variable $X(0)$, serving as an initial condition for the equation. The corresponding *stochastic differential equation* is

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t),\ t \in [0,T], \qquad [9.1]$$

with the initial condition $X(0)$. It may be written in coordinate form as

$$dX_i(t) = a_i(t, X(t))dt + \sum_{j=1}^{k} b_{ij}(t, X(t))dW_j(t),\ t \in [0,T], i = 1, \ldots, d,$$

so it is in fact a system of (stochastic differential) equations. Nevertheless, we will follow the tradition, calling it an equation. It is worth mentioning the similarity of this equation and [8.3], obtained by heuristic reasoning. The functions a and b are called the *drift* and *diffusion* coefficients, respectively.

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

DEFINITION 9.1.– *A (strong)* solution *to equation* [9.1] *is an Itô process* $\{X(t), t \geq 0\}$ *in* $\mathbb{R}^d$ *such that its stochastic differential satisfies* [9.1]. *In other words, it is a progressively measurable process satisfying*

$$X(t) = X(0) + \int_0^t a(s, X(s))dt + \int_0^t \sigma(s, X(s))dW(s),\ t \geq 0, \qquad [9.2]$$

almost surely for any $t \in [0, T]$.

Since we have agreed on assuming Itô processes to be continuous, the same agreement will be in force for solutions of stochastic differential equations.

9.1.1. *Existence and uniqueness of solution*

Let us turn now to the question of the solvability of stochastic differential equations. We will assume that the coefficients a, b are measurable and satisfy the following conditions with some non-random constant $K > 0$:

– *linear growth*: for any $t \in [0, T]$, $x \in \mathbb{R}^d$

$$|a(t, x)| + |b(t, x)| \leq K(1 + |x|). \qquad [9.3]$$

– *Lipschitz continuity*: for any $t \in [0, T]$, $x, y \in \mathbb{R}^d$

$$|a(t, x) - a(t, y)| + |b(t, x) - b(t, y)| \leq K\, |x - y|\,. \qquad [9.4]$$

As in the previous section, $|\cdot|$ denotes absolute value, vector norm or matrix norm, depending on the context. We will also use the symbol C for a generic constant; its value might change between lines.

We start with a result establishing *a priori* estimates for the solution.

THEOREM 9.1.– *Let* X *be a solution to equation* [9.1] *satisfying* [9.3] *with square-integrable initial condition:* $\mathrm{E}|X(0)|^2 < \infty$. *Then,*

$$\mathrm{E} \sup_{t\in[0,T]} |X(t)|^2 \leq C\left(1 + \mathrm{E}|X(0)|^2\right)$$

with constant C *depending only on* K *and* T.

REMARK 9.1.– The assumption of square integrability of $X(0)$ is not essential and is made just for technical simplicity. It is possible to prove a similar estimate for conditional expectation:

$$\mathrm{E}\left(\sup_{t\in[0,T]} |X(t)|^2 \,\middle|\, \mathcal{F}_0 \right) \leq C\big(1 + |X(0)|^2\big),$$

which is sufficient for further development.

PROOF.– We denote

$$X^*(t) = \sup_{s\in[0,t]} |X(s)|$$

and set $\tau_N = \inf\{t \geq 0 : X^*(t) \geq N\} \wedge T$, $N \geq 1$. We estimate

$$X^*(t) \leq |X(0)| + \sup_{s\in[0,t]} \int_0^s |a(u, X(u))|\, du + \sup_{s\in[0,t]} \left| \int_0^s b(u, X(u)) dW(u) \right|$$

$$\leq |X(0)| + K \int_0^t \big(1 + |X(u)|\,\big) du + \sup_{s\in[0,t]} \left| \int_0^s b(u, X(u)) dW(u) \right|$$

$$\leq |X(0)| + KT + K \int_0^t X^*(u) du + \sup_{s\in[0,t]} \left| \int_0^s b(u, X(u)) dW(u) \right|,$$

whence

$$\mathrm{E}X^*(t \wedge \tau_N)^2 \leq 4 \left(\mathrm{E}\,|X(0)|^2 + K^2T^2 + K^2\mathrm{E} \left(\int_0^{t\wedge\tau_N} X^*(u) du \right)^2 \right.$$
$$\left. + \,\mathrm{E} \sup_{s\in[0,t]} \left| \int_0^{s\wedge\tau_N} b(u, X(u)) dW(u) \right|^2 \right).$$

Using the Cauchy–Schwarz inequality, we get

$$\left(\int_0^{t\wedge\tau_N} X^*(u) du \right)^2 \leq \left(\int_0^t X^*(u \wedge \tau_N) du \right)^2 \leq t \int_0^t X^*(u \wedge \tau_N)^2 du.$$

Theorem 8.9 through [8.4] implies

$$\mathrm{E} \sup_{s\in[0,t]} \left| \int_0^{s\wedge\tau_N} b(u, X(u)) dW(u) \right|^2$$

$$= \mathrm{E} \sup_{s\in[0,t]} \left| \int_0^s b(u, X(u)) \mathbb{1}_{u\leq\tau_N} dW(u) \right|^2 \leq \mathrm{E} \left(\int_0^t |b(u, X(u))|^2 \mathbb{1}_{u\leq\tau_N} du \right]$$

$$\leq 2K^2 \int_0^t \mathrm{E}\left(\left(1 + |X(u)|^2\right) \mathbb{1}_{u\leq\tau_N}\right) du \leq C \left(1 + \int_0^t \mathrm{E}X^*(u \wedge \tau_N)^2 du \right).$$

Combining these estimates, we get

$$\mathrm{E}X^*(t \wedge \tau_N)^2 \leq C \left(1 + \mathrm{E}\,|X(0)|^2 + \int_0^t \mathrm{E}X^*(u \wedge \tau_N)^2 du \right),$$

whence by theorem A1.12,

$$\mathrm{E}X^*(T\wedge\tau_N)^2 \le C\left(1+\mathrm{E}\left|X(0)\right|^2\right).$$

The proof is concluded by letting $N\to\infty$, noting that $X^*(T)<\infty$ thanks to continuity, and appealing to the Fatou lemma. □

With the *a priori* estimates at hand, we are now in a position to prove the result about the unique solvability of a stochastic differential equation.

THEOREM 9.2.– *Let the coefficients of equation* [9.1] *satisfy* [9.3], [9.4] *and* $X(0)$ *be such that* $\mathrm{E}|X(0)|^2<\infty$. *Then, the equation has a unique solution* X. *Moreover,* $X\in\mathcal{H}_2([0,T])$.

PROOF.– The idea is to use the Banach fixed-point theorem. Take some $\lambda>0$ (to be chosen later) and introduce the norm

$$\|Y\|_\lambda^2 := \int_0^T e^{-\lambda t}\mathrm{E}|Y(t)|^2dt,\ Y\in\mathcal{H}_2([0,T]).$$

It is easy to see that $e^{-\lambda T}\left\|Y\right\|^2_{\mathcal{H}_2([0,T])}\le\left\|Y\right\|^2_\lambda\le\left\|Y\right\|^2_{\mathcal{H}_2([0,T])}$, so this norm is equivalent to $\left\|\cdot\right\|_{\mathcal{H}_2([0,T])}$; however, this will not play a crucial role in our argument.

For $Y\in\mathcal{H}_2([0,T])$, we define the process $F(Y)$ by

$$F(Y)(t) = X(0)+\int_0^t a(s,Y(s))ds+\int_0^t b(s,Y(s))dW(s),\ t\in[0,T].$$

Our first aim is to show that $F(Y)\in\mathcal{H}_2([0,T])$. Using the Cauchy–Schwarz inequality and the Itô isometry, we estimate

$$\begin{aligned}\mathrm{E}|F(Y)(t)|^2 &\le 3\left(\mathrm{E}|X(0)|^2+\mathrm{E}\left|\int_0^t a(s,Y(s))ds\right|^2\right.\\ &\qquad\left.+\,\mathrm{E}\left|\int_0^t b(s,Y(s))dW(s)\right|^2\right)\\ &\le 3\left(\mathrm{E}|X(0)|^2+t\mathrm{E}\int_0^t|a(s,Y(s))|^2ds+\mathrm{E}\int_0^t|b(s,Y(s))|^2ds\right)\\ &\le C\left(1+\int_0^t\mathrm{E}(1+|Y(s)|^2)ds\right)\le C(1+\|Y\|^2_{\mathcal{H}_2([0,T])}).\end{aligned}$$

Hence, $\|F(Y)\|^2_{\mathcal{H}_2([0,T])} < \infty$, as announced.

Similarly, for $Y', Y'' \in \mathcal{H}_2([0,T])$,

$$\mathrm{E}|F(Y')(t) - F(Y'')(t)|^2 \leq 2\left(\mathrm{E}\left|\int_0^t \big(a(s,Y'(s)) - a(s,Y''(s))\big)ds\right|^2\right.$$

$$\left. + \mathrm{E}\left|\int_0^t \big(b(s,Y'(s)) - b(s,Y''(s))\big)dW(s)\right|^2\right)$$

$$\leq 2\left(t\mathrm{E}\int_0^t |a(s,Y'(s)) - a(s,Y''(s))|^2\, ds\right.$$

$$\left. + \mathrm{E}\int_0^t |b(s,Y'(s)) - b(s,Y''(s))|^2\, ds\right)$$

$$\leq 2K^2(T+1)\int_0^t \mathrm{E}\,|Y'(s) - Y''(s)|^2\, ds.$$

Therefore,

$$\|F(Y') - F(Y'')\|^2_\lambda \leq 2K^2(T+1)\int_0^T e^{-\lambda t}\int_0^t \mathrm{E}\,|Y'(s) - Y''(s)|^2\, ds\, dt$$

$$= 2K^2(T+1)\int_0^T \mathrm{E}\,|Y'(s) - Y''(s)|^2 \int_s^T e^{-\lambda t} dt\, ds$$

$$\leq \frac{2K^2(T+1)}{\lambda}\int_0^T e^{-\lambda s}\mathrm{E}\,|Y'(s) - Y''(s)|^2\, ds = \frac{2K^2(T+1)}{\lambda}\,\|Y' - Y''\|^2_\lambda\,.$$

Setting $\lambda = 4K^2(T+1)$, we get that F is a contractive map on $\mathcal{H}_2([0,T])$ with respect to $\|\cdot\|_\lambda$. Therefore, by the Banach fixed-point theorem, there exists a unique process $X \in \mathcal{H}_2([0,T])$ satisfying $X = F(X)$ and thus solving [9.1]. It remains to note that, thanks to theorem 9.1, any solution to [9.1] must belong to $\mathcal{H}_2([0,T])$. □

REMARK 9.2.– A similar existence and uniqueness result may be shown for an equation with random coefficients. Specifically, let the coefficients $a\colon [0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}^d$ and $b\colon [0,T]\times\mathbb{R}^d\times\Omega\to\mathbb{R}^{d\times k}$ be "adapted" in the sense that for any progressively measurable process $Y = \{Y(t), t\in[0,T]\}$ the processes $\{a(t,Y(t,\omega),\omega), t\in[0,T]\}$ and $\{b(t,Y(t,\omega),\omega), t\in[0,T]\}$ are also progressively measurable. Then, assuming that [9.3] and [9.4] hold with non-random constant K, the corresponding stochastic differential equation has a unique solution. The linear growth assumption [9.3] may be further relaxed to the requirement that $\int_0^T \mathrm{E}\big(|a(t,0)|^2 + |b(t,0)|^2\big)dt < \infty$ or even that $\int_0^T \big(|a(t,0)| + |b(t,0)|^2\big)dt < \infty$ almost surely; however, in the latter case, the proof will be more involved.

9.1.2. *Some special stochastic differential equations*

Itô process. An Itô process with $dX(t) = a(t)dt + b(t)dW(t)$ may be viewed as a solution to a stochastic differential equation with coefficients independent of x.

Linear equation. Let $d = k = 1$ and $a(t,x) = \mu(t)x$, $b(t,x) = \sigma(t)x$ be linear. The functions α and β may be non-deterministic or, more precisely, progressively measurable processes with $\int_0^T \left(|\mu(t)| + \sigma(t)^2\right) dt < \infty$. The corresponding equation reads

$$dX(t) = X(t)\big(\mu(t)dt + \sigma(t)dW(t)\big). \qquad [9.5]$$

In the deterministic case, this can be solved by dividing over $X(t)$ and noting that $\frac{dX(t)}{X(t)} = d(\log|X(t)|)$. This will not work in the stochastic case, as the chain rule (Itô formula) is different. However, the Itô formula is the correct approach, as we can write (not bothering for the moment about non-differentiability of $\log|x|$ at 0)

$$\begin{aligned} d\log|X(t)| &= \frac{dX(t)}{X(t)} + \frac{1}{2}\cdot\frac{-1}{X(t)^2}\sigma(t)^2X(t)^2dt \\ &= \Big(\mu(t) - \frac{\sigma(t)^2}{2}\Big)dt + \sigma(t)dW(t), \end{aligned}$$

thus obtaining, as in the previous example, an equation with coefficients independent of x. Clearly, this is solved by

$$\log|X(t)| = \log|X(0)| + \int_0^t \Big(\mu(s) - \frac{\sigma(s)^2}{2}\Big)ds + \int_0^t \sigma(s)dW(s),$$

whence, thanks to continuity

$$X(t) = X(0)\exp\left\{\int_0^t \Big(\mu(s) - \frac{\sigma(s)^2}{2}\Big)ds + \int_0^t \sigma(s)dW(s)\right\}. \qquad [9.6]$$

We can apply the Itô formula to prove that the process given by [9.6] solves [9.5]. In the case of constant μ and σ, the solution is further simplified to

$$X(t) = X(0)e^{(\mu - \frac{\sigma^2}{2})t + \sigma W(t)};$$

such a process is called a *geometric Brownian motion*.

In the case where $k > 1$, a formula similar to [9.6] is valid, with $\sigma(t)^2$ replaced by $|\sigma(t)|^2$. However, for $d > 1$, there is no hope, in general, to get such a nice expression.

See also the discussion in section 9.8 concerning the diffusion model of financial markets, where linear stochastic differential equations arise.

Semi-linear equation. Consider now a scalar (i.e. $d = k = 1$) equation, called the *Langevin equation*:

$$dX(t) = \theta\big(X(t) - \mu\big)dt + \sigma\, dW(t), \qquad [9.7]$$

where θ, μ and σ are fixed parameters. As in the previous example, it is possible to consider time-dependent (even random) coefficients with minimal changes, but we will stick to the simplest situation, which is already sufficiently enlightening.

In a deterministic setting, an equation like [9.7] is usually solved using the variation of constants. Fortunately, this is also possible in the stochastic setting without any substantial difference. That said, the solution to a homogeneous version $dZ(t) = \theta Z(t)$ of [9.7] is $Z(t) = Ce^{\theta t}$. Letting the constant C vary, we look for a solution to [9.7] in the form $X(t) = C(t)e^{\theta t}$. Since $e^{\theta t}$ has the usual differential, differentiation of the product does not differ from the deterministic setting:

$$dX(t) = \theta C(t)e^{\theta t}dt + e^{\theta t}dC(t) = \theta X(t)dt + e^{\theta t}dC(t).$$

Substituting this into [9.7] yields

$$e^{\theta t}dC(t) = -\theta\mu\, dt + \sigma\, dW(t),$$

whence, noting that $C(0) = X(0)$,

$$C(t) = X(0) - \mu + \mu e^{-\theta t} + \sigma \int_0^t e^{-\theta s} dW(s).$$

As a result,

$$X(t) = \mu + (X(0) - \mu)e^{\theta t} + \sigma \int_0^t e^{\theta(t-s)} dW(s).$$

This Gaussian process is called the *Ornstein–Uhlenbeck process*. It is often considered only for $\theta < 0$. In this case, it is easy to see that the mean of X converges to μ and the variance to $\frac{\sigma^2}{2\theta}$. Therefore, in the long run, the process tends to oscillate around μ, which is called the *mean-reverting property*. This property is of particular interest in financial mathematics, where the Ornstein–Uhlenbeck process is used to model interest rates and stochastic volatility; this is the so-called *Vasicek model*.

If the initial condition is $X(0) = 0$, then the resulting process is the one-sided Ornstein–Uhlenbeck process from Definition 3.9; if $X(0)$ is a random variable having the Gaussian distribution $\mathcal{N}(\mu, \frac{\sigma^2}{2\theta})$, then the resulting process coincides on the positive half-line with the two-sided Ornstein–Uhlenbeck process from Definition 3.10.

Brownian bridge. Consider again a scalar equation

$$dX(t) = -\frac{X(t)}{T-t}dt + dW(t),\ t \in [0,T),$$

with the initial condition $X(0) = 0$. Despite the fact that its drift coefficient does not satisfy [9.3] and [9.4] near T, it still does satisfy these assumptions on $[0,T']$ for any $T' \in (0,T)$, which implies that a unique solution on $[0,T)$ exists. Let us look at the differential of $X(t)/(T-t)$. By the product differentiation rule [8.20],

$$d\left(\frac{X(t)}{T-t}\right) = \frac{X(t)}{(T-t)^2}dt - \frac{X(t)}{(T-t)^2}dt + \frac{1}{T-t}dW(t),\ t \in [0,T),$$

whence

$$X(t) = (T-t)\int_0^t \frac{1}{T-u}dW(u),\ t \in [0,T).$$

Thanks to the results of section 3.5, this is a centered Gaussian process with the covariance

$$\begin{aligned}\mathrm{E}X(t)X(s) &= (T-t)(T-s)\mathrm{E}\left(\int_0^t \frac{1}{T-u}dW(u)\int_0^s \frac{1}{T-u}dW(u)\right)\\ &= (T-t)(T-s)\int_0^s \frac{1}{(T-u)^2}du\\ &= (T-t)(T-s)\left(\frac{1}{T-s} - \frac{1}{T}\right) = s - \frac{ts}{T},\ s \le t.\end{aligned}$$

In view of symmetry,

$$\mathrm{E}\big(X(t)X(s)\big) = t \wedge s - \frac{ts}{T},\ t,s \in [0,T),$$

so X is the Brownian bridge between points 0 and T in time and points 0 and 0 in space, considered in section 3.4.4; in particular, we can define $X(T) = 0$.

9.2. Properties of solutions to stochastic differential equations

As mentioned previously, a solution to stochastic differential equation [9.1] is continuous and square integrable. Let us derive some further properties. We will always assume that [9.3] and [9.4] are satisfied.

THEOREM 9.3.– *Let* $\mathrm{E}\,|X(0)|^{2p} < \infty$ *for some* $p \geq 1$. *Then*

$$\mathrm{E} \sup_{t\in[0,T]} |X(t)|^{2p} \leq C \left(1 + \mathrm{E}|X(0)|^{2p}\right)$$

and for any $t, s \in [0, T]$

$$\mathrm{E}|X(t) - X(s)|^{2p} \leq C\,|t - s|^{p} \left(1 + \mathrm{E}|X(0)|^{2p}\right)$$

with constant C *depending only on* K, T, *and* p.

PROOF.– The first inequality is derived similarly to theorem 9.1: we define $X^*(t) = \sup_{s\in[0,t]} |X(s)|$, $\tau_N = \inf\{t \geq 0 : X^*(t) \geq N\} \wedge T$, $N \geq 1$ and estimate

$$\mathrm{E}X^*(t \wedge \tau_N)^{2p} \leq 4^{p-1} \left(\mathrm{E}\,|X(0)|^{2p} + K^{2p}T^{2p} + K^{2p}\mathrm{E}\left(\int_0^t X^*(u \wedge \tau_N)du\right)^{2p} \right.$$
$$\left. + \,\mathrm{E} \sup_{s\in[0,t]} \left|\int_0^s b(u, X(u))\mathbb{1}_{\tau_N \leq u} dW(u)\right|^{2p}\right).$$

The Hölder inequality gives

$$\left(\int_0^t X^*(u \wedge \tau_N)du\right)^{2p} \leq t^{2p-1} \int_0^t X^*(u \wedge \tau_N)^{2p}du.$$

Using theorem 8.9 and Hölder inequalities, we get

$$\mathrm{E} \sup_{s\in[0,t]} \left|\int_0^s b(u, X(u))\mathbb{1}_{u\leq\tau_N} dW(u)\right|^{2p}$$
$$\leq \mathrm{E}\left(\int_0^t |b(u, X(u))|^2 \mathbb{1}_{u\leq\tau_N} du\right)^{p}$$
$$\leq K^p t^{p-1} \int_0^t \mathrm{E}\left(\left(1 + |X(u)|\right)^{2p} \mathbb{1}_{u\leq\tau_N}\right) du$$
$$\leq C\left(1 + \int_0^t \mathrm{E}X^*(u \wedge \tau_N)^{2p}du\right).$$

Summing up, we obtain

$$\mathrm{E}X^*(t\wedge\tau_N)^{2p} \le C\left(1+\mathrm{E}\left|X(0)\right|^{2p} + \int_0^t \mathrm{E}X^*(u\wedge\tau_N)^{2p}du\right),$$

which, as in theorem 9.1, leads to the first inequality through the application of the Grönwall and Fatou lemmas.

The second inequality is proved similarly using the Hölder inequality and theorem 8.9:

$$\begin{aligned}\mathrm{E}|X(t)-X(s)|^{2p} &\le 2^{2p-1}\left(\mathrm{E}\left|\int_s^t a(u,X(u))du\right|^{2p}\right.\\ &\quad \left. + \mathrm{E}\left|\int_s^t b(u,X(u))dW(u)\right|^{2p}\right)\\ &\le C\left((t-s)^{2p-1}\mathrm{E}\int_s^t \left|a(u,X(u))\right|^{2p}du + \mathrm{E}\left(\int_s^t \left|b(u,X(u))\right|^2 du\right)^p\right)\\ &\le C\left((t-s)^{2p-1}\int_s^t \mathrm{E}\left(1+\left|X(u)\right|^{2p}\right)du\right.\\ &\quad \left. + (t-s)^{p-1}\int_s^t \mathrm{E}\left(1+\left|X(u)\right|^{2p}\right)du\right)\\ &\le C|t-s|^p \sup_{u\in[0,T]} \mathrm{E}\left|X(u)\right|^{2p} \le C\left|t-s\right|^p\left(1+\mathrm{E}|X(0)|^{2p}\right). \qquad \square\end{aligned}$$

Further, we will focus on the regularity of the solution with respect to the initial data. Consider equation [9.1] on a smaller interval:

$$dX(s) = a(s,X(s))ds + b(s,X(s))dW(s),\ s\in[t,T],$$

with a non-random initial condition $X(t)=x\in\mathbb{R}^d$. We denote the unique solution of this equation by $X_{t,x}=\{X_{t,x}(s), s\in[t,T]\}$.

THEOREM 9.4.– *1) For any $t',t''\in[0,T], x\in\mathbb{R}^d$, $p\ge 1$,*

$$\mathrm{E}\sup_{s\in[t'\vee t'',T]}\left|X_{t',x}(s)-X_{t'',x}(s)\right|^{2p} \le C\left|t'-t''\right|^p\left(1+\left|x\right|^{2p}\right)$$

with constant C depending only on K, T and p.

2) For any $t\in[0,T], x',x''\in\mathbb{R}^d$, $p\ge 1$,

$$\mathrm{E}\sup_{s\in[t,T]}\left|X_{t,x'}(s)-X_{t,x''}(s)\right|^{2p} \le C\left|x'-x''\right|^{2p}$$

with constant C depending only on K, T and p.

PROOF.– 1) Assume without any loss of generality that $t' \leq t''$ and denote $\Delta(t) = \sup_{s\in[t'',t]} |X_{t',x}(s) - X_{t'',x}(s)|$, $t \in [t'', T]$. We have

$$\mathrm{E}\Delta(t'')^{2p} = \mathrm{E}\left|X_{x,t'}(t'') - x\right|^{2p}$$
$$= \mathrm{E}\left|X_{x,t'}(t'') - X_{x,t'}(t')\right|^2 \leq C|t' - t''|^p \left(1 + |x|^{2p}\right)$$

by theorem 9.3 and

$$\mathrm{E}\Delta(t)^{2p} \leq 3^{2p-1} \Bigg(\mathrm{E}\left|X_{x,t'}(t'') - x\right|^{2p}$$
$$+ \mathrm{E} \sup_{s\in[t'',t]} \left|\int_{t''}^{s} \big(a(u, X_{t',x}(u)) - a(u, X_{t'',x}(u))\big)du\right|^{2p}$$
$$+\mathrm{E} \sup_{s\in[t'',t]} \left|\int_{t''}^{s} \big(b(u, X_{t',x}(u)) - b(u, X_{t'',x}(u))\big)dW(u)\right|^{2p}\Bigg)$$
$$\leq C\Bigg(|t' - t''|^p \left(1 + |x|^{2p}\right)$$
$$+ \mathrm{E}\int_{t''}^{t} |a(u, X_{t',x}(u)) - a(u, X_{t'',x}(u))|^{2p}\, du$$
$$+ \mathrm{E}\int_{t''}^{t} |b(u, X_{t',x}(u)) - b(u, X_{t'',x}(u))|^{2p}\, du\Bigg)$$
$$C\left(|t' - t''|^p \left(1 + |x|^{2p}\right) + \int_{t''}^{t} \mathrm{E}\left|X_{t',x}(u) - X_{t'',x}(u)\right|^{2p} du\right)$$
$$\leq C\left(|t' - t''|^p \left(1 + |x|^{2p}\right) + \int_{t''}^{t} \mathrm{E}\Delta(u)^{2p} du\right)$$

for any $t \in [t'', T]$. Applying the Grönwall lemma, we arrive at the desired inequality.

2) Similarly, denoting $\Delta_1(s) = \sup_{u\in[t,s]} |X_{t,x'}(u) - X_{t,x''}(u)|$ so that $\Delta_1(t) = |x' - x''|$, we get

$$\mathrm{E}\Delta_1(s)^{2p} \leq C\left(|x' - x''|^{2p} + \int_{t}^{s} \mathrm{E}\Delta_1(u)^{2p} du\right)$$

and conclude by using the Grönwall lemma. □

This statement allows us to deduce the existence of a modification of $X_{t,x}(s)$, jointly continuous in t, x, s. We give the formulation below, omitting the proof, which uses the same idea as that of theorem 6.9.

THEOREM 9.5.– *There exists a modification of* $\{X_{t,x}(s), x \in \mathbb{R}^d, 0 \le t \le s \le T\}$*, jointly continuous in* t, x, s*. Moreover, for any* $\alpha \in (0, 1/2)$ *and* $\beta \in (0, 1)$*, this modification satisfies*

$$|X_{t',x'}(s') - X_{t'',x''}(s'')| \le C(R, \omega) \left(|t' - t''|^\alpha + |s' - s''|^\alpha + |x' - x''|^\beta \right)$$

for any $R > 0$*,* $x', x'' \in \mathbb{R}^d$ *such that* $|x'| \le R$*,* $|x''| < R$ *and* $t', t'', s', s'' \in [0, T]$ *such that* $t' \vee t'' \le s' \wedge s''$*. In other words, this modification is Hölder continuous of an order up to* $1/2$ *in* t, s *and locally Hölder continuous of an order up to* 1 *in* x*.*

Let us now turn to the Markov property, discussed previously in Chapter 7. In layman's terms, it means that the future evolution of a process is independent, given the present state, of its past. In the context of stochastic differential equations, it means that if we use the state as some moment of a solution to the stochastic differential equation as an initial condition at that moment, then we will reproduce its future path.

Similarly to above, for $t \in [0, T]$ and an $\mathcal{F}_t$-measurable random vector ξ in $\mathbb{R}^d$, let $X_{t,\xi}$ denote the solution to $dX(s) = a(s, X(s))ds + b(s, X(s))dW(s)$, $s \in [t, T]$, with $X(t) = \xi$.

THEOREM 9.6.– *For any* $s \in [t, T]$*,*

$$X_{t,X_{0,x}(t)}(s) = X_{0,x}(s)$$

almost surely.

REMARK 9.3.– It is possible to prove a similar result when $t = \tau$ is a stopping time. This requires a lot of technical work: proving the existence of a version of $X_{t,x}$, jointly measurable in all variables, including ω, and proving the substitution rule for the Itô integral: $\int_a^b f(\zeta, t)dW(t) = \int_a^b f(x, t)dW(t)|_{x=\zeta}$ for $\mathcal{F}_a$-measurable variable ζ. To avoid these technicalities, we establish only the simplest version of the Markov property, as stated above.

PROOF.– From the definition of $X_{0,x}$, we have

$$X_{0,x}(u) = x + \int_0^u a(z, X_{0,x}(z))dz + \int_0^u b(z, X_{0,x}(x))dW(z),\ u \in [0, T].$$

Substituting t and $s \ge t$ for u, and subtracting, we get

$$X_{0,x}(s) = X_{0,x}(t) + \int_t^s a(z, X_{0,x}(z))dz + \int_t^s b(z, X_{0,x}(x))dW(z),\ s \in [t, T],$$

which exactly means that $X_{0,X_{0,x}(t)}(s) = X_{0,x}(s)$, $s \in [t, T]$. □

9.3. Continuous dependence of solutions on coefficients

It is not rare that we have to consider not a single stochastic differential equation, but a whole family of such equations, with coefficients depending on parameter. Most notably, this manifests in modeling, where we have to calibrate a model so that it better describes the reality. In such cases, we have to ensure that small changes in parameters will not lead to a significant change in the behavior of a model. In this section, we study this important question of the continuous dependence of a modeled process on parameters for stochastic differential equations.

Consider a sequence of stochastic differential equations, indexed by integer $n \geq 1$:

$$dX^n(t) = a^n(t, X^n(t))dt + b^n(t, X^n(t))dW(t),\ t \in [0, T], \qquad [9.8]$$

with the initial condition $X^n(0)$; as before, the coefficients $a^n\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $b^n\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times k}$ are jointly measurable, and the initial condition $X^n(0)$ is an $\mathcal{F}_0$-measurable random variable.

Let the coefficients a^n, b^n satisfy assumptions [9.3] and [9.4] with a constant K independent of n, i.e. for any $n \geq 1$, $t \in [0, T]$, $x, y \in \mathbb{R}^d$ we have

$$\begin{aligned}&|a^n(t,x)| + |b^n(t,x)| \leq K(1 + |x|),\\ &|a^n(t,x) - a^n(t,y)| + |b^n(t,x) - b^n(t,y)| \leq K\,|x - y|\,. \end{aligned} \qquad [9.9]$$

For simplicity, we will assume that the initial conditions are square integrable, i.e. $\mathrm{E}|X^n(0)|^2 < \infty$, $n \geq 1$. Then, thanks to theorem 9.2, each of the stochastic differential equations [9.8] has a unique solution.

Further, assume pointwise convergence of the coefficients: for all $t \in [0, T]$, $x \in \mathbb{R}^d$,

$$a^n(t,x) \to a(t,x), \quad b^n(t,x) \to b(t,x), \quad n \to \infty, \qquad [9.10]$$

and the mean-square convergence of initial conditions:

$$X^n(0) \xrightarrow{\mathcal{L}_2(\Omega)} X(0), \quad n \to \infty. \qquad [9.11]$$

Note that a, b satisfy [9.9] as well, so the stochastic differential equation

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t),\ t \in [0, T], \qquad [9.12]$$

with the initial condition $X(0)$ has a unique solution.

THEOREM 9.7.– *Assume* [9.9]–[9.11]. *Then, the solutions of equations* [9.8] *converge to that of* [9.12]*; moreover,*

$$\mathrm{E}\sup_{t\in[0,T]}|X^n(t)-X(t)|^2\to 0,\quad n\to\infty.$$

PROOF.– Denote $\Delta^n(t)=\mathrm{E}\sup_{s\in[0,t]}|X^n(s)-X(s)|^2$ and estimate

$$\begin{aligned}\Delta^n(t)\le 3\Bigg(&\mathrm{E}|X^n(0)-X(0)|^2\\&+\mathrm{E}\sup_{s\in[0,t]}\left|\int_0^s\big(a^n(u,X^n(u))-a(u,X(u))\big)du\right|^2\\&+\mathrm{E}\sup_{s\in[0,t]}\left|\int_0^s\big(b^n(u,X^n(u))-b(u,X(u))\big)dW(u)\right|^2\Bigg).\end{aligned}\tag{9.13}$$

Using theorem 8.9, we have

$$\begin{aligned}&\mathrm{E}\sup_{s\in[0,t]}\left|\int_0^s\big(b^n(u,X^n(u))-b(u,X(u))\big)dW(u)\right|^2\\&\le C\int_0^t\mathrm{E}\big|b^n(u,X^n(u))-b(u,X(u))\big|^2du\\&\le C\int_0^t\mathrm{E}\big|b^n(u,X^n(u))-b^n(u,X(u))\big|^2du\\&+C\int_0^t\mathrm{E}\big|b^n(u,X(u))-b(u,X(u))\big|^2du\\&\le C\int_0^t\mathrm{E}\big|X^n(u)-X(u)\big|^2du+C\int_0^t\mathrm{E}\big|b^n(u,X(u))-b(u,X(u))\big|^2du\\&\le C\int_0^t\Delta^n(u)du+C\int_0^t\mathrm{E}\big|b^n(u,X(u))-b(u,X(u))\big|^2du.\end{aligned}$$

Similarly, using the Cauchy–Schwarz inequality, we have

$$\begin{aligned}&\mathrm{E}\sup_{s\in[0,t]}\left|\int_0^s\big(a^n(u,X^n(u))-a(u,X(u))\big)du\right|^2\\&\le C\int_0^t\Delta^n(u)du+C\int_0^t\mathrm{E}\big|a^n(u,X(u))-a(u,X(u))\big|^2du.\end{aligned}$$

Substituting these two inequalities into [9.7] and using the Grönwall lemma, we get

$$\Delta^n(T) \le C\Bigg(\mathrm{E}\,|X^n(0) - X(0)|^2$$

$$+ \int_0^T \mathrm{E}\left(\left|a^n(u,X(u)) - a(u,X(u))\right|^2 + \left|b^n(u,X(u)) - b(u,X(u))\right|^2\right) du\Bigg).$$

The first term on the right-hand side converges to zero as $n \to \infty$; in the second term, the expression under expectation vanishes and is bounded, thanks to [9.9] and [9.10], by $16K^2(1 + \sup_{t\in[0,T]} |X(t)|^2)$, which is integrable by theorem 9.1. Therefore, we get $\Delta^n(T) \to 0$, $n \to \infty$, as required. □

9.4. Weak solutions to stochastic differential equations

Let $a\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $b\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times k}$ be some measurable functions and W be a standard Wiener process in $\mathbb{R}^k$. Consider a stochastic differential equation

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t),\ t \in [0,T], \qquad [9.14]$$

with $\mathcal{F}_0$-measurable initial condition $X(0)$.

DEFINITION 9.2.– *A weak solution to stochastic differential equation* [9.14] *is a triple, consisting of:*

– a stochastic basis $(\Omega', \mathcal{F}', \{\mathcal{F}'_t, t \ge 0\}, \mathrm{P}')$*;*

– a Wiener process W' *on this basis;*

– an adapted process $\{X'(t), t \in [0,T]\}$ *on this basis such that* $X'(0) \overset{d}{=} X(0)$ *and*

$$dX'(t) = a(t, X'(t))dt + b(t, X'(t))dW'(t),\ t \in [0,T].$$

The difference from the notion of strong solution is that the former is constructed for *a given* Wiener process; in fact, a strong solution is a function of initial conditions and the path of the underlying Wiener process. In contrast, a weak solution is constructed *for some* Wiener process and in general it is not measurable with respect to this Wiener process. The following classical example illustrates the difference between the notions quite well.

EXAMPLE 9.1.– *Consider the following scalar stochastic differential equation, called the Tanaka equation:*

$$dX(t) = \text{sign } X(t)dW(t) \qquad [9.15]$$

with the initial condition $X(0) = 0$*; here, we denote* $\text{signx} = \mathbb{1}_{\text{x}\geq 0} - \mathbb{1}_{\text{x}<0}$ *to avoid zero values. Now take a standard Wiener process* $\{B(t), t \geq 0\}$ *on some stochastic basis. Define*

$$W'(t) = \int_0^t \text{sign } B(t)dB(t).$$

It can be seen that $W'(t)$ *is a standard Wiener process on the same stochastic basis (the simplest argument comes through the Lévy characterization theorem; see section 8.8). Since*

$$dW'(t) = \text{sign } B(t)dB(t),$$

we have

$$dB(t) = \text{sign } B(t)dW'(t),$$

so $B(t)$*, accompanied by the corresponding stochastic basis and the Wiener process* W'*, is a weak solution to* [9.15]. *Moreover,* $-B(t)$ *is a weak solution too, since*

$$d\big(-B(t)\big) = \text{sign }\big(-B(t)\big)dW'(t).$$

Therefore, there are at least two solutions corresponding to the same Wiener process W'*; in such a case, we say that* pathwise uniqueness *fails for the equation. As a result, there can be no strong solution, see e.g. [CHE 01]. It is also worth mentioning that in this example, the Wiener process* W' *is expressed as a function of the solution* B*, but not vice versa.*

EXAMPLE 9.2.– *As a contrasting example, we can consider a scalar equation*

$$dX(t) = \sigma(X(t))dW(t), \quad t \geq 0,$$

with $X(0) = 0$*, where* $\sigma(x) = \sigma_+\mathbb{1}_{x\geq 0} + \sigma_-\mathbb{1}_{x<0}$ *and* $\sigma_+, \sigma_- > 0$*. This equation has a unique strong solution by the Nakao theorem, see e.g. [JEA 09, theorem 1.5.5.1 (iii)]. Therefore, we see a remarkable phenomenon: the properties of the equations are very different when* σ_+ *and* σ_- *have the same sign, as here, and when they have different signs, as in the previous example.*

The concept of weak solution is important because in many situations (for some examples, see sections 9.7 and 9.8), we need to consider only some functionals of a solution to stochastic differential equations. Then, it is irrelevant whether the solution is given for the particular underlying Wiener process. What matters is the probabilistic distribution of the solution, which will be unique under weaker assumptions than those required for the existence of strong solutions.

One of the useful tools to construct weak solutions is the Girsanov theorem (see section 8.9).

THEOREM 9.8.– *Let X be a weak solution to* [9.14]. *Let also measurable functions $c\colon [0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ and $d\colon [0,T]\times\mathbb{R}^k\to\mathbb{R}^d$ be such that $c(t,x)=b(t,x)d(t,x)$ for all $t\in[0,T]$, $x\in\mathbb{R}^d$ and*

$$\mathrm{E}\exp\left\{\frac{1}{2}\int_0^T |d(t,X(t))|^2\,dt\right\}<\infty.$$

Then, the stochastic differential equation

$$dY(t)=\big(a(t,Y(t))+c(t,Y(t))\big)dt+b(t,Y(t))dW(t)$$

with the initial condition $Y(0)=X(0)$ has a weak solution given by the triple $(\Omega,\mathcal{F},\{\mathcal{F}_t,t\in[0,T]\},\mathrm{Q})$,

$$W'(t)=W(t)-\int_0^t d(s,X(s))ds, t\in[0,T],$$

and $\{X(t),t\in[0,T]\}$, where the probability measure Q is given by

$$\frac{d\mathrm{Q}}{d\mathrm{P}}=\exp\left\{\int_0^T\big(d(t,X(t)),dW(t)\big)-\frac{1}{2}\int_0^T|d(t,X(t))|^2\,dt\right\}.$$

PROOF.– By theorem 8.14, W' is a Wiener process under Q. Moreover, similarly to the proof of theorem 8.13, it can be shown that

$$\mathrm{E}\left(\frac{d\mathrm{Q}}{d\mathrm{P}}\,\Big|\,\mathcal{F}_0\right)=1$$

almost surely. As a result, $\mathrm{Q}|_{\mathcal{F}_0}=\mathrm{P}|_{\mathcal{F}_0}$, in particular, $X(0)$ has the same distribution under Q as under P. Finally, it follows from assumptions that

$$\begin{aligned}dX(t)&=\big(a(t,X(t))+c(t,X(t))\big)dt-c(t,X(t))dt+b(t,X(t))dW(t)\\&=\big(a(t,X(t))+c(t,X(t))\big)dt-b(t,X(t))d(t,X(t))dt+b(t,X(t))dW(t)\\&=\big(a(t,X(t))+c(t,X(t))\big)dt+b(t,X(t))dW'(t),\end{aligned}$$

thus concluding the proof. □

9.5. Solutions to SDEs as diffusion processes

Consider a stochastic differential equation in $\mathbb{R}^d$

$$dX(t)=a(t,X(t))dt+b(t,X(t))dW(t) \qquad [9.16]$$

where the coefficients $a\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ and $b\colon [0,T]\times\mathbb{R}^d \to \mathbb{R}^{d\times k}$ are continuous and satisfy the assumptions [9.3] and [9.4]. Assuming that the solution is a measurable function of initial condition, we get from theorem 9.6 that X is a Markov process. It turns out to be a diffusion process.

THEOREM 9.9.– *Under the above assumptions, the solution X to* [9.16] *is a diffusion process with drift $a(t,x)$ and diffusion matrix $\sigma(t,x) = b(t,x)b(t,x)^\top$.*

PROOF.– We are going to check the conditions of proposition 7.3. Denote by $X_{s,x}(t)$ the solution to [9.16] with the initial condition $X(s) = x$. By definition, the transition probability is $P(s,x,t,A) = \mathrm{P}\{X_{s,x}(t) \in A\}$. Then, by theorem 9.3,

$$\int_{\mathbb{R}^d} |y-x|^4 P(s,x,s+h,dy) = \mathrm{E}\,|X_{s,x}(s+h) - x|^4$$
$$\leq Ch^2(1+|x|^4) = o(h),\ h \to 0+,$$

whence the first assumption of proposition 7.3 follows. Further,

$$\int_{\mathbb{R}^3} (y-x)P(s,x,s+h,dy) = \mathrm{E}\,(X_{s,x}(s+h) - x)$$
$$= \mathrm{E}\left(\int_s^{s+h} a(u, X_{s,x}(u))du + \int_s^{s+h} b(u, X_{s,x}(u))dW(u)\right)$$
$$= \mathrm{E}\int_s^{s+h} a(u, X_{s,x}(u))du.$$

Therefore, using the continuity of a,

$$\left|\frac{1}{h}\int_{\mathbb{R}^3} (y-x)P(s,x,s+h,dy) - a(s,x)\right|$$
$$= \left|\frac{1}{h}\mathrm{E}\int_s^{s+h} a(u, X_{s,x}(u))du - a(s,x)\right|$$
$$\leq \frac{1}{h}\left|\mathrm{E}\int_s^{s+h} \big(a(u, X_{s,x}(u)) - a(u,x)\big)du\right|$$
$$\leq \frac{1}{h}\int_s^{s+h} \mathrm{E}\,|a(u, X_{s,x}(u)) - a(u,x)|\,du + \frac{1}{h}\int_s^{s+h} |a(u,x) - a(s,x)|\,du$$
$$\leq \frac{C}{h}\int_s^{s+h} \mathrm{E}\,|X_{s,x}(u) - x|\,du + \sup_{u\in[s,s+h]} |a(u,x) - a(s,x)|$$

$$\leq \frac{C}{h}\int_s^{s+h} |u-s|^{1/2}\, du + \sup_{u\in[s,s+h]} |a(u,x)-a(s,x)|$$

$$\leq Ch^{1/2} + \sup_{u\in[s,s+h]} |a(u,x)-a(s,x)| \to 0,\ h\to 0+,$$

which is the second assumption of proposition 7.3. Concerning the third one, for any $z \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^3} (y-x,z)^2 P(s,x,s+h,dy) = \mathrm{E}\big(X_{s,x}(s+h)-x,z\big)^2$$

$$= \mathrm{E}\left(I_1(h)+I_2(h)\right)^2 = \mathrm{E}I_1(h)^2 + \mathrm{E}I_2(h)^2 + 2\mathrm{E}\left(I_1(h)I_2(h)\right),$$

where

$$I_1(h) = \int_s^{s+h} \big(a(u,X_{s,x}(u)),z\big)du,$$

$$I_2(h) = \int_s^{s+h} \big(b(u,X_{s,x}(u))dW(u),z\big).$$

By the Cauchy–Schwarz inequality,

$$\mathrm{E}I_1(h)^2 \leq h\int_s^{s+h} \mathrm{E}\big(a(u,X_{s,x}(u)),z\big)^2 du$$

$$\leq C|z|^2 h\int_s^{s+h} \mathrm{E}\big(1+|X_{s,x}(u)|^2\big)du \leq C|z|^2h^2 = o(h),\ h\to 0+\,.$$

Thanks to the Itô isometry,

$$\mathrm{E}I_2(h)^2 = \mathrm{E}\left(\int_s^{s+h} \big(b(u,X_{s,x}(u))^\top z, dW(u)\big)\right)^2$$

$$\leq \int_s^{s+h} \mathrm{E}\left|b^\top(u,X_{s,x}(u))z\right|^2 du.$$

Similarly to the proof of the second assumption of proposition 7.3, using the continuity of b,

$$\frac{1}{h}I_2(h) \to \left|b(s,x)^\top z\right|^2 = \big(b(s,x)b(s,x)^\top z, z\big) = \big(\sigma(s,x)z,z\big).$$

Finally,

$$\frac{1}{h}\left|\mathrm{E}\left(I_1(h)I_2(h)\right)\right| \le \left(\frac{1}{h}\mathrm{E}\big(I_1(h)^2\big)\cdot\frac{1}{h}\mathrm{E}\big(I_2(h)^2\big)\right)^{1/2} \to 0,\ h\to 0+\,.$$

This gives

$$\frac{1}{h}\int_{\mathbb{R}^3}(y-x,z)^2P(s,x,s+h,dy)\to\big(\sigma(s,x)z,z\big),\ h\to 0+,$$

finishing the proof. □

9.6. Viability, comparison and positivity of solutions to stochastic differential equations

Consider a stochastic differential equation in $\mathbb{R}^d$

$$dX(t) = a(t,X(t))dt + b(t,X(t))dW(t),\ t\ge 0, \qquad [9.17]$$

where the coefficients $a\colon \mathbb{R}_+\times\mathbb{R}^d \to \mathbb{R}^d$ and $b\colon \mathbb{R}_+\times\mathbb{R}^d\to\mathbb{R}^{d\times k}$ satisfy assumptions [9.3] and [9.4], and W is a Wiener process in $\mathbb{R}^k$. As shown in section 9.5, it is a diffusion process with generator

$$\begin{aligned}\mathcal{L}_t f(x) &= \sum_{i=1}^d a_i(t,x)\frac{\partial}{\partial x_i}f(x) + \frac{1}{2}\sum_{i,j=1}^d\sum_{k=1}^m b_{ik}(t,x)b_{jk}(t,x)\frac{\partial^2}{\partial x_i\partial x_j}f(x)\\ &= \big(a(t,x),D_xf(x)\big) + \frac{1}{2}\mathrm{tr}\big(b(t,x)b(t,x)^\top D^2_{xx}f(x)\big),\quad f\in C^2(\mathbb{R}^d).\end{aligned}$$

In this section, we first address the question of the *viability* of process X in some subset of $\mathbb{R}^d$.

DEFINITION 9.3.– *Solution X to* [9.17] *is called* viable *in a set $A\subset\mathbb{R}^d$ if* $\mathrm{P}\{X(t)\in A$ *for all* $t\ge 0\} = 1$ *provided that $X(0)\in A$ almost surely.*

Let $\varphi : \mathbb{R}^d \to \mathbb{R}$, $\varphi\in C^2(\mathbb{R}^d)$ be a function such that $D_x\varphi(x)\ne 0$ when $\varphi(x)=0$. Assume that the set

$$A = \{x : \varphi(x)\ge 0\}$$

is non-empty and denote $\partial A = \{x : \varphi(x) = 0\}$ its boundary.

THEOREM 9.10.– *Assume the following conditions:*

1) for any $t \geq 0$, $x \in \partial A$,

$$\mathcal{L}_t \varphi(x) \geq 0;$$

2) for any $t \geq 0$, $x \in \partial A$,

$$\beta(t,x) := b(t,x)^\top D_x \varphi(x) = 0.$$

Then, X *is viable in* A.

REMARK 9.4.– It is well known that the gradient $D_x\varphi(x)$ is a vector orthogonal to the hypersurface $\partial A = \{\varphi(x) = 0\}$. Therefore, the second assumption means that near the boundary of the set A, the process diffuses mainly in directions along the boundary; there is no diffusion toward the boundary (and hence there is viability).

We start by establishing some auxiliary results.

LEMMA 9.1.– *For* $x \in \mathbb{R}^d$*, let* $z(x)$ *be one of the closest points to* x *points from* ∂A*, i.e. such that* $\varphi(z(x)) = 0$ *(it exists since the set* ∂A *is closed). Then, for any* $R > 0$*, there exists* $C_R > 0$ *such that*

$$|x - z(x)| \leq C_R\, |\varphi(x)|$$

for all $x \in \mathbb{R}^d$ *with* $|x| \leq R$.

PROOF.– Let us choose arbitrary $a \in \partial A$. Clearly, $|x - z(x)| \leq |x - a| \leq |R| + |a|$ when $|x| \leq R$. In particular, the points $z(x)$ lie in some bounded set when $|x| \leq R$.

The point $z(x)$ is a minimizer of $|x-y|^2$ given that $\varphi(x) = 0$. Therefore, from the Lagrange multiplier method, we know that either $D_x\varphi(z(x)) = 0$ (which is impossible by our assumption) or $D_y|y-x|^2\big|_{y=z(x)} = 2(z(x) - x)$ is collinear to $D_x\varphi(z(x))$. Using this collinearity, by Taylor's formula, we have

$$\begin{aligned}\varphi(x) &= \varphi(z(x)) + (D_x\varphi(z(x)), x - z(x)) + R(x)|x - z(x)|^2 \\ &= \pm |D_x\varphi(z(x))| \cdot |x - z(x)| + R(x)|x - z(x)|^2, \end{aligned} \quad [9.18]$$

where $R(x)$ is bounded since $\varphi \in C^2(\mathbb{R}^d)$ and x, $z(x)$ are bounded.

Now assume the contrary and let, for each $n \geq 1$, $x_n \in \mathbb{R}^d$ be such that $|x_n| \leq R$ and $|x_n - z(x_n)| \geq n\varphi(x_n) > 0$. Since $\{x_n, n \geq 1\}$ is bounded, there exists a convergent subsequence; without any loss of generality, let $x_n \to x_0$, $n \to \infty$. As $|x_n - z_n(x)|$ is bounded, we must have $\varphi(x_0) = 0$; therefore, $z(x_n) \to x_0$, $n \to \infty$. Thus, from [9.18], we have

$$\begin{aligned} 0 &= \lim_{n\to\infty} \frac{|\varphi(x_n)|}{|x_n - z(x_n)|} = \lim_{n\to\infty} \big| D_x\varphi(z(x_n)) \pm R(x_n)|x_n - z(x_n)| \big| \\ &= |D_x\varphi(x_0)|, \end{aligned}$$

contradicting the assumption that $D_x\varphi(x_0) \neq 0$. □

LEMMA 9.2.– *Assume the following:*

1) for any $t \geq 0$, $x \in \partial A$,

$$\mathcal{L}_t\varphi(x) > 0;$$

2) for any $t \geq 0$, $x \in \partial A$,

$$b(t,x)^\top D_x\varphi(x) = 0.$$

If τ *is a bounded stopping time such that* $X(\tau) \in A$ *and* $X(\tau)$ *is bounded, then there exists a stopping time* $\theta > \tau$ *a.s. such that* $X(t) \in A$ *for all* $t \in [\tau, \theta]$.

PROOF.– If $\omega \notin B := \{X(\tau) \in \partial A\}$, then define $\theta_1 = \inf\{t \geq \tau : X(t) \in \partial A\} \wedge (\tau + 1)$, thanks to the continuity of X, $\theta_1 > \tau$.

Now let $\omega \in B$. Fix some positive $R > |X(\tau)|$ and define stopping times

$$\tau' = \inf\{s \geq \tau : \mathcal{L}_s\varphi(X(s)) < 0\}, \qquad \tau_R = \min\{s \geq \tau : |X(s)| \geq R\}.$$

As usual, we suppose that a stopping time equals ∞ if the corresponding set is empty. In view of continuity, $\tau' > \tau$ and $\tau_R > \tau$ almost surely on B.

For any non-random $u \geq \tau$ put $\theta_u = u \wedge \tau' \wedge \tau_R$ and apply the Itô formula to the process $\varphi(X(\cdot))$:

$$\varphi(X(\theta_u))\mathbb{1}_B = \int_\tau^{\theta_u} \Big(\mathcal{L}_s\varphi(X(s))ds + \big(D_x\varphi(X(s)), b(s, X(s))dW(s)\big)\Big)\mathbb{1}_B.$$

Since X is bounded on $[\tau, \theta_u]$, the above Itô integral has zero expectation given $\mathcal{F}_\tau$, so

$$\mathrm{E}(\varphi(X(t))\mathbb{1}_B) = \mathrm{E}\left(\int_\tau^{\theta_u} \mathcal{L}_s\varphi(X(s))ds\,\mathbb{1}_B\right).$$

For a non-negative function $\psi \in C(\mathbb{R})$ such that $\int_{\mathbb{R}} \psi(x)dx = 1$ and $\psi(x) = 0$, $x \notin [0,1]$, we define

$$\psi_n(x) = n\int_0^{|x|}\int_0^y \psi(nz)dz\,dy.$$

Obviously, $\psi_n(x) \uparrow |x|$ as $n \to \infty$ and $|\psi_n'(x)| \le 1$, $n \ge 1$.

Applying the Itô formula to $\psi_n(\varphi(X(\cdot)))$, we get

$$\psi_n(\varphi(X(\theta_u)))\mathbb{1}_B = \int_\tau^{\theta_u} \psi_n'(\varphi(X(s)))\mathcal{L}_s\varphi(X(s))ds\, \mathbb{1}_B$$

$$+ \int_\tau^{\theta_u} \psi_n'(\varphi(X(s)))\big(\beta(s,X(s)), dW(s)\big)\mathbb{1}_B$$

$$+\frac{1}{2}\int_\tau^{\theta_u} \psi_n''(\varphi(X(s)))\,|\beta(s,X(s))|^2\,ds\,\mathbb{1}_B,$$

whence

$$\begin{aligned} \mathrm{E}\left(\psi_n(\varphi(X(\theta_u)))\mathbb{1}_B\right) &= \mathrm{E}\left(\int_0^{\theta_u} \psi_n'(\varphi(X(s)))\mathcal{L}_s\varphi(X(s))ds\,\mathbb{1}_B\right) \\ &\quad +\frac{1}{2}\mathrm{E}\left(\int_\tau^{\theta_u} \psi_n''(\varphi(X(s)))\,|\beta(s,X(s))|^2\,ds\,\mathbb{1}_B\right). \end{aligned} \quad [9.19]$$

Recall that $\mathcal{L}_s\varphi(X_s) \ge 0$ for $s < \theta_u$, and $|\psi_n'(x)| \le 1$, so the first term in the right-hand side of [9.19] does not exceed $\mathrm{E}\left(\int_0^{\theta_u} \mathcal{L}_s\varphi(X(s))ds\,\mathbb{1}_B\right)$. We will prove now that the second term vanishes.

Let $z(x)$ be one of the closest points to x points from ∂A. For $|x| \le R$,

$$\begin{aligned} &|\beta(s,x)| = |\beta(s,x) - \beta(s,z(x))| \\ &\le \left|b(s,x)^\top\big(D_x\varphi(x) - D_x\varphi(z(x))\big)\right| + \left|\big(b(s,x) - b(s,z(x))\big)^\top D_x\varphi(z(x))\right| \\ &\le C_R\big(\,|x - z(x)| + |x - z(x)|\,\big) \le C_R\,|x - z(x)|\,. \end{aligned}$$

Therefore, using lemma 9.1 and recalling that $\theta_u \le \tau + u$, we have

$$\begin{aligned} &\mathrm{E}\left(\int_\tau^{\theta_u} \psi_n''(\varphi(X(s)))\,|\beta(s,X(s))|^2\,ds\,\mathbb{1}_B\right) \\ &\le C_R n \mathrm{E}\left(\int_\tau^{\theta_u} \psi(n\varphi(X(s)))\varphi(X(s))^2\right) ds \\ &= \frac{C_R}{n}\mathrm{E}\left(\int_\tau^{\theta_u} \psi(n\varphi(X(s)))\big(\varphi(X(s))\big)^2 ds\right) \\ &\le \frac{C_R u}{n}\sup_{x\in\mathbb{R}} x^2\psi(x) \to 0, \quad n \to \infty. \end{aligned}$$

Taking $n \to \infty$, we get from [9.19],

$$\mathrm{E}\left(|\varphi(X(\theta_u))|\,\mathbb{1}_B\right) \le \liminf_{n\to\infty} \mathrm{E}\left(\psi_n(\varphi(X(\theta_u)))\,\mathbb{1}_B\right)$$
$$\le \mathrm{E}\left(\int_0^{\theta_u} \mathcal{L}_s\varphi(X(s))ds\,\mathbb{1}_B\right) = \mathrm{E}\left(\varphi(X(\theta_u))\,\mathbb{1}_B\right),$$

and hence, $\varphi(X(\theta_u)) \ge 0$ almost surely on B. Since $\theta_u = u \wedge \tau' \wedge \tau_R$ and u is arbitrary, we get the desired claim with $\theta = \theta_u \mathbb{1}_B + \theta_1 \mathbb{1}_{B^c} > \tau$ almost surely. □

PROOF (Proof of theorem 9.10).– Let $X(0) \in A$. It is enough to prove that $\mathbb{P}(X(t) \in A) = 1$ for all $t \ge 0$. Indeed, this would imply $\mathbb{P}\big(X(t) \in A \text{ for all } t \in \mathbb{R}_+ \cap \mathbb{Q}\big) = 1$, yielding $\mathrm{P}\big\{X(t) \in A \text{ for all } t \ge 0\big\} = 1$ thanks to the closedness of A and the continuity of X.

First assume that $\mathcal{L}_t\varphi(x) > 0$ for all $t \ge 0$, $x \in \partial A$.

Define $\tau_A = \inf\{s \ge 0 : X(s) \notin A\}$. Since X is continuous, we have $X(\tau_A) \in A$ whenever $\tau_A < \infty$. Assume that $\mathrm{P}\{\tau_A < \infty\} > 0$. Thanks to the continuity of the probability measure, there exist some $r > 0$ and $t > 0$ such that $\mathrm{P}\{\tau_A \le t, |X(\tau_A)| < r\} > 0$. Therefore, defining $\tau = \tau_A \wedge t \wedge \inf\{t \ge 0 : |X(t)| \ge r\}$, we have $\mathrm{P}\{\tau_A \le \tau\} > 0$. Applying lemma 9.2, we get the existence of $\theta > \tau$ such that $X(t) \in A$, $t \in [\tau, \theta]$, almost surely, which contradicts the definition of τ_A and the fact that $\mathrm{P}\{\tau_A \le \tau\} > 0$.

Now we prove the statement in its original form. Let $\{a^n(t,x), n \ge 1\}$ be a sequence of coefficients such that for all $t \ge 0$, $x, y \in \mathbb{R}^d$ $|a^n(t,x)| \le C(1 + |x|)$, $|a^n(t,x) - a^n(t,y)| \le C\,|x-y|$, $n \ge 1$, $a^n(t,x) \to a(t,x)$, $n \to \infty$ and

$$\mathcal{L}_t^n\varphi(x) := \big(a^n(t,x), D_x\varphi(x)\big) + \frac{1}{2}\mathrm{tr}\big(b(t,x)b(t,x)^\top D_{xx}^2\varphi(x)\big) > 0.$$

We can take, for example, $a_n(t,x) = a(t,x) + n^{-1}D_x\varphi(x)G(x)$ with a positive smooth function $G\colon \mathbb{R}^d \to R$, which does not vanish on ∂A and decays on infinity sufficiently rapidly so that $D_x\varphi(x)G(x)$ is bounded together with its derivative.

Let X^n be the solution of

$$dX^n(t) = a^n(t, X^n(t))dt + b(t, X^n(t))dW(t),\ t \ge 0, \qquad [9.20]$$

with the initial condition $X^n(0) = X(0)$. From theorem 9.7, we have

$$\sup_{t\in[0,T]} |X^n(t) - X(t)| \xrightarrow{\mathrm{P}} 0,\ n \to \infty.$$

We have shown $\mathrm{P}\{X^n(t) \in A \text{ for all } t \ge 0\} = 1$, whence $\mathrm{P}\{X(t) \in A \text{ for all } t \ge 0\} = 1$ thanks to the convergence of X^n to X and the closedness of A. □

9.6.1. *Comparison theorem for one-dimensional projections of stochastic differential equations*

Let us now formulate a stochastic version of a comparison theorem. While we formulate it in a multi-dimensional setting, it is basically about a pathwise comparison in one-dimensional cases (hence the name of this section).

Let X^i, $i=1,2$, be solutions to stochastic differential equations

$$dX^i(t) = a^i(t, X^i(t))dt + b^i(t, X^i(t))dW(t),\ t \geq 0,$$

where the coefficients a^i, $i=1,2$, and b satisfy the assumptions [9.3] of linear growth and [9.4] of Lipschitz continuity; the initial conditions $X^i(0) = (X^i_1(0), \dots, X^i_d(0))$, $i=1,2$, are $\mathcal{F}_0$-measurable random vectors. Fix some $l \in \{1, \dots, d\}$ (which will be the index of the coordinate we compare).

THEOREM 9.11.– *Assume that*

1) $X^1_l(0) \leq X^2_l(0)$ *almost surely;*

2) for any $t \geq 0$ *and any* $x^1, x^2 \in \mathbb{R}^d$ *such that* $x^1_l = x^2_l$, $a^1_l(t, x^1) \leq a^2_l(t, x^2)$*;*

3) for any $j = 1, \dots, k$, *the coefficients* b^1_{jl} *and* b^2_{jl} *coincide and depend only on the* l*th coordinate of* x, *i.e. there exists some* $b_{jl}\colon \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ *such that for any* $t \geq 0$ *and any* $x \in \mathbb{R}^d$, $b^1_{jl}(t,x) = b^2_{jl}(t,x) = b_{jl}(t, x_l)$*.*

Then, $\mathrm{P}\{X^1_l(t) \leq X^2_l(t), t \geq 0\} = 1$.

PROOF.– Consider the process

$$X(t) = (X^1(t), X^2(t)) \in \mathbb{R}^{2d} = \left\{(x^1, x^2) : x^1, x^2 \in \mathbb{R}^d\right\}$$

and set $\varphi(x) = x^2_l - x^1_l$. Then, in the notation of theorem 9.10, we have $A = \left\{x \in \mathbb{R}^{2d} : x^1_l \leq x^2_l\right\}$, $\partial A = \left\{x \in \mathbb{R}^{2d} : x^1_l = x^2_l\right\}$ and

$$\mathcal{L}_t \varphi(x) = \left(a^2(t, x^1) - a^1(t, x^2), D_x \varphi(x)\right).$$

Further, the diffusion coefficient of X is

$$\tilde{b} = \begin{pmatrix} b^1(t, x^1) \\ b^2(t, x^2) \end{pmatrix},$$

so $\tilde{b}(t,x)^\top D_x \varphi(x) = \left(b^2_{lj}(t, x^2) - b^1_{lj}(t, x^1), j = 1, \dots, d\right)$. By the assumption, $X_0 = (X^1(0), X^2(0)) \in A$ almost surely, $\mathcal{L}_t \varphi(x) \geq 0$ and $\tilde{b}(t,x)^\top D_x \varphi(x) = 0$ for $x \in \partial A$. Thus, we get the desired statement from theorem 9.10. $\square$

9.6.2. *Non-negativity of solutions to stochastic differential equations*

Non-negativity is an important feature in modeling. For example, the prices of ordinary stocks cannot be negative in view of limited liability, which in layman's terms means that an investor in stock of some company has no liability for the misfortunes of the company. Therefore, it is important to establish the non-negativity criteria.

Consider again equation [9.17] and assume that its coefficients satisfy the existence and uniqueness assumptions [9.3], [9.4]. Let $S \subset \{1, \dots, d\}$ be a fixed non-empty set of coordinates. We are interested in the non-negativity of X_i, $i \in S$, which is exactly the viability in the set $A = \big\{x = (x_1, \dots, x_d) \in \mathbb{R}^d : x_i \geq 0\big\}$.

THEOREM 9.12.– *Let the following hold:*

1) $X_{i(0)} \geq 0$, $i \in S$;

2) if $x \in \mathbb{R}^d$ *is such that* $x_i = 0$ *for some* $i \in S$ *and* $x_l \geq 0$ *for any* $l \in S$ *then* $a_i(t, x) \geq 0$ *and* $b(t, x) = 0$.

Then, $\mathrm{P}\{X_i(t) \geq 0$ *for all* $i \in S$ *and* $t \geq 0\} = 1$.

PROOF.– Let $f_i(x) = |x|\mathbb{1}_{i\in S} + x\mathbb{1}_{i\notin S}$, $i = 1, \dots, d$, $x \in \mathbb{R}$. Consider the equation

$$d\widetilde{X}(t) = \widetilde{a}(t, \widetilde{X}(t))dt + \widetilde{b}(t, \widetilde{X}(t))dW(t),\ t \geq 0,$$

with the initial condition $\widetilde{X}(0) = X(0)$, where

$$\widetilde{a}(t, x) = a\big(t, f_1(x_1), \dots, f_d(x_d)\big),\ \widetilde{b}(t, x) = b\big(t, f_1(x_1), \dots, f_d(x_d)\big).$$

For arbitrary $i \in \mathcal{S}$, it is easy to check that $\widetilde{a}$ and $\widetilde{b}$ satisfy the assumptions of theorem 9.10 with $\varphi(x) = x_i$. Therefore, $\widetilde{X}(t)$ is viable in each $A_i = \{x_i \geq 0\}$, $i \in S$, so $\mathbb{P}(\widetilde{X}_i(t) \geq 0$ for all $i \in S$ and $t \geq 0) = 1$. Hence, $\widetilde{X}(t)$ solves [9.17] and, by uniqueness, we obtain $\mathrm{P}\{X(t) = \widetilde{X}(t), t \geq 0\} = 1$, as claimed. □

9.7. Feynman–Kac formula

This section is devoted to the remarkable connection between solutions of stochastic differential equations and those of partial differential equations. Let $a : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{d\times d}$ be continuous functions. For $t \in [0, T]$, consider a second-order partial differential operator

$$\begin{aligned}\mathcal{L}_t f(x) &= \sum_{i=1}^{d} a_i(t, x)\frac{\partial}{\partial x_i} f(x) + \frac{1}{2}\sum_{i,j=1}^{d} \sigma_{ij}(t, x)\frac{\partial^2}{\partial x_i \partial x_j} f(x) \\ &= \big(a(t, x), D_x f(x)\big) + \frac{1}{2}\mathrm{tr}\big(\sigma(t, x) D^2_{xx} f(x)\big), \quad f \in C^2(\mathbb{R}^d),\end{aligned}$$

which is the infinitesimal generator of a diffusion process with drift $a(t,x)$ and diffusion matrix $\sigma(t,x)$. Under the assumption that $\sigma(t,x)$ is non-negatively definite, $\mathcal{L}_t$ is of parabolic type. Consider the boundary value problem

$$\frac{\partial}{\partial t}u(t,x) + \mathcal{L}_t u(t,x) - r(t,x)u(t,x) + f(t,x) = 0, \quad t \in [0,T), x \in \mathbb{R}^d$$
$$u(T,x) = g(x), \quad x \in \mathbb{R}^d, \qquad [9.21]$$

for a backward parabolic partial differential equation; here, $r, f\colon [0,T] \times \mathbb{R}^d \to \mathbb{R}$ and $g\colon \mathbb{R}^d \to R$ are some continuous functions. In order to associate this equation to a probabilistic object, note that $\mathcal{L}_t$ resembles a generator of a diffusion process. Indeed, the non-negativity of $\sigma(t,x)$ implies that there exists some $d \times d$-matrix $b(t,x)$ with $\sigma(t,x) = b(t,x)b(t,x)^\top$, for example, we may use the so-called Cholesky decomposition of $\sigma(t,x)$; we will assume that b is continuous as well. Consider a stochastic differential equation

$$dX(s) = a(s, X(s))dt + b(s, X(s))dW(s),\ s \in [t,T], \qquad [9.22]$$

with the initial condition $X(t) = x$. Assume that it has a weak solution, which we denote by $X^{t,x}$; for notational simplicity, we will use usual symbols for the stochastic basis and the Wiener process corresponding to this solution.

The following theorem establishes a probabilistic representation of a solution to [9.21] as a functional of a solution to [9.22]. This representation is called the *Feynman–Kac formula*.

THEOREM 9.13.– *Assume that a, b satisfy the linear growth assumption* [9.3]*, f, g satisfy a quadratic growth assumption $|f(t,x)| + |g(x)| \le C(1+|x|^2)$ for all $t \in [0,T]$, $x \in \mathbb{R}^d$ and r is bounded from below. Let $u \in C^{1,2}([0,T)\times\mathbb{R}^d) \cap C([0,T]\times\mathbb{R}^d)$ be a solution to* [9.21] *such that $u(t,x) \le C(1+|x|^2)$ for all $t \in [0,T]$, $x \in \mathbb{R}^d$. Then, for all $t \in [0,T]$, $x \in \mathbb{R}^d$,*

$$u(t,x) = \mathrm{E}\left(\int_t^T \nu(t,s) f(s, X_{t,x}(s))ds + \nu(t,T)g(X_{t,x}(T))\right), \qquad [9.23]$$

where $\nu(t,s) = \exp\left\{-\int_t^s r(u, X_{t,x}(u))du\right\}$, $s \in [t,T]$.

REMARK 9.5.– The term $\nu(t,s)$ frequently plays a role of discounting factor, which explains the negative sign used above.

REMARK 9.6.– The terminal condition g is often non-differentiable; for example, in financial mathematics, $g(x) = (x_i - K)^+$ may be used. This explains why u is assumed to be differentiable only on $[0, T)$. Also it is well known that, in general, a solution to [9.21] is not unique but there are some rapidly growing extraneous solutions, that is why the quadratic growth assumption on u is imposed.

PROOF.– Take any $T' \in (t, T)$ and write by the Itô formula

$$\begin{aligned}\nu(t,T')u(T',X_{t,x}(T')) &= \nu(t,t)u(t,X_{t,x}(t)) \\ &+ \int_t^{T'} \nu(t,s)\left(\frac{\partial}{\partial t}u(s,X_{t,x}(s)) + \mathcal{L}_t u(s,X_{t,x}(s))\right. \\ &\left.-r(s,X_{t,x}(s))u(s,X_{t,x}(s))\right) ds \\ &+ \int_t^{T'} \nu(t,s)\big(D_x u(s,X_{t,x}(s)), b(s,X_{t,x}(s))dW(s)\big) \\ = u(t,x) &- \int_t^{T'} \nu(t,s)f(s,X_{t,x}(s))ds \\ &+ \int_t^{T'} \nu(t,s)\big(D_x u(s,X_{t,x}(s)), b(s,X_{t,x}(s))dW(s)\big).\end{aligned}$$

We can assume that the integrand in the Itô integral is bounded, otherwise a standard localization argument may be used. Then, taking expectations,

$$u(t,x) = \mathrm{E}\left(\int_t^{T'} \nu(t,s)f(s,X_{t,x}(s))ds + \nu(t,T')u(T',X_{t,x}(T'))\right).$$

Thanks to our assumptions, the expression under expectation is up to constant bounded by $\sup_{s\in[t,T]} |X_{t,x}(s)|^2$, which is integrable by theorem 9.1 (the latter clearly holds for weak solutions as well). Therefore, using the dominated convergence theorem and the terminal condition $u(T,x) = g(x)$, we arrive at [9.23] as $T' \to T-$. □

9.8. Diffusion model of financial markets

Consider the following continuous time financial market model. Let $T > 0$ be a finite time horizon and $(\Omega, \mathcal{F}, \{\mathcal{F}_t, t \in [0,T]\}, \mathrm{P})$ be a stochastic basis. Here, the filtration $\{\mathcal{F}_t, t \in [0,T]\}$ is interpreted as the information flow: at each time $t \in [0,T]$, $\mathcal{F}_t$ is the information available to the market up to this time. For convenience, we will assume that $\mathcal{F}_0$ is trivial, i.e. it contains only P-null sets and their complements.

There are $d+1$ traded assets in our model: a bond (bank account) and d risky assets, stocks. Concerning the bond, we assume that it has the following dynamics:

$$S_0(t) = \exp\left\{\int_0^t r(s)ds\right\}, \quad t \in [0,T],$$

where the progressively measurable bounded process $r(s) \geq 0$ is the instantaneous interest rate. The progressive measurability requirement implies that S_0 is *locally riskless*: at each moment, there is a complete knowledge of its immediate return. The risky assets are assumed to have diffusion dynamics driven by a d-dimensional standard Wiener process W. The coincidence of dimension is in a sense a non-redundancy assumptions: each risky asset has its own source of randomness (possibly correlated with our sources). Precisely, the dynamics are

$$dS_i(t) = \mu_i(t)S_i(t)dt + \sum_{j=1}^{d} \sigma_{ij}(t)S_i(t)dW_j(t),\ t \in [0,T], i = 1,\dots,d, \qquad [9.24]$$

where μ and σ are bounded progressively measurable processes in $\mathbb{R}^d$ and $d \times d$, respectively. The drift term μ plays the role of a deterministic trend in the price evolution, while the diffusion term σ describes the stochastic part of the evolution, which is called *volatility*. The presence of the Wiener process makes the immediate return on a stock unpredictable, thus the term "risky asset".

We will assume that σ is non-singular, again for non-redundancy reasons. Using the Itô formula, we can see that the process

$$S_i(t) = S_i(0)\exp\left\{\int_0^t \left(\mu_i(t) - \frac{1}{2}\sum_{j=1}^{d}\sigma_{ij}(s)^2\right)ds + \sum_{j=1}^{d}\int_0^t \sigma_{ij}(s)dW_j(s)\right\}$$

solves [9.24]. Therefore, the stock prices are positive provided that their initial prices are also positive.

For brevity, the vector of risky asset prices will be denoted by $S(t) = (S_1(t),\dots, S_d(t))$, and equation [9.24] can be abbreviated as

$$dS(t) = S(t) * \big(\mu(t)dt + \sigma(t)dW(t)\big), \quad t \in [0,T],$$

where $*$ is the coordinatewise product of vectors. The whole vector of asset prices will be denoted by $\overline{S}(t) = (S_0(t), S(t)) = (S_0(t), S_1(t),\dots, S_d(t))$.

A *portfolio*, or *strategy*, is a progressively measurable process

$$\overline{\gamma}(t) = (\gamma_0(t), \gamma_1(t), \ldots, \gamma_d(t)),\ t \in [0, T],$$

where $\gamma_i(t)$ denotes the quantity of ith asset in the portfolio. The progressive measurability of the portfolio means that investor's decisions at time t are based only on information available to the market at that time (in other words, trading based on *insider* information about future price changes is prohibited). There can be negative numbers in the portfolio: for the bond, this means borrowing money, for a stock, its short sale.

The *value*, or *wealth*, of the portfolio at time $t \in [0, T]$ is the total price of assets in the portfolio, i.e.

$$C^\gamma(t) = \big(\overline{\gamma}(t), \overline{S}(t)\big) = \sum_{i=0}^{d} \gamma_k(t) S_k(t).$$

It is important to distinguish portfolios which are conservative in the sense that they do not use external money and there is no money going outside them. Such portfolios are called *self-financing* (or self-financed). The changes in the capital of such portfolios are only due to changes in the asset prices, not due to some external inflows or outflows of capital; this leads to the following equation:

$$dC^\gamma(t) = \big(\overline{\gamma}(t), d\overline{S}(t)\big), \quad t \in [0, T].$$

For many reasons, it is convenient to deal with discounted values using the bond S_0 as a *numéraire*. Namely, we define discounted price processes as $X_i(t) = S_i(t)/S_0(t)$, $t \in [0, T]$, $i = 1, \ldots, d$. The corresponding dynamics are

$$dX_i(t) = X_i(t)\big(\mu_i(t) - r(t))dt + \sum_{j=1}^{d} \sigma_{ij}(t) dW_j(t)\big), \quad t \in [0, T], i = 1, \ldots, d.$$

Denoting by $\mathbf{1}$ the vector of ones, we can abbreviate the last system as

$$dX(t) = X(t) * \big((\mu_i(t) - r(t)\mathbf{1})dt + \sigma(t) dW(t)\big), \quad t \in [0, T].$$

This is often expressed in terms of the so-called *risk premium process* $\lambda(t) = \sigma(t)^{-1}(\mu_i(t) - r(t)\mathbf{1})$:

$$dX(t) = X(t) * \sigma(t)\big(\lambda(t)\, dt + dW(t)\big), \quad t \in [0, T].$$

In the case where $d = 1$, $\lambda(t) = (\mu(t) - r(t))/\sigma(t)$ is the *Sharpe ratio*, a popular tool to measure the performance of a risky investment.

The *discounted value* of a portfolio is

$$V^\gamma(t) = C^\gamma(t)/S_0(t) = \gamma_0(t) + \big(\gamma(t), X(t)\big), \quad t \in [0,T],$$

where $\gamma(t) = (\gamma_1(t), \dots, \gamma_d(t))$ is the risky component of the portfolio.

In the following, we will only consider self-financing portfolios. With this at hand, the discounted value satisfies

$$dV^\gamma(t) = \big(\gamma(t), dX(t)\big), \quad t \in [0,T],$$

equivalently,

$$V^\gamma(t) = V^\gamma(0) + \int_0^t \big(\gamma(s), dX(s)\big), \quad t \in [0,T]. \qquad [9.25]$$

Hence, the amount of bond in the portfolio can be determined from its risky component and the initial capital:

$$\gamma_0(t) = V^\gamma(0) + \int_0^t \big(\gamma(s), dX(s)\big) - \big(\gamma(t), X(t)\big), \quad t \in [0,T]. \qquad [9.26]$$

9.8.1. *Admissible portfolios, arbitrage and equivalent martingale measure*

So far, we have not imposed any restrictions on the portfolio except its adaptedness. In reality, we cannot trade arbitrary amounts of bonds and stocks, and the price will certainly depend on the amount traded. Reflecting all possible restrictions and transaction costs in a model would be impossible but we can consider some well-behaved portfolios. Note that in order to write formulas like [9.25], we already have to assume a certain integrability of the portfolio. There are many other possible sets of restrictions (see [SHI 99, BJÖ 04] for more information). We take one of the simplest assumptions: that the discounted capital is bounded from below. This can be interpreted as a fixed credit line provided by a broker.

DEFINITION 9.4.– *A portfolio γ is* admissible *if $\int_0^T |\gamma(t)|^2\, dt < \infty$ almost surely (so that the integral in* [9.25] *is well defined), and there exists some non-random constant $a \in \mathbb{R}$ such that*

$$V^\gamma(t) \geq a, \quad t \in [0,T],$$

almost surely.

Next, we define the fundamental notion of arbitrage.

DEFINITION 9.5.– *An* arbitrage opportunity *is an admissible portfolio* γ *with*

$$V^\gamma(0) \le 0,\ V^\gamma(T) \ge 0 \text{ almost surely and } \mathrm{P}(V^\gamma(T) > 0) > 0.$$

The market model is called arbitrage-free *if there are no arbitrage opportunities.*

In layman's terms, an arbitrage opportunity is a possibility to make a certain profit without any risk, a "free lunch". In real markets, arbitrage opportunities are possible (and some financial companies have special departments to hunt for them) but they quickly disappear due to increasing demand for arbitrage positions.

There is a well-known "martingale" betting strategy in a game, where we can put a stake on some outcome, like flipping tails with a fair coin: we can double the bet in the case of loss and exit the game in the case of win. This strategy guarantees the gambler his initial bet provided that he has unbounded capital at his disposal. However, if we limit possible losses, as in definition 9.4, then the arbitrage opportunity will no longer be available.

This means that the non-arbitrage property depends on the definition of the admissible portfolio used. In [BJÖ 04, SHI 99] we can find discussions of different notions of arbitrage.

Let us move on to another important concept of financial mathematics.

DEFINITION 9.6.– *A measure* Q *on* $(\Omega, \mathcal{F})$ *is an* equivalent martingale measure *if* $\mathrm{Q} \sim \mathrm{P}$ *and the discounted price process* $\{X(t), t \in [0,T]\}$ *is a* Q*-martingale, i.e.*

$$\mathrm{E}^{\mathrm{Q}}(X_i(t) \mid \mathcal{F}_s) = X_i(s)$$

almost surely for all $s \le t \le T$*,* $i = 1, \dots, d$*.*

It turns out that the existence of an equivalent martingale measure is closely related to the absence of arbitrage. We first prove the martingale property for the discounted value of an admissible portfolio.

THEOREM 9.14.– *Let* Q *be an equivalent martingale measure. Then for any admissible portfolio* γ*, the discounted capital process* $\{V^\gamma(t), t \in [0,T]\}$ *is a* Q*-supermartingale.*

PROOF.– Similarly to the proof of lemma 8.1, it can be shown that there is a sequence of simple processes of the form

$$\gamma^n(t) = \sum_{k=1}^{k_n} \gamma_k^n \mathbb{1}_{[t_{k-1}^n, t_k^n)}(t), \quad t \in [0,T], n \ge 1$$

where γ_k^n is an $\mathcal{F}_{t_{k-1}^n}$-measurable random vector in $\mathbb{R}^d$ such that

$$\int_0^T |\gamma^n(t) - \gamma(t)|^2 \, dt \to 0, \; n \to \infty,$$

almost surely.

Consider now the corresponding self-financing portfolios with the initial capital $V^\gamma(0)$. Then

$$V^{\gamma^n}(t) = V^\gamma(0) + \sum_{k=1}^{k_n} \left(\gamma_k^n, X(t_k^n \wedge t) - X(t_{k-1}^n \wedge t)\right)$$

is a Q-martingale. Indeed, each summand is easily seen to be a martingale thanks to the Q-martingale property of X. On the other hand, in view of the boundedness of μ, r and σ, we have from theorem 8.6

$$\sup_{t \in [0,T]} \left| V^{\gamma^n}(t) - V^\gamma(t) \right| \xrightarrow{\mathrm{P}} 0, \; n \to \infty.$$

Therefore, defining $\tau_n = \inf\left\{t \geq 0 : \left|V^{\gamma^n}(t) - V^\gamma(t)\right| \geq 1\right\} \wedge T$ and $\widetilde{\gamma}^n(t) = \gamma^n(t)\mathbb{1}_{t \leq \tau_n}$ and considering the corresponding self-financing portfolios with the initial capital $V^\gamma(0)$, we have

$$\sup_{t \in [0,T]} \left| V^{\widetilde{\gamma}^n}(t) - V^\gamma(t) \right| \xrightarrow{\mathrm{P}} 0, \; n \to \infty.$$

Extracting, if necessary, a subsequence, we can assume that the convergence is almost sure. For every $t \in [0,T]$, we have $V^\gamma(t) \geq a$ with some $a \in \mathbb{R}$, so it follows from the definition of $\widetilde{\gamma}^n$ that $V^{\widetilde{\gamma}^n}(t) \geq a - 1$, $t \in [0,T]$. Therefore, using the martingale property of $V^{\widetilde{\gamma}^n}$ and the Fatou lemma (which can be used thanks to the boundedness from below), we have for any $s < t \leq T$,

$$\begin{aligned} V^\gamma(s) &= \liminf_{n \to \infty} V^{\widetilde{\gamma}^n}(s) = \liminf_{n \to \infty} \mathrm{E}^{\mathrm{Q}} \left(V^{\widetilde{\gamma}^n}(t) \mid \mathcal{F}_s \right) \\ &\geq \mathrm{E}^{\mathrm{Q}} \left(\liminf_{n \to \infty} V^{\widetilde{\gamma}^n}(t) \mid \mathcal{F}_s \right) = \mathrm{E}^{\mathrm{Q}} \left(V^\gamma(t) \mid \mathcal{F}_s \right), \end{aligned}$$

which is the desired supermartingale property. □

The following fact is a simplified version of the so-called *first fundamental theorem of asset pricing* relating the absence of arbitrage to the existence of a martingale measure.

THEOREM 9.15.– *If there exists an equivalent martingale measure, then the market model is arbitrage-free.*

PROOF.– Let Q be the equivalent martingale measure, and an admissible portfolio γ be such that $V^\gamma(0) \le 0$ and $\mathrm{P}(V^\gamma(T) \ge 0) = 1$. Since $\mathrm{Q} \sim \mathrm{P}$, also $\mathrm{Q}(V^\gamma(T) \ge 0) = 1$. However, by the previous theorem, $V^\gamma(T)$ is a Q-supermartingale, therefore

$$\mathrm{E}^{\mathrm{Q}} V^\gamma(T) \le V^\gamma(0)$$

a.s., whence $\mathrm{Q}(V^\gamma(T) = 0) = 1$. The equivalence of measures then implies $\mathrm{P}(V^\gamma(T) = 0) = 1$, so there are indeed no arbitrage opportunities. □

REMARK 9.7.– In some cases, the statement of theorem 9.15 may be reverted; see [BJÖ 04, SHI 99] for details.

A sufficient condition for the existence of an equivalent martingale measure is given by the Girsanov theorem.

THEOREM 9.16.– *Let the risk-premium process λ satisfy*

$$\mathrm{E} \exp \left\{ \frac{1}{2} \int_0^T |\lambda(t)|^2 \, dt \right\} < \infty.$$

Then, the measure Q defined by

$$\frac{d\mathrm{Q}}{d\mathrm{P}} = \exp \left\{ - \int_0^T \big(\lambda(t), dW(t)\big) - \frac{1}{2} \int_0^T |\lambda(t)|^2 \, dt \right\}$$

is an equivalent martingale measure. Consequently, the market model is arbitrage-free.

PROOF.– Arguing as in the proof of corollary 8.1, we get

$$X(t) = X(0) + \int_0^t X(s) * \sigma(s) dW^\lambda(s), \; t \in [0, T], \qquad [9.27]$$

and W^λ is a Wiener process with respect to Q. From theorem 9.3 applied to equation [9.27], it follows that $\mathrm{E}^{\mathbb{Q}} \sup_{t \in [0,T]} |X(t)|^2 < \infty$, so the integrand in [9.27] is square integrable. Thus, X is a Q-martingale. □

9.8.2. *Contingent claims, pricing and hedging*

The bond S_0 and stocks $S_1, \ldots, S_d$ are usually called primary (financial) assets to distinguish them from other instruments traded in the market: options, futures, swaps, warrants, etc. We will consider only instruments maturing at time T and paying some random quantity, which is called *payoff* of the claim. For brevity, we will identify contingent claims with their payoffs.

DEFINITION 9.7.– *A* contingent claim *is a non-negative $\mathcal{F}_T$-measurable random variable C. The contingent claim is called a derivative of primary assets $S_0, S_1, \dots, S_d$ if it pays an amount depending on the primary assets, i.e. $C \in \sigma\left(\overline{S}(t), t \in [0,T]\right)$. A* European claim*, or European option, is a derivative depending only on the ultimate prices, i.e. $C = f(S_0(T), S_1(T) \dots, S_d(T))$; here, f is called the* payoff function *of the claim.*

EXAMPLE 9.3.– *A European call option with strike price K and the maturity date T on the unit of stock S_k is a contract which gives its buyer (holder of the option) the right to buy the designated (here, unit) amount of the underlying asset (here, S_k) at the moment T for the agreed price K. Note that it is up to the holder to decide whether he would use this right, exercising the option at time T to buy the stock (in which case the option writer has the obligation to sell the stock). It is natural to assume that the holder acts rationally, exercising the option if and only if its current price $S_k(T)$ exceeds the strike price K. The virtual profit of the option holder in this situation, called* in the money*, is $S_k(T) - K$, the amount he would realize from selling the stock at current price. When $S_k(T) < K$, the option is useless (*out of the money*), since the holder can buy the underlying asset cheaper. Thus, the amount $C_{\text{call}} = (S_k(T) - K)^+ = \max(S_k(T) - K, 0)$ is naturally identified with the payoff of the option. Moreover, very often an option is* cash-settled*, i.e. the amount C_{call} is paid to the holder instead of delivery of the asset; sometimes the delivery is even impossible, as the underlying asset is a stock index or interest rate.*

EXAMPLE 9.4.– *Similarly, a European put option gives its holder the right to sell the underlying asset at time T for the strike price K. The corresponding payoff is $C_{\text{put}} = (K - S_k(T))^+$.*

EXAMPLE 9.5.– *An exchange option allows one to exchange a unit of asset S_k at time T for K units of asset S_j. Its payoff is $(S_k(T) - KS_j(T))^+$. It is often called a* Margrabe *option, named after William Margrabe, who derived the formula for its price in a simple diffusion model.*

EXAMPLE 9.6.– *A* basket *option is written on a portfolio of assets. Say, an option allowing one to buy a portfolio containing a_i of asset S_i, $i = 1, \dots, d$, for a strike price K has payoff $\left(\sum_{i=1}^d a_i S_i(T)\right)^+$. Options on stock indices are examples of basket options. Basket options are also called* rainbow *options; the latter allow negative weights a_i so that, for example, the exchange option is a rainbow option with $a_k = 1$, $a_j = -K$, $a_i = 0$, $i \neq j, k$.*

EXAMPLE 9.7.– *The payoff of* Asian *derivative security depends on the average price* $A_k(T) = \frac{1}{T}\int_0^T S_k(t)dt$ *of some asset* S_k *over time interval* $[0,T]$*. Therefore, the payoff of Asian security depends on the whole trajectory of the price process, not only on the terminal value, as that of a European option. Such financial instruments are thus called* path dependent. *Examples of Asian options:* average price call *option paying* $(A_k(T) - K)^+$, average price put *paying* $(K - A_k(T))^+$, average strike call *paying* $(S_k(T) - A_k(T))^+$.

EXAMPLE 9.8.– *Another class of path-dependent options are so-called* lookback *options depending on the minimum price* $m_k(T) = \min_{t\in[0,T]} S_k(t)$ *and/or maximum price* $m_k(T) = \min_{t\in[0,T]} S_k(t)$ *of asset* S_k*. One example is the* fixed strike lookback call *option with payoff* $(M_k(T) - K)^+$*. It can be interpreted as a European call option, where the maturity* $t \in [0,T]$ *is chosen retrospectively at time* T*. A floating strike lookback call has the payoff* $(M_k(T) - S_k(T))^+$*. Other examples are* barrier options*, where the payoff is void if the price of the underlying asset hits (for* knock-in *options) or does not hit (for* knock-out *options) the prescribed barrier. Depending on the position of the barrier, there are* up *and* down *options. There is also the call/put ambivalence, which gives in total eight types of barrier options. For example, up-and-in call option pays* $(S_k(T) - K)^+ \mathbb{1}_{M_k(T)\geq H}$*, down-and-out put,* $(K - S_k(T))^+ \mathbb{1}_{m_k(T)>L}$.

Pricing contingent claims is a fundamental task of financial mathematics; there are numerous approaches based on different concepts. The most established of them is the arbitrage pricing theory, which we briefly describe here.

Let us start with the case where a contingent claim is a final value of some portfolio.

DEFINITION 9.8.– *A contingent claim* C *is* attainable *(or* marketable*) if* $C = C^\gamma(T) = \big(\overline{\gamma}(T), \overline{S}(T)\big)$ *for some admissible portfolio* γ*, called a* replicating *portfolio. The initial capital of the replicating portfolio is called a* fair price *of* C*:* $\pi(C) = C^\gamma(0)$.

One claim can be replicated by different portfolios, and *a priori* there can be several fair prices. When there are no arbitrage opportunities in the market model, the price is normally unique, and this is the so-called *law of one price*. We will prove it under the additional assumption that there exists an equivalent martingale measure.

THEOREM 9.17.– *Let there exist an equivalent martingale measure* Q*. Then, every attainable contingent claim* C *has a unique fair price. Moreover, it is equal to the discounted expected value with respect to* Q*:*

$$\pi(C) = \mathrm{E}^{\mathrm{Q}} \frac{C}{S_0(T)}.$$

PROOF.– Let γ be a replicating portfolio for C. By theorem 9.14, the discounted capital V^γ of γ is a Q-martingale; therefore,

$$\mathrm{E}^{\mathrm{Q}} \frac{C}{S_0(T)} = \mathrm{E}^{\mathrm{Q}} V^\gamma(T) = V^\gamma(0) = \pi(C),$$

as claimed. □

The second conclusion of this theorem allows us to extend the definition of fair prices to a larger class of contingent claims. We denote by $\mathcal{P}$ the set of all equivalent martingale measures.

DEFINITION 9.9.– *A* fair price *of contingent claim C is a finite discounted expected value of C with respect to an equivalent martingale measure, that is, an element of*

$$\Pi(C) = \left\{ \mathrm{E}^{\mathrm{Q}} \frac{C}{S_0(T)} \;\middle|\; \mathrm{Q} \in \mathcal{P}, \mathrm{E}^{\mathrm{Q}} \frac{C}{S_0(T)} < \infty \right\}.$$

For an attainable claim, $\Pi(C) = \{\pi(C)\}$, provided that an equivalent martingale measure exists; for a non-attainable claim, this set is usually non-empty and is an open interval. We will prove that each fair price is *non-arbitrage*.

THEOREM 9.18.– *Any $\pi \in \Pi(C)$ is a non-arbitrage price for C, i.e. there exists a non-negative adapted process $\{S_{d+1}(t), t \in [0,T]\}$ such that $S_{d+1}(0) = \pi$, $S_{d+1}(T) = C$, and the extended model with traded assets $S_0, S_1, \ldots, S_d, S_{d+1}$ is arbitrage-free.*

PROOF.– By definition, $\pi = \mathrm{E}^{\mathrm{Q}}\big(C/S_0(T)\big)$ for some $\mathrm{Q} \in \mathcal{P}$. We define

$$S_{d+1}(t) = S_0(t)\, \mathrm{E}^{\mathrm{Q}} \left(\frac{C}{S_0(T)} \;\middle|\; \mathcal{F}_t \right), \ t \in [0,T]. \qquad [9.28]$$

Then, $S_{d+1}(0) = \pi$, $S_{d+1}(T) = C$, and it is easy to see from [9.28] that Q is an equivalent martingale measure in the extended market. Then, the extended model is arbitrage-free thanks to theorem 9.15. □

PART 2

Statistics of Stochastic Processes

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

10

Parameter Estimation

In this chapter, we consider the simplest models consisting of the drift (regular) component and the diffusion (non-regular) component. We propose different estimators of unknown scaling parameters of the models under consideration and study their strong consistency. Recall the definition of strong consistency for discrete time and continuous time observations.

DEFINITION 10.1.–

i) An estimator $\widehat{Y}_n$ of the unknown real-valued parameter Y, constructed by the finite number $\{Y_1, \ldots, Y_n\}$ of observations is called strongly consistent *if $\widehat{Y}_n \to Y$ a.s. as $n \to \infty$.*

ii) An estimator $\widehat{Y}_t$ of the unknown real-valued parameter Y, constructed from the observations $\{Y_s, 0 \le s \le t\}$ is called strongly consistent *if $\widehat{Y}_t \to Y$ a.s. as $t \to \infty$.*

10.1. Drift and diffusion parameter estimation in the linear regression model with discrete time

Consider the linear regression model of the form

$$X_n = X_0 + \theta b_n + R_n,\ n \ge 1,$$

where $\{b_n, n \ge 1\}$ is a known sequence of real numbers, not all of them being equal to zero, $X_0 \in \mathbb{R}$ and $\theta \in \mathbb{R}$ are the parameters to be estimated, θb_n is a *drift* component (in other words, regular part, or signal), and $\{R_n, n \ge 1\}$ is a sequence of random variables that is treated as the *diffusion* component (in other words, irregular part, or noise). Initial value X_0 is also called an *intercept* term and θ is called a *slope* parameter. Throughout the section, we assume that the noise is centered, i.e. $\mathrm{E}R_n = 0$, $n \ge 1$.

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

10.1.1. *Drift estimation in the linear regression model with discrete time in the case when the initial value is known*

Assume that $X_0 \in \mathbb{R}$ is known. Consider two estimators of the parameter $\theta \in \mathbb{R}$ constructed by the observations $\{X_0, X_1, \ldots, X_n\}$. One of them can be constructed under the additional assumption that $b_n \neq 0$, $n \geq 1$, and has the form $\widehat{\theta}_n^{(1)} = \frac{X_n}{b_n}$ (the estimator constructed from the last observation). Another one, $\widehat{\theta}_n^{(2)}$, is the *least squares estimator* (LSE). To construct the LSE, we consider $\Delta X_n = X_n - X_{n-1}$, $n \geq 1$, and minimize the quadratic form $\sum_{i=1}^{n}(\Delta X_i - \theta b_i)^2$ in $\theta \in \mathbb{R}$, getting $\widehat{\theta}_n^{(2)} = \frac{\sum_{i=1}^{n} \Delta X_i b_i}{\sum_{i=1}^{n} b_i^2}$. These estimators admit, respectively, stochastic representations of the form

$$\widehat{\theta}_n^{(1)} = \theta + \frac{R_n}{b_n} \quad \text{and} \quad \widehat{\theta}_n^{(2)} = \theta + \frac{\sum_{i=1}^{n} \varepsilon_i b_i}{\sum_{i=1}^{n} b_i^2},$$

where $\varepsilon_i = R_i - R_{i-1}$, $i \geq 1$.

Evidently, both estimators are unbiased, that is $\mathrm{E}\widehat{\theta}_n^{(1)} = \mathrm{E}\widehat{\theta}_n^{(2)} = \theta$. Their consistency properties depend on both the properties of the drift $\{b_n, n \geq 1\}$ and the noise $\{R_n, n \geq 1\}$, as well as on the relationship between them.

THEOREM 10.1.–

1) Let $\{\varepsilon_n, n \geq 1\}$ be a sequence of iid random variables, $\mathrm{E}|\varepsilon_1| < \infty$ and $|b_n| \geq bn$ for some $b > 0$ and for all $n \geq n_0$, where $n_0 \in \mathcal{N}$ is some integer number. Then the estimator $\widehat{\theta}_n^{(1)}$ is strongly consistent as $n \to \infty$.

2) Let $\{\varepsilon_n, n \geq 1\}$ be a sequence of independent random variables, $\varepsilon_n \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$, $n \geq 1$, and $\sum_{n=1}^{\infty} \frac{\mathrm{E}\varepsilon_n^2}{n^2} < \infty$. Let $|b_n| \geq bn$ for some $b > 0$ and for all $n \geq n_0$, where $n_0 \in \mathcal{N}$ is some integer number. Then the estimator $\widehat{\theta}_n^{(1)}$ is strongly consistent as $n \to \infty$.

3) Let $\{\varepsilon_n, n \geq 1\}$ be a sequence of iid random variables, $\varepsilon_n \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathrm{P})$, $n \geq 1$, for some $1 < p < 2$. Let $|b_n| \geq bn^{1/p}$ for some $b > 0$ and for all $n \geq n_0$, where $n_0 \in \mathcal{N}$ is some integer number. Then the estimator $\widehat{\theta}_n^{(1)}$ is strongly consistent.

Theorem 10.1 is an immediate consequence of theorems A2.18, A2.19 and A2.20 from section A2.7.

EXAMPLE 10.1.– *The simplest illustrations of the assumptions of theorem 10.1 are as follows, respectively:*

1) $\varepsilon_n \sim N(0,1)$ are iid random variables, $b_n = n$;

2) ε_n are independent random variables with $\mathrm{E}\varepsilon_i^2 \le Ci^{1-\delta}$ for some $0 < \delta < 1$, $b_n = n$. In this case, $\sum_{n=1}^{\infty} \frac{\mathrm{E}\varepsilon_n^2}{n^2} \le C \sum_{n=1}^{\infty} \frac{n^{1-\delta}}{n^2} < \infty$.

3) ε_n are iid random variables with $\mathrm{E}|\varepsilon_i|^{3/2} < \infty$, $b_n = n^{2/3}$.

In the next theorem, we use the class Ψ of functions introduced in section A2.7.

THEOREM 10.2.– *Let $\{\varepsilon_n, n \ge 1\}$ be a sequence of centered orthogonal random variables from $\mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$. The estimator $\widehat{\theta}_n^{(1)}$ is strongly consistent under either of two the assumptions:*

1) $\{|b_n|, n \ge 1\}$ is a non-decreasing sequence, $|b_n| \to \infty$ as $n \to \infty$, and $\sum_{n=1}^{\infty} \frac{\mathrm{E}\varepsilon_n^2}{b_n^2} \log^2 n < \infty$.

2) $|b_n| \ge bn$ for some $b > 0$ and $n \ge n_0$, and $\sum_{i=1}^{n} \mathrm{E}\varepsilon_i^2 = O\left(\frac{n^2}{\psi(n)\log^2 n}\right)$ for some $\psi \in \Psi$.

PROOF.– These statements follow immediately from theorems A2.21 and A2.22. □

EXAMPLE 10.2.–

1) Let $\mathrm{E}\varepsilon_n^2 \le c$, $b_n = n^{\frac{1}{2}+\alpha}$ for some $\alpha > 0$ and for $n \ge 1$, and ε_n be centered and uncorrelated (for example, independent). Then $\sum_{n=1}^{\infty} \frac{\mathrm{E}\varepsilon_n^2}{b_n^2} \log^2 n \le c \sum_{n=1}^{\infty} \frac{\log^2 n}{n^{1+2\alpha}} < \infty$, therefore, $\frac{R_n}{b_n} = \frac{\sum_{i=1}^{n} \varepsilon_i}{n^{\frac{1}{2}+\alpha}} \to 0$ a.s. as $n \to \infty$, and $\widehat{\theta}_n^{(1)}$ is strongly consistent.

2) Let $\mathrm{E}\varepsilon_n^2 \le cn^{\alpha}$ for some $\alpha \in (0,1)$ for $n \ge 1$, and $\{\varepsilon_n, n \ge 1\}$ as above, be centered and uncorrelated, $b_n = n$. Then for $\psi(x) = x^{\beta} \in \Psi$, $0 < \beta < 1 - \alpha$ we have that

$$\sum_{i=1}^{n} \mathrm{E}\varepsilon_i^2 \le cn^{\alpha+1} \le cn^{2-\beta} = O\left(\frac{n^2}{\psi(n)\log^2 n}\right),$$

and $\frac{R_n}{b_n} = \frac{R_n}{n} \to 0$ a.s.

Now consider the cases of martingale and stationary sequences.

THEOREM 10.3.– *Let $\{R_n, \mathcal{F}_n, n \ge 1\}$ be a square-integrable martingale, $\mathcal{F}_n = \sigma\{\varepsilon_1, \dots, \varepsilon_n\} = \sigma\{R_1, \dots, R_n\}$, $n \ge 1$, $\mathcal{F}_0 = \{\emptyset, \Omega\}$.*

1) Under any assumption 1) or 2) from theorem 10.2 it holds that $\widehat{\theta}_n^{(1)}$ is strongly consistent.

2) Let $\sum_{i=1}^{n} \mathrm{E}(\varepsilon_i^2|\mathcal{F}_{i-1}) \to \infty$, $n \to \infty$ *and there exists a random variable* $\xi > 0$ *such that* $\sum_{i=1}^{n} \mathrm{E}(\varepsilon_i^2|\mathcal{F}_{i-1}) \le \xi |b_n|$, $n \ge n_0$. *Then* $\widehat{\theta}_n^{(1)}$ *is strongly consistent.*

PROOF.– The first statement follows from the orthogonality of ε_i: if R_n is $\mathcal{F}_n$-martingale, then for $i > k$,

$$\mathrm{E}\varepsilon_i\varepsilon_k = \mathrm{E}(\mathrm{E}(\varepsilon_i|\mathcal{F}_k)\varepsilon_k) = \mathrm{E}(\mathrm{E}(R_i - R_{i-1}|\mathcal{F}_k)\varepsilon_k) = 0.$$

Therefore, we can apply theorem 10.2. The second statement follows from SLLN for martingales, theorem 5.21:

$$\left|\frac{R_n}{b_n}\right| = \frac{|R_n|}{\langle R\rangle_n} \cdot \frac{\langle R\rangle_n}{|b_n|} \le \frac{|R_n|}{\langle R\rangle_n} \cdot \xi \to 0$$

a.s. if

$$\langle R\rangle_n = \sum_{i=1}^{n} \mathrm{E}(\varepsilon_i^2|\mathcal{F}_{i-1}) \to \infty$$

as $n \to \infty$. □

THEOREM 10.4.– *Let* $\{\varepsilon_i, i \ge 1\}$ *be a stationary Gaussian sequence and* $\mathrm{E}\varepsilon_0\varepsilon_n \to 0$ *as* $n \to \infty$. *Also, let* $|b_n| \ge bn$ *for* $b > 0$ *and* $n \ge n_0$. *Then* $\widehat{\theta}_n^{(1)}$ *is strongly consistent.*

PROOF.– Convergence $\left|\frac{R_n}{b_n}\right| = \frac{|R_n|}{n} \cdot \frac{n}{|b_n|} \to 0$ a.s. follows immediately from theorem A2.15. □

Conditions of strong consistency of $\widehat{\theta}_n^{(2)}$ can be formulated in a similar way if we replace ε_n for $\varepsilon_n \Delta b_n$ and the condition $|b_n| \ge bn$, $n \ge n_0$ for the condition $\sum_{i=1}^{n} (\Delta b_i)^2 \ge bn$, $n \ge n_0$. In the case when $b_n = n$, assumptions supplying strong consistency of $\widehat{\theta}_n^{(1)}$ and $\widehat{\theta}_n^{(2)}$ coincide as well as the corresponding examples.

REMARK 10.1.– Let the sequence b_n, $n \ge 1$, be non-negative and non-decreasing. Assume that the limit $\Delta := \lim_{n\to\infty} \Delta b_n$ exists and $\Delta \in (0, +\infty)$. Then it follows from the Stolz-Cèsaro theorem that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} (\Delta b_i)^2}{n} = \lim_{n\to\infty} \Delta b_n = \Delta,$$

and

$$\lim_{n\to\infty} \frac{b_n}{n} = \lim_{n\to\infty} \Delta b_n = \Delta,$$

which means that both conditions hold: $\sum_{i=1}^{n} (\Delta b_i)^2 \ge bn$, $n \ge n_0$ and $|b_n| \ge bn$, $n \ge n_0$.

10.1.2. *Drift estimation in the case when the initial value is unknown*

Assume that both $X_0 \in \mathbb{R}$ and $\theta \in \mathbb{R}$ are unknown parameters to be estimated. Consider the LSE estimator of the parameters (X_0, θ) constructed by the observations $\{X_1, \ldots, X_n\}$. In this case, we should minimize the quadratic form $\sum_{i=1}^n (\Delta X_i - \theta b_i)^2$ in $(X_0, \theta) \in \mathbb{R}^2$, obtaining

$$\widehat{X}_0 = X_1 - \widehat{\theta}_n b_1, \widehat{\theta}_n = \frac{\sum_{i=1}^n \Delta X_i b_i}{\sum_{i=1}^n b_i^2},$$

or, finally,

$$\widehat{X}_0 = X_1 - \frac{\sum_{i=1}^n \Delta X_i b_i}{\sum_{i=1}^n b_i^2} b_1, \widehat{\theta}_n = \frac{\sum_{i=1}^n \Delta X_i b_i}{\sum_{i=1}^n b_i^2}.$$

Strong consistency of these estimators is studied in the same way as that of $\widehat{\theta}_n^{(2)}$.

REMARK 10.2.– The LSE for the multidimensional linear regression scheme is treated in detail in the book [SEB 03].

10.2. Estimation of the diffusion coefficient in a linear regression model with discrete time

Consider the model of the form

$$X_n = X_0 + b_n + \sigma R_n, n \geq 1$$

where $X_0 \in \mathbb{R}$, $\{b_n, n \geq 1\}$ is a sequence of real numbers, $\{R_n, n \geq 1\}$ is a sequence of random variables of the form $R_n = \sum_{i=1}^n \varepsilon_i$, where $\{\varepsilon_i, i \geq 1\}$ are independent centered random variables, and $\sigma > 0$ is a parameter to be estimated. In order to estimate σ, we define the differences $\Delta X_n = X_n - X_{n-1} = \Delta b_n + \sigma \varepsilon_n$, where $\Delta b_n = b_n - b_{n-1}$ and $\varepsilon_n = R_n - R_{n-1}$, and construct an estimator of the form

$$(\widehat{\sigma})_n^2 = \frac{\sum_{i=1}^n (\Delta X_i)^2}{n} = \frac{\sum_{i=1}^n (\Delta b_i)^2}{n} + \frac{2\sigma \sum_{i=1}^n \varepsilon_i \Delta b_i}{n} + \frac{\sigma^2 \sum_{i=1}^n \varepsilon_i^2}{n}.$$

THEOREM 10.5.– *Let the following conditions hold: ε_i are iid random variables, $\varepsilon_i \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$, $i \geq 1$, and*

$$\frac{\sum_{i=1}^n (\Delta b_i)^2}{n} \to b, \ \text{as } n \to \infty, \qquad [10.1]$$

where $b \geq 0$ is some constant. Then $(\widehat{\sigma})_n^2 \to \sigma^2 \mathrm{E}\varepsilon_1^2 + b$ a.s. as $n \to \infty$.

PROOF.– If ε_i are iid random variables with $\mathrm{E}\varepsilon_1^2 < \infty$, then by the version of SLLN formulated in theorem A2.18, $\frac{\sum_{i=1}^n \varepsilon_i^2}{n} \to \mathrm{E}\varepsilon_1^2$ a.s. as $n \to \infty$. Moreover, $\varepsilon_i \Delta b_i$ are independent centered random variables with $\sum_{i=1}^\infty \frac{\mathrm{E}\varepsilon_i^2(\Delta b_i)^2}{i^2} = \mathrm{E}\varepsilon_1^2 \sum_{i=1}^\infty \frac{(\Delta b_i)^2}{i^2}$. Further, under condition [10.1], there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ it holds that

$$\sum_{i=1}^n (\Delta b_i)^2 \leq (b+1)n.$$

Therefore, for such n that $2^{n+1} \geq n_0$, the following inequalities hold:

$$\sum_{i=2^n}^{2^{n+1}} \frac{(\Delta b_i)^2}{i^2} \leq 2^{-2n} \sum_{i=1}^{2^{n+1}} (\Delta b_i)^2 \leq 2^{-2n}(b+1)2^{n+1} \leq 2(b+1)2^{-n}.$$

It means that the series $\sum_{i=1}^\infty \frac{(\Delta b_i)^2}{i^2}$ converges. Therefore, according to the version of SLLN formulated in theorem A2.19, $\frac{\sum_{i=1}^\infty \varepsilon_i \Delta b_i}{n} \to 0$ a.s. as $n \to \infty$, and the proof follows. □

REMARK 10.3.–

i) Let $\Delta b_n \to b$ as $n \to \infty$. Then $\frac{\sum_{i=1}^n (\Delta b_i)^2}{n} \to b$ as $n \to \infty$, according to the Stolz-Cesáro theorem, and it follows as in the proof of theorem 10.5 that $\sum_{i=1}^\infty \frac{(\Delta b_i)^2}{i^2} < \infty$. In this case theorem 10.5 holds.

ii) Let $b_n = n$. Then condition [10.1] and consequently theorem 10.5 hold with $b = 1$.

10.3. Drift and diffusion parameter estimation in the linear model with continuous time and the Wiener noise

Consider the model of the form

$$X_t = X_0 + \theta f(t) + \sigma W_t,$$

where $X_0 \in \mathbb{R}$ is a known initial value, $\theta \in \mathbb{R}$, $\sigma > 0$ are parameters, $f = f(t)$ is a positive measurable function that we assume to know, and $W = \{W_t, t \geq 0\}$ is a Wiener process. As before, θ is called a drift parameter, while σ is called a diffusion parameter.

10.3.1. *Drift parameter estimation*

We suppose that X_t is observable for all $t \geq 0$, and propose two estimators for θ constructed by its observations on the interval $[0, T]$, assuming that $\sigma > 0$ is known. The first one is the estimator constructed by the last observation, it has a form $\widehat{\theta}_T^{(1)} = \frac{X_T}{f(T)}$. The second one is the least square estimator (LSE) constructed under additional assumption $f \in C^{(1)}(\mathbb{R}_+)$ and f' does not equal to zero identically, as a result of minimization of the quadratic form that can be formally written as $\int_0^T (dX_s - \theta f'(s)ds)^2$. Finally, the estimator has a form

$$\widehat{\theta}_T^{(2)} = \frac{\int_0^T f'(s)dX_s}{\int_0^T (f'(s))^2 ds}.$$

THEOREM 10.6.–

i) Let there exist $t_0 \geq 0$ and $C > 0$ such that $|f(t)| \geq Ct$ for all $t \geq t_0$. Then $\widehat{\theta}_T^{(1)}$ is a strongly consistent estimator of θ.

ii) Let $f \in C^{(1)}(\mathbb{R}_+)$, and $\int_0^T (f'(s))^2 ds \to \infty$ as $T \to \infty$. Then $\widehat{\theta}_T^{(2)}$ is a strongly consistent estimator of θ.

PROOF.– i) According to SLLN for martingales with continuous time (theorem 8.10), $\frac{W_T}{T} \to 0$ a.s. as $T \to \infty$, whence the proof of (i) follows.

ii) Note that

$$\widehat{\theta}_T^{(2)} = \theta + \frac{\int_0^T f'(s)dW_s}{\int_0^T (f'(s))^2 ds}.$$

Due to condition (ii), we can apply SLLN here as well for martingales with continuous time, according to which $\frac{\int_0^T f'(s)dW_s}{\int_0^T (f'(s))^2 ds} \to 0$ a.s. as $T \to \infty$. Hence, the proof follows. □

REMARK 10.4.– The conditions (i) and (ii) are not completely overlapping. For example, let $f(t) = |f_1(t)| + t$, where the function f_1 is nowhere differentiable. Then we can not apply $\widehat{\theta}_T^{(2)}$, however, $\widehat{\theta}_T^{(1)}$ works. Conversely, let $f(t) = \sqrt{t}\mathbb{1}_{t \geq 1}$. Then $\int_0^\infty (f'(s))^2 ds = \frac{1}{4} \int_1^\infty \frac{ds}{s} = +\infty$, so $\widehat{\theta}_T^{(2)}$ works, but $\frac{W_T}{\sqrt{T}} \sim \mathcal{N}(0, 1)$, therefore $\frac{W_T}{\sqrt{T}}$ does not converge to zero as $T \to \infty$.

REMARK 10.5.– As is established in theorem A2.23, $\frac{W_T}{T^\alpha} \to 0$ a.s. as $T \to \infty$ for any $\alpha > \frac{1}{2}$. Therefore, condition (i) can be relaxed to $|f(t)| \geq Ct^\alpha$ for some $C > 0$, $\alpha > \frac{1}{2}$ and all $t \geq t_0 > 0$.

10.3.2. *Diffusion parameter estimation*

In order to estimate σ assuming θ to be known, we use the properties of quadratic variation of the Wiener process. Namely, we fix some $T > 0$ and consider the equidistant partition of $[0, T]$, e.g. $\pi_n = \left\{\frac{Tk}{n} = t_k^{(n)}, 0 \leq k \leq n\right\}$. According to remark 4.2, $\sum_{k=1}^n (W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}})^2 \to T$ a.s. as $n \to \infty$. Therefore, we can formulate and prove the following result.

THEOREM 10.7.– *Let for any* $T > 0$, $f \in C([0, T]) \cap BV([0, T])$. *Then the estimator*

$$(\widehat{\sigma})^2_{T,n} := \sum_{k=1}^n (X_{t_k^{(n)}} - X_{t_{k-1}^{(n)}})^2 \to \sigma^2 T \ \textit{a.s. as} \ n \to \infty,$$

i.e., $\frac{(\widehat{\sigma})^2_{T,n}}{T} \to \sigma^2$ *a.s. as* $n \to \infty$.

PROOF.– We have that

$$(\widehat{\sigma})^2_{T,n} = \sum_{k=1}^n \left(\theta \left[f_{t_k^{(n)}} - f_{t_{k-1}^{(n)}}\right] + \sigma \left(W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}}\right)^2\right).$$

Denote $\Delta f_k = f_{t_k^{(n)}} - f_{t_{k-1}^{(n)}}$ and $\Delta W_k = W_{t_k^{(n)}} - W_{t_{k-1}^{(n)}}$. Then

$$(\widehat{\sigma})^2_{T,n} = \theta^2 \sum_{k=1}^n (\Delta f_k)^2 + 2\theta\sigma \sum_{k=1}^n \Delta f_k \Delta W_k + \sigma^2 \sum_{k=1}^n (\Delta W_k)^2.$$

As we mentioned before, $\sum_{k=1}^n (\Delta W_k)^2 \to T$ a.s. as $n \to \infty$. Since $f \in C([0, T]) \cap BV([0, T])$, we can conclude that

$$\sum_{k=1}^n (\Delta f_k)^2 \leq \max_{1 \leq k \leq n} |\Delta f_k| \sum_{k=1}^n |\Delta f_k| \leq \max_{1 \leq k \leq n} |\Delta f_k| \text{Var}_{[a,b]} f \to 0, \ \ n \to \infty,$$

(see also theorem A1.11), and similarly

$$\left|\sum_{k=1}^n \Delta f_k \Delta W_k\right| \leq \max_{1 \leq k \leq n} |\Delta W_k| \text{Var}_{[a,b]} f \to 0, \ \text{ a.s. as } \ n \to \infty,$$

and hence the proof follows. □

REMARK 10.6.– Some related properties of functions of bounded variation are considered in section A1.5.

10.4. Parameter estimation in linear models with fractional Brownian motion

Consider a model

$$X_t = X_0 + \theta f(t) + \sigma B_t^H,\ t \geq 0, \qquad [10.2]$$

where $X_0 \in \mathbb{R}$, $f \in C^{(1)}(\mathbb{R})$, $\sigma > 0$ and $\theta \in \mathbb{R}$ are the parameters to be estimated, $B^H = \{B_t^H, t \geq 0\}$ is a fractional Brownian motion with Hurst index $H \in (0,1)$.

10.4.1. *Estimation of Hurst index*

Assume that the index $H \in (0,1)$ is unknown. Fix any $T > 0$ and suppose that the process X is observable on $[0,T]$. Consider a sequence of uniform partitions

$$\overline{\pi}_n = \left\{T\delta_k,\ \ \delta_k^{(n)} = \frac{k}{2^n}, 0 \leq k \leq 2^n\right\},\ n \geq 1. \qquad [10.3]$$

According to theorem 4.6, for any $H \in (0,1)$

$$\frac{2^{n(2H-1)}}{T^{2H}} \sum_{k=1}^{2^n} \left(B^H_{T\delta_k^{(n)}} - B^H_{T\delta_{k-1}^{(n)}}\right)^2 \to \frac{2}{\sqrt{\pi}} \ \text{ a.s. as } \ n \to \infty. \qquad [10.4]$$

Replacing n with $2n$, we get from [10.4] that

$$\frac{2^{2n(2H-1)}}{T^{2H}} \sum_{k=1}^{2^{2n}} \left(B^H_{T\delta_k^{(2n)}} - B^H_{T\delta_{k-1}^{(2n)}}\right)^2 \to \frac{2}{\sqrt{\pi}} \ \text{ a.s. as } \ n \to \infty. \qquad [10.5]$$

Dividing the left-hand side of [10.5] by the left-hand side of [10.4], we get that

$$2^{n(2H-1)} \frac{\sum_{k=1}^{2^{2n}} \left(B^H_{T\delta_k^{(2n)}} - B^H_{T\delta_{k-1}^{(2n)}}\right)^2}{\sum_{k=1}^{2^n} \left(B^H_{T\delta_k^{(n)}} - B^H_{T\delta_{k-1}^{(n)}}\right)^2} \to 1 \ \text{ a.s. as } \ n \to \infty. \qquad [10.6]$$

Denote

$$R_{T,n}(B^H) = \frac{\sum_{k=1}^{2^{2n}} \left(B^H_{T\delta_k^{(2n)}} - B^H_{T\delta_{k-1}^{(2n)}}\right)^2}{\sum_{k=1}^{2^n} \left(B^H_{T\delta_k^{(n)}} - B^H_{T\delta_{k-1}^{(n)}}\right)^2}. \qquad [10.7]$$

Then it follows from [10.6] that

$$\log_2 R_{T,n}(B^H) + n(2H-1) \to 0, \ \text{ a.s., or}$$
$$-\frac{\log_2 R_{T,n}(B^H)}{2n} + \frac{1}{2} \to H \ \text{ a.s. as } \ n \to \infty. \qquad [10.8]$$

Now, let $R_{T,n}(X)$ be the right-hand side of [10.7], where we substitute X instead of B^H. Taking [10.8] into account, we can prove the following result.

THEOREM 10.8.– *For any $T > 0$*

$$\widehat{H}_{T,n} := -\frac{\log_2 R_{T,n}(X)}{2n} + \frac{1}{2}$$

is a strongly consistent estimator of H.

PROOF.– It is sufficient to prove that

$$r_n = \frac{\log_2 R_{T,n}(X)}{2n} - \frac{\log_2 R_{T,n}(B^H)}{2n} \to 0 \ \text{ a.s. as } \ n \to \infty.$$

However, $r_n = \frac{1}{2n} \log_2 \frac{R_{T,n}(X)}{R_{T,n}(B^H)}$, and

$$R_n(X) = \frac{\sum_{k=1}^{2^{2n}} \left(X_{T\delta_k^{(2n)}} - X_{T\delta_{k-1}^{(2n)}} \right)^2}{\sum_{k=1}^{2^n} \left(X_{T\delta_k^{(n)}} - X_{T\delta_{k-1}^{(n)}} \right)^2} = \left[\sum_{k=1}^{2^{2n}} \theta^2 \left(f_{T\delta_k^{(2n)}} - f_{T\delta_{k-1}^{(2n)}} \right)^2 \right.$$
$$+ 2\theta\sigma \sum_{k=1}^{2^{2n}} \left(f_{T\delta_k^{(2n)}} - f_{T\delta_{k-1}^{(2n)}} \right) \left(B^H_{T\delta_k^{(2n)}} - B^H_{T\delta_{k-1}^{(2n)}} \right)$$
$$\left. + \sigma^2 \sum_{k=1}^{2^{2n}} \left(B^H_{T\delta_k^{(2n)}} - B^H_{T\delta_{k-1}^{(2n)}} \right)^2 \right] \left[\sum_{k=1}^{2^n} \theta^2 \left(f_{T\delta_k^{(n)}} - f_{T\delta_{k-1}^{(n)}} \right)^2 \right.$$
$$+ 2\theta\sigma \sum_{k=1}^{2^n} \left(f_{T\delta_k^{(n)}} - f_{T\delta_{k-1}^{(n)}} \right) \left(B^H_{T\delta_k^{(n)}} - B^H_{T\delta_{k-1}^{(n)}} \right)$$
$$\left. + \sigma^2 \sum_{k=1}^{2^n} \left(B^H_{T\delta_k^{(n)}} - B^H_{T\delta_{k-1}^{(n)}} \right)^2 \right]^{-1}.$$

According to theorem 10.8 and remark A1.1,

$$\lim_{n\to\infty}\sum_{k=1}^{2^n}\left(f_{T\delta_k^{(n)}}-f_{T\delta_{k-1}^{(n)}}\right)^2=0, \text{ and}$$

$$\lim_{n\to\infty}\sum_{k=1}^{2^n}\left(f_{T\delta_k^{(n)}}-f_{T\delta_{k-1}^{(n)}}\right)\left(B^H_{T\delta_k^{(n)}}-B^H_{T\delta_{k-1}^{(n)}}\right)=0 \text{ a.s.} \qquad [10.9]$$

The same is true if we replace n in [10.9] for $2n$. Therefore $\frac{R_{T,n}(X)}{R_{T,n}(B^H)}\to 1$ a.s. as $n\to\infty$ and consequently $r_n\to 0$ a.s. as $n\to\infty$. □

10.4.2. *Estimation of the diffusion parameter*

Consider the model [10.2] in which we assume the Hurst parameter $H\in(0,1)$ to be known. Let us estimate the parameter $\sigma>0$. Consider the sequence [10.3] of uniform partitions of $[0,T]$.

THEOREM 10.9.– *For any $T>0$ and for any $H\in(0,1)$,*

$$(\widehat{\sigma})^2_{H,T,n}:=\frac{2^{n(2H-1)}\sqrt{\pi}}{2T^{2H}}\sum_{k=1}^{2^n}\left(X_{T\delta_k^{(n)}}-X_{T\delta_{k-1}^{(n)}}\right)^2$$

is a strongly consistent estimator of σ^2.

PROOF.– We start with evident equality

$$(\widehat{\sigma})^2_{H,T,n}=\frac{2^{n(2H-1)}\sqrt{\pi}}{2T^{2H}}\left(\sum_{k=1}^{2^n}\left(f_{T\delta_k^{(n)}}-f_{T\delta_{k-1}^{(n)}}\right)^2\right.$$
$$+2\sum_{k=1}^{2^n}\left(f_{T\delta_k^{(n)}}-f_{T\delta_{k-1}^{(n)}}\right)\left(B^H_{T\delta_k^{(n)}}-B^H_{T\delta_{k-1}^{(n)}}\right) \qquad [10.10]$$
$$\left.+\sigma^2\sum_{k=1}^{2^n}\left(B_{T\delta_k^{(n)}}-B_{T\delta_{k-1}^{(n)}}\right)^2\right).$$

Let us analyze the asymptotic behavior of all terms in the right-hand side of [10.10]. Denote $C_f=\max_{0\le t\le T}|f'(t)|$. Then

$$2^{n(2H-1)}\sum_{k=1}^{2^n}\left(f_{T\delta_k^{(n)}}-f_{T\delta_{k-1}^{(n)}}\right)^2\le C_f^2 2^{n(2H-1)}2^n\left(\frac{T}{2^n}\right)^2$$
$$=C_f^2T^2 2^{2nH-n+n-2n}=C_f^2T^2 2^{2n(H-1)}\to 0 \text{ as } n\to\infty. \qquad [10.11]$$

To estimate the middle term, recall that for any $\omega \in \Omega'$ with $\mathrm{P}\{\Omega'\} = 1$ and for any $0 < \delta < H$ there exists such $C(\omega, T, \delta)$ that $|B_t^H - B_s^H| \le C(\omega, T, \delta)|t-s|^{H-\delta}$ for any $0 \le s, t \le T$ (see example 6.5). Therefore, for any $\omega \in \Omega'$

$$\begin{aligned} S_n(\omega) &:= 2^{n(2H-1)} \sum_{k=1}^{2^n} \left| f_{T\delta_k^{(n)}} - f_{T\delta_{k-1}^{(n)}} \right| \left| B^H_{T\delta_k^{(n)}} - B^H_{T\delta_{k-1}^{(n)}} \right| \\ &\le 2^{n(2H-1)} 2^n C_f \frac{T}{2^n} C(\omega, T, \delta) \left(\frac{T}{2^n} \right)^{H-\delta} \\ &= C_f C(\omega, T, \delta) T^{1+H-\delta} 2^{2Hn-n+n-n-nH+n\delta} \\ &= C_f C(\omega, T, \delta) T^{1+H-\delta} 2^{(H-1)n+n\delta}. \end{aligned} \qquad [10.12]$$

Choosing $0 < \delta < H \wedge (1-H)$ we get that $S_n(\omega) \to 0$ a.s. as $n \to \infty$. Finally, it follows from [10.10]–[10.12] that

$$\frac{2^{n(2H-1)}\sqrt{\pi}}{2T^{2H}} \sigma^2 \sum_{k=1}^{2^n} \left(B^H_{T\delta_k^{(n)}} - B^H_{T\delta_{k-1}^{(n)}} \right)^2 \to \sigma^2 \ \text{ a.s. as } \ n \to \infty,$$

and hence the proof follows. □

10.5. Drift parameter estimation

Consider the model [10.2] in which we assume the Hurst parameter $H \in (0, 1)$ to be known. The Diffusion parameter $\sigma > 0$ is also assumed to be known, and we can simply put $\sigma = 1$. And so, let us estimate the parameter $\theta \in \mathbb{R}$ in the model

$$X_t = X_0 + \theta f(t) + B_t^H, \ \ t \ge 0,$$

where $f \in C^{(1)}(\mathbb{R})$. Note that both integrals $\int_0^t (t-s)^{\frac{1}{2}-H} s^{\frac{1}{2}-H} f'(s) ds$ and $\int_0^t (t-s)^{\frac{1}{2}-H} s^{\frac{1}{2}-H} dB_s^H$ are well defined, and moreover, the process $M_t^H = C'_H \int_0^t (t-s)^{\frac{1}{2}-H} s^{\frac{1}{2}-H} dB_s^H$ is a continuous martingale with quadratic variation $[M^H]_t = t^{2-2H}$, see section A2.8. Therefore, we can consider the process $Y_t^H = \theta F_t^H + M_t^H$, where $Y_t^H = C'_H \int_0^t (t-s)^{\frac{1}{2}-H} s^{\frac{1}{2}-H} dX_s$ and $F_t^H = C'_H \int_0^t (t-s)^{\frac{1}{2}-H} s^{\frac{1}{2}-H} f'(s) ds$.

THEOREM 10.10.–

i) Let there exist $t_0 > 0$ such that $|F_t^H| \ge C t^{2-2H}$ for some $C > 0$ and any $t \ge t_0$. Then $\widehat{\theta}_T^{(1)} = \frac{Y_T^H}{F_T^H}$ is a strongly consistent estimator of θ.

ii) Let there exist $t_0 > 0$ and $\delta > 0$ such that $|f(t)| \geq Ct^{H+\delta}$, for some $C > 0$ and any $t \geq t_0$. Then $\widehat{\theta}_T^{(2)} = \frac{X_T}{f(T)}$, considered for $T \geq t_0$, is a strongly consistent estimator of θ.

PROOF.– i) Obviously, $\frac{Y_T^H}{F_T^H} = \theta + \frac{M_T^H}{F_T^H}$, and the proof immediately follows from the SLLN for martingales.

ii) In this case $\widehat{\theta}_T^2 = \frac{X_0}{f(T)} + \theta + \frac{B_T^H}{f(T)}$, and the proof follows from theorem A2.24.□

10.6. Drift parameter estimation in the simplest autoregressive model

Consider the linear autoregressive model of the form

$$X_n = \theta X_{n-1} + \varepsilon_n, \ n \geq 1. \tag{10.13}$$

Assume that $\{\varepsilon_n, n \geq 1\}$ are i.i.d. $\mathcal{N}(0,1)$-random variables representing the noise, $\theta \in \mathbb{R}$ is a parameter to be estimated. The values $\{X_0, X_1, \ldots, X_n, \ldots\}$ are assumed to be observable, $X_0 \in \mathbb{R}$. And so, our goal is to construct the statistics that will be an estimator of θ, based on the observations $\{X_k, k \geq 1\}$.

For this purpose, we construct the least square estimator, which minimizes the value $\sum_{i=1}^n (X_i - \theta X_{i-1})^2$. A minimum value is achieved at the point $\widehat{\theta}_n = \frac{\sum_{i=1}^n X_i X_{i-1}}{\sum_{i=1}^n X_{i-1}^2}$. To establish the asymptotic properties of $\widehat{\theta}_n$, we need the following lemma.

LEMMA 10.1.– *For any $\theta \in \mathbb{R}$ it holds that $\sum_{i=1}^n X_{i-1}^2 \to \infty$ a.s. as $n \to \infty$.*

PROOF.– a) Let $\theta > 1$. Note that $X_n = \theta^n X_0 + \theta^{n-1}\varepsilon_1 + \ldots + \theta\varepsilon_{n-1} + \varepsilon_n$. Further, it follows from the Cauchy-Schwartz inequality that

$$\sum_{i=1}^n X_{i-1}^2 \geq \frac{\left(\sum_{i=1}^n X_{i-1}\right)^2}{n}.$$

Therefore, for any $\lambda > 0$

$$\mathrm{E}\exp\left\{-\lambda \sum_{i=1}^n X_{i-1}^2\right\} \leq \mathrm{E}\exp\left\{-\frac{\lambda}{n}\left(\sum_{i=1}^n X_{i-1}\right)^2\right\}.$$

Denote

$$m_n = \mathrm{E}\sum_{i=1}^{n} X_{i-1} = \sum_{i=1}^{n} \theta^{i-1} X_0 = \frac{\theta^n - 1}{\theta - 1} X_0,$$

and

$$\sigma_n^2 = \mathrm{E}\left(\sum_{i=1}^{n} X_{i-1} - m_n\right)^2.$$

Since ε_i are pairwise uncorrelated, we get that

$$\begin{aligned}\sigma_n^2 &= \mathrm{E}\left(\varepsilon_1 \frac{\theta^{n-1}-1}{\theta-1} + \varepsilon_2 \frac{\theta^{n-2}-1}{\theta-1} + \ldots + \varepsilon_{n-1}\right)^2 \\ &= (\theta-1)^{-2}((\theta^{n-1}-1)^2 + (\theta^{n-2}-1)^2 + \ldots + (\theta-1)^2) \\ &= (\theta-1)^{-2}(\theta^{2(n-1)} + \theta^{2(n-2)} \\ &\quad + \ldots + \theta^2 - 2\theta^{n-1} - 2\theta^{n-2} - \ldots - 2\theta + n - 1) \\ &= (\theta-1)^{-2}\left(\frac{\theta^{2n}-1}{\theta^2-1} - 2\frac{\theta^n-1}{\theta-1} + n\right) \\ &= (\theta-1)^{-3}\left(\frac{\theta^{2n}-1}{\theta+1} - 2(\theta^n-1) + n(\theta-1)\right).\end{aligned} \qquad [10.14]$$

Now, on one hand, it follows from [10.14] that $\sigma_n = O(\theta^n)$ as $n \to \infty$. On the other hand, if we have a $\mathcal{N}(m_n, \sigma_n^2)$-random variable ξ_n, then

$$\begin{aligned}\mathrm{E}\exp\left\{-\frac{\lambda}{n}\xi_n^2\right\} &= \left(\sqrt{2\pi}\sigma_n\right)^{-1}\int_{\mathbb{R}} e^{-\frac{\lambda}{n}x^2 - \frac{(x-m_n)^2}{2\sigma_n^2}}dx \\ &\le \sqrt{\frac{n}{2\lambda}}\sigma_n^{-1}\int_{\mathbb{R}}\frac{e^{-\frac{\lambda x^2}{n}}}{\sqrt{2\pi}\sqrt{\frac{n}{2\lambda}}}dx = \sqrt{\frac{n}{2\lambda}}\sigma_n^{-1}.\end{aligned}$$

Therefore, $\mathrm{E}\exp\left\{-\frac{\lambda}{n}\left(\sum_{i=1}^{n} X_{i-1}\right)^2\right\} \to 0$ as $n \to \infty$, and consequently

$$\mathrm{E}\exp\left\{-\lambda\sum_{i=1}^{n} X_{i-1}^2\right\} \to 0$$

as $n \to \infty$ for any $\lambda > 0$. Since $\sum_{i=1}^{n} X_{i-1}^2$ is a non-decreasing sequence, we have from lemma A2.10 that $\sum_{i=1}^{n} X_{i-1}^2 \to \infty$ a.s. as $n \to \infty$.

b) Let $\theta = 1$. Then

$$X_n = X_0 + \sum_{k=1}^{n} \varepsilon_k, \quad m_n = \mathrm{E} \sum_{i=1}^{n} X_{i-1} = nX_0,$$

$$\sigma_n^2 = \mathrm{E}\left(\sum_{i=1}^{n} X_{i-1} - nX_0\right)^2 = \mathrm{E}\left(\sum_{i=1}^{n}\left(\sum_{k=1}^{i-1}\varepsilon_k\right)\right)^2$$
$$= \mathrm{E}\left(\sum_{k=1}^{i-1}\varepsilon_k(n-k)\right)^2$$
$$= \sum_{k=1}^{n-1}(n-k)^2 = \sum_{j=1}^{n-1} j^2 = \frac{(n-1)n(2n-1)}{6} = O(n^3), \text{ as } n \to \infty,$$

and we can conclude as in (a) because $\sqrt{n}\sigma_n^{-1} \to 0$ as $n \to \infty$.

c) Let $\theta < -1$. In this case $\sigma_n = O(|\theta|^n)$, therefore as in (a), $\sqrt{n}\sigma_n^{-1} \to 0$ as $n \to \infty$ and hence we get the statement.

d) Let $\theta = -1$. Then

$$X_n = (-1)^n X_0 + \sum_{k=1}^{n}(-1)^{n-k}\varepsilon_k,$$

therefore

$$(-1)^n X_n = X_0 + \sum_{k=1}^{n}(-1)^k\varepsilon_k.$$

Operating with $Y_n := (-1)^n X_n$ instead of X_n and $\epsilon_k = (-1)^k\varepsilon_k$ instead of ε_k, we get that their moments $m_n = nX_0$ and

$$\sigma_n^2 = \mathrm{E}\left(\sum_{k=1}^{n-1}(-1)^k\epsilon_k(n-k)\right)^2 = \sum_{k=1}^{n-1}(n-k)^2 = O(n^3) \text{ as } n \to \infty.$$

Additionally, $\sum_{i=1}^{n} X_{i-1}^2 = \sum_{i=1}^{n} Y_{i-1}^2$, and we can conclude as in (b).

e) Finally, let $\theta \in (-1, 1)$. Then

$$\sum_{i=0}^{n} X_i^2 = X_0^2 + \theta^2 \sum_{i=0}^{n-1} X_i^2 + 2\theta \sum_{i=0}^{n-1} X_i\varepsilon_{i+1} + \sum_{i=1}^{n}\varepsilon_i^2. \quad [10.15]$$

Denote $M_n = \sum_{i=0}^{n-1} X_i \varepsilon_{i+1}$. This is square-integrable martingale (more precisely, M_n has the moments of any order), and

$$\langle M \rangle_n = \sum_{i=0}^{n-1} \mathrm{E}\left((X_i \varepsilon_{i+1})^2 | \mathcal{F}_i\right) = \sum_{i=1}^{n-1} X_i^2.$$

According to theorem 5.22 $\{\langle M \rangle_\infty < \infty\} \subset \{M \to\}$. For such $\omega \in \Omega$, where $\langle M \rangle_\infty < \infty$, we can pass to the limit as $n \to \infty$ in [10.15] and get the equality

$$\langle M \rangle_\infty = X_0^2 + \theta^2 \langle M \rangle_\infty + 2\theta M_\infty + \lim_{n\to\infty} \sum_{i=1}^{n} \varepsilon_i^2,$$

or

$$(1-\theta^2) \langle M \rangle_\infty = X_0^2 + 2\theta M_\infty + \lim_{n\to\infty} \sum_{i=1}^{n} \varepsilon_i^2. \qquad [10.16]$$

Now, note that according to the strong law of large numbers,

$$\frac{\sum_{i=1}^{n} \varepsilon_i^2}{n} \to \mathrm{E}\varepsilon_1^2 = 1 \text{ a.s.},$$

which means that $\sum_{i=1}^{n} \varepsilon_i^2 \to \infty$ a.s. It means that [10.16] is contradictory, which means in turn that the probability of the event $\{\langle M \rangle_\infty < \infty\}$ is zero and hence the proof follows. □

THEOREM 10.11.– *The estimator*

$$\widehat{\theta}_n = \frac{\sum_{i=1}^{n} X_i X_{i-1}}{\sum_{i=1}^{n} X_{i-1}^2}$$

is a strongly consistent estimator of parameter θ in the linear regression scheme [10.13].

PROOF.– We can transform $\widehat{\theta}_n$ as

$$\widehat{\theta}_n = \theta + \frac{\sum_{i=1}^{n} X_{i-1}\varepsilon_i}{\sum_{i=1}^{n} X_{i-1}^2}.$$

Denote by M_n martingale $M_n = \sum_{i=1}^{n} X_{i-1}\varepsilon_i$. Then its quadratic characteristics equal $\langle M \rangle_n = \sum_{i=1}^{n} X_{i-1}^2$. As it follows from lemma 10.1, $\langle M \rangle_\infty = \infty$ a.s. Then it follows from the strong law of large numbers for martingales (theorem 5.21) that $\frac{M_n}{\langle M \rangle_n} \to 0$ as $n \to \infty$ with probability 1, and hence the proof follows. □

REMARK 10.7.– An autoregressive scheme [10.13] is a particular case of autoregressive-moving-average (ARMA) models of the form

$$X_n = \sum_{i=1}^{p} \theta_{n-i} X_{n-i} + \sum_{i=1}^{p} \varphi_{n-i} \varepsilon_{n-i} + \varepsilon_n,$$

which are treated in detail within the framework of time series analysis, see [BRO 06] for example.

10.7. Drift parameters estimation in the homogeneous diffusion model

Consider the diffusion process $X = \{X_t, t \geq 0\}$ which is a solution to the stochastic differential equation with homogeneous, i.e. not depending on t, coefficients:

$$dX_t = \theta_0 a(X_t)dt + b(X_t)dW_t. \quad [10.17]$$

We assume that $X|_{t=0} = X_0 \in \mathbb{R}$ is non-random, coefficients $a, b : \mathbb{R} \to \mathbb{R}$ are measurable functions, $b(x) > 0$ for all $x \in \mathbb{R}$, $W = \{W_t, t \geq 0\}$ is a Wiener process, $\theta_0 \in \mathbb{R}$ is an unknown drift parameter to be estimated. We also assume that the SDE [10.17] has a unique strong solution X_t, $t \geq 0$. It follows from theorem 9.2 that the existence and uniqueness of a strong solution is supplied in a homogeneous case by the Lipschitz condition: (A) there exists $K > 0$ such that for any $x, y \in \mathbb{R}$ $|a(x) - a(y)| + |b(x) - b(y)| \leq K|x - y|$, since in a homogeneous case it implies linear growth condition. Let us construct a maximum likelihood estimator of θ. To this end we test the hypothesis $H_{\theta_0} : \theta = \theta_0$ against the alternative $H_0 : \theta = 0$. The measure that corresponds to H_{θ_0} is P_{θ_0}, the measure that corresponds to H_0 is P_0, up to our notation.

Let us transform the right-hand side of [10.17] as follows: $dX_t = b(X_t) \left[\theta \frac{a(X_t)}{b(X_t)} dt + dW_t\right]$. And so, under H_0 the process $W_t + \theta \int_0^t \frac{a(X_s)}{b(X_s)} ds$ is a new Wiener process $\widetilde{W}_t$, so that $\widetilde{W}_t = W_t + \theta \int_0^t \frac{a(X_s)}{b(X_s)} ds$. According to the Girsanov theorem,

$$\left.\frac{d\mathrm{P}_0}{d\mathrm{P}_{\theta_0}}\right|_{\mathcal{F}_T} = \exp\left\{-\theta_0 \int_0^T \frac{a(X_s)}{b(X_s)} dW_s - \frac{1}{2}\theta_0^2 \int_0^t \frac{a^2(X_s)}{b^2(X_s)} ds\right\}.$$

Since we are interested in the true value θ_0 of an unknown parameter, we consider the inverse likelihood function $\frac{d\mathrm{P}_0}{d\mathrm{P}_{\theta_0}}\Big|_{\mathcal{F}_T}$ and get that

$$\frac{d\mathrm{P}_{\theta_0}}{d\mathrm{P}_0}\Big|_{\mathcal{F}_T} = \exp\left\{\theta_0 \int_0^T \frac{a(X_s)}{b(X_s)} dW_s + \frac{1}{2}\theta_0^2 \int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds\right\}$$
$$= \exp\left\{\theta_0 \int_0^T \frac{a(X_s)}{b^2(X_s)} dX_s - \frac{1}{2}\theta_0^2 \int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds\right\}. \quad [10.18]$$

Obviously, the following condition is needed for the right-hand side of [10.18] to be well-defined.

(B) for any $T > 0$, $\int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds < \infty$ a.s.

THEOREM 10.12.– *Let the conditions (A) and (B) hold. Then the maximum likelihood function for the estimation of parameter θ_0 has a form*

$$\widehat{\theta}_T := \left(\int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds\right)^{-1} \int_0^T \frac{a(X_s)}{b^2(X_s)} dX_s. \quad [10.19]$$

PROOF.– As we have just established, the likelihood ratio for the equivalent measures P_0 is presented by the right-hand side [10.18]. To maximize it, we take the quadratic in θ_0 function

$$\theta_0 \int_0^T \frac{a(X_s)}{b^2(X_s)} dX_s - \frac{1}{2}\theta_0^2 \int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds,$$

obtaining that the maximum is achieved in $\widehat{\theta}_T$, represented by the right-hand side of [10.19]. □

Now consider the denominator of $\widehat{\theta}_T$, $\int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds$. Its asymptotic behavior is based on the following result from [MIS 14]. Since the proof of this result involves some notions (local time, recurrent and transient diffusion processes) which are beyond the scope of the present book, we omit it.

THEOREM 10.13.– *Let the following conditions hold: (A), (B) and (C) function $b(x) \neq 0$ for all $x \in \mathbb{R}$, function $a(x)$ is not identically zero, functions $\frac{1}{b^2(x)}$, $\frac{a^2(x)}{b^2(x)}$ and $\frac{a^2(x)}{b^4(x)}$ are locally integrable, i.e. integrals*

$$\int_{[-N,N]} b^{-2}(x)dx, \int_{[-N,N]} a^2(x)b^{-2}(x)dx \text{ and } \int_{[-N,N]} a^2(x)b^{-4}(x)dx$$

exist for any $N > 0$. Then $\int_0^\infty \frac{a^2(X_s)}{b^2(X_s)} ds = +\infty$ a.s.

Taking into account this theorem, we can establish our main result.

THEOREM 10.14.– *Let the condition* (A)*,* (B) *and* (C) *hold. Then*

$$\widehat{\theta}_T = \left(\theta_0^2 \int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds\right)^{-1} \int_0^T \frac{a(X_s)}{b^2(X_s)} dX_s$$

is a strongly consistent estimator of the unknown drift parameter θ_0*.*

PROOF.– Let us present $\widehat{\theta}_T$ as

$$\widehat{\theta}_T = \theta_0 + \frac{\int_0^T \frac{a(X_s)}{b(X_s)} dW_s}{\int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds}.$$

Under condition (B) the process $M_t = \int_0^t \frac{a(X_s)}{b(X_s)} dW_s$ is a local martingale with quadratic variation $[M]_t = \int_0^t \frac{a^2(X_s)}{b^2(X_s)} ds$.

According to condition (C) and theorem 10.13, $\int_0^T \frac{a^2(X_s)}{b^2(X_s)} ds \to \infty$ a.s. as $T \to \infty$. Therefore, the proof follows from the SLLN for local martingales with continuous time, established in theorem 8.10. □

EXAMPLE 10.3.– *Consider the Langevin equation with the unknown drift parameter*

$$dX_t = \theta X_t dt + dW_t, \;\; t \geq 0.$$

In this case, coefficient $a(x) = x$ is not identically zero, $b(x) = 1 \neq 0$, conditions (B) and (C) obviously are fulfilled. It means that

$$\int_0^\infty \frac{a^2(X_s)}{b^2(X_s)} ds = \int_0^\infty X_s^2 ds = +\infty \;\text{ a.s.,}$$

therefore the maximum likelihood estimator $\widehat{\theta}_T = \left(\int_0^T X_s^2 ds\right)^{-1} \int_0^T X_s dX_s$ is a strongly consistent estimator of θ. Another more particular approach involving the calculation of the Laplace transform

$$\mathrm{E}\exp\left\{-\int_0^T X_s^2 ds\right\}$$

$$= \left(\frac{e^{-\theta T}\sqrt{2+\theta^2}}{\sqrt{\theta^2+2}\cosh(T\sqrt{\theta^2+2}) - \theta\sinh(T\sqrt{\theta^2+2})}\right)^{\frac{1}{2}},$$

which evidently tends to zero as $T \to \infty$ implying that $\int_0^T X_s^2 ds \to \infty$ a.s. as $T \to \infty$, was considered in [LIP 01].

11

Filtering Problem. Kalman-Bucy Filter

We considered the problem of the construction of an optimal filter in the linear two-dimensional partially observed Gaussian model and reduced it to the solution of two equations, one of them being a Riccati differential equation and the other one being a linear stochastic differential equation. For technical simplicity, we consider the proofs only for the case where the initial equations have constant coefficients; however, final formulas are presented for the general case when the coefficients depend on time.

11.1. General setting

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathrm{P})$ be a complete probability space with filtration. Also, let $W(t) = (W_1(t), W_2(t))$ be a two-dimensional Wiener process, i.e. Wiener processes W_1 and W_2 are independent.

Consider a two-dimensional Gaussian process $(X_1, X_2) = \{(X_1(t), X_2(t)), 0 \leq t \leq T\}$, which is a unique solution of the following system of stochastic differential equations

$$\begin{cases} dX_1(t) = a(t)X_1(t)dt + b(t)dW_1(t), \\ dX_2(t) = A(t)X_1(t)dt + B(t)dW_2(t), \end{cases}$$

where $a, A : \mathbb{R}^+ \to \mathbb{R}$, $b, B : \mathbb{R}^+ \to \mathbb{R}^+$, $X_1(0) = x_1 \in \mathbb{R}$, $X_2(0) = x_2 \in \mathbb{R}$. The coefficients will be assumed to satisfy

$$\int_0^T |a(t)|dt + \int_0^T |A(t)|dt + \int_0^T b^2(t)dt + \int_0^T B^2(t)dt < \infty. \qquad [11.1]$$

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

Note that the equation for X_1 is the Langevin equation, so that X_1 is the Ornstein-Uhlenbeck process. We assume that X_1 is a non-observable process which can be interpreted as the input signal, while X_2 is observable and can be interpreted as the output signal.

11.2. Auxiliary properties of the non-observable process

From now on, we consider a model with constant coefficients, that is,

$$\begin{cases} dX_1(t) = a\,X_1(t)dt + b\,dW_1(t), \\ dX_2(t) = A\,X_1(t)dt + B\,dW_2(t). \end{cases} \qquad [11.2]$$

Let us study in more detail the non-observable process $X_1(t)$, which can be presented in the integral form as follows:

$$X_1(t) = x_1 + a\int_0^t X_1(s)ds + bW_1(t). \qquad [11.3]$$

It is an Ornstein-Uhlenbeck process, the unique solution of Langevin equation [11.3]. It can be presented as follows:

$$X_1(t) = x_1 e^{at} + be^{at}\int_0^t e^{-au}dW_1(u).$$

Evidently, $\mathrm{E}X_1(t) = x_1 e^{at}$ and $\mathrm{E}X_1(t) = 0,\ t \geq 0$ if $x_1 = 0$. In the latter case $\mathrm{E}X_2(t) = x_2$. Further, the covariance $K(t,s)$ of X_1 is equal to the following expression:

$$K(t,s) = \frac{b^2}{2a}\left(e^{a(t+s)} - e^{a|t-s|}\right),$$

see sections 3.4.5 and 9.1.2.

If $x_1 = 0$, then

$$K(t,s) = \frac{b^2}{2a}\left(e^{a(t+s)} - e^{a|t-s|}\right). \qquad [11.4]$$

Function $K(t,s)$ is continuous and consequently bounded on any rectangle $[0,T]^2$. Moreover, it is differentiable on the interval $[s,T]$ (at the point s, the right derivative is considered) for any $s \in [0,T]$ and

$$\frac{\partial K(t,s)}{\partial t} = aK(t,s),\ t \in [s,T]. \qquad [11.5]$$

11.3. What is an optimal filter

The Kalman-Bucy problem of optimal linear filtering is the following: recall that we assume the process X_1 to be not observable, while the process X_2 is observable. The problem is to reconstruct (to estimate and to filter) X_1 in an optimal way by the observations of X_2. More exactly, if we wish to be non-anticipative, we want to reconstruct $X_1(t)$ by the observations of $\{X_2(s), 0 \le s \le t\}$. Obviously, $X_1(t)$ has the moments of any order. Furthermore, if we wish to be within the framework of $\mathcal{L}_2$-theory then the goal is to find such $\widehat{X}_1(t)$ that

$$\mathrm{E}\left(X(t) - \widehat{X}_1(t)\right)^2 = \inf_{\xi \in \mathcal{L}_2(\Omega, \mathcal{F}_t^{X_2}, \mathrm{P})} \mathrm{E}(X(t) - \xi)^2, \qquad [11.6]$$

where $\left\{\mathcal{F}_t^{X_2}, t \in [0, T]\right\}$ is the filtration generated by X_2.

LEMMA 11.1.– *Random variable $\widehat{X}_1(t) := \mathrm{E}(X_1(t)|\mathcal{F}_t^{X_2})$ is the optimal estimator of $X_1(t)$ in the sense of equation* [11.6]. *More exactly,*

$$\mathrm{E}\left(X(t) - \mathrm{E}(X_1(t)|\mathcal{F}_t^{X_2})\right)^2 = \min_{\xi \in \mathcal{L}_2(\Omega, \mathcal{F}_t^{X_2}, \mathrm{P})} \mathrm{E}(X(t) - \xi)^2.$$

PROOF.– Indeed, for any $\zeta \in \mathcal{L}_2(\Omega, \mathcal{F}_t^{X_2}, \mathrm{P})$

$$\begin{aligned}\mathrm{E}(X_1(t) - \zeta)^2 &= \mathrm{E}\left(X_1(t) - \widehat{X}_1(t) - \zeta + \widehat{X}_1(t)\right)^2 \\ &= \mathrm{E}\left(X_1(t) - \widehat{X}_1(t)\right)^2 + 2\mathrm{E}\left(X_1(t) - \widehat{X}_1(t)\right)\left(\zeta - \widehat{X}_1(t)\right) \\ &\quad + \mathrm{E}\left(\widehat{X}_1(t) - \zeta\right)^2.\end{aligned}$$

Further,

$$\begin{aligned}&\mathrm{E}\left(X_1(t) - \widehat{X}_1(t)\right)\left(\zeta - \widehat{X}_1(t)\right) \\ &= \mathrm{E}\left(\left(\zeta - \widehat{X}_1(t)\right)\mathrm{E}\left(X_1(t) - \widehat{X}_1(t)|\mathcal{F}_t^{X_2}\right)\right) \\ &= \mathrm{E}\left(\left(\zeta - \widehat{X}_1(t)\right)\left(\mathrm{E}\left(X_1(t)|\mathcal{F}_t^{X_2}\right) - \widehat{X}_1(t)\right)\right) = 0.\end{aligned}$$

Therefore,

$$\mathrm{E}(X_1(t) - \zeta)^2 = \mathrm{E}(X_1(t) - \widehat{X}_1(t))^2 + \mathrm{E}(\widehat{X}_1(t) - \zeta)^2 \ge \mathrm{E}(X_1(t) - \widehat{X}_1(t))^2$$

with the equality achieved at the unique point $\zeta = \widehat{X}_1(t)$, and hence the proof follows. □

And so, we know what an optimal filter is at the point t: it is equal to $\widehat{X}_1(t) = \mathrm{E}(X_1(t)|\mathcal{F}_t^{X_2})$. Now our goal is to present $\widehat{X}_1(t)$ in terms of the observable process X_2.

11.4. Representation of an optimal filter via an integral equation with respect to an observable process

Now we present $\widehat{X}_1(t)$ as the integral w.r.t. observable Gaussian process $\{X_2(s), 0 \le s \le t\}$ with some non-random kernel.

For technical simplicity, assume from now on that $x_1 = 0$ so that $\mathrm{E}X_1(t) = 0$ (see section 11.2).

LEMMA 11.2.– *There exists a kernel $G = G(t,s) : [s,T] \times [0,T] \to \mathbb{R}$ such that for any $t \in [0,T]$ $\int_0^t G^2(t,s)ds < \infty$, and the optimal filter $\widehat{X}_1(t)$ admits the representation*

$$\widehat{X}_1(t) = \int_0^t G(t,s)dX_2(s), \;\; t \in [0,T].$$

REMARK 11.1.– Integral $\int_0^t G(t,s)dX_2(s)$ can be decomposed as follows:

$$\int_0^t G(t,s)dX_2(s) = A\int_0^t G(t,s)X_1(s)ds + B\int_0^t G(t,s)dW_2(s),$$

and both integrals exist under condition $\int_0^t G^2(t,s)ds < \infty$.

PROOF.– Let us fix any $t \in [0,T]$ and consider the sequence of dyadic partitions $\pi_n = \left\{t\delta_k^{(n)}, 0 \le k \le 2^n, \delta_k^{(n)} = \frac{k}{2^n}\right\}$, $n \ge 1$. Consider the Gaussian vector $\Big\{X_1(t), X_2(t\delta_1^{(n)}) - X_2, X_2(t\delta_2^{(n)}) - X_2(t\delta_1^{(n)}), \ldots, X_2(t\delta_k^{(n)}) - X_2(t\delta_{k-1}^{(n)}), \ldots,$ $X_2(t) - X_2(t\delta_{2^n}^{(n)})\Big\}$, and denote the sequence of σ-fields

$$\mathcal{F}_t^{(n)} = \sigma\left\{X_2(t\delta_1^{(n)}), X_2(t\delta_2^{(n)}) - X_2(t\delta_1^{(n)}), \ldots, X_2(t) - X_2(t\delta_{2^n}^{(n)})\right\}.$$

Note that under condition $x_1 = 0$ we have that $\mathrm{E}X_2(t) = x_2$; therefore, $\mathrm{E}(X_2(t\delta_k^{(n)}) - X_2(t\delta_{k-1}^{(n)})) = 0, 1 \le k \le 2^n$.

According to the theorem on normal correlation (theorem 3.1), we have the representation

$$\mathrm{E}(X_1(t)|\mathcal{F}_t^{(n)}) = \sum_{k=1}^{n} \alpha_k^{(n)}(t)\left(X_2(t\delta_k^{(n)}) - X_2(t\delta_{k-1}^{(n)})\right). \qquad [11.7]$$

In this relation, coefficients $\alpha_k(t)$ are non-random and depend both on point t and the number k. Also, they depend on partition π_n. Further, note that σ-fields $\mathcal{F}_t^{(n)}$ are non-decreasing: $\mathcal{F}_t^{(1)} \subset \mathcal{F}_t^{(2)} \subset \ldots \subset \mathcal{F}_t^{(n)} \subset \ldots$, and $\mathcal{F}_t^{X_2} = \sigma\left\{\bigcup_{n=1}^{\infty} \mathcal{F}_t^{(n)}\right\}$. Therefore, according to the Lévy convergence theorem (theorem 5.8)

$$\mathrm{E}(X_1(t)|\mathcal{F}_t^{(n)}) \to \mathrm{E}(X_1(t)|\mathcal{F}_t^{X_2}) \text{ a.s. as } n \to \infty. \tag{11.8}$$

Moreover, since $X_1(t)$ is a Gaussian random variable, it has moments of any order, consequently, the sequence of random variables $\left\{\mathrm{E}(X_1(t)|\mathcal{F}_t^{(n)}), n \geq 1\right\}$ is uniformly integrable. Now, denote $G_n(t,s) = \alpha_k^{(n)}(t)$ for $s \in [t\delta_{k-1}^{(n)}, t\delta_k^{(n)})$. We can rewrite the representation [11.7] as follows:

$$\mathrm{E}(X_1(t)|\mathcal{F}_t^{(n)}) = \int_0^t G_n(t,s)dX_2(s). \tag{11.9}$$

It follows from [11.8] and uniform integrability that

$$\mathrm{E}(X_1(t)|\mathcal{F}_t^{(n)}) \to \mathrm{E}(X_1(t)|\mathcal{F}_t^{X_2}) \text{ in } \mathcal{L}_2(\Omega,\mathcal{F},\mathrm{P}) \text{ as } n \to \infty. \tag{11.10}$$

In turn, this means that the integrals $\int_0^t G_n(t,s)dX_2(s)$ in the right-hand side of [11.9] create a Cauchy sequence in $\mathcal{L}_2(\Omega,\mathcal{F},\mathrm{P})$, so that

$$\mathrm{E}\left(\int_0^t (G_n(t,s) - G_m(t,s))dX_2(s)\right)^2 \to 0 \text{ as } m,n \to \infty.$$

Let us transform the latter expectation in the following manner:

$$\begin{aligned}
&\mathrm{E}\left(\int_0^t (G_n(t,s) - G_m(t,s))dX_2(s)\right)^2 \\
&= A^2\mathrm{E}\left(\int_0^t (G_n(t,s) - G_m(t,s))X_1(s)ds\right)^2 \\
&\quad + 2AB\,\mathrm{E}\left(\int_0^t (G_n(t,s) - G_m(t,s))X_1(s)ds \int_0^t (G_n(t,s)\right. \\
&\quad \left. - G_m(t,s))dW_2(s)\right) \\
&\quad + B^2\mathrm{E}\left(\int_0^t (G_n(t,s) - G_m(t,s))dW_2(s)\right)^2.
\end{aligned} \tag{11.11}$$

Since the process X_1 does not depend on W_2, we get that

$$\begin{aligned}
&\mathrm{E}\int_0^t (G_n(t,s) - G_m(t,s))X_1(s)ds \int_0^t (G_n(t,s) - G_m(t,s))dW_2(s) \\
&= \mathrm{E}\int_0^t (G_n(t,s) - G_m(t,s))X_1(s)ds \cdot \mathrm{E}\int_0^t (G_n(t,s) \\
&\quad - G_m(t,s))dW_2(s) \\
&= \mathrm{E}\int_0^t (G_n(t,s) - G_m(t,s))X_1(s)ds \cdot 0 = 0. \qquad [11.12]
\end{aligned}$$

Moreover, it follows from the isometry property that

$$\begin{aligned}
&\mathrm{E}\left(\int_0^t (G_n(t,s) - G_m(t,s))dW_2(s)\right)^2 \\
&= \int_0^t (G_n(t,s) - G_m(t,s))^2 ds. \qquad [11.13]
\end{aligned}$$

Substituting [11.12] and [11.13] into [11.11], we get that

$$\begin{aligned}
&\mathrm{E}\left(\int_0^t (G_n(t,s) - G_m(t,s))dX_2(s)\right)^2 \\
&= A^2\mathrm{E}\left(\int_0^t (G_n(t,s) - G_m(t,s))X_1(s)ds\right)^2 \\
&\quad + B^2\int_0^t (G_n(t,s) - G_m(t,s))^2 ds \to 0 \text{ as } m,n\to\infty.
\end{aligned}$$

For the last term, this means that it tends to zero as $m, n \to \infty$; therefore, $G_n(t,\cdot)$ is a Cauchy sequence in $\mathcal{L}_2([0,t],\lambda_1)$ where there exists a function $G(t,s)$ such that $G_n(t,\cdot) \to G(t,\cdot)$ in $\mathcal{L}_2([0,t],\lambda_1)$ as $n\to\infty$. Since the value $\mathrm{E}X_1^2(s)$ is bounded on $[0,T]$,

$$\begin{aligned}
&\mathrm{E}\left(\int_0^t (G_n(t,s) - G(t,s))X_1(s)ds\right)^2 \\
&\le \int_0^t (G_n(t,s) - G(t,s))^2 ds \cdot \mathrm{E}\int_0^t X_1^2(s)ds \to 0 \text{ as } n\to\infty.
\end{aligned}$$

Finally, we get from the previous relations that

$$\int_0^t G_n(t,s)dX_2(s) = A\int_0^t G_n(t,s)X_1(s)ds + B\int_0^t G_n(t,s)dW_2(s) \to$$
$$\to \int_0^t G(t,s)dX_2(s) \text{ in } \mathcal{L}_2(\Omega,\mathcal{F},\mathrm{P}) \text{ as } n\to\infty. \quad [11.14]$$

It follows from [11.9], [11.10] and [11.14] that

$$\mathrm{E}(X_1(t)|\mathcal{F}_t^{X_2}) = \int_0^t G(t,s)dX_2(s),$$

and the theorem is proved. □

REMARK 11.2.– It is reasonable to consider $\mathrm{E}(X_t - \widehat{X}_t)^2$ as the error in the solution of the problem of optimal filtering. In what follows, we shall use the notations $m_t = \widehat{X}_t$ and $\sigma_t^2 = \mathrm{E}(X_t - m_t)^2$.

11.5. Integral Wiener-Hopf equation

Now we establish integral equation for the kernel $G(t,s)$.

LEMMA 11.3.– *The kernel* $G(t,s) : [s,T]\times[0,T] \to \mathbb{R}$ *from lemma 11.2 satisfies for any* $t\in[0,T]$ *the following integral Wiener-Hopf equation:*

$$AK(t,u) = A^2\int_0^t G(t,s)K(s,u)ds + B^2G(t,u). \quad [11.15]$$

PROOF.– Note that for any $t\in[0,T]$

$$\left|\int_0^t G(t,s)K(s,u)ds\right| \le \max_{0\le u,\,s\le t}|K(s,u)|\int_0^t |G(t,s)|ds < \infty,$$

since K is a continuous bounded kernel on $[0,T]^2$, and $G(t,\cdot)\in\mathcal{L}_2([0,t],\lambda_1)$. For any $t\in[0,T]$ consider a measurable bounded function $f(t,s)$ and define the integral

$$\int_0^t f(t,s)dX_2(s) = A\int_0^t f(t,s)X_1(s)ds + B\int_0^t f(t,s)dW_2(s).$$

Evidently two latter integrals are square-integrable, and we can consider the following expectation:

$$\mathrm{E}(X_1(t) - \widehat{X}_1(t))\int_0^t f(t,s)dX_2(s).$$

Note that this expectation equals zero:

$$
\begin{aligned}
&\mathrm{E}(X_1(t) - \widehat{X}_1(t)) \int_0^t f(t,s)dX_2(s) \\
&= \mathrm{E}\left(\mathrm{E}(X_1(t) - \widehat{X}_1(t))|\mathcal{F}_t^{X_2}\right) \int_0^t f(t,s)dX_2(s) \qquad [11.16] \\
&= \mathrm{E}(\widehat{X}_1(t) - \widehat{X}_1(t)) \int_0^t f(t,s)dX_2(s) = 0.
\end{aligned}
$$

At the same time, the left-hand side of [11.16] can be rewritten as follows:

$$
\begin{aligned}
0 &= \mathrm{E}X_1(t) \int_0^t f(t,s)dX_2(s) - \mathrm{E}\int_0^t G(t,s)dX_2(s) \int_0^t f(t,s)dX_2(s) \\
&= A\mathrm{E}X_1(t) \int_0^t f(t,s)X_1(s)ds + B\mathrm{E}X_1(t) \int_0^t f(t,s)dW_2(s) \\
&\quad - \mathrm{E}\left(A\int_0^t G(t,s)X_1(s)ds + B\int_0^t G(t,s)dW_2(s)\right) \times \\
&\quad \times \left(A\int_0^t f(t,s)X_1(s)ds + B\int_0^t f(t,s)dW_2(s)\right). \qquad [11.17]
\end{aligned}
$$

Taking into account independence of the processes X_1 and W_2, we get that

$$
\mathrm{E}X_1(t) \int_0^t f(t,s)dW_2(s) = \mathrm{E}X_2(t)\,\mathrm{E}\int_0^t f(t,s)dW_2(s) = 0,
$$

and similarly the following expectations equal zero:

$$
\mathrm{E}\int_0^t G(t,s)X_1(s)ds \int_0^t f(t,s)dW_2(s) = 0
$$

and

$$
\mathrm{E}\int_0^t G(t,s)dW_2(s) \int_0^t f(t,s)X_1(s)ds = 0.
$$

So, we get from [11.17] that

$$\begin{aligned}0 &= A\mathrm{E}\int_0^t f(t,s)X_1(t)X_1(s)ds \\ &\quad - A^2\mathrm{E}\int_0^t\int_0^t G(t,u)f(t,s)X_1(s)X_1(u)dsdu \\ &\quad - B^2\int_0^t G(t,s)f(t,s)ds = A\int_0^t f(t,s)K(t,s)ds \qquad [11.18] \\ &\quad - A^2\int_0^t f(t,s)\int_0^t G(t,u)K(s,u)duds - B^2\int_0^t f(t,s)G(t,s)ds \\ &= \int_0^t f(t,s)\left[AK(t,s) - A^2\int_0^t G(t,u)K(s,u)du - B^2G(t,s)\right]ds.\end{aligned}$$

Since the bounded measurable function f is arbitrary, we can get from [11.18] by standard approximation methods that

$$AK(t,s) - A^2\int_0^t G(t,u)K(s,u)du - B^2G(t,s) = 0,$$

and the proof follows. □

LEMMA 11.4.– *For any* $t \in [0,T]$*, equation* [11.15] *has a unique solution* $G = G(t,s)$ *in the space* $\mathcal{L}_2([0,t],\lambda_1)$*, and this solution satisfies the relations*

$$G(s,s) = \frac{A}{B^2}\sigma_s^2, \quad G(t,s) = g(t,s)G(s,s) = \frac{A}{B^2}\sigma_s^2 G(s,s), \quad \textit{and}$$

$$\frac{\partial g(t,s)}{\partial t} = \left(a - \frac{A^2}{B^2}\sigma_t^2\right)g(t,s), \quad g(s,s) = 1. \qquad [11.19]$$

PROOF.– i) Let $G_i(t,s)$, $i = 1,2$ be two solutions of [11.15] in the space $\mathcal{L}_2([0,t],\lambda_1)$. For their difference $\Delta_G(t,s) = G_1(t,s) - G_2(t,s)$ we have the equation

$$A^2\int_0^t \Delta_G(t,s)K(s,u)ds + B^2\Delta_G(t,u) = 0. \qquad [11.20]$$

Multiply [11.20] by $\Delta_G(t,u)$ and integrate from 0 to t:

$$A^2\int_0^t\int_0^t \Delta_G(t,u)\Delta_G(t,s)K(s,u)dsdu + B^2\int_0^t \Delta_G^2(t,u)du = 0.$$

Note that the kernel $K(s,u)$ is non-negatively definite, as any covariance function:

$$\int_0^t\int_0^t \Delta_G(t,u)\Delta_G(t,s)K(s,u)dsdu = \mathrm{E}\left(\int_0^t \Delta_G(t,u)X_1(s)ds\right)^2 \geq 0.$$

Therefore,

$$\int_0^t \Delta_G^2(t,u)du = 0,$$

where

$$\Delta_G(t,u) = 0,\;\; 0 \leq u \leq t,\;\; \text{a.e. w.r.t. } \lambda_1.$$

And so, we proved the uniqueness of the solution of [11.15] in the space $\mathcal{L}_2([0,t],\lambda_1)$ for all $t \in [0,T]$.

ii) Note that the function σ is bounded on $[0,T]$. Indeed, it follows from [11.4] that

$$\begin{aligned}\sigma_t^2 &= \mathrm{E}\left(X_1(t) - \mathrm{E}((X_1(t)|\mathcal{F}_t^{X_2})\right)^2 \leq 2\mathrm{E}X_1^2(t) + 2\mathrm{E}\left(\mathrm{E}(X_1(t)|\mathcal{F}_t^{X_2})\right)^2\\ &\leq 4\mathrm{E}X_1^2(t) = K(t,t) = \frac{b^2}{2a}\left(e^{2at}-1\right) \leq \frac{b^2}{2|a|}\left(e^{2|a|T}-1\right).\end{aligned}$$

Therefore, equation [11.19] has the unique solution of the form

$$g(t,s) = \exp\left\{\int_s^t \left(a - \frac{A^2}{B^2}\sigma_u^2\right)du\right\}.$$

iii) It follows from Wiener-Hopf equation [11.15] that

$$AK(t,t) = A^2\int_0^t G(t,s)K(s,t)ds + B^2G(t,t).$$

Therefore,

$$\begin{aligned}B^2G(t,t) &= AK(t,t) - A^2\int_0^t G(t,s)K(s,t)ds\\ &= A\mathrm{E}X_1^2(t) - A^2\mathrm{E}\int_0^t G(t,s)X_1(t)X_1(s)ds\\ &= A\mathrm{E}X_1(t)\left(X_1(t) - A\int_0^t G(t,s)X_1(s)ds\right)\end{aligned}$$

$$= A\mathrm{E}X_1(t)\left(X_1(t) - A\int_0^t G(t,s)X_1(s)ds - B\int_0^t G(t,s)dW_2(s)\right)$$
$$= A\mathrm{E}X_1(t)\left(X_1(t) - \int_0^t G(t,s)dX_2(s)\right) = A\mathrm{E}X_1(t)\left(X_1(t) - \widehat{X}_1(t)\right).$$

Further,

$$\mathrm{E}\widehat{X}_1(t)\left(X_1(t) - \widehat{X}_1(t)\right) = \mathrm{E}\left(\widehat{X}_1(t)\mathrm{E}\left(X_1(t) - \widehat{X}_1(t)|\mathcal{F}_t^{X_2}\right)\right)$$
$$= \mathrm{E}\left(\widehat{X}_1(t)\left(\widehat{X}_1(t) - \widehat{X}_1(t)\right)\right) = 0.$$

Therefore,

$$B^2G(t,t) = A\mathrm{E}\left(X_1(t) - \widehat{X}_1(t)\right)^2 = A\sigma_t^2. \qquad [11.21]$$

iv) Assume that $G(t,s)$ is a.e. differentiable in t. If we find such a solution of [11.15] and it is in the space $\mathcal{L}_2([0,t],\lambda_1)$ for all $t \in [0,T]$, then it will be the unique solution, due to (i). Under this assumption, we can differentiate left- and right-hand sides of [11.15] in t:

$$A\frac{\partial K(t,u)}{\partial t} = A^2G(t,t)K(t,u) + A^2\int_0^t \frac{\partial G(t,s)}{\partial t}K(s,u)ds + B^2\frac{\partial G(t,u)}{\partial t}.$$

Substituting $\frac{\partial K(t,u)}{\partial t}$ from [11.5] and $G(t,t)$ from [11.21], we get

$$aAK(t,u) = \frac{A^3}{B^2}\sigma_t^2K(t,u) + A^2\int_0^t \frac{\partial G(t,s)}{\partial t}K(s,u)ds + B^2\frac{\partial G(t,u)}{\partial t}. \qquad [11.22]$$

Now we substitute the value of $K(t,u)$ from [11.15] into [11.22] and obtain that

$$a\left(A^2\int_0^t G(t,s)K(s,u)ds + B^2G(t,u)\right)$$
$$= \frac{A^2}{B^2}\sigma_t^2\left(A^2\int_0^t G(t,s)K(s,u)ds + B^2G(t,u)\right)$$
$$+ A^2\int_0^t \frac{\partial G(t,s)}{\partial t}K(s,u)ds + B^2\frac{\partial G(t,u)}{\partial t},$$

or

$$A^2 \int_0^t \left(aG(t,s) - \frac{A^2}{B^2}\sigma_t^2 G(t,s) - \frac{\partial G(t,s)}{\partial t} \right) K(s,u)ds$$
$$+B^2 \left(aG(t,u) - \frac{A^2}{B^2}\sigma_t^2 G(t,u) - \frac{\partial G(t,u)}{\partial t} \right) = 0. \qquad [11.23]$$

Equality [11.23] means that any function G satisfying equation

$$\frac{\partial G(t,s)}{\partial t} = \left(a - \frac{A^2}{B^2}\sigma_t^2 \right) G(t,u) \qquad [11.24]$$

transforms [11.23] into identity; therefore said function G satisfies equation [11.15]. However, as was stated in (i), [11.15] can have only one solution. Therefore, this unique solution satisfies [11.24]. Denote $g(t,s) = \frac{G(t,s)}{G(s,s)}$. Then, $g(s,s) = 1$ and

$$\frac{\partial g(t,s)}{\partial t} = \frac{\partial G(t,s)}{\partial t} \cdot \frac{1}{G(s,s)} = \left(a - \frac{A^2}{B^2}\sigma_t^2 \right) g(t,s).$$

Lemma is proved. □

REMARK 11.3.– As it was mentioned before, function $g(t,s)$ has a form

$$g(t,s) = \exp\left\{ \int_s^t \left(a - \frac{A^2}{B^2}\sigma_u^2 \right) du \right\}.$$

This means, in particular, that

$$g(t,s) = \frac{g(t,0)}{g(s,0)}.$$

THEOREM 11.1.– *The process* $m_t = \widehat{X}_t = \mathrm{E}\left(X_t | \mathcal{F}_t^{X_2} \right)$ *is the unique solution of the stochastic differential equation*

$$dm_t = \left(a - \frac{A^2}{B^2}\sigma_t^2 \right) m_t dt + \frac{A}{B^2}\sigma_t^2 dX_2(t), \qquad [11.25]$$

while σ_t^2 *satisfies Riccati equation*

$$(\sigma_t^2)' = b^2 + 2a\sigma_t^2 - \frac{A^2}{B^2}(\sigma_t^2)^2, \quad m_0 = \sigma_0^2 = 0. \qquad [11.26]$$

PROOF.– i) Taking into account remark 11.3, we get the relations

$$\begin{aligned} m_t &= \int_0^t G(t,s)dX_2(s) = \int_0^t g(t,s)G(s,s)dX_2(s) \\ &= \int_0^t \frac{g(t,0)}{g(s,0)}G(s,s)dX_2(s) = \frac{A}{B^2}g(t,0)\int_0^t \frac{\sigma_s^2}{g(s,0)}dX_2(s). \end{aligned}$$

Applying Itô formula to the product of $g(t,0)$ and $\int_0^t \frac{\sigma_s^2}{g(s,0)}dX_2(s)$, we get

$$\begin{aligned} dm_t &= \frac{A}{B^2}\frac{dg(t,0)}{dt}\int_0^t \frac{\sigma_s^2}{g(s,0)}dX_2(s)dt + \frac{A}{B^2}\frac{g(t,0)}{g(t,0)}\sigma_t^2 dX_2(t) \\ &= \frac{A}{B^2}\left(a - \frac{A^2}{B^2}\sigma_t^2\right)g(t,0)\int_0^t \frac{\sigma_s^2}{g(s,0)}dX_2(s)dt + \frac{A}{B^2}\sigma_t^2 dX_2(t) \\ &= \frac{A}{B^2}\left(a - \frac{A^2}{B^2}\sigma_t^2\right)\int_0^t g(t,s)\sigma_s^2 dX_2(s)dt + \frac{A}{B^2}\sigma_t^2 dX_2(t) \\ &= \left(a - \frac{A^2}{B^2}\sigma_t^2\right)\int_0^t G(t,s)dX_2(s)dt + \frac{A}{B^2}\sigma_t^2 dX_2(t) \\ &= \left(a - \frac{A^2}{B^2}\sigma_t^2\right)m_t dt + \frac{A}{B^2}\sigma_t^2 dX_2(t). \end{aligned}$$

And so, we get equation [11.25] for m_t.

ii) In order to get [11.26], we denote $R_t = X_1(t) - m_t$, so that $\mathrm{E}R_t^2 = \sigma_t^2$, and write dR_t as

$$\begin{aligned} dR_t &= dX_1(t) - dm_t \\ &= aX_1(t)dt + bdW_1(t) - \left(a - \frac{A^2}{B^2}\sigma_t^2\right)m_t dt \\ &\quad - \frac{A^2}{B^2}\sigma_t^2 X_1(t)dt - \frac{A}{B}\sigma_t^2 dW_2(t) \\ &= a(X_1(t) - m_t)dt - \frac{A^2}{B^2}\sigma_t^2(X_1(t) - m_t)dt \\ &\quad + bdW_1(t) - \frac{A}{B}\sigma_t^2 dW_2(t) \\ &= \left(a - \frac{A^2}{B^2}\sigma_t^2\right)R_t dt + bdW_1(t) - \frac{A}{B}\sigma_t^2 dW_2(t). \end{aligned}$$

Therefore, due to the independence of the processes W_1 and W_2, their mutual quadratic characteristics are equal to zero, and

$$\begin{aligned} dR_t^2 &= 2R_t dR_t + \left(b^2 + \frac{A^2}{B^2}\sigma_t^2\right) dt \\ &= 2\left(a - \frac{A^2}{B^2}\sigma_t^2\right) R_t^2 dt + R_t \left(b dW_1(t) - \frac{A}{B}\sigma_t^2 dW_2(t)\right) \\ &\quad + \left(b^2 + \frac{A^2}{B^2}(\sigma_t^2)^2\right) dt, \end{aligned}$$

where

$$\begin{aligned} R_t^2 &= 2a\int_0^t R_s^2 ds - 2\frac{A^2}{B^2}\int_0^t \sigma_s^2 R_s^2 ds + b\int_0^t R_s dW_1(s) \\ &\quad - \frac{A}{B}\int_0^t \sigma_s^2 R_s dW_2(s) + \int_0^t \left(b^2 + \frac{A^2}{B^2}(\sigma_s^2)^2\right) ds. \end{aligned} \quad [11.27]$$

Taking expectation of both sides of [11.27], we get

$$\mathrm{E}R_t^2 = 2a\int_0^t \mathrm{E}R_s^2 ds - 2\frac{A^2}{B^2}\int_0^t \sigma_s^2 \mathrm{E}R_s^2 ds + \int_0^t \left(b^2 + \frac{A^2}{B^2}(\sigma_s^2)^2\right) ds,$$

or

$$\sigma_t^2 = b^2 t + 2a\int_0^t \sigma_s^2 ds - \frac{A^2}{B^2}\int_0^t (\sigma_s^2)^2 ds,$$

and we get equation [11.26]. □

REMARK 11.4.– Consider the so-called innovation process $Z_t = B^{-1}X_2(t) - \frac{A}{B}\int_0^t m_u du$. It is $\mathcal{F}_t^{X_2}$-adapted and

$$\begin{aligned} \mathrm{E}\left(Z_t - Z_s | \mathcal{F}_s^{X_2}\right) &= \mathrm{E}\left(\frac{A}{B}\int_0^t X_1(u) du + W_2(t) - W_2(t)\right. \\ &\qquad \left. - \frac{A}{B}\int_s^t m_u du | \mathcal{F}_s^{X_2}\right) \\ &= \frac{A}{B}\mathrm{E}\left(\int_s^t \mathrm{E}\left(X_1(u)|\mathcal{F}_u^{X_2}\right) du\right. \\ &\qquad \left. - \int_s^t \mathrm{E}\left(X_1(u)|\mathcal{F}_u^{X_2}\right) du | \mathcal{F}_s^{X_2}\right) = 0. \end{aligned}$$

Thus, Z_t is $\mathcal{F}_t^{X_2}$-martingale. Furthermore, Z is a continuous process, it is square-integrable martingale, and $\langle Z \rangle_t = [Z]_t = [W_2]_t = t$. According to the one-parameter version of the Lévy representation theorem (theorem 8.11), Z_t is the $\mathcal{F}_t^{W_2}$-Wiener process. Now, let $f = f(t,s) : [s,T] \times [0,T] \to \mathbb{R}$ be a bounded measurable function. Then

$$\mathrm{E}(X_1(t) - m_t) \int_0^t f(t,s) dZ_s = 0.$$

It follows from the one-parameter version of the Itô representation theorem (theorem 8.15) that there exists a measurable function $m(t,s) : [s,T] \times [0,T] \to \mathbb{R}$ such that $\int_0^t m^2(t,s)ds < \infty$ and $m_t = \int_0^t m(t,s)dZ_s$. Therefore,

$$\mathrm{E}X_1(t) \int_0^t f(t,s)dZ_s = \int_0^t m(t,s)f(t,s)ds. \qquad [11.28]$$

Since

$$Z_t = \frac{A}{B}\int_0^t X_1(u)du - \frac{A}{B}\int_0^t m_u du + W_2(t),$$

the left-hand side of [11.28] can be rewritten as follows:

$$\begin{aligned}\mathrm{E}X_1(t)\left(\int_0^t f(t,s)dZ_s\right) &= \frac{A}{B}\mathrm{E}X_1(t)\int_0^t f(t,u)X_1(u)du \\ &- \frac{A}{B}\mathrm{E}X_1(t)\int_0^t f(t,u)m_u du + \mathrm{E}X_1(t)\int_0^t f(t,u)dW_2(u) \\ &= \frac{A}{B}\int_0^t f(t,u)\mathrm{E}X_1(t)(X_1(u) - m_u)du,\end{aligned}$$

because it follows from the independence of the processes X_1 and W_2 that $\mathrm{E}X_1(t)\int_0^t f(t,u)dW_2(u) = 0$. Further, $X_1(t) = e^{at}\int_0^t e^{-au}dW_1(u)$, therefore,

$$\mathrm{E}(X_1(t)|\mathcal{F}_s) = e^{at}\int_0^s e^{-au}dW_1(u) = e^{a(t-s)}X_1(s)$$

for any $s \le t$, and for any $u \le t$

$$\begin{aligned}\mathrm{E}X_1(t)(X_1(u) - m_u) &= \mathrm{E}\left(\mathrm{E}(X_1(t)|\mathcal{F}_u)\right)(X_1(u) - m_u) \\ &= e^{a(t-u)}\mathrm{E}X_1(u)(X_1(u) - m_u) \\ &= e^{a(t-u)}\mathrm{E}(X_1(u) - m_u)^2 = e^{a(t-u)}\sigma_u^2.\end{aligned}$$

This means that [11.28] can be rewritten as follows:

$$\frac{A}{B}\int_0^t f(t,u)e^{a(t-u)}\sigma_u^2 du = \int_0^t m(t,u)f(t,u)du.$$

Since f is arbitrary, we get that

$$m(t,u) = \frac{A}{B}e^{a(t-u)}\sigma_u^2,$$

and

$$\begin{aligned} m_t &= \int_0^t m(t,u)dZ_u = \frac{A}{B}e^{at}\int_0^t e^{-au}\sigma_u^2 dZ_u \\ &= \frac{A}{B}e^{at}\int_0^t e^{-au}\sigma_u^2\left(\frac{1}{B}dX_2(u) - \frac{A}{B}m_u du\right). \end{aligned} \quad [11.29]$$

Now we can apply to m_t the Itô formula for the product of two processes and get that

$$dm_t = am_t dt + \frac{A}{B^2}\sigma_t^2 dX_2(t) - \frac{A^2}{B^2}\sigma_t^2 m_t dt,$$

and this formula coincides with [11.25].

REMARK 11.5.– Denote $k = A^2/B^2$,

$$y_1 = \frac{1}{k}\left(a + \sqrt{a^2+kb^2}\right),\ y_2 = \frac{1}{k}\left(a - \sqrt{a^2+kb^2}\right)$$

and

$$c = (y_1 - y_2)^{-1} = \frac{k}{2\sqrt{a^2+kb^2}}.$$

Then the solution of Riccati equation with zero initial condition has the following form:

$$\left|\frac{y-y_1}{y-y_2}\right| = -\frac{y_1}{y_2}e^{x/c}.$$

Furthermore, equation [11.29] has the following unique solution

$$m_t = \frac{A}{B^2}\int_0^t \sigma_s^2 \exp\left\{\int_0^s \left(k\sigma_u^2 - a\right)du\right\} dX_2(s)\cdot\exp\left\{\int_0^t \left(a - k\sigma_s^2\right)ds\right\}.$$

For a general scheme,

$$dX_1(t) = a(t)X_1(t)dt + b(t)dW_1(t),$$

$$dX_2(t) = A(t)X_1(t)dt + B(t)dW_2(t),$$

$$X_1(0) \in \mathbb{R}, \ \ X_2(0) \in \mathbb{R}$$

with non-random measurable functions $a(t)$, $b(t)$, $A(t)$ and $B(t)$ satisfying [11.1]; all previous calculations can be repeated with evident modifications to get the following result.

THEOREM 11.2.– *Let* $\int_0^T A^2(t)dt < \infty$ *and* $|B(t)| \geq B > 0$. *Then* $m_t = \mathrm{E}\left(X_1(t)|\mathcal{F}_t^{X_2}\right)$ *and* $\sigma_t^2 = \mathrm{E}\left(X_1(t) - m_t\right)^2$ *satisfy the following system of equations:*

$$dm_t = \left(a(t) - \frac{A^2(t)}{B^2(t)}\sigma_t^2\right) m_t dt + \frac{A(t)}{B^2(t)}\sigma_t^2 dX_2(t),$$

$$(\sigma_t^2)' = b(t) + 2a(t)\sigma_t^2 - \frac{A^2(t)}{B^2(t)}(\sigma_t^2)^2,$$

$$m_0 = X_1(0), \ \ \sigma_0^2 = 0.$$

Appendices

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

Appendix 1

Selected Facts from Calculus, Measure Theory and the Theory of Operators

A1.1. Some facts from the theory of metric spaces

Let (S, ρ) be a metric space.

DEFINITION A1.1.– *The set $A \subset S$ is called compact if any sequence $\{x_n, n \geq 1\}$ from A has a convergence subsequence.*

According to the Hausdorff criteria, in a complete metric space, the set A is compact if and only if any ε-covering of A by open balls of radius $\varepsilon > 0$ has a finite sub-covering. We shall denote by $B(a, r) = \{x \in S : \rho(x, a) < \varepsilon\}$ and $\overline{B}(a, r) = \{x \in S : \rho(x, a) \leq \varepsilon\}$ open and closed balls, respectively. Recall that the set $A \subset S$ is closed if it contains all limit points and is open if any point $a \in A$ admits an open ball $B(a, \varepsilon) \subset A$ for some $\varepsilon > 0$.

THEOREM A1.1.– *Let (S, ρ) be a complete separable metric space. Then, the space $(S^{(k)}, \rho_k)$ is also a complete separable metric space, where*

$$S^{(k)} = \underbrace{S \times \ldots \times S}_{k},$$

and ρ_k is defined by [1.2].

PROOF.– Let $M \subset S$ be a countable separant, i.e. for any $\varepsilon > 0$ and any $x \in S$ there exists $x(\varepsilon) \in M$ such that $\rho(x, x(\varepsilon)) < \varepsilon$. Consider the set $M^{(k)} = \underbrace{M \times \ldots \times M}_{k}$.

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

Let $\varepsilon > 0$ be fixed. For any $x^{(k)} = (x_1, \dots, x_k) \in S^{(k)}$, where $x_i \in S$ for $1 \le i \le k$, we can find $x_i\left(\frac{\varepsilon}{k}\right) \in M$. Then,

$$\rho_k(x^{(k)}, x^k\left(\frac{\varepsilon}{k}\right)) < \varepsilon,$$

where $x^k\left(\frac{\varepsilon}{k}\right) = \left(x_1\left(\frac{\varepsilon}{k}\right), \dots, x_k\left(\frac{\varepsilon}{k}\right)\right)$. Evidently, $M^{(k)}$ is countable, and we can understand that $M^{(k)}$ is separable. Now, let $x_n^{(k)}$ be a Cauchy sequence in $S^{(k)}$, i.e. $\rho_k(x_n^{(k)}, x_m^{(k)}) \to 0$ as $n, m \to \infty$. Then obviously any coordinate $x_{n,i}^{(k)}$ is a Cauchy sequence in metric ρ for any $1 \le i \le k$. Therefore, $x_{n,i}^{(k)} \to x_i^{(k)}$, where $x_i^{(k)} \in S$ in metric ρ, when $x_n^{(k)} \to x^{(k)} = (x_1^{(k)}, \dots, x_k^{(k)})$ in metric ρ_k. □

THEOREM A1.2.– *Let (S, ρ, Σ) be a complete separable metric space with Borel σ-field Σ, and let P be a probability measure on Σ. Then, for any $\varepsilon > 0$ and any $A \in \Sigma$, there exists a compact set $K \subset A$ such that $\mathrm{P}\{A \setminus K\} < \varepsilon$.*

PROOF.– Any probability measure on Σ is regular, that is, for any $A \in \Sigma$,

$$\mathrm{P}\{A\} = \sup\left\{\mathrm{P}\{C\} : C \subset A, C \text{ closed}\right\} = \inf\left\{\mathrm{P}\{O\} : O \supset A, O \text{ open}\right\}.$$

Therefore assume that A is closed. In this case, we can consider A as a complete separable metric subspace of S. Let $M_A = \{a_i, i \ge 1\}$ be a countable dense subset of A. Then for any $\delta > 0$, $\bigcup_{k=1}^{\infty}(B(a_k, \delta) \cap A) = A$. Therefore, $\mu(A) = \lim_{n\to\infty} \mathrm{P}\left\{\bigcup_{k=1}^{\infty}(B(a_k, \delta) \cap A)\right\}$ for any $\delta > 0$. Let $\varepsilon > 0$. Then for any $m \ge 1$ there exists $n_m \ge 1$ such that

$$\mathrm{P}\left\{\bigcup_{k=1}^{n_m}\left(B\left(a_k, \frac{1}{m}\right) \cap A\right)\right\} > \mu(A) - \varepsilon \cdot 2^{-m}.$$

Let $K = \bigcap_{m=1}^{\infty} \bigcup_{k=1}^{n_m}\left(\overline{B}\left(a_k, \frac{1}{m}\right) \cap A\right)$. Then K is closed and for any $\delta > 0$

$$K \subset \bigcup_{k=1}^{n_m}\left(\overline{B}\left(a_k, \frac{1}{m}\right) \cap A\right) \subset \bigcup_{k=1}^{n_m} B(a_k, \delta)$$

for $m > \frac{1}{\delta}$. Therefore K is compact. Moreover,

$$\mathrm{P}\{A \setminus K\} = \mathrm{P}\left\{\bigcup_{m=1}^{\infty} A \setminus \bigcup_{k=1}^{n_m}\left(\overline{B}\left(a_k, \frac{1}{m}\right) \cap A\right)\right\}$$
$$\le \sum_{m=1}^{\infty} \mathrm{P}\left\{A \setminus \bigcup_{k=1}^{n_m}\left(\overline{B}\left(a_k, \frac{1}{m}\right) \cap A\right)\right\} < \sum_{m=1}^{\infty} \varepsilon \cdot 2^{-m} = \varepsilon. \quad \square$$

A1.2. Marcinkiewicz interpolation theorem

Let $r > 1$. Consider the map T from $\mathcal{L}_r(\Omega, \mathcal{F}, \mathrm{P})$ into $\mathcal{L}_r(\Omega, \mathcal{F}, \mathrm{P})$.

DEFINITION A1.2.– *The map T is called subadditive if for any $\xi, \eta \in \mathcal{L}_r(\Omega, \mathcal{F}, \mathrm{P})$*

$$|T(\xi + \eta)(\omega)| \le |T(\xi)(\omega)| + |T(\eta)(\omega)|.$$

THEOREM A1.3.– *Let $T : \mathcal{L}_r(\Omega, \mathcal{F}, \mathrm{P}) \to \mathcal{L}_r(\Omega, \mathcal{F}, \mathrm{P})$ be subadditive, and for any $\kappa > 0$ and $\xi \in \mathcal{L}_r(\Omega, \mathcal{F}, \mathrm{P})$*

$$\mathrm{P}\{T(\xi)(\omega) > x\} \le \left(\frac{C_1 \|\xi\|_{\mathcal{L}_1(\Omega,\mathcal{F},\mathrm{P})}}{x} \right) \wedge \left(\frac{C_2 \|\xi\|^r_{\mathcal{L}_r(\Omega,\mathcal{F},\mathrm{P})}}{x^r} \right). \qquad [A1.1]$$

Then for any $1 < p < r$

$$\|T(\xi)\|_{\mathcal{L}_p(\Omega,\mathcal{F},\mathrm{P})} \le C_p \|\xi\|_{\mathcal{L}_p(\Omega,\mathcal{F},\mathrm{P})},$$

where C_1, C_2, C_p are some constants, C_p depends on C_1, C_2, p and r.

PROOF.– Let $\xi \in \mathcal{L}_r(\Omega, \mathcal{F}, \mathrm{P})$ and $x > 0$ be fixed. Denote ξ_1 and ξ_2 as the random variables $\xi_1 = \xi \mathbb{1}_{|\xi|>x}$ and $\xi_2 = \xi \mathbb{1}_{|\xi|\le x}$. It follows from subadditivity and [A1.1] that

$$G(x) := \mathrm{P}\{\omega : |T(\xi)(\omega)| > x\} \le \mathrm{P}\left\{\omega : |T(\xi_1)(\omega)| > \frac{x}{2}\right\}$$
$$+ \mathrm{P}\left\{\omega : |T(\xi_2)(\omega)| > \frac{x}{2}\right\} \le \frac{2C_1}{x}\mathrm{E}(|\xi|\mathbb{1}_{|\xi|>x}) + \frac{2^r C_2}{x^r}\mathrm{E}(|\xi|^r \mathbb{1}_{|\xi|\le x}).$$

Furthermore, for any $1 < p < r$

$$\|T(\xi)\|^p_{\mathcal{L}_p(\Omega,\mathcal{F},\mathrm{P})} = -\int_0^\infty x^{-p} dG(x) = p\int_0^\infty x^{p-1}G(x)dx$$
$$\le 2C_1 p \int_0^\infty x^{p-2}\mathrm{E}(|\xi|\mathbb{1}_{|\xi|>x})dx + 2^r C_2 p \int_0^\infty x^{p-r-1}\mathrm{E}(|\xi|^r \mathbb{1}_{|\xi|\le x})dx$$
$$= 2C_1 p\mathrm{E}\left(|\xi| \int_0^\infty x^{p-2}\mathbb{1}_{x<|\xi|}dx\right) + 2^r C_2 p\mathrm{E}\left(|\xi|^r \int_0^\infty x^{p-r-1}\mathbb{1}_{x\ge|\xi|}dx\right)$$
$$= 2C_1 p\mathrm{E}\left(|\xi| \int_0^{|\xi|} x^{p-2}dx\right) + 2^r C_2 p\mathrm{E}\left(|\xi|^r \int_{|\xi|}^\infty x^{p-r-1}dx\right)$$
$$= \frac{2C_1 p}{p-1}\mathrm{E}|\xi|^p + \frac{2^r C_2 p}{r-p}\mathrm{E}|\xi|^p = (C_p)^p \mathrm{E}|\xi|^p,$$

where $C_p = \left(\frac{2C_1 p}{p-1} + \frac{2^r C_2 p}{r-p}\right)^{1/p}$. □

A1.3. Approximation of integrable functions

Let $f\colon \mathbb{R} \to \mathbb{R}$ be a (Borel) measurable function and $p \geq 1$ be a fixed real number. Define

$$\|f\|_p := \left(\int_{\mathbb{R}} |f(t)|^p \, dt \right)^{1/p}$$

and

$$\mathcal{L}_p(\mathbb{R}) = \left\{ f\colon \mathbb{R} \to \mathbb{R} \;\middle|\; f \text{ is measurable and } \|f\|_p < \infty \right\}.$$

This is a linear space and $\|\cdot\|_p$ is a seminorm.

Call a function $h\colon \mathbb{R} \to \mathbb{R}$ *elementary* if it has a form

$$h(t) = \sum_{j=1}^{m} c_j \mathbb{1}_{[a_j, b_j)}(t),$$

where $a_j, b_j, c_j, j = 1, \ldots, m$ are some real numbers where $a_j < b_j$.

PROPOSITION A1.1.– *The set of elementary functions is dense in $\mathcal{L}_p(\mathbb{R})$, i.e. for any $f \in \mathcal{L}_p(\mathbb{R})$ and any $\delta > 0$ there exists an elementary function h such that $\|f - h\|_p < \delta$.*

PROOF.– Define for $n \geq 1$

$$\varphi_n(x) = \sum_{k=1}^{n^2} \frac{k}{n} \left(\mathbb{1}_{[\frac{k}{n}, \frac{k+1}{n})}(x) - \mathbb{1}_{[-\frac{k+1}{n}, -\frac{k}{n})}(x) \right), \; x \in \mathbb{R}.$$

Then $|\varphi_n(x)| \leq |x|$, $x \in \mathbb{R}$, and $\varphi_n(x) \to x$, $n \to \infty$. Therefore, defining $g_n(t) = \varphi_n(f(t)) \to f(t)$, $t \in \mathbb{R}$, we have $|g_n(t)| \leq |f(t)|$, and $g_n(t) = \varphi_n(f(t)) \to f(t)$, $n \to \infty$. Hence, by the dominated convergence theorem, $\|f - g_n\|_p \to 0$, $n \to \infty$, so there exists $N \geq 1$ such that $\|f - g_N\|_p < \delta/2$.

Denote

$$A_k = \left\{ t \in \mathbb{R} : f(t) \in \left[\frac{k}{N}, \frac{k+1}{N} \right) \right\}, A_{-k}$$
$$= \left\{ t \in \mathbb{R} : f(t) \in \left[-\frac{k+1}{N}, -\frac{k}{N} \right) \right\}, k = 1, \ldots, N^2,$$

so that

$$g_N(t) = \sum_{k=1}^{N^2} \frac{k}{N} \left(\mathbb{1}_{A_k}(t) - \mathbb{1}_{A_{-k}}(t) \right).$$

The sets A_k and A_{-k} are measurable. Moreover, they have finite measure, as otherwise $\int_{\mathbb{R}} |g_N(t)|^p \, dt$ would be infinite. Since the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is generated by the semiring of half-open intervals of the form $[a, b)$, then by the Caratheodory approximation theorem (see [BIL 95, theorem 11.4]), for any $k = 1, 2, \ldots, N^2$ there exist disjoint intervals $[a_i, b_i)$, $i = 1, \ldots, m_k$, such that

$$\lambda_1\left(A_k \triangle \bigcup_{i=1}^{m_k} [a_i, b_i) \right) < \frac{\delta^p}{(4N)^{p+2}},$$

and disjoint intervals $[a_i', b_i')$, $i = 1, \ldots, m_k'$, such that

$$\lambda_1\left(A_{-k} \triangle \bigcup_{i=1}^{m_k'} [a_i', b_i') \right) < \frac{\delta^p}{(4N)^{p+2}}.$$

Define elementary function

$$h(t) = \sum_{k=1}^{N^2} \frac{k}{N^2} \left(\sum_{i=1}^{m_k} \mathbb{1}_{[a_i, b_i)}(t) - \sum_{i=1}^{m_k'} \mathbb{1}_{[a_i', b_i')}(t) \right).$$

Then

$$\begin{aligned}
&\lambda_1(\{t \in \mathbb{R} : h(t) \neq g_N(t)\}) \\
&\leq \sum_{k=1}^{N^2} \left(\lambda_1\left(A_k \triangle \bigcup_{i=1}^{m_k} [a_i, b_i) \right) + \lambda_1\left(A_{-k} \triangle \bigcup_{i=1}^{m_k'} [a_i', b_i') \right) \right) \\
&< \sum_{k=1}^{N^2} \frac{\delta^p}{2^{2p+1} N^{p+2}} < \frac{\delta^p}{(4N)^p}.
\end{aligned}$$

Therefore, taking into account that $|g_N(t)| \leq N$ and $|h(t)| \leq N$, we get

$$\begin{aligned}
&\|g_N - h\|_p \leq \left(\int_{\{t \in \mathbb{R} : h(t) \neq g_N(t)\}} \big(|g_N(t)| + |h(t)| \big)^p dt \right)^{1/p} \\
&\leq 2N \cdot \lambda_1(\{t \in \mathbb{R} : h(t) \neq g_N(t)\})^{1/p} < 2N \cdot \frac{\delta}{4N} = \frac{\delta}{2}.
\end{aligned}$$

Using the triangle inequality, we obtain $\|f-h\|_p \le \|f-g_N\|_p + \|g_N-h\|_p < \delta$, as required. □

PROPOSITION A1.2.– *The set $C_{fin}(\mathbb{R})$ of compactly supported continuous functions is dense in $\mathcal{L}_p(\mathbb{R})$, i.e. for any $f \in \mathcal{L}_p(\mathbb{R})$ and any $\delta > 0$, there exists a function $g \in C_{fin}(\mathbb{R})$ such that $\|f-g\|_p < \delta$.*

PROOF.– Using proposition A1.1, it is enough to prove the statement for elementary functions. In turn, by the triangle inequality, it suffices to prove it for indicator functions of the form $h(t) = \mathbb{1}_{[a,b)}(t)$, $a, b \in \mathbb{R}$, $a < b$. Defining

$$h_n(t) = \begin{cases} 1, & t \in [a, b - \frac{1}{n}], \\ n(t-a), & t \in [a - \frac{1}{n}, a), \\ n(b-t), & t \in (b - \frac{1}{n}, b], \\ 0, & t \notin [a - \frac{1}{n}, b], \end{cases}$$

we have $h_n \in C(\mathbb{R})$ and $h_n(t) \to \mathbb{1}_{[a,b)}(t)$, $n \to \infty$. Moreover, $|h_n(t)| \le \mathbb{1}_{[a-1,b]}(t)$; so, by the dominated convergence theorem $\|h_n - h\|_p \to 0$, $n \to \infty$, the statement follows. □

It turns out that one may construct a continuous function, which approximates $f \in \mathcal{L}_p(\mathbb{R})$, explicitly. Namely, define for $\varepsilon > 0$

$$f_\varepsilon(t) = \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} f(x)dx, \; t \in \mathbb{R}.$$

THEOREM A1.4.– *For any $f \in \mathcal{L}_p(\mathbb{R})$, $f_\varepsilon \in \mathcal{L}_p(\mathbb{R})$ and $\|f - f_\varepsilon\|_p \to 0$, $\varepsilon \to 0+$.*

PROOF.– By Jensen's inequality,

$$\begin{aligned} \|f_\varepsilon\|_p^p &= \int_{\mathbb{R}} \left| \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} f(x)dx \right|^p dt \le \int_{\mathbb{R}} \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} |f(x)|^p \, dx \, dt \\ &= \int_{\mathbb{R}} \frac{1}{\varepsilon} |f(x)|^p \int_{x}^{x+\varepsilon} dt \, dx = \int_{\mathbb{R}} |f(x)|^p \, dx = \|f\|_p^p < \infty, \end{aligned} \quad \text{[A1.2]}$$

so $f_\varepsilon \in \mathcal{L}_p(\mathbb{R})$. Further, fix some $\delta > 0$ and take $g \in C_{fin}(\mathbb{R})$ such that $\|f-g\|_p < \delta$. Then, using the triangle inequality,

$$\|f - f_\varepsilon\|_p \le \|f-g\|_p + \|g-g_\varepsilon\|_p + \|g_\varepsilon - f_\varepsilon\|_p . \quad \text{[A1.3]}$$

Similarly to [A1.2], $\|g_\varepsilon - f_\varepsilon\|_p \le \|f-g\| < \delta$. Further, thanks to continuity, $g_\varepsilon(t) \to g(t)$, $\varepsilon \to 0+$, for any $t \in \mathbb{R}$. Moreover, if $f(x) = 0$, $x \notin [-T, T]$, then $|g(t)| \le \max_{x \in [-T,T]} g(x) \mathbb{1}_{[-T-1,T+1]}(t)$ for any $\varepsilon \in (0,1)$. Thus, by the dominated convergence theorem, $\|g - g_\varepsilon\|_p \to 0$, $\varepsilon \to 0+$. Hence, we obtain the following relation from [A1.3]

$$\limsup_{\varepsilon \to 0+} \|f - f_\varepsilon\|_p \leq 2\delta.$$

Letting $\delta \to 0+$, we arrive at the statement. □

PROOF.– We will also need an approximation result with respect to the arbitrary probability measure μ on $\mathbb{R}^d$. It can be understood as the distribution of a random vector ζ in $\mathbb{R}^d$. One can define the spaces of integrable functions w.r.t. μ in the same way as for the Lebesgue measure, viz.

$$\mathcal{L}(\mathbb{R}^d, \mu) = \left\{ f \colon \mathbb{R}^d \to \mathbb{R} \;\middle|\; \int_{\mathbb{R}^d} |f(x)|^p \, \mu(dx) = \mathrm{E}\, |f(\zeta)|^p < \infty \right\}.$$

Call a function $g \colon \mathbb{R}^d \to \mathbb{R}$ a *trigonometric polynomial* if it has a form

$$g(t) = \sum_{k=0}^{n} \left(a_k \cos(\theta^k, t) + b_k \sin(\theta^k, t) \right), \ t \in \mathbb{R}^d,$$

where $a_k, b_k \in \mathbb{R}$, $\theta^k \in \mathbb{R}^d$.

THEOREM A1.5.– *The set of trigonometric polynomials is dense in* $\mathcal{L}_p(\mathbb{R}, \mu)$*, i.e. for any* $f \in \mathcal{L}_p(\mathbb{R}, \mu)$ *and any* $\delta > 0$ *there exists a trigonometric polynomial* g *such that* $\mathrm{E}\, |f(\zeta) - g(\zeta)|^p < \delta$.

PROOF.– Similar to theorem A1.4, the set $C_{fin}(\mathbb{R}^d)$ is dense in $\mathcal{L}_p(\mathbb{R}^d, \mu)$, and so it is enough to consider the case where f is a continuous compactly supported function. Evidently, we can assume that the support of f is in $[-R, R]^d$, where R is sufficiently large.

Consider the set $\mathcal{T}$ of trigonometric polynomials of the form

$$g(t) = a_0 + \sum_{\theta_1, \ldots, \theta_d = 1}^{n} \left(a_k \cos \frac{\pi(\theta, t)}{R} + b_k \sin \frac{\pi(\theta, t)}{R} \right)$$

with some $n \geq 1$, $a_k, b_k \in \mathbb{R}$. These are $2R$-periodic functions in each variable, and so they can be understood as functions on a d-dimensional torus $T_d := [-R, R]^d$ (where endpoints $-R$ and R of the segment $[-R, R]$ are identified). The set $\mathcal{T}$ is an algebra (a linear set closed under multiplication) and it separates the points of T_d: for any points $t_1 \neq t_2$ in T_d there is a function $g \in \mathcal{T}$, such that $g(t_1) \neq g(t_2)$. Moreover, $f \in C(T_d)$, since f vanishes on the boundary of T_d. Therefore, by the

Stone-Weierstrass theorem, for any $\varepsilon \in (0,1)$ there is some $g \in \mathcal{T}$ with $\sup_{t\in T_d} |f(t) - g(t)| < \varepsilon$. Then

$$\begin{aligned} \mathrm{E}\,|f(\zeta) - g(\zeta)|^p &= \mathrm{E}\big(\,|f(\zeta) - g(\zeta)|^p\, \mathbb{1}_{[-R,R]^d}(\zeta)\big) \\ &\quad + \mathrm{E}\big(\,|g(\zeta)|^p\, \mathbb{1}_{\mathbb{R}^d\setminus[-R,R]^d}(\zeta)\big) \\ &\le \varepsilon^p + \sup_{t\in\mathbb{R}^d} |g(t)|\, \mathrm{P}\left\{\zeta \in \mathbb{R}^d \setminus [-R,R]^d\right\}. \end{aligned}$$

Thanks to $2R$-periodicity of g,

$$\sup_{t\in\mathbb{R}^d} |g(t)| = \sup_{t\in[-R,R]^d} |g(t)| \le \sup_{t\in[-R,R]^d} |f(t)| + \varepsilon < \sup_{t\in\mathbb{R}^d} |f(t)| + 1.$$

By continuity of probability, $\mathrm{P}\left\{\zeta \in \mathbb{R}^d \setminus [-R,R]^d\right\} \to 0$, $R \to +\infty$. Consequently, $\mathrm{E}\,|f(\zeta) - g(\zeta)|^p$ can be made arbitrarily small, which concludes the proof. □

A1.4. Moduli of continuity

Let $f : [a,b] \to \mathbb{R}$ be a measurable function.

DEFINITION A1.3.–

1) Function f is continuous at the point $t_0 \in (a,b)$ if for any $\varepsilon > 0$ there exists $\delta > 0$ such that for $t \in (t_0 - \delta, t_0 + \delta) \cap [a,b]$ $|f(t) - f(t_0)| < \varepsilon$. Or, in other words, $\lim_{t\uparrow t_0} f(t) = \lim_{t\to t_0+} f(t) = f(t_0)$.

2) Function f is continuous on $[a,b]$ if it is continuous at any point $t \in (a,b)$ and $\lim_{t\to a+} f(t) = f(a)$, $\lim_{t\uparrow b} f(t) = f(b)$. We denote $C([a,b])$ as the space of continuous functions on $[a,b]$

3) Function f is uniformly continuous on $[a,b]$ if for any $\varepsilon > 0$, there exists $\delta > 0$, such that for any $t_1, t_2 \in [a,b]$ such that $|t_1 - t_2| < \delta$, we have that $|f(t_1) - f(t_2)| < \varepsilon$.

Introduce the following modulus of continuity

$$\Delta_c(f, [a,b], \delta) = \sup_{\substack{|t_1 - t_2| < \delta,\\ t_1, t_2 \in [a,b]}} |f(t_1) - f(t_2)|.$$

THEOREM A1.6.– Cantor-Heine theorem on uniform continuity. *Function $f : [a,b] \to \mathbb{R}$ is continuous on $[a,b]$ if and only if it is uniformly continuous on $[a,b]$.*

The next result immediately follows from the definition of a uniformly continuous function and the Cantor-Heine theorem on uniform continuity.

THEOREM A1.7.– *Function* $f : [a, b] \to \mathbb{R}$ *is continuous on* $[a, b]$ *if and only if*

$$\lim_{\delta \to 0} \Delta_c(f, [a, b], \delta) = 0.$$

DEFINITION A1.4.–

1) Function $f : [a, b] \to \mathbb{R}$ *has no discontinuities of the second kind on* $[a, b]$ *if there exist limits* $\lim_{t \uparrow t_0} f(t) =: f(t_0-)$ $\lim_{t \to t_0+} f(t) = f(t_0+)$ *at any point* $t \in (a, b)$ *and there exist limits* $\lim_{t \to a+} f(t)$ *and* $\lim_{t \to b-} f(t)$.

2) Function $f : [a, b] \to \mathbb{R}$ *is* D-regular *on* $[a, b]$ *if it has no discontinuities of the second kind on* $[a, b]$ *and at any point* $t \in (a, b)$

$$f(t) = f(t-) \ \textit{or} \ f(t) = f(t+),$$

and $f(a+) = f(a)$, $f(b-) = f(b)$.

3) Function $f : [a, b] \to \mathbb{R}$ *is a càdlàg function if it has no discontinuities of the second kind, is continuous from the right at any point* $t \in [a, b)$ *and continuous from the left at point* b.

4) Function $f : [a, b] \to \mathbb{R}$ *has on* $[a, b]$ *at least* k ε*-oscillations if there exists points* $\{t_0, t_1, \ldots, t_k\} \subset [a, b]$ *such that* $|f(t_i) - f(t_{i-1})| \geq \varepsilon$, $1 \leq i \leq k$.

THEOREM A1.8.– *Function* $f : [a, b] \to \mathbb{R}$ *has no discontinuities of the second kind on* $[a, b]$ *if and only if it has for any* $\varepsilon > 0$ *only a finite number of* ε*-oscillations.*

PROOF.– $\Rightarrow$ Let function f have no discontinuities of the second kind on $[a, b]$. Assume that for some $\varepsilon > 0$, we have an infinite number of ε-oscillations, so that we have an infinite (one- or two-sided) sequence of points such that $|f(t_i) - f(t_{i-1})| \geq \varepsilon$. Let, for example, the sequence be two-sided, $a \leq \ldots < t_{-k} < t_{-k+1} < \ldots < t_{-1} < t_0 < t_1 < \ldots < t_k < t_{k+1} < \ldots \leq b$, and $|f(t_i) - f(t_{i-1})| \geq \varepsilon$ for any $i \in \mathbb{Z}$. Then $t_{-k} \to t_-+ \in [a, b]$ and $t_k \uparrow t_+ \in [a, b]$. This means that the limits $f(t_-+)$ and $f(t_+-)$ do not exist, which is a contradiction. In the case of a one-sided sequence, for example, the sequence of the form $a \leq t_0 < t_1 < \ldots < t_k < t_{k+1} < \ldots \leq b$ can be considered similarly.

$\Leftarrow$ Let us assume that for any $\varepsilon > 0$ we have only a finite number of ε-oscillations. Fix some $\varepsilon > 0$, consider any point $t_0 \in (a, b)$ (points a and b can be considered similarly), and consider any increasing sequence $t_1 < t_2 < \ldots < t_n < \ldots < t_0$ such that $t_n \uparrow t_0, n \to \infty$. Then, there exists a number $n_0 \in \mathbb{N}$ such that $|f(t) - f(t_{n_0})| < \varepsilon$ for $t_{n_0} < t < t_0$. Then, for any $m, n > n_0$ we have that

$$|f(t_m) - f(t_n)| \leq |f(t_m) - f(t_{n_0})| + |f(t_m) - f(t_{n_0})| < 2\varepsilon.$$

It means that $f(t_n), n \geq 1$ is a Cauchy sequence, and therefore there exists a limit $f(t_0-)$. Existence of $f(t_0+)$ can be proved similarly. □

Introduce the following moduli of continuity

$$\Delta_d(f, [a,b], \delta) = \sup_{a \leq t \leq a+\delta} |f(t) - f(0)| + \sup_{b-\delta \leq t \leq b} |f(t) - f(b)| \\ + \sup_{\substack{a \leq t_1 < t_2 < t_3 \leq b, \\ t_3 - t_1 \leq \delta}} \min\big(|f(t_2) - f(t_1)|, |f(t_3) - f(t_2)|\big). \quad [A1.4]$$

THEOREM A1.9.– *The following conditions are equivalent:*

i) function $f : [a,b] \to \mathbb{R}$ *is D-regular on* $[a,b]$.

ii) $\lim_{\delta \to 0} \Delta_d(f, [a,b], \delta) = 0$.

PROOF.– Let function f be D-regular on $[a,b]$. Then, it has a right-hand limit at point a, therefore,

$$\sup_{a < t_1 < t_2 \leq a+\delta} |f(t_2) - f(t_1)| \to 0$$

as $\delta \to 0$. Similarly, f has a left-hand limit at point b. Therefore,

$$\sup_{b-\delta \leq t_1 < t_2 < b} |f(t_2) - f(t_1)| \to 0$$

as $\delta \to 0$. Assume that

$$\Delta_d(f, [a,b], \delta) \to \beta > 0 \text{ as } \delta \to 0$$

(the limit exists because $\Delta_d(f, [a,b], \delta)$ is nondecreasing in δ). It means that there exist three sequences $a \leq t_1^{(n)} < t_2^{(n)} < t_3^{(n)} \leq b$, $t_2^{(n)} - t_1^{(n)} < \frac{1}{n}$, $t_3^{(n)} - t_2^{(n)} < \frac{1}{n}$ such that $|f(t_3^{(n)}) - f(t_2^{(n)})| \geq \frac{\beta}{2}$, $|f(t_2^{(n)}) - f(t_1^{(n)})| \geq \frac{\beta}{2}$. If we take convergent subsequence $t_1^{(n_k)}$, $t_1^{n_k} \to t_0$, say, then $t_i^{(n_k)} \to t_0$, $i = 2, 3$. Assume that $t_0 \in (a,b)$. And then, for an infinite number of n_k there can be one of the following possibilities: $t_1^{(n_k)} \uparrow t_0$, $t_2^{(n_k)} \uparrow t_0$, or $t_1^{(n_k)} \uparrow t_0$, $t_2^{(n_k)} = t_0$, $t_3^{(n_k)} \to t_0+$, or $t_2^{(n_k)} \to t_0+$, $t_3^{(n_k)} \to t_0+$. In the first case $|f(t_2^{(n_k)}) - f(t_1^{(n_k)})| \to 0$ because f has the limit

$f(t_0-)$, in the third case $|f(t_3^{(n_k)}) - f(t_2^{(n_k)})| \to 0$ because f has the limit $f(t_0+)$, and in second case $|f(t_2^{(n_k)}) - f(t_1^{(n_k)})| = |f(t_1^{(n_k)}) - f(t_0)| \to 0$ if $f(t_0) = f(t_0-)$ and $|f(t_3^{(n_k)}) - f(t_2^{(n_k)})| = |f(t_3^{(n_k)}) - f(t_0)| \to 0$ if $f(t_0) = f(t_0+)$. We get contradiction which means that f satisfies (ii).

Let f satisfy (ii). Then, f has right-hand limit at point a and has left-hand limit at point b. Consider any point $t_0 \in (a, b)$. Then,

$$\sup_{a\vee(t_0-\delta)\le t_1<t_2<t_3\le(t_0+\delta)\wedge b} (\min(|f(t_2) - f(t_1)|, |f(t_3) - f(t_2)|)) \to 0 \,\text{as}\, \delta \to 0.$$

Therefore, if we take any sequences $a \vee (t_0 - \frac{1}{n}) \le t_1^{(n)} < t_2^{(n)} = t_0 < t_3^{(n)} \le (t_0 + \frac{1}{n}) \wedge b$, we get that either there exists a subsequence $\{n_k, k \ge 1\}$ such that $|f(t_0) - f(t_1^{(n_k)})| \to 0$ which means that f is continuous at t_0 from the left, or $|f(t_3^{(n_k)}) - f(t_0)| \to 0$ which means that f is continuous at t_0 from the right. Let f be continuous at t_0 from the left. Then, if we take any sequence $a \vee (t_0 - \frac{1}{n}) \le t_1^{(n)} < t_2^{(n)} < t_3^{(n)} \le (t_0 + \frac{1}{n}) \wedge b$, for example, $t_1^{(n)} < t_0 < t_2^{(n)} < t_3^{(n)}$, then either there exists subsequence $\{n_k, k \ge 1\}$ such that $|f(t_2^{(n_k)}) - f(t_1^{(n_k)})| \to 0$ which means that f is continuous at t_0, or $|f(t_3^{(n_k)}) - f(t_2^{(n_k)})| \to 0$ which means that f has right-hand limit at t_0. The case when f is continuous at t_0 from the right is considered similarly.□

A1.5. Functions of bounded variation

Consider the measurable function $f : [a, b] \to \mathbb{R}$ and any partition $\pi([a, b]) = \{a = t_0 < t_1 < \ldots < t_n = T\}$.

DEFINITION A1.5.– *Function f has bounded variation on the interval $[a, b]$ if $\mathrm{Var}_{[a,b]} f = \sup_{\pi[a,b]} \sum_{k=1}^{n} |f(t_k) - f(t_{k-})| < \infty$. In this case, we write $f \in BV([a, b])$ and its total variation on $[a, b]$ equals $\mathrm{Var}_{[a,b]} f$.*

THEOREM A1.10.– *Let $f \in BV([a, b])$. Then, it can be decomposed as $f_t = f_t^+ - f_t^-$, where both functions $f^\pm$ are non-decreasing on $[a, b]$.*

PROOF.– We can put $f_t^+ = \mathrm{Var}_{[a,t]}(f)$, $t \in [a, b]$. Then, it is easy to see that both functions f^+ and $f^- = f^+ - f$ are non-decreasing. □

THEOREM A1.11.– *Let $f \in C([a, b]) \cap BV([a, b])$. Create a sequence of partitions $\pi_n([a, b]) = \left\{a = t_0^{(n)} < t_1^{(n)} < \ldots < t_{k_n}^{(n)} = b\right\}$. If $|\pi_n| = \max_{1\le k\le k_n} |t_k^{(n)} - t_{k-1}^{(n)}| \to 0$, $n \to \infty$, then for any $p > 1$ $\sum_{k=1}^{k_n} \left| f_{t_k^{(n)}} - f_{t_{k-1}^{(n)}} \right|^p$ as $n \to \infty$.*

PROOF.– We have a very simple upper bound:

$$\sum_{k=1}^{k_n} \left| f_{t_k^{(n)}} - f_{t_{k-1}^{(n)}} \right|^p \le \max_{1\le k\le k_n} \left| f_{t_k^{(n)}} - f_{t_{k-1}^{(n)}} \right|^{p-1} \sum_{k=1}^{k_n} \left| f_{t_k}^{(n)} - f_{t_{k-1}}^{(n)} \right|$$

$$\le \max_{1\le k\le k_n} \left| f_{t_k^{(n)}} - f_{t_{k-1}^{(n)}} \right| \mathrm{Var}_{[a,b]} f \to 0 \text{ as } n \to \infty,$$

where we applied the Cantor-Heine theorem in the last relation. □

REMARK A1.1.– By similar calculations we can obtain the following result: let $f \in BV_{[a,b]}$, $g \in C[a,b]$. Then for any sequence $(\pi_n, n \ge 1)$ of partitions with $|\pi_n| \to 0$, $n \to \infty$, we have that

$$\lim_{n\to\infty} \sum_{k=1}^{k_n} \left(f_{t_k^{(n)}} - f_{t_{k-1}^{(n)}} \right) \left(g_{t_k^{(n)}} - g_{t_{k-1}^{(n)}} \right) = 0.$$

A1.6. Grönwall inequality

The following simple inequality, called the Grönwall inequality, or the Grönwall-Bellman lemma (it was formulated by Grönwall in a differential form; the integral form below was shown by Bellman) is a very efficient tool to study both deterministic and stochastic differential equations.

THEOREM A1.12.– *Let a function $f\colon [0,T] \to \mathbb{R}^+$ be integrable such that*

$$f(t) \le a + b \int_0^t f(s)ds, \; t \in [0,T]. \qquad \text{[A1.5]}$$

Then the function f admits an estimate

$$f(t) \le a e^{bt}, \; t \in [0,T].$$

PROOF.– Repeatedly plugging the inequality for f into [A1.5], we get

$$f(t) \le a + b \int_0^t \left(a + b \int_0^s f(u)du \right) ds = a + abt + \int_0^t f(u) \int_u^t ds\, du$$

$$= a(1 + bt) + \int_0^t f(u)(t-u)du$$

$$\le a\Big(1 + bt + \frac{(bt)^2}{2}\Big) + \int_0^t \frac{(t-v)^2}{2} f(v)dv$$

$$\leq \cdots \leq a\sum_{k=0}^{n}\frac{(bt)^k}{k!}+\int_0^t\frac{(t-z)^n}{n!}f(z)dz.$$

Estimate

$$\int_0^t\frac{(t-z)^n}{n!}f(z)dz \leq \frac{T^n}{n!}\int_0^t f(z)dz \to 0,\ n\to\infty.$$

Therefore, letting $n\to\infty$, we get

$$f(t)\leq a\sum_{k=0}^{\infty}\frac{(bt)^k}{k!}=ae^{bt}.$$ □

REMARK A1.2.– The Grönwall inequality should be applied carefully: one should ensure *a priori* that the function f in question is integrable, as any function f such that $\int_0^t f(s)ds=+\infty$, $t\in(0,T]$, clearly satisfies [A1.5].

REMARK A1.3.– It can be shown similarly that if a non-negative function f satisfies $f(t)\leq a+\int_0^t b(s)f(s)ds$, $t\in[0,T]$, for some non-negative integrable function b such that bf is integrable as well, then

$$f(t)\leq ae^{\int_0^t b(s)ds},\ t\in[0,T].$$

A1.7. Normed spaces, linear operators and semigroups

In this section, we present some basic notions and basic statements from operator theory without proofs.

DEFINITION A1.6.– *The set* $\mathbb{Z}$ *is called a (real)* linear space *if for any* $x,y\in\mathbb{Z}$, $x+y\in\mathbb{Z}$, *and for any* $x\in\mathbb{Z}$, $\alpha\in\mathbb{R}$ $\alpha x\in\mathbb{Z}$, *and the following properties hold:*

1) $\forall x,y\in\mathbb{Z}\quad x+y=y+x$*;*

2) $\forall x,y,z\in\mathbb{Z}\quad x+(y+z)=(x+y)+z$*;*

3) there exists the unique zero element $0\in\mathbb{Z}$ *such that for any* $z\in\mathbb{Z}\quad z+=z$*;*

4) $\forall z\in\mathbb{Z}$, *there exists the unique* $(-z)\in\mathbb{Z}$ *such that* $z+(-z)=0$*;*

5) $\forall\alpha,\beta\in\mathbb{R}\ \forall z\in\mathbb{Z}\quad \alpha(\beta z)=(\alpha\beta)z$ *and* $(\alpha+\beta)z=\alpha z+\beta z$*;*

6) $1\cdot z=z$*;*

7) $\forall\alpha\in\mathbb{R}$, $\forall x,y\in\mathbb{Z}\quad \alpha(x+y)=\alpha x+\alpha y$.

DEFINITION A1.7.– *Let* $\mathbb{Z}$ *be a real linear space. A function* $z \in \mathbb{Z} \to \|z\| \geq 0$ *is called a* norm*, if it satisfies the following conditions:*

1) $\|z\| = 0 \Leftrightarrow z = 0$;

2) $\forall \alpha \in \mathbb{R}$ *and* $\forall z \in \mathbb{Z} \quad \|\alpha z\| = |\alpha| \, \|z\|$;

3) $\forall x, y \in \mathbb{Z} \quad \|x + y\| \leq \|x\| + \|y\|$.

A linear space with norm is called a linear normed space.

DEFINITION A1.8.– *A set* $\widetilde{\mathbb{Z}} \subset \mathbb{Z}$ *is* closed *if it contains all limit points, i.e. for any* $\|z_n - z\| \to 0$, $z_n \in \widetilde{\mathbb{Z}}$, $n \geq 1$ *we have that* $z \in \widetilde{\mathbb{Z}}$. *A linear closed set* $\widetilde{\mathbb{Z}} \subset \mathbb{Z}$ *is called a* subspace.

DEFINITION A1.9.– *Let* $\mathbb{Z}_1$ *and* $\mathbb{Z}_2$ *be two linear normed spaces with norms* $\|\cdot\|_1$ *and* $\|\cdot\|_2$, *respectively. The map* $A : \mathbb{Z}_1 \to \mathbb{Z}_2$ *is called*

i) a linear operator *if* $\forall x, y \in \mathbb{Z}_1$ *and* $\forall \alpha, \beta \in \mathbb{R}$ $A(\alpha x + \beta y) = \alpha Ax + \beta Ay$;

ii) a continuous operator *if for any* $x_n, n \geq 1$, $x \in \mathbb{Z}$ *such that* $\|x_n - x\|_1 \to 0$ *as* $n \to \infty$ *we have that* $\|Ax_n - Ax\|_2 \to 0$.

iii) a bounded operator, *if there exists* $C > 0$ *such that* $\forall z \in \mathbb{Z}_1$ $\|Az\|_2 \leq C \|z\|_2$.

LEMMA A1.1.–

1) A linear operator $A : \mathbb{Z}_1 \to \mathbb{Z}_2$ *is bounded if and only if it is continuous.*

$$2) \quad \|A\| := \frac{\|Az\|_2}{\|z\|_1} = \sup_{z \in \mathbb{Z}_1, \|z\|_1 \leq 1} \|Az\|_2$$
$$= \inf \left\{ C > 0 : \forall z \in \mathbb{Z}_1 \quad \|Az\|_2 \leq C \|z\|_1 \right\},$$

and $\|A\| < \infty$ *if and only if* A *is a bounded operator.*

The number $\|A\|$ *is called a norm of the linear bounded operator* A.

DEFINITION A1.10.– *The family of linear bounded operators* $\{T_t, t \geq 0\} : \mathbb{Z} \to \mathbb{Z}$, *where* $\mathbb{Z}$ *is a normed linear space, is called a semigroup, if* $T_0 = I$ *(identical operator), and for any* $t, s \geq 0$

$$T_{t+s} = T_t T_s (= T_s T_t). \qquad [A1.6]$$

DEFINITION A1.11.– *Operator*

$$Ax = \lim_{t \to 0+} \frac{T_t x - x}{t},$$

defined on such $x \in \mathbb{Z}$ for which the limit exists in norm of the space $\mathbb{Z}$ is called a generator *(*infinitesimal operator*) of semigroup* $\{T_t, t \geq 0\}$.

Denote D_A the domain of A.

THEOREM A1.13.–

1) Generator A is a closed linear operator from D_A into $\mathbb{Z}$, i.e. if $x_n \to x$ in the norm of $\mathbb{Z}$, $x_n \in D_A$, $n \geq 1$, and $Ax_n \to y$ in the norm of $\mathbb{Z}$, then $y \in D_A$ and $y = Ax$.

2) For any $t \geq 0$ and $z \in D_A$

$$\frac{dT_t z}{dt} = AT_t z = T_t A z,$$

where $\left.\frac{dT_t}{dt}\right|_{t=0}$ is understood as the right-hand side derivable.

3) For any $0 \leq s \leq t$ and $z \in D_A$

$$T_t z - T_s z = \int_s^t AT_u z \, du = \int_s^t T_u A z \, du.$$

The next theorem states that the linear bounded operator can be added under the sign of the integral and removed from the sign of the integral.

THEOREM A1.14.– *Let us have a linear bounded operator $B : \mathbb{Z} \to \mathbb{Z}$ and let $\{P_t, t \geq 0\}$ be such family of linear bounded operators that $\int_0^\infty P_t dt$ exists as the limit of Riemann sums, in the sense that*

$$\int_0^\infty P_t dt = \lim_{n \to \infty} \sum_{k=1}^{k_n} P_{t_k^{(n)}} \Delta t_{k+1}^{(n)}$$

for any sequence of partitions

$$\pi_n = \{0 = t_0^{(n)} < t_1^{(n)} < \ldots < t_{k_n}^{(n)}\}$$

such that $t_{k_n}^{(n)} \to \infty$ and $|\pi_n| \to 0$ as $n \to \infty$, and let $\int_0^\infty \|P_t\| \, dt < \infty$. Then $B \int_0^\infty P_t dt = \int_0^\infty BP_t dt$.

PROOF.– If condition $\int_0^\infty \|P_t\| \, dt < \infty$ is fulfilled the sums $\{\sum BP_{t_k^n} \Delta t_k^{(n)}, n \geq 1\}$ create a Cauchy sequence and therefore $\int_0^\infty BP_t dt$ is also a limit of the Riemann sums, in the same sense. And then, the proof follows immediately from two observations: first, the linear operator can be added and removed from the sign of any sum, in particular

$$B\left(\sum P_{t_k^{(n)}}\Delta t_{k+1}^{(n)}\right)=\sum BP_{t_k^{(n)}}\Delta t_{k+1}^{(n)}, \quad \text{[A1.7]}$$

and second, the linear bounded operator is continuous and therefore we can go to the limit in [A1.7]. □

Appendix 2

Selected Facts from Probability Theory and Auxiliary Computations for Stochastic Processes

A2.1. Families of sets and monotone class theorems

Consider families of subsets of some universal set Ω.

DEFINITION A2.1.– *A non-empty family $\mathcal{P}$ of sets is called a* π-system *if it is closed under an intersection, i.e. for any $A, B \in \mathcal{P}$, $A \cap B \in \mathcal{P}$.*

DEFINITION A2.2.– *A family $\mathcal{L}$ of sets is called a* λ-system *if it satisfies the following conditions:*

1) $\Omega \in \mathcal{L}$;

2) if $A, B \in \mathcal{L}$ and $A \subset B$, then $B \setminus A \in \mathcal{L}$;

3) if $\{A_n, n \geq 1\} \subset \mathcal{L}$ are such that for any $n \geq 1$, $A_n \subset A_{n+1}$, then $\bigcup_{n\geq 1} A_n \in \mathcal{L}$.

REMARK A2.1.– λ-systems are also called d-systems and Dynkin systems, named after Eugene Dynkin, who introduced them. It is easy to see that the following conditions are equivalent to the definition of a λ-system:

1) $\Omega \in \mathcal{L}$;

2) if $A \in \mathcal{L}$, then $A^c := \Omega \setminus A \in \mathcal{L}$;

3) if $\{A_n, n \geq 1\} \subset \mathcal{L}$ are disjoint, then $\bigcup_{n\geq 1} A_n \in \mathcal{L}$.

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

Recall also that a family $\mathcal{F}$ is called σ-algebra (or a σ-field) if it satisfies

1) $\Omega \in \mathcal{F}$;

2) if $A \in \mathcal{F}$, then $\Omega \setminus A \in \mathcal{F}$;

3) if $\{A_n, n \geq 1\} \subset \mathcal{F}$, then $\bigcup_{n\geq 1} A_n \in \mathcal{F}$.

LEMMA A2.1.– *A family $\mathcal{F}$ is a σ-algebra if and only if it is a π-system and a λ-system.*

PROOF.– A σ-algebra is clearly a π-system and a λ-system, and so we only need to prove sufficiency. The first two conditions from the definition of σ-algebra coincide with those from the equivalent definition of a λ-system. For any $A, B \in \mathcal{F}$, $A \cup B = (A^c \cap B^c)^c \in \mathcal{F}$, therefore, $\mathcal{F}$ is closed under taking finite unions. Consequently, for any $\{A_n, n \geq 1\} \subset F$, we have $U_n := \bigcup_{k=1}^n A_k \in \mathcal{F}$ and $U_n \subset U_{n+1}$, $n \geq 1$, and so $\bigcup_{n\geq 1} A_n = \bigcup_{n\geq 1} U_n \in \mathcal{F}$. □

For any family $\mathcal{C} \subset 2^\Omega$, denote

$$\lambda(\mathcal{C}) = \bigcap_{\lambda\text{-system } \mathcal{L}\supset\mathcal{C}} \mathcal{L}$$

the smallest λ-system which contains $\mathcal{C}$. Recall also that $\sigma(\mathcal{C})$ is the smallest σ-algebra containing $\mathcal{C}$.

THEOREM A2.1.– (*Dynkin's π-λ theorem*) *For any π-system $\mathcal{P}$,*

$$\lambda(\mathcal{P}) = \sigma(\mathcal{P}).$$

In particular, if $\mathcal{L} \supset \mathcal{P}$ is a λ-system, then $\mathcal{L} \supset \sigma(\mathcal{P})$.

PROOF.– Since any σ-algebra is a λ-system, we have $\lambda(\mathcal{P}) \subset \sigma(\mathcal{P})$. Thanks to lemma A2.1, to prove the opposite inclusion, we need to show that $\lambda(\mathcal{P})$ is a π-system. Define the class

$$\mathcal{L}' = \{A \in \lambda(\mathcal{P}) \mid \text{for any } B \in \mathcal{P}, A \cap B \in \lambda(\mathcal{P})\}.$$

Obviously, $\mathcal{P} \subset \mathcal{L}'$. Let us check that $\mathcal{L}'$ is a λ-system. Obviously, $\Omega \in \mathcal{L}$. If $A_1, A_2 \in \mathcal{L}'$ are such that $A_1 \subset A_2$, then for any $B \in \mathcal{P}$, $A_i \cap B \in \lambda(\mathcal{P})$, $i = 1, 2$ and $A_1 \cap B \subset A_2 \cap B$, so

$$(A_2 \setminus A_1) \cap B = (A_2 \cap B) \setminus (A_1 \cap B) \in \lambda(\mathcal{P}).$$

Consequently, $A_2 \setminus A_1 \in \mathcal{L}'$. Further, if $\{A_n, n \geq 1\} \subset \mathcal{L}'$ are such that for any $n \geq 1$, $A_n \subset A_{n+1}$, then for any $B \in \mathcal{P}$ and any $n \geq 1$, $A_n \cap B \in \lambda(\mathcal{P})$ and $A_n \cap B \subset A_{n+1} \cap B$, so

$$\left(\bigcup_{n\geq 1} A_n\right) \cap B = \bigcup_{n\geq 1} (A_n \cap B) \in \lambda(\mathcal{P}).$$

As a result, $\bigcup_{n\geq 1} A_n \in \mathcal{L}'$, which shows that $\mathcal{L}'$ is indeed a λ-system. Since $\mathcal{L}' \supset \mathcal{P}$, we have $\mathcal{L}' = \lambda(\mathcal{P})$. With this at hand, the same argument shows that

$$\{A \in \lambda(\mathcal{P}) \mid \text{for any } B \in \lambda(\mathcal{P}), A \cap B \in \lambda(\mathcal{P})\} = \lambda(\mathcal{P}),$$

hence, $\lambda(\mathcal{P})$ is a π-system, as required. □

As an immediate corollary, we have the following result.

THEOREM A2.2.– *Let* P *and* Q *be two probability measures defined on some* σ*-algebra* $\mathcal{F}$*. If* $\mathrm{P}\{A\} = \mathrm{Q}\{A\}$ *for any set* $A \in \mathcal{P}$*, and* A *is a* π*-system, then* $\mathrm{P}\{A\} = \mathrm{Q}\{A\}$ *for any* $A \in \sigma(\mathcal{P})$*.*

PROOF.– Define

$$\mathcal{G} = \{A \in \mathcal{F} \mid \mathrm{P}\{A\} = \mathrm{Q}\{A\}\}.$$

It follows easily from the properties of probability measures that $\mathcal{G}$ is a λ-system. Also $\mathcal{G} \supset \mathcal{P}$ by assumption, and so $\mathcal{G} \supset \lambda(\mathcal{P})$. The statement then follows from theorem A2.1. □

Theorem A2.2 is a very efficient tool to prove equality of probability measures, as it assumes merely that $\mathcal{P}$ is closed under intersections. An even more powerful result is its functional counterpart. It is often called a *functional monotone class theorem* due to the fact that general measure-theoretic results of this kind are concerned with monotone classes.

THEOREM A2.3.– *Let* $\mathcal{P}$ *be a* π*-system, and a family* $\mathfrak{F}$ *of real-valued functions defined on* Ω *satisfy*

i) for any $A \in \mathcal{P}$*,* $\mathbb{1}_A \in \mathfrak{F}$*;*

ii) $\mathbb{1} \in \mathfrak{F}$*, where* $\mathbb{1}(x) \equiv 1$*;*

iii) $\mathfrak{F}$ *is linear, i.e. for any* $f_1, f_2 \in \mathfrak{F}$ *and* $a_1, a_2 \in \mathbb{R}$*,* $a_1 f_1 + a_2 f_2 \in \mathfrak{F}$*;*

iv) if $\{f_n, n \geq 1\} \subset \mathfrak{F}$ *is such that* $0 \leq f_n(x) \leq f_{n+1}(x)$*,* $n \geq 1$*,* $x \in \Omega$*,* $f_n(x) \to f(x)$*,* $n \to \infty$*, and* f *is bounded, then* $f \in \mathfrak{F}$*.*

Then $\mathfrak{F}$ contains all bounded $\sigma(\mathcal{P})$-$\mathcal{B}(\mathbb{R})$-measurable functions.

PROOF.– Consider first the family

$$\mathcal{F} = \{A \in \sigma(\mathcal{P}) \mid \mathbb{1}_A \in \mathfrak{F}\}.$$

The condition (i) means that $\mathcal{F} \supset \mathcal{P}$, and the conditions (ii)–(iv) immediately imply that $\mathcal{F}$ is a λ-system. Therefore, by theorem A2.1, $\mathcal{F} \supset \sigma(\mathcal{P})$. By linearity, $\mathfrak{F}$ contains all simple $\mathcal{F}$-measurable functions. Since all bounded non-negative $\mathcal{F}$-$\mathcal{B}(\mathbb{R})$-measurable functions are limits of increasing sequences of simple functions, they belong to $\mathfrak{F}$ as well. Finally, any bounded $\mathcal{F}$-$\mathcal{B}(\mathbb{R})$-measurable function f is a difference of two bounded non-negative $\mathcal{F}$-$\mathcal{B}(\mathbb{R})$-measurable functions, e.g. $f = f^+ - f^-$, and so by linearity $f \in \mathfrak{F}$, as required. □

A2.2. Some calculations related to Gaussian and Poisson distributions

In spite of the fact that the following result is well known, its proof is interesting by itself; therefore, we present it here.

LEMMA A2.2.– *Let a random variable ξ have Gaussian distribution with* $\mathrm{E}\xi = m$ *and* $\mathrm{Var}\,\xi = \sigma^2$, *i.e., $\xi \sim \mathcal{N}(m, \sigma^2)$. Then its characteristic function has a form: for any $\lambda \in \mathbb{R}$*

$$\mathrm{E}\exp\{i\lambda\xi\} = \exp\left\{i\lambda m - \frac{\lambda^2\sigma^2}{2}\right\}. \qquad [A2.1]$$

PROOF.– The following equalities are evident:

$$\mathrm{E}\exp\{i\lambda\xi\} = \frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}} \exp\{i\lambda x\}\exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}dx.$$

Now, $i\lambda x - \frac{(x-m)^2}{2\sigma^2} = -\frac{1}{2\sigma^2}\left(x - (m + i\lambda\sigma^2)\right)^2 + i\lambda m - \frac{\lambda^2\sigma^2}{2}$. Therefore,

$$\mathrm{E}\exp\{i\lambda\xi\} = \exp\left\{i\lambda m - \frac{\lambda^2\sigma^2}{2}\right\}\frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left\{-\frac{(x-(m+i\lambda\sigma^2))^2}{2\sigma^2}\right\}dx.$$

It remains to prove that

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\mathbb{R}} \exp\left\{-\frac{(x-(m+i\lambda\sigma^2))^2}{2\sigma^2}\right\}dx = 1.$$

To this end, we note that

$$\int_{\mathbb{R}} \exp\left\{-\frac{(x-(m+i\lambda\sigma^2))^2}{2\sigma^2}\right\} dx = \int_{-\infty-i\lambda\sigma^2}^{+\infty+i\lambda\sigma^2} e^{-\frac{y^2}{2\sigma^2}}\, dy, \qquad \text{[A2.2]}$$

where integration in the last integral is over a straight line $0x'$ in the complex plane, parallel to $0x$. Define a closed contour K_R, as is shown in Figure A2.1.

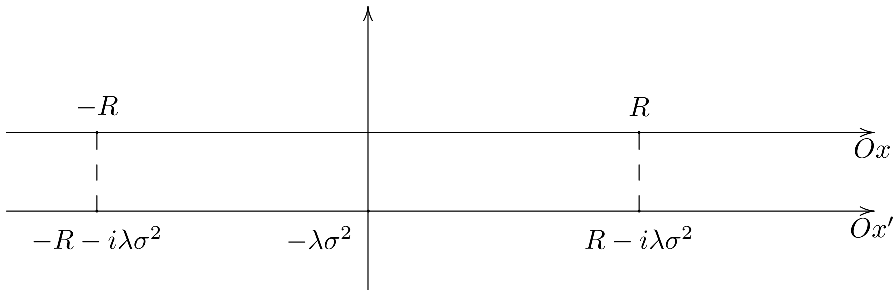


Figure A2.1. *Parallel axes and contour of integration*

Since the function $p(y) = e^{-\frac{y^2}{2\sigma^2}}$ is analytic in the complex plane, we have that

$$\int_{K_R} e^{-\frac{y^2}{2\sigma^2}}\, dy = 0.$$

Therefore,

$$\int_{-R-i\lambda\sigma^2}^{R-i\lambda\sigma^2} e^{-\frac{y^2}{2\sigma^2}}\, dy = \int_{-R}^{R} e^{-\frac{y^2}{2\sigma^2}}\, dy + \int_{-\lambda\sigma^2}^{0} e^{-\frac{(-R+iy)^2}{2\sigma^2}}\, dy - \int_{-\lambda\sigma^2}^{0} e^{-\frac{(R+iy)^2}{2\sigma^2}}\, dy. \qquad \text{[A2.3]}$$

The last two integrals in [A2.3] obviously vanish as $R \to \infty$. So,

$$\int_{-\infty-i\lambda\sigma^2}^{+\infty-i\lambda\sigma^2} e^{-\frac{y^2}{2\sigma^2}}\, dy = \int_{-\infty}^{+\infty} e^{-\frac{y^2}{2\sigma^2}}\, dy = \sigma\sqrt{2\pi}. \qquad \text{[A2.4]}$$

Now the equality [A2.1], and consequently the statement of the lemma, follow from [A2.2] and [A2.4]. □

LEMMA A2.3.– *Let* $\xi \sim \mathcal{N}(0,1)$. *Then,*

$$\mathrm{E}|\xi|^p = \left(\frac{2^p}{\pi}\right)^{\frac{1}{2}} \Gamma\left(\frac{p+1}{2}\right) \text{ for any } p > 0.$$

PROOF.– Evidently,

$$\begin{aligned}\mathrm{E}|\xi|^p &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x|^p e^{-\frac{x^2}{2}} dx = \sqrt{\frac{2}{\pi}} \int_0^\infty x^p e^{-\frac{x^2}{2}} dx \\ &= \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \int_0^\infty y^{\frac{p-1}{2}} e^{-y} dy = \frac{2^{\frac{p}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{p+1}{2}\right). \qquad \square\end{aligned}$$

LEMMA A2.4.– *Let the random variable* ξ *have a Poisson distribution with parameter* $\lambda > 0$. *Then its characteristic function has a form: for any* $\beta \in \mathbb{R}$

$$\mathrm{E}\exp\{i\beta\xi\} = e^{\lambda(e^{i\beta}-1)}.$$

PROOF.– Indeed,

$$\begin{aligned}\mathrm{E}\exp\{i\beta\xi\} &= \sum_{n=0}^{\infty} e^{i\beta n} \mathrm{P}\{\xi = n\} = \sum_{n=0}^{\infty} e^{i\beta n} e^{-\lambda} \frac{\lambda^n}{n!} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^{i\beta})^n}{n!} = e^{-\lambda} e^{\lambda e^{i\beta}} = e^{\lambda(e^{i\beta}-1)}. \qquad \square\end{aligned}$$

DEFINITION A2.3.– *Let* $\{\xi_n, n \geq 1, \xi\}$ *be random variables with cumulative distribution functions* $\{F_n, n \geq 1\}$, F *respectively. We say that* ξ_n *weakly converges to* ξ *as* $n \to \infty$, *with the notation* $\xi_n \xrightarrow{W} \xi$, *if for any point* x *where* F *is continuous, we have that* $F_n(x) \to F(x)$, $n \to \infty$.

Equivalent definitions of weak convergence and properties of weakly convergent random variables are described in detail in [BIL 99]. We mention here only that the weak convergence of the random variables is equivalent to the convergence of their characteristic functions, more precisely, to the point-wise convergence $\varphi_n(\lambda) \to \varphi(\lambda)$ at any point $\lambda \in \mathbb{R}$.

LEMMA A2.5.–

1) Let $\{\xi_n, n \geq 1\}$ *be a sequence of Gaussian random variables with* $\mathrm{E}\xi_n = m_n$ *and* $\mathrm{Var}\xi_n = \sigma_n^2 > 0$. *Let* $m_n \to m$ *and* $\sigma_n^2 \to \sigma^2 \in (0, \infty)$. *Then* $\xi_n \xrightarrow{W} \xi$, *where* ξ *is a Gaussian r.v. with* $\mathrm{E}\xi = m$ *and* $\mathrm{Var}\,\xi = \sigma^2$.

2) Let $\{\xi_n, n \geq 1\}$ be a sequence of Gaussian random variables with $\mathrm{E}\xi_n = m_n$ and $\mathrm{Var}\,\xi_n = \sigma_n^2$. Also, let $\xi_n \xrightarrow{W} \xi$, $n \to \infty$. Then there exist limits $\lim_{n\to\infty} m_n =: m$, $\lim_{n\to\infty} \sigma_n =: \sigma$ and $\xi = \mathcal{N}(m, \sigma^2)$.

PROOF.– 1) Consider

$$\mathrm{E}\exp\{i\lambda\xi_n\} = \exp\left\{i\lambda m_n - \frac{1}{2}\lambda^2\sigma_n^2\right\} \to \exp\left\{i\lambda m - \frac{1}{2}\lambda^2\sigma^2\right\},\ n \to \infty,$$

for any $\lambda \in \mathbb{R}$. Since the convergence of characteristic functions at every point $\lambda \in \mathbb{R}$ is equivalent to the weak convergence of random variables, we get that $\xi_n \xrightarrow{W} \xi = \mathcal{N}(m, \sigma^2)$, $n \to \infty$.

2) We first show that the sequences $\{m_n, n \geq 1\}$ and $\{\sigma_n, n \geq 1\}$ are bounded. Denote F and F_n, $n \geq 1$, the cumulative distribution functions of ξ and ξ_n, $n \geq 1$, respectively. Let a, b be points of continuity of the cumulative distribution function F, such that $F(a) < 1/3$, $F(b) > 2/3$. By the definition of weak convergence, $F_n(a) \to F(a)$ and $F_n(b) \to F(b)$, $n \to \infty$, so there exists some $n_0 \geq 1$ such that for all $n \geq n_0$, $F_n(a) < 1/3$ and $F_n(b) > 2/3$. Since $\xi_n = \mathcal{N}(m_n, \sigma_n^2)$, we have $F_n(a) = \Phi((a - m_n)/\sigma_n) < 1/3$ for $n \geq n_0$, where Φ is the standard normal cumulative distribution function. Hence $a - m_n < \sigma_n\Phi^{-1}(1/3)$, so $m_n > a - \sigma_n\Phi^{-1}(1/3) \geq a$. Similarly,

$$m_n < b - \sigma_n\Phi^{-1}(2/3) \leq b,\ n \geq n_0, \qquad \text{[A2.5]}$$

which implies the boundedness of $\{m_n, n \geq 1\}$. Further, from [A2.5] we have

$$\sigma_n < \frac{b - m_n}{\Phi^{-1}(2/3)},$$

so $\{\sigma_n, n \geq 1\}$ is bounded as well.

Now let $\left\{(m_{n_k}, \sigma_{n_k}^2), k \geq 1\right\}$ be a subsequence of $\left\{(m_n, \sigma_n^2), n \geq 1\right\}$ such that $(m_{n_k}, \sigma_{n_k}^2) \to (m, \sigma^2)$, $k \to \infty$. It follows from 1) and uniqueness of weak limit that $\xi = \mathcal{N}(m, \sigma^2)$. Moreover, appealing again to the uniqueness of weak limit, each convergent subsequence of $\left\{(m_n, \sigma_n^2), n \geq 1\right\}$ must converge to (m, σ^2), which means that (m_n, σ_n^2) itself converges to (m, σ^2), concluding the proof. □

A2.3. Notion and properties of the uniform integrability

DEFINITION A2.4.– *A family of random variables $\{X_t, t \in \mathbb{T}\}$ is uniformly integrable if*

$$\lim_{C\to\infty} \sup_{t\in\mathbb{T}} \mathrm{E}|X_t|\mathbb{1}_{|X_t|\geq C} = 0.$$

THEOREM A2.4.– *Let the family* $\{X_n, n \geq 1\}$ *be uniformly integrable. Then,*

1) $\sup_{n\geq 1} \mathrm{E}|X_n| < \infty$*;*

2) If additionally $X_n \to X_\infty$ *in probability, then* $\mathrm{E}|X_n - X_\infty| \to 0$*,* $n \to \infty$*;*

3) If $X_n \to X_\infty$*,* $n \to \infty$ *in probability,* $\mathrm{E}|X_n - X_\infty| \to 0$*,* $n \to \infty$*, and* $\{X_n, X_\infty\} \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$*, then* $\{X_n, n \geq 0\}$ *is uniformly integrable.*

PROOF.– 1) Obviously,

$$\sup_{n\geq 1} \mathrm{E}|X_n| \leq C + \sup_{n\geq 1} \mathrm{E}|X_n| \mathbb{1}_{|X_n|\geq C}.$$

Choosing $C_0 > 0$ such that $\sup_{n\geq 1} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C_0} \leq 1$, we get that $\sup_{n\geq 1} \mathrm{E}|X_n| \leq C_0 + 1$.

2) It is clear that $\mathrm{E}|X_\infty| \leq \limsup_{n\to\infty} \mathrm{E}|X_n| < \infty$ by Fatou's lemma. And then, for any $0 < \varepsilon < C$

$$\mathrm{E}|X_n - X_\infty| \leq \varepsilon + C\mathrm{P}\left\{|X_n - X_\infty| > \varepsilon\right\} + \mathrm{E}|X_n - X_\infty| \mathbb{1}_{|X_n - X_\infty|>C}. \quad \text{[A2.6]}$$

Consider the last term and bound it from above as follows:

$$A_{n,C} := \mathrm{E}|X_n - X_\infty| \mathbb{1}_{|X_n - X_\infty|>C} \leq \mathrm{E}|X_n| \left(\mathbb{1}_{|X_n|>\frac{C}{2}} + \mathbb{1}_{|X_\infty|>\frac{C}{2}}\right) + \mathrm{E}|X_\infty| \left(\mathbb{1}_{|X_n|>\frac{C}{2}} + \mathbb{1}_{|X_\infty|>\frac{C}{2}}\right).$$

We know that

$$\sup_{n\geq 1} \mathrm{E}|X_n| \mathbb{1}_{|X_n|>\frac{C}{2}} \to 0, \ C \to \infty \quad \text{[A2.7]}$$

and

$$\mathrm{E}|X_\infty| \mathbb{1}_{|X_\infty|>\frac{C}{2}} \to 0, \ C \to \infty. \quad \text{[A2.8]}$$

Now,

$$\sup_{n\geq 1} \mathrm{E}|X_n| \mathbb{1}_{|X_\infty|>\frac{C}{2}} \leq b\mathrm{P}\left\{|X_\infty| > \frac{C}{2}\right\} + \sup_{n\geq 1} \mathrm{E}|X_n| \mathbb{1}_{|X_n|\geq b} \to \sup_{n\geq 1} \mathrm{E}|X_n| \mathbb{1}_{|X_n|\geq b} \ \text{as} \ C \to \infty.$$

Choosing $b > 0$ such that

$$\sup_{n\geq 1} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq b} < \varepsilon,$$

we get that

$$\lim_{C\to\infty} \sup_{n\geq 1} \mathrm{E}|X_\infty|\mathbb{1}_{|X_n|>\frac{C}{2}} < \varepsilon, \qquad \text{[A2.9]}$$

for any $\varepsilon > 0$, whence this limit is zero. Finally,

$$\mathrm{E}|X_\infty|\mathbb{1}_{|X_n|>\frac{C}{2}} \leq b\mathrm{P}\left\{|X_n| > \frac{C}{2}\right\} + \mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|>b}$$

$$\leq \frac{2b}{C}\mathrm{E}|X_n| + \mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|\geq b} \to \mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|\geq b} \text{ as } C\to\infty.$$

Choosing any $\varepsilon > 0$ and then $b > 0$ sufficiently large so that $\mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|\geq b} < \varepsilon$, we get that

$$\mathrm{E}|X_\infty|\mathbb{1}_{|X_n|\geq\frac{C}{2}} < \varepsilon.$$

Combining [A2.7]–[A2.9], we get that $\lim_{C\to\infty}\sup_{n\geq 0} A_{n,C} = 0$. Returning to [A2.6], note that

$$\lim_{n\to\infty} \mathrm{E}|X_n - X_\infty| \leq \varepsilon + \sup_{n\geq 1} A_{n,C}.$$

Therefore, letting $C \to \infty$ and then $\varepsilon \to 0$, we get the proof.

3) Let $X_n \xrightarrow{\mathrm{P}} X_\infty$, $n \to \infty$, $\mathrm{E}|X_n - X_\infty| \to 0$ and $\{X_n, n \geq 1, X_\infty\} \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$.

Choose $n_0 \in \mathbb{N}$ such that for $n \geq n_0$, $\mathrm{E}|X_n - X_\infty| < \varepsilon$. Then,

$$\sup_{n\geq 0} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} \leq \max_{0\leq n\leq n_0} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} + \sup_{n\geq n_0} \mathrm{E}|X_n - X_\infty|$$

$$+ \sup_{n\geq n_0} \mathrm{E}|X_\infty|\mathbb{1}_{|X_n|\geq C}.$$

Now, choose $C > 0$ such that $\max_{0\leq n\leq n_0} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} < \varepsilon$. Then,

$$\sup_{n\geq 0} \mathrm{E}|X_n|\mathbb{1}_{|X_n|\geq C} \leq 2\varepsilon + b\sup_{n\geq 0} \mathrm{P}\left\{|X_n| \geq C\right\} + \mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|>b}$$

$$\leq 2\varepsilon + \frac{b}{C}\sup_{n\geq 0} \mathrm{E}|X_n| + \mathrm{E}|X_\infty|\mathbb{1}_{|X_\infty|>b}.$$

Since $\mathrm{E}|X_n - X_\infty| \to 0$, then $\sup_{n\geq 0} \mathrm{E}|X_n| < \infty$. Therefore, we can choose $b > 0$ such that $\mathrm{E}|X_\infty| \mathbb{1}_{|X_\infty > b|} < \varepsilon$ and we can choose $C > 0$ such that $\frac{b}{C} \sup_{n\geq 0} \mathrm{E}_n |X_n| < \varepsilon$. Finally, we get that $\sup_{n\geq 0} \mathrm{E}|X_n| \mathbb{1}_{|X_n|\geq C} \leq 4\varepsilon$, and the proof follows. □

The following theorem is a criterion for uniform integrability; it is usually used as a simple sufficient condition.

THEOREM A2.5.– (*de la Vallée-Poussin*) *A family* $\{X_t, t \in \mathbb{T}\}$ *is uniformly integrable if and only if there exists a non-decreasing function* $V : \mathbb{R}_+ \to \mathbb{R}_+$ *such that* $V(x)/x \to \infty$, $x \to \infty$, *and* $\sup_{t\in\mathbb{T}} \mathrm{E}V(X_t) < \infty$.

PROOF.– *Necessity*. Let $\{C_n, n \geq 1\}$ be such that $\sup_{t\in\mathbb{T}} \mathrm{E}\left(|X_t| \mathbb{1}_{|X_t|\geq C_n}\right) \leq 2^{-n}$; without loss of generality we can assume that $C_{n+1} > C_n$ for all $n \geq 1$. Setting $V(x) = x \sum_{n=1}^{\infty} n \mathbb{1}_{[C_n, C_{n+1})}(x)$, we have for any $t \in \mathbb{T}$,

$$\mathrm{E}V(X_t) = \sum_{n=1}^{\infty} n\, \mathrm{E}\left(|X_t|\, \mathbb{1}_{[C_n, C_{n+1})}(X_t)\right) \leq \sum_{n=1}^{\infty} n 2^{-n},$$

so $\sup_{t\in\mathbb{T}} \mathrm{E}V(X_t) \leq \sum_{n=1}^{\infty} n2^{-n} < \infty$, as required.

Sufficiency. It follows from $V(x)/x \to +\infty$, $x \to \infty$, that

$$m(C) := \sup_{x\geq C} \frac{x}{V(x)} \to 0,\ C \to \infty,$$

so

$$\begin{aligned}\sup_{t\in\mathbb{T}} \mathrm{E}\left(|X_t|\, \mathbb{1}_{|X_t|\geq C}\right) &\leq \sup_{t\in\mathbb{T}} m(C) \mathrm{E}\left(V(|X_t|) \mathbb{1}_{|X_t|\geq C}\right) \\ &\leq m(C) \sup_{t\in\mathbb{T}} \mathrm{E}V(|X_t|) \to 0,\ C \to \infty.\end{aligned}$$

□

A2.4. Measurability, conditional expectation and conditional probability

A2.4.1. *Measurability with respect to a generated σ-field*

Let $(\Omega, \mathcal{F})$ and $(\mathcal{S}, \Sigma)$ be measurable spaces and $\xi \colon \Omega \to \mathcal{S}$ be measurable. Recall that the σ-field generated by ξ is defined as $\sigma(\xi) = \left\{\xi^{-1}(A), A \in \Sigma\right\}$.

THEOREM A2.6.– *A random variable* $\eta : \Omega \to \mathbb{R}$ *is* $\sigma(X)$-$\mathcal{B}(\mathbb{R})$*-measurable if and only if there exists a* Σ-$\mathcal{B}(\mathbb{R})$*-measurable function* $h \colon \mathcal{S} \to \mathbb{R}$ *such that* $\eta = h \circ \xi$, *i.e.* $\eta(\omega) = h(\xi(\omega))$ *for all* $\omega \in \Omega$.

PROOF.– Let first η be simple, i.e. $\eta(\omega) = \sum_{i=1}^{k} a_i \mathbb{1}_{A_i}(\omega)$, where the numbers $a_i \in \mathbb{R}$ are distinct, and $A_i \in \mathcal{F}$. The $\sigma(X)$-$\mathcal{B}(\mathbb{R})$-measurability of η implies that for any i, $A_i = \eta^{-1}(\{a_i\}) \in \sigma(X)$, and so there exists some $B_i \in \mathcal{B}(\mathbb{R})$ such that $A_i = \xi^{-1}(B_i)$. Therefore, $\eta = h \circ \xi$ with $h = \sum_{i=1}^{k} a_i \mathbb{1}_{B_i}$.

Let now η be arbitrary $\sigma(X)$-$B(\mathbb{R})$-measurable function. There exists a sequence $\{\eta_n, n \geq 1\}$ of simple $\sigma(X)$-$B(\mathbb{R})$-measurable functions such that $\eta_n(\omega) \to \eta(\omega)$, $n \to \infty$, for any $\omega \in \Omega$. As we have shown already, there are some Σ-$\mathcal{B}(\mathbb{R})$-measurable functions $h_n \colon \mathcal{S} \to \mathbb{R}$ such that $\eta_n = h_n \circ \xi$. The set

$$B = \left\{ x \in \mathcal{S} : \text{the limit } \lim_{n\to\infty} h_n(x) \text{ exists} \right\} \in \Sigma,$$

and the limit of measurable functions is measurable, therefore the function

$$h(x) = \begin{cases} \lim_{n\to\infty} h_n(x), & x \in B, \\ 0, & x \notin B, \end{cases}$$

is Σ-$\mathcal{B}(\mathbb{R})$-measurable. Since, clearly, $h \circ \xi = \eta$, the proof is complete. □

REMARK A2.2.– It is known that any complete separable metric space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is Borel-isomorphic to a Borel subset $B \in \mathcal{B}(\mathbb{R})$, i.e. there exists a measurable bijection $f \colon \mathcal{X} \to B$, such that its inverse is measurable as well. Consequently, theorem A2.6 is valid for any complete separable metric space in place of $\mathbb{R}$. More detail on this matter is available in [KAL 02].

A2.4.2. *Conditional expectation*

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a complete probability space, and $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field.

DEFINITION A2.5.– *Let ξ be an integrable random variable, i.e.* $\mathrm{E}|\xi| < \infty$. *An integrable random variable η is called a* conditional expectation with respect to $\mathcal{G}$ *if it satisfies*

– $\mathrm{E}(\xi \mid \mathcal{G})$ *is $\mathcal{G}$-measurable.*

– *For any set $A \in \mathcal{G}$*

$$\int_A \xi \, d\mathrm{P} = \int_A \eta \, d\mathrm{P},$$

or, equivalently,

$$\mathrm{E}(\xi \mathbb{1}_A) = \mathrm{E}(\eta \mathbb{1}_A).$$

The conditional expectation is denoted by $\mathrm{E}(\xi \mid \mathcal{G})$.

THEOREM A2.7.– *For any integrable random variable ξ, its conditional expectation $\mathrm{E}(\xi \mid \mathcal{G})$ exists. Moreover, it is unique up to a set of measure zero.*

PROOF.– Define a set function

$$\mathrm{M}\{A\} = \mathrm{E}(\xi \mathbb{1}_A) = \int_A \xi \, d\mathrm{P}, \ A \in \mathcal{G}.$$

Due to the properties of Lebesgue integral, this is a σ-finite signed measure on $\mathcal{G}$, absolutely continuous with respect to P. Therefore, it has a Radon-Nikodym derivative

$$\eta = \frac{d\mathrm{M}}{d\mathrm{P}},$$

i.e. a $\mathcal{G}$-measurable random variable η such that for any $A \in \mathcal{G}$,

$$\mathrm{M}(A) = \int_A \eta \, d\mathrm{P},$$

which establishes the existence.

Further, if $\eta' = \mathrm{E}(\xi \mid \mathcal{G})$ is another conditional expectation, define $B = \{\eta' > \eta\}$. Thanks to $\mathcal{G}$-measurability, $B \in \mathcal{G}$; therefore,

$$\mathrm{E}(\eta' \mathbb{1}_B) = \mathrm{E}(\xi \mathbb{1}_B) = \mathrm{E}(\eta \mathbb{1}_B).$$

Obviously, $\eta' \mathbb{1}_B \geq \eta \mathbb{1}_B$, and so the above equality implies $\mathrm{P}\{\eta' \mathbb{1}_B = \eta \mathbb{1}_B\} = 1$, when $\mathrm{P}\{\eta' > \eta\} = 0$. Similarly, $\mathrm{P}\{\eta' < \eta\} = 0$, yielding the uniqueness. □

For the sake of brevity, through the end of this section, equations and inequalities concerning random variables are understood in the almost sure sense without additional notice.

THEOREM A2.8.– *Let ξ, ζ be integrable random variables. Then the following properties hold.*

i) If $\xi \leq \zeta$, then $\mathrm{E}(\xi \mid \mathcal{G}) \leq \mathrm{E}(\zeta \mid \mathcal{G})$.

ii) If ξ is $\mathcal{G}$-measurable, then $\mathrm{E}(\xi \mid \mathcal{G}) = \xi$.

iii) (Locality) if $\xi = \zeta$ on some set $A \in \mathcal{G}$, then $\mathrm{E}(\xi \mid \mathcal{G}) = \mathrm{E}(\zeta \mid \mathcal{G})$ on A.

iv) (Linearity) for any $a, b \in \mathbb{R}$, $\mathrm{E}(a\xi + b\zeta \mid \mathcal{G}) = a\mathrm{E}(\xi \mid \mathcal{G}) + b\mathrm{E}(\zeta \mid \mathcal{G})$.

v) If ξ is independent of $\mathcal{G}$ (i.e. events A and $\{\xi \in B\}$ are independent for any $A \in \mathcal{G}$, $B \in \mathcal{B}(\mathbb{R})$), then $\mathrm{E}(\xi \mid \mathcal{G}) = \mathrm{E}\xi$; in particular,

$$\mathrm{E}(\xi \mid \{\varnothing, \Omega\}) = \mathrm{E}\xi.$$

vi) (Tower properties) if $\mathcal{H} \subset \mathcal{G}$ is a sub-σ-field, then

$$\mathrm{E}(\mathrm{E}(\xi \mid \mathcal{G}) \mid \mathcal{H}) = \mathrm{E}(\mathrm{E}(\xi \mid \mathcal{H}) \mid \mathcal{G}) = \mathrm{E}(\xi \mid \mathcal{H}).$$

in particular, taking $\mathcal{H} = \{\varnothing, \Omega\}$

$$\mathrm{E}(\mathrm{E}(\xi \mid \mathcal{G}) \mid \mathcal{H}) = \mathrm{E}(\mathrm{E}(\xi \mid \mathcal{G})) = \mathrm{E}\xi.$$

vii) If ξ is $\mathcal{G}$-measurable and bounded, then $\mathrm{E}(\xi\eta \mid \mathcal{G}) = \xi\mathrm{E}(\eta \mid \mathcal{G})$.

viii) (Jensen's inequality) if $f : \mathbb{R} \to \mathbb{R}$ is a convex function, and $f(\xi)$ is integrable, then

$$f(\mathrm{E}(\xi \mid \mathcal{G})) \le \mathrm{E}(f(\xi) \mid \mathcal{G}).$$

PROOF.– Properties (i)–(iii) follow immediately from definition.

Concerning (iv), for any $A \in \mathcal{G}$,

$$\begin{aligned}\mathrm{E}\left(\big(a\mathrm{E}(\xi \mid \mathcal{G}) + b\mathrm{E}(\zeta \mid \mathcal{G})\big)\mathbb{1}_A\right) &= a\mathrm{E}\left(\mathrm{E}(\xi \mid \mathcal{G})\mathbb{1}_A\right) + b\mathrm{E}\left(\mathrm{E}(\zeta \mid \mathcal{G})\mathbb{1}_A\right) \\ &= a\mathrm{E}\left(\xi\mathbb{1}_A\right) + b\mathrm{E}\left(\zeta\mathbb{1}_A\right) = \mathrm{E}\left((a\xi + b\zeta)\mathbb{1}_A\right),\end{aligned}$$

whence by definition we get $a\mathrm{E}(\xi \mid \mathcal{G}) + b\mathrm{E}(\zeta \mid \mathcal{G}) = \mathrm{E}(a\xi + b\zeta \mid \mathcal{G})$.

For (v), write for any $A \in \mathcal{G}$, in view of independence,

$$\mathrm{E}(\xi\mathbb{1}_A) = \mathrm{E}\xi \cdot \mathrm{E}\mathbb{1}_A = \mathrm{E}\big((\mathrm{E}\xi)\mathbb{1}_A\big),$$

so $\mathrm{E}\xi = \mathrm{E}(\xi \mid \mathcal{G})$.

Further, for $\mathcal{H} \subset \mathcal{G}$, $\mathrm{E}(\mathrm{E}(\xi \mid \mathcal{H}) \mid \mathcal{G}) = \mathrm{E}(\xi \mid \mathcal{H})$ follows from the $\mathcal{H}$-measurability, and hence, $\mathcal{G}$-measurability of $\mathrm{E}(\xi \mid \mathcal{H})$. To establish another tower property, write for arbitrary $A \in \mathcal{H}$

$$\mathrm{E}(\mathrm{E}(\mathrm{E}(\xi \mid \mathcal{G}) \mid \mathcal{H})\mathbb{1}_A) = \mathrm{E}(\mathrm{E}(\xi \mid \mathcal{G})\mathbb{1}_A) = \mathrm{E}(\xi\mathbb{1}_A),$$

since $A \in \mathcal{G}$, so $\mathrm{E}(\mathrm{E}(\xi \mid \mathcal{G}) \mid \mathcal{H}) = \mathrm{E}(\xi \mid \mathcal{H})$.

To prove (vii), we need to show that for any $A \in \mathcal{G}$,

$$\mathrm{E}(\xi\mathrm{E}(\eta \mid \mathcal{G})\mathbb{1}_A) = \mathrm{E}(\xi\eta\,\mathbb{1}_A). \qquad [A2.10]$$

For fixed η and A, let $\mathfrak{F}$ be the family of $\mathcal{G}$-measurable functions ξ satisfying [A2.10]. From (iii), we have $\mathrm{E}(\eta \mathbb{1}_B \mid \mathcal{G}) = \mathrm{E}(\eta \mid \mathcal{G})\mathbb{1}_B$, so $\mathbb{1}_B \in \mathfrak{F}$. Due to linearity of expectation, $\mathfrak{F}$ is linear. Finally, if ξ is bounded and $\xi_n \in \mathfrak{F}$ such that $0 \leq \xi_n \leq \xi_{n+1}$, and $\xi_n \to \xi$, $n \to \infty$, then by the dominated convergence theorem

$$\mathrm{E}(\xi \mathrm{E}(\eta \mid \mathcal{G})\mathbb{1}_A) = \lim_{n\to\infty} \mathrm{E}(\xi_n \mathrm{E}(\eta \mid \mathcal{G})\mathbb{1}_A) = \lim_{n\to\infty} \mathrm{E}(\xi_n \eta \, \mathbb{1}_A) = \mathrm{E}(\xi \eta \, \mathbb{1}_A),$$

which means that $\xi \in \mathfrak{F}$. Using theorem A2.3, we get that $\mathfrak{F}$ contains all bounded $\mathcal{G}$-measurable variables, finishing the proof. □

THEOREM A2.9.– *Let $\{\xi_n, n \geq 1\}$ be a sequence of random variables and η be an integrable random variable such that $|\xi_n| \leq \eta$ for all $n \geq 1$. Then*

1) if $\xi_n \to \xi$, $n \to \infty$, almost surely, then $\mathrm{E}(\xi_n \mid \mathcal{G}) \to \mathrm{E}(\xi \mid \mathcal{G})$, $n \to \infty$, almost surely;

2) if $\xi_n \xrightarrow{\mathrm{P}} \xi$, $n \to \infty$, then $\mathrm{E}(\xi_n \mid \mathcal{G}) \xrightarrow{\mathrm{P}} \mathrm{E}(\xi \mid \mathcal{G})$, $n \to \infty$.

PROOF.– Let first $\xi_n \to \xi$, $n \to \infty$, almost surely. Then $\zeta_k = \sup_{k \geq n} |\xi_k - \xi| \to 0$, $n \to \infty$, almost surely. By Jensen's inequality,

$$|\mathrm{E}(\xi_n \mid \mathcal{G}) - \mathrm{E}(\xi \mid \mathcal{G})| \leq \mathrm{E}(|\xi_n - \xi| \mid \mathcal{G}) \leq \mathrm{E}(\zeta_n \mid \mathcal{G}). \qquad \text{[A2.11]}$$

Since the sequence $\{\zeta_n, n \geq 1\}$ is non-increasing, the conditional expectations $\mathrm{E}(\zeta_n \mid \mathcal{G})$ are non-increasing as well, so there exists some limit $\zeta = \lim_{n\to\infty} \mathrm{E}(\zeta_n \mid \mathcal{G})$. By the Lebesgue-dominated convergence theorem, for any $A \in \mathcal{G}$,

$$\mathrm{E}(\zeta \mathbb{1}_A) = \lim_{n\to\infty} \mathrm{E}\left(\mathrm{E}(\zeta_n \mid \mathcal{G})\mathbb{1}_A\right) = \lim_{n\to\infty} \mathrm{E}\left(\zeta_n \mathbb{1}_A\right) = 0,$$

which implies that $\eta = 0$ almost surely in view of its $\mathcal{G}$-measurability. As a result,

$$\mathrm{E}(\xi_n \mid \mathcal{G}) \to \mathrm{E}(\xi \mid \mathcal{G}), \; n \to \infty,$$

almost surely.

To prove the second statement, write from [A2.11] and the tower property of conditional expectation

$$\mathrm{E}\big(|\mathrm{E}(\xi_n \mid \mathcal{G}) - \mathrm{E}(\xi \mid \mathcal{G})|\big) \leq \mathrm{E}(|\xi_n - \xi|) \to 0, \; n \to \infty,$$

by the dominated convergence theorem. This means that $\mathrm{E}(\xi_n \mid \mathcal{G}) \to \mathrm{E}(\xi \mid \mathcal{G})$, $n \to \infty$, in $\mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$, consequently, in probability. □

A2.4.3. *Conditional probability*

Let, as above, $(\Omega, \mathcal{F}, \mathrm{P})$ be a complete probability space, and $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field.

DEFINITION A2.6.– *The conditional probability with respect to $\mathcal{G}$ is*

$$\mathrm{P}\{A \mid \mathcal{G}\} = \mathrm{E}(\mathbb{1}_A \mid \mathcal{G}), \quad A \in \mathcal{F}.$$

From the properties of conditional expectation it follows that $\mathrm{P}\{A \mid \mathcal{G}\} \geq 0$ a.s. $\mathrm{P}\{\Omega \mid \mathcal{G}\} = 1$ a.s. Also for any disjoint $A_1, A_2, \ldots$ we get from linearity and theorem A2.9 that

$$\mathrm{P}\left\{ \bigcup_{n\geq 1} A_n \,\middle|\, \mathcal{G} \right\} = \mathrm{E}\left(\mathbb{1}_{\bigcup_{n\geq 1} A_n} \,\middle|\, \mathcal{G}\right) = \lim_{N\to\infty} \mathrm{E}\left(\mathbb{1}_{\bigcup_{n=1}^{N} A_n} \,\middle|\, \mathcal{G}\right)$$

$$= \lim_{N\to\infty} \mathrm{E}\left(\sum_{n=1}^{N} \mathbb{1}_{A_n} \,\middle|\, \mathcal{G}\right) = \lim_{N\to\infty} \sum_{n=1}^{N} \mathrm{E}\left(\mathbb{1}_{A_n} \,\middle|\, \mathcal{G}\right) = \sum_{n\geq 1} \mathrm{P}\{A_n \mid \mathcal{G}\}$$

almost surely. However, the zero-probability event may depend on the sequence of events, and so the countable additivity of $\mathrm{P}\{A_n \mid \mathcal{G}\}$ may fail even for every $\omega \in \Omega$. However, one might be able to adjust conditional probabilities of individual sets in order to make it countably additive. This motivates the following definition.

DEFINITION A2.7.– *The function* $\mathrm{P}_{\mathcal{G}} \colon \mathcal{F} \times \Omega \to \mathbb{R}$ *is called a* regular conditional probability *with respect to $\mathcal{G}$ if*

– *for any $A \in \mathcal{F}$, $\mathrm{P}_{\mathcal{G}}(A, \omega) = \mathrm{P}\{A \mid \mathcal{G}\}(\omega)$ almost surely;*

– *for any $\omega \in \Omega$, $\mathrm{P}_{\mathcal{G}}(\cdot, \omega)$ is a probability measure on $\mathcal{F}$.*

A regular conditional probability allows us to compute the expectation as the usual Lebesgue integral.

THEOREM A2.10.– *Let $\mathrm{P}_{\mathcal{G}}$ be a regular conditional probability relation with respect to $\mathcal{G}$. And then, for any integrable ξ,*

$$\mathrm{E}(\xi \mid \mathcal{G})(\omega) = \int_{\Omega} \xi(\omega')\mathrm{P}_{\mathcal{G}}(d\omega', \omega) \qquad [A2.12]$$

almost surely.

PROOF.– Consider the family $\mathfrak{F}$ of integrable random variables ξ that satisfy [A2.12] almost surely. By the definition of regular conditional probability, $\mathbb{1}_A \in \mathfrak{F}$ for any A. Also, $\mathfrak{F}$ is linear and, due to the Lebesgue-dominated convergence theorem and theorem A2.9, is closed under taking limits of dominated sequences. Arguing as in the proof of theorem A2.3, we get that $\mathfrak{F}$ contains all integrable random variables. □

DEFINITION A2.8.– *Let* $\xi\colon \Omega \to \mathbb{R}$ *be a random variable. A* regular conditional distribution *of* ξ *with respect to* $\mathcal{G}$ *is a function* $P_{\xi|\mathcal{G}}\colon \mathcal{B}(\mathbb{R}) \times \Omega \to \mathbb{R}$ *such that*

– *for any* $A \in \mathcal{B}(\mathbb{R})$, $P_{\xi|\mathcal{G}}(B,\omega) = \mathrm{P}\{\xi \in B \mid \mathcal{G}\}(\omega)$ *almost surely;*

– *for any* $\omega \in \Omega$, $P_{\xi|\mathcal{G}}(\cdot,\omega)$ *is a probability measure on* $\mathcal{B}(\mathbb{R})$.

A regular conditional cumulative distribution function *of* ξ *with respect to* $\mathcal{G}$ *is a function* $F_{\xi|\mathcal{G}}\colon \mathbb{R} \times \Omega \to \mathbb{R}$ *such that*

– *for any* $x \in \mathbb{R}$, $F_{\xi|\mathcal{G}}(x,\omega) = \mathrm{P}\{\xi \le x \mid \mathcal{G}\}(\omega)$ *almost surely;*

– *for any* $\omega \in \Omega$, $F_{\xi|\mathcal{G}}(\cdot,\omega)$ *is a cumulative distribution function on* $\mathcal{B}(\mathbb{R})$.

THEOREM A2.11.– *Let* $\xi\colon \Omega \to \mathbb{R}$ *be a random variable. If* $P_{\xi|\mathcal{G}}$ *is a regular conditional distribution of* ξ *with respect to* $\mathcal{G}$, *then* $F_{\xi|\mathcal{G}}(x,\omega) = P_{\xi|\mathcal{G}}((-\infty,x],\omega)$ *is a regular conditional cumulative distribution function of* ξ *with respect to* $\mathcal{G}$. *Vice versa, if* $F_{\xi|\mathcal{G}}(x,\omega)$ *is a regular conditional cumulative distribution function of* ξ *with respect to* $\mathcal{G}$, *then the Lebesgue-Stieltjes measure* $P_{\xi|\mathcal{G}}(\cdot,\omega)$ *generated by* $F_{\xi|\mathcal{G}}(\cdot,\omega)$ *is a regular conditional distribution of* ξ *with respect to* $\mathcal{G}$. *If, additionally,* ξ *is integrable, then*

$$\mathrm{E}(\xi \mid \mathcal{G})(\omega) = \int_{\mathbb{R}} x P_{\xi|\mathcal{G}}(dx,\omega) = \int_{\mathbb{R}} x F_{\xi|\mathcal{G}}(dx,\omega)$$

almost surely.

PROOF.– The first statement is obvious. To prove the second, consider the family $\mathcal{A}$ of Borel sets A satisfying

$$P_{\xi|\mathcal{G}}(A,\omega) = \mathrm{P}\{\xi \in A \mid \mathcal{G}\}(\omega)$$

almost surely. Since $F_{\xi|\mathcal{G}}$ is a regular conditional cumulative distribution function of ξ, we have $(-\infty,x] \in \mathcal{A}$ for all $x \in \mathbb{R}$. Furthermore, from the properties of conditional expectation and of probability measure it follows that $\mathcal{A}$ is a λ-system. Since the intervals $(-\infty,x]$ form a π-system $\mathcal{P}$, by theorem A2.1, $\mathcal{A} \supset \sigma(\mathcal{P}) = \mathcal{B}(\mathbb{R})$.

The third statement is proved similarly to theorem A2.10. □

THEOREM A2.12.– *For any random variable* ξ, *a regular conditional distribution of* ξ *with respect to* $\mathcal{G}$ *exists.*

PROOF.– By theorem A2.11, it suffices to show the existence of a regular conditional distribution function. Let $F(q,\omega) = \mathrm{P}\{\xi \le q\}$, $q \in \mathbb{Q}$. Define

$$\Omega_+ = \left\{ \lim_{n\to+\infty} F(n,\omega) = 1 \right\},\ \Omega_- = \left\{ \lim_{n\to-\infty} F(n,\omega) = 0 \right\},$$

$$\Omega_{q_1,q_2} = \{F(q_1,\omega) \le F(q_2,\omega)\},\ q_1,q_2 \in \mathbb{Q},\ q_1 < q_2,$$

$$\Omega_q = \left\{ \lim_{\substack{r\to q\\ r\in\mathbb{Q}, r>q}} F(r,\omega) = F(q,\omega) \right\},\ q \in \mathbb{Q}.$$

From theorem A2.9 it follows that $\mathrm{P}\{\Omega_+\} = \mathrm{P}\{\Omega_-\} = \mathrm{P}\{\Omega_q\} = 1$ for all $q \in \mathbb{Q}$, from theorem A2.8, that $\mathrm{P}\{\Omega_{q_1,q_2}\} = 1$ for all rational $q_1 < q_2$. Set

$$\Omega' = \Omega_+ \cap \Omega_- \cap \left(\bigcap_{q\in\mathbb{Q}} \Omega_q \right) \cap \left(\bigcap_{\substack{q_1,q_2\in\mathbb{Q}\\ q_1<q_2}} \Omega_{q_1,q_2} \right)$$

and

$$F_{\xi|\mathcal{G}}(x,\omega) = \begin{cases} \inf_{q\ge x, q\in\mathbb{Q}} F(q,\omega), & \omega \in \Omega' \\ \mathbb{1}_{[0,+\infty)}(x), & \omega \notin \Omega'. \end{cases}$$

It is easy to see that $F_{\xi|\mathcal{G}}$ is a cumulative distribution function. From theorem A2.9 it follows that for any $x \in \mathbb{R}$,

$$F_{\xi|\mathcal{G}}(x,\omega) = \lim_{\substack{q\to x\\ q\in\mathbb{Q}, q\ge x}} \mathrm{P}\{\xi \le q \mid \mathcal{G}\} = \mathrm{P}\{\xi \le x \mid \mathcal{G}\}$$

almost surely, concluding the proof. □

REMARK A2.3.– As it was mentioned in remark A2.2, any separable metric space is Borel isomorphic to a Borel subset of $\mathbb{R}$. Therefore, the above theorem is also valid for a random variable taking values in a separable metric space.

A2.5. Stationary sequences and ergodic theorems

Consider the probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Let F be a measurable transformation $(\Omega, \mathcal{F})$ in $(\Omega, \mathcal{F})$. For any $A \in \mathcal{F}$ denote $F^{-1}A$ the pre-image of A transformation F.

DEFINITION A2.9.– *Transformation F preserves measure* P *if for any $A \in \mathcal{F}$*

$$\mathrm{P}\{F^{-1}A\} = \mathrm{P}\{A\}.$$

A set A is called F-invariant if $\mathrm{P}\left\{(F^{-1}A) \triangle A\right\} = 0$. It follows from the properties of pre-images that the family $\mathcal{G}$ of invariant sets is a σ-field.

For $k \in \mathbb{N}$ we denote F^k the kth power of transformation F.

THEOREM A2.13.– **(Ergodic Birkhoff–Khinchin theorem)** *Let F be a measurable transformation $(\Omega, \mathcal{F})$ into $(\Omega, \mathcal{F})$ preserving measure* P. *Also let $\xi(\omega)$ be an integrable random variable. Then, there exists a limit with probability* 1,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(F^k\omega) = \overline{\xi}(\omega),$$

where $\overline{\xi} = \mathrm{E}(\xi \mid \mathcal{G})$, and where $\mathcal{G}$ is a σ-field of invariant sets.

DEFINITION A2.10.– *Probability measure* P *is said to be* ergodic *for transformation F, if any F-invariant set has probability measure* 0 *or* 1.

DEFINITION A2.11.– *Stochastic process $X = \{X_t, t \in \mathbb{T}\}$, where $\mathbb{T} = \mathbb{Z}^+$ or $\mathbb{R}^+$, is called stationary (in the narrow sense) if for any $t_1, t_2, \dots, t_n \in \mathbb{T}$ and $h > 0$ such that $t_1 + h, t_2 + h, \dots, t_n + h \in \mathbb{T}$*

$$(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}) \overset{\mathrm{d}}{=} (X_{t_1}, X_{t_2}, \dots, X_{t_n}),$$

where $\overset{\mathrm{d}}{=}$ means the equality in distribution.

DEFINITION A2.12.– *Square-integrable real-valued process $X = \{X_t, t \in \mathbb{T}\}$ is called stationary (in the wide sense) if $\mathrm{E}X_t = m$ (some constant value), and*

$$\mathrm{E}X_tX_s = \mathrm{E}X_{t+h}X_{s+h} \textit{ for any } s, t, s+h, t+h \in \mathbb{T},\ h > 0.$$

REMARK A2.4.– If X is stationary in the wide sense, then its covariance function

$$r(s,t) := \mathrm{E}(X_s - \mathrm{E}X_s)(X_t - \mathrm{E}X_t) = \mathrm{E}(X_0 - m)(X_{t-s} - m).$$

Therefore, we can introduce the function $R(t) = r(0,t)$ so that $R(t) = \mathrm{E}(X_s - m)(X_{t+s} - m)$ for any $t, s, t+s \in \mathbb{T}$.

REMARK A2.5.– Let the process $X = \{X_t, t \in \mathbb{T}\}$ be Gaussian. Since finite-dimensional distributions are uniquely determined by the mean $\mathrm{E}X_t$ and the covariance function $r(t,s) = \mathrm{E}(X_s - \mathrm{E}X_s)(X_t - \mathrm{E}X_t)$, the wide-sense and the narrow-sense stationarities of X are equivalent. Thus, we shall call the Gaussian process X stationary if $\mathrm{E}X_t = m$ and $r(t,s) = r(t+h, s+h)$, $t, s, t+h, s+h \in \mathbb{T}$.

Let us return to the ergodic theorem. Consider a Gaussian stationary sequence $X = \{X_n, n \geq 0\}$. We can consider it on the canonical probability space $(\Omega, \mathcal{F}, \mathrm{P})$, where $\Omega = \mathbb{R}^{\mathbb{N}}$, $\mathcal{F} = \mathbb{K}_{\text{cyl}}$, the σ-algebra of cylinder sets. In this case $X_n(\omega) = \omega(n)$, and we can consider the transformation $F : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$, which is the unit time shift, $\omega(n) \to \omega(n+1)$, and so F^k if the shift $\omega(n) \to \omega(n+k)$. The stationarity of the process is reflected in the invariance of P with respect to F, i.e. $\mathrm{P}F^{-1} = \mathrm{P}$.

Assume that the measure P on $(\mathbb{R}^{\mathbb{N}}, \mathbb{K}_{\text{cyl}})$ has a mixing property in the following form: for any $A, B \in \mathbb{K}_{\text{cyl}}$ with $\mathrm{P}\{B\} > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left\{F^{-n}A \mid B\right\} = \mathrm{P}\{A\}. \qquad \text{[A2.13]}$$

LEMMA A2.6.– *If condition* [A2.13] *holds, then probability measure* P *is ergodic for transformation* F.

PROOF.– Let C be an F-invariant set with $\mathrm{P}\{C\} \neq 0$. Then, we put $A = B = C$ in [A2.13] and get that

$$\lim_{n\to\infty} \mathrm{P}\left\{F^{-n}A \mid B\right\} = \lim_{n\to\infty} \mathrm{P}\left\{F^{-n}C \mid C\right\} = \mathrm{P}\left\{C \mid C\right\} = \mathrm{P}\left\{C\right\}.$$

But $\mathrm{P}\left\{C \mid C\right\} = 1$, so $\mathrm{P}\{C\} = 1$, as required. □

Now let F preserve the measure P.

THEOREM A2.14.– *Probability measure is* F*-ergodic if and only if for any integrable random variable* ξ

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(F^k\omega) = \mathrm{E}\xi. \qquad \text{[A2.14]}$$

PROOF.– *Necessity*. Let probability measure P be F-ergodic. Since F preserves P, it follows from Birkhoff-Khinchin theorem that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \xi(F^k\omega) = \overline{\xi}(\omega),$$

$\mathrm{E}\overline{\xi}(\omega) = \mathrm{E}\xi(\omega)$. Since the random variable $\overline{\xi}(\omega)$ is F-invariant, symmetric difference of the sets

$$F^{-1}\left\{\omega : \overline{\xi}(\omega) < x\right\} = \left\{\omega : \overline{\xi}(F\omega) < x\right\}$$

and $\{\omega : \overline{\xi}(\omega) < x\}$ has P-measure 0 for any $x \in \mathbb{R}$. This means that any set $\{\omega : \overline{\xi}(\omega) < x\}$ is F-invariant therefore it follows from ergodicity of P that $\mathrm{P}\{\omega : \overline{\xi}(\omega) < x\} = 0$ otherwise 1. This means that $\overline{\xi}$ is a constant, and from Birkhoff-Khinchin theorem $\overline{\xi} = \mathrm{E}\xi$.

Sufficiency. Let [A2.14] hold, and substitute $\xi(\omega) = \mathbb{1}_A(\omega)$ for some event A. Then,

$$\lim_{n\to\infty} \sum_{k=0}^{n-1} \mathbb{1}_A(F^k\omega) = \mathrm{P}\{A\}.$$

Now, assume that there exists F-invariant set B with $0 < \mathrm{P}\{B\} < 1$. Then, the sets B, FB, F^2B, ... differ from each other on the sets of P-measure 0, and consequently

$$\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_A(F^k\omega) = \mathbb{1}_A(\omega)$$

and is not a constant. The obtained contradiction implies that P is ergodic for the transformation F. □

Now, let $X = \{X_n, n \geq 0\}$ be a Gaussian stationary sequence, $\mathrm{E}X_n = m$, $\mathrm{E}(X_0 - m)(X_n - m) = R(n)$.

THEOREM A2.15.– *If* $R(n) \to 0$ *as* $n \to \infty$, *then* $\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} X_k = m$ *a.s.*

PROOF.– According to lemma A2.6 and theorem A2.13, it is sufficient to establish the mixing condition [A2.13].

Equality [A2.13] is a partial case of the following equality

$$\lim_{n\to\infty} \mathrm{E}Y_n\eta = \mathrm{E}Y_0\mathrm{E}\eta, \quad \text{[A2.15]}$$

where $Y_n = \xi(F^n\omega)$, and ξ, η are square-integrable random variables. To prove [A2.15], it is sufficient to establish that for bounded functions $f(x_1, \ldots, x_p)$ and $g(x_1, \ldots, x_p)$, for any $p \geq 1$

$$\mathrm{E}f(X_n, \ldots, X_{n+p})g(X_0, \ldots, X_p) \to \mathrm{E}f(X_0, \ldots, X_p)g(X_0, \ldots, X_p), \quad \text{[A2.16]}$$

and then apply the approximation procedure. Establish [A2.16] for $p = 1$; for $p > 1$ the proof is similar. For technical simplicity, let $m = 0$. We have that the joint density of X_n and X_0 equals

$$\begin{aligned} p_{n,0}(x,y) &= \frac{1}{2\pi\sqrt{D}} \exp\left\{-\frac{1}{2}\left(\frac{R(0)}{D}(x^2+y^2) - \frac{2R(n)}{D}xy\right)\right\} \\ &\to \frac{1}{2\pi R(0)} \exp\left\{-\frac{1}{2}\frac{x^2}{R(0)} - \frac{1}{2}\frac{y^2}{R(0)}\right\}, \end{aligned} \quad \text{[A2.17]}$$

where $D = R^2(0) - R^2(n)$. It follows from [A2.17] that $p_{n,0}(x,y) \to p_0(x)p_0(y)$, when [A2.16] follows for $p = 1$. □

A2.6. Auxiliary martingale inequalities and decompositions

THEOREM A2.16.– (**Krickeberg decomposition for martingales**) *Let $X = \{X_n, \mathcal{F}_n, n \geq 0\}$ be a martingale with $\sup_{n\geq 0} \mathrm{E}|X_n| < \infty$. Then X can be decomposed as follows:*

$$X_n = X'_n - X''_n, \quad \text{[A2.18]}$$

where X' and X'' are non-negative martingales and

$$\sup_{n\geq 0} \mathrm{E}|X_n| \leq \mathrm{E}X'_n + \mathrm{E}X''_n = \mathrm{E}X'_0 + \mathrm{E}X''_0.$$

PROOF.– Condition $\sup_{n\geq 0} \mathrm{E}|X_n| < \infty$ together with theorem 5.5 supplies existence of the limit $X_\infty = \lim_{n\to\infty} X_n$ a.s. and $\mathrm{E}|X_\infty| < \infty$, while $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$ according to theorem 5.7. If we decompose $X_\infty = X_\infty^+ - X_\infty^-$, then X_∞^+ and X_∞^- are integrable random variables, $X_n = \mathrm{E}(X_\infty^+|\mathcal{F}_n) - \mathrm{E}(X_\infty^-|\mathcal{F}_n)$, and $X'_n = \mathrm{E}(X_\infty^+|\mathcal{F}_n)$, $X''_n = \mathrm{E}(X_\infty^-|\mathcal{F}_n)$ both are non-negative martingales. Therefore, we have decomposition [A2.18]. Moreover,

$$\begin{aligned} \mathrm{E}|X_n| &\leq \mathrm{E}|X'_n| + \mathrm{E}|X''_n| = \mathrm{E}X'_n + \mathrm{E}X''_n = \mathrm{E}X'_0 + \mathrm{E}X''_0 \\ &= \mathrm{E}X_\infty^+ + X_\infty^- = \mathrm{E}|X_\infty|, \end{aligned}$$

and the proof follows. □

REMARK A2.6.– We can apply a similar reasoning to the martingale $X = \{X_n, \mathcal{F}_n, 0 \leq n \leq N\}$ and get the same decomposition but with X_N instead of X_∞. We shall say that such a decomposition is obtained with respect to the moment N.

THEOREM A2.17.– *Let $\{X_n, n \geq 0\}$ be a martingale, $\{\varphi_n, n \geq 1\}$ be a predictable process, $\varphi_0 = 0$, and $|\varphi_n| \leq 1$ a.s. Consider the martingale transformation of the following form: $S_0^X = 0$, $S_n^X = \sum_{k=1}^n \varphi_k(X_k - X_{k-1}), n \geq 1$. Then* *i) For any $n \geq 1$ and any $\lambda > 0$*

$$\lambda \mathrm{P}\left\{\max_{0\leq k\leq n} \left|S_k^X\right| > \lambda\right\} \leq 22 \max_{0\leq k\leq n} \mathrm{E}|X_k|. \qquad \text{[A2.19]}$$

ii) For any $n \geq 1$ and $p \geq 1$

$$\mathrm{E}\left|S_n^X\right|^p \leq c_p \mathrm{E}|X_n|^p, \qquad \text{[A2.20]}$$

where c_p depends only on p and does not depend on n.

PROOF.– i) Let X be a non-negative martingale. Consider the non-negative bounded supermartingale $Z_n = X_n \wedge \lambda$, $\lambda > 0$ (see example 5.8, (ii)). Since Z_k and X_k, $1 \leq k \leq n$, coincide on the set $\{\omega : \max_{0\leq k\leq n} X_k \leq \lambda\}$, we have that

$$\lambda \mathrm{P}\left\{\max_{0\leq k\leq n} |S_k^X| > \lambda\right\} \leq \lambda \mathrm{P}\left\{\max_{0\leq k\leq n} X_k > \lambda\right\} + \lambda \mathrm{P}\left\{\max_{0\leq k\leq n} |S_k^Z| > \lambda\right\}.$$

According to remark 5.9,

$$\lambda \mathrm{P}\left\{\max_{0\leq k\leq n} |X_k| > \lambda\right\} \leq \mathrm{E}X_0. \qquad \text{[A2.21]}$$

Further, S_n^Z is a supermartingale as an integral transformation of a supermartingale. Consider the Doob decomposition of Z: $Z_n = M_n - A_n$, where M is a martingale, and A is a non-decreasing predictable bounded process with $A_0 = 0$. Note that consequently, A and M are non-negative processes with bounded increments,

$$A_n - A_{n-1} = -\mathrm{E}(Z_n - Z_{n-1}|\mathcal{F}_{n-1}) \leq 2\lambda.$$

So, all processes involved are square integrable. Furthermore,

$$S_n^Z = S_n^M - S_n^A, \text{ and } |S_n^A| \leq \sum_{k=1}^n |\varphi_k|(A_k - A_{k-1}) \leq A_n.$$

Therefore, $\max_{0\leq k\leq n} |S_k^Z| \leq \max_{0\leq k\leq n} |S_k^M| + A_n$, and for any $\alpha \in (0,1)$

$$\lambda \mathrm{P}\left\{\max_{0\leq k\leq n} |S_k^Z| > \lambda\right\} \leq \lambda \mathrm{P}\{A_n > \alpha\lambda\} + \lambda \mathrm{P}\left\{\max_{0\leq k\leq n} |S_k^M| > (1-\alpha)\lambda\right\}.$$

Let $\alpha = \frac{1}{2}$. Since $\mathrm{E}A_n = \mathrm{E}M_n - \mathrm{E}Z_n \leq \mathrm{E}M_n = \mathrm{E}M_0$ we have that

$$\lambda\,\mathrm{P}\left\{A_n > \frac{\lambda}{2}\right\} \leq 2\mathrm{E}M_0 = 2\mathrm{E}Z_0 \leq 2\mathrm{E}X_0. \qquad [A2.22]$$

Further, since M is a square integrable martingale, S^M is a square integrable martingale as well. Therefore, according to theorem 5.11,

$$\lambda\,\mathrm{P}\left\{\max_{0\leq k\leq n} \left|S_k^M\right| > \frac{\lambda}{2}\right\} \leq \frac{4}{\lambda}\mathrm{E}\left|S_n^M\right|^2,$$

and

$$\mathrm{E}|S_n^M|^2 = \mathrm{E}\sum_{k=0}^{n-1} \varphi_k^2 (M_{k+1} - M_k)^2 \leq \mathrm{E}[M]_n.$$

Note that

$$\mathrm{E}(M_{k+1} - M_k)^2 = \mathrm{E}(Z_{k+1} - Z_k)^2 + 2\mathrm{E}(Z_{k+1} - Z_k)(A_{k+1} - A_k)$$
$$+\mathrm{E}(A_{k+1} - A_k)^2 = \mathrm{E}(Z_{k+1} - Z_k)^2 - \mathrm{E}(A_{k+1} - A_k)^2,$$

because

$$\mathrm{E}(Z_{k+1} - Z_k)(A_{k+1} - A_k) = \mathrm{E}\left(\mathrm{E}(Z_{k+1} - Z_k)|\mathcal{F}_k\right)(A_{k+1} - A_k)$$
$$= -\mathrm{E}(A_{k+1} - A_k)^2.$$

So,

$$\mathrm{E}|S_n^M|^2 \leq \mathrm{E}[Z]_n - \mathrm{E}[A]_n \leq \mathrm{E}[Z]_n.$$

Furthermore,

$$(Z_k - Z_{k-1})^2 = Z_k^2 - Z_{k-1}^2 + 2Z_{k-1}(Z_{k-1} - Z_k),$$

therefore,

$$\mathrm{E}[Z]_n \leq \mathrm{E}Z_n^2 + 2\sum_{k=1}^{n-1} \mathrm{E}\left(Z_{k-1}\mathrm{E}(Z_{k-1} - Z_k|\mathcal{F}_{k-1})\right)$$
$$= \mathrm{E}Z_n^2 + 2\sum_{k=1}^{n-1} \mathrm{E}Z_{k-1}(A_k - A_{k-1}) \leq 2\lambda\,\mathrm{E}(Z_n + A_n) \qquad [A2.23]$$
$$= 2\lambda\,\mathrm{E}M_n = 2\lambda\,\mathrm{E}M_0 = 2\lambda\,\mathrm{E}Z_0 \leq 2\,\lambda\,\mathrm{E}X_0.$$

Finally, we get from [A2.21]–[A2.23] that for any $\lambda > 0$

$$\lambda\,\mathrm{P}\left\{\max_{0\le k\le n}|S_k^X| > \lambda\right\} \le \mathrm{E}X_0 + 2\mathrm{E}X_0 + 8\mathrm{E}X_0 = 11\mathrm{E}X_0.$$

Therefore, for non-negative martingales we have proved a stronger result than [A2.19], namely, with constant 11 instead of 22.

Now, let X be an arbitrary martingale, and let $X_n = X_n' - X_n''$ be its Krickeberg decomposition constructed with respect to the moment N (theorem A2.16 and remark A2.6). Then $\mathrm{E}|X_n| = \mathrm{E}X_n' + \mathrm{E}X_n'' = \mathrm{E}X_0' + \mathrm{E}X_0''$, and

$$\lambda\,\mathrm{P}\left\{\max_{0\le k\le n}|S_k^X| > \lambda\right\} \le \lambda\,\mathrm{P}\left\{\max_{0\le k\le n}|S_k^{X'}| > \frac{\lambda}{2}\right\}$$
$$+\,\lambda\,\mathrm{P}\left\{\max_{0\le k\le n}|S_k^{X''}| > \frac{\lambda}{2}\right\}$$
$$\le 22(\mathrm{E}X_0' + \mathrm{E}X_0'') = 22\mathrm{E}|X_n| \le 22\sup_{0\le k\le n}\mathrm{E}|X_k|.$$

ii) Now we use the Marcinkiewicz interpolation theorem (theorem A2.3). Assume that the right-hand side of [A2.20] is bounded in $n \ge 1$. Then, X is a uniformly integrable martingale and $X_n = \mathrm{E}(X_\infty|\mathcal{F}_n)$ (see theorem 5.7). Denote by $T_n(X_\infty)$ the transformation of the form $T_n(X_\infty) = S_n^X$ with some fixed predictable process $\{\varphi_k, k \ge 1\}$, $\varphi_0 = 0$, $|\varphi_k| \le 1$, $k \ge 1$ a.s. Then T_n is a linear, and consequently, sub-additive transformation. Moreover, note that $\{|X_k|, k \ge 1\}$ is a submartingale, therefore $\mathrm{E}|X_\infty| \ge \mathrm{E}|X_k|$, $k \ge 1$. Then, it follows from [A2.19] that

$$\mathrm{P}\left\{|T_n(X_\infty)| > \lambda\right\} \le \frac{1}{\lambda}\mathrm{E}|X_\infty|.$$

Furthermore,

$$\mathrm{P}\left\{|T_n(X_\infty)| > \lambda\right\} \le \frac{1}{\lambda^2}\mathrm{E}|T_n(X_\infty)|^2 = \frac{1}{\lambda^2}\mathrm{E}|S_n^X|^2$$
$$= \frac{1}{\lambda^2}\mathrm{E}\sum_{k=1}^n \varphi_k^2(X_k - X_{k-1})^2 \le \frac{1}{\lambda^2}\mathrm{E}\sum_{k=1}^n (X_k - X_{k-1})^2 \le \frac{1}{\lambda^2}\mathrm{E}X_n^2.$$

This means that all assumptions of theorem A2.3 are fulfilled with $C_1 = 22$ and $C_2 = 1$, and it follows that the inequality [A2.20] holds for any $1 < p \le 2$. Now consider $p > 2$. Let $p^{-1} + q^{-1} = 1$. Then $1 < q < 2$, consequently, for any random variable η

$$\mathrm{E}|X_\infty T_n(\eta)| \le \left(\mathrm{E}|X_\infty|^p\right)^{1/p}\left(\mathrm{E}|T_n(\eta)|^q\right)^{1/q}$$
$$\le C_q\left(\mathrm{E}|X_\infty|^p\right)^{1/p}\left(\mathrm{E}|\eta|^q\right)^{1/q}. \qquad \text{[A2.24]}$$

Further, assume that $\mathrm{E}|X_\infty T_n(\eta)| < \infty$. Then for $X_k = \mathrm{E}(X_\infty|\mathcal{F}_k)$ and $Y_k = \mathrm{E}(\eta|\mathcal{F}_k)$ we have that

$$\begin{aligned}\mathrm{E}(X_\infty T_n(\eta)) &= \mathrm{E}\left(X_\infty \sum_{k=1}^n \varphi_k(Y_k - Y_{k-1})\right) \\ &= \mathrm{E}\left(X_n \sum_{k=1}^n \varphi_k(Y_k - Y_{k-1})\right) = \sum_{k=1}^n \mathrm{E}\varphi_k X_k(Y_k - Y_{k-1}) \\ &= \sum_{k=1}^n (\mathrm{E}\varphi_k X_k \eta - \mathrm{E}\varphi_k X_k Y_{k-1}) = \sum_{k=1}^n (\mathrm{E}\varphi_k X_k \eta - \mathrm{E}\varphi_k X_{k-1} Y_{k-1}) \\ &= \sum_{k=1}^n (\mathrm{E}\varphi_k X_k \eta - \mathrm{E}\varphi_k X_{k-1}\eta) = \sum_{k=1}^n \mathrm{E}\varphi_k (X_k - X_{k-1})\eta \\ &= \mathrm{E}(T_n(X_\infty)\eta). \end{aligned} \qquad \text{[A2.25]}$$

We get from [A2.24] and [A2.25] that

$$|\mathrm{E}(T_n(X_\infty)\eta)| \le C_q \big(\mathrm{E}|X_\infty|^p\big)^{1/p} \big(\mathrm{E}|\eta|^q\big)^{1/q}. \qquad \text{[A2.26]}$$

Therefore, taking supremum over η with $\mathrm{E}|\eta|^q = 1$ in both sides of [A2.26], we get

$$\sup_{\eta:\mathrm{E}|\eta|^q=1} |\mathrm{E}(T_n(X_\infty)\eta)| \le C_q \big(\mathrm{E}|X_\infty|^p\big)^{1/p}. \qquad \text{[A2.27]}$$

It follows from lemma A2.6 that

$$\sup_{\eta:\mathrm{E}|\eta|^q=1} \big|\mathrm{E}\big(T_n(X_\infty)\eta\big)\big| = \big(\mathrm{E}|T_n(X_\infty)|^p\big)^{1/p}. \qquad \text{[A2.28]}$$

The proof follows now from [A2.27] and [A2.28]. □

For any process X and any $n \ge 0$ denote $X_n^* := \max_{0\le k\le n} |X_k|$.

LEMMA A2.7.– *Let $M = \{M_n, \mathcal{F}_n, n \ge 0\}$ be a martingale, for which $|M_{n+1} - M_n| \le Q_n$, where Q_n is a $\mathcal{F}_n$-measurable random variable, $n \ge 1$. Then, for any $n \ge 1$*

$$\mathrm{E}[M]_n^{1/2} \le 3\mathrm{E} \max_{0\le k\le n} |M_k| + \mathrm{E} \max_{0\le k\le n} Q_k = 3\mathrm{E}|M_n|^* + \mathrm{E}Q_n^*, \qquad \text{[A2.29]}$$

$$\mathrm{E}\max_{0\le k\le n}|M_k| \le 3\mathrm{E}[M]_n^{1/2} + \mathrm{E}\max_{0\le k\le n} Q_k = 3\mathrm{E}[M]_n^{1/2} + \mathrm{E}Q_n^*. \qquad \text{[A2.30]}$$

PROOF.– We only prove [A2.29], and [A2.30] is proved similarly. Denote for any $x > 0$

$$\tau_x = \inf\{n \ge 0 : |M_n| + Q_n > x\}.$$

Then on the set $\{\omega : \tau_x(\omega) > k\}$ we have that $|M_{\tau_x\wedge k}| = |M_k| \le x$, and on the set $\{\omega : \tau_x(\omega) = k\}$ we have that

$$|M_k| \le |M_{k-1}| + |M_k - M_{k-1}| \le |M_{k-1}| + W_{k-1} \le x.$$

Therefore, on the set $\{\omega : \tau_x > 0\}$ we have for any $k \ge 1$ that

$$|M_{\tau_x\wedge k}| \le x. \qquad \text{[A2.31]}$$

Obviously, we can estimate the following probability from above:

$$\mathrm{P}\left\{[M]_n > x^2\right\} \le \mathrm{P}\{\tau_x \le n\} + \mathrm{P}\left\{\tau_x > n, [M]_n > x^2\right\}.$$

Now, on one hand,

$$\mathrm{P}\{\tau_x \le n\} = \mathrm{P}\left\{\max_{0\le k\le n}(|M_k| + W_k) > x\right\},$$

and on the other hand, it follows from Doob's optional theorem that

$$\mathrm{E}\big(M_{k\wedge\tau_x} \,\big|\, \mathcal{F}_{(k-1)\wedge\tau_x}\big) = M_{(k-1)\wedge\tau_x}.$$

Therefore,

$$\begin{aligned}
&\mathrm{P}\left\{\tau_x > n, [M]_n > x^2\right\} \le \mathrm{P}\left\{\tau_x > n, [M]_{\tau_x\wedge n} > x^2\right\}\\
&\le x^{-2}\mathrm{E}\mathbb{1}_{\tau_x>n}[M]_{\tau_x\wedge n} \le x^{-2}\mathrm{E}\mathbb{1}_{\tau_x>0}[M]_{\tau_x\wedge n}\\
&= x^{-2}\mathrm{E}\mathbb{1}_{\tau_x>0}\sum_{k=1}^{\tau_x\wedge n}(M_k - M_{k-1})^2\\
&= x^{-2}\mathrm{E}\mathbb{1}_{\tau_x>0}\sum_{k=1}^{n}(M_{k\wedge\tau_x} - (M_{(k-1)\wedge\tau_x})^2\\
&= x^{-2}\mathrm{E}\mathbb{1}_{\tau_x>0}\sum_{k=1}^{n}(M_{k\wedge\tau_x}^2 - 2M_{k\wedge\tau_x}M_{(k-1)\wedge\tau_x} + M_{(k-1)\wedge\tau_x}^2)\\
&\le x^{-2}\mathrm{E}\mathbb{1}_{\tau_x>0}M_{n\wedge\tau_x}^2 \le x^{-2}\mathrm{E}(|M_{n\wedge\tau_x}|\wedge x)^2 \le x^{-2}\mathrm{E}((\max_{0\le k\le n}|M_k|)\wedge x)^2,
\end{aligned}$$

where the inequality $|M_{n\wedge\tau_x}| \leq |M_{n\wedge\tau_x}| \wedge x$ follows from [A2.31].

Therefore, if we set $0 \cdot \infty = 0$, then

$$\begin{aligned}
&\mathrm{E}[M]_n^{1/2} = \int_0^\infty \mathrm{P}\left\{[M]_n > x^2\right\} dx \\
&\leq \int_0^\infty \mathrm{P}\left\{\max_{0\leq k\leq n}(|M_k| + Q_k) > x\right\} dx + \mathrm{E}\int_0^\infty \max_{0\leq k\leq n}\frac{(|M_k|\wedge x)^2}{x^2} dx \\
&\leq \mathrm{E}\max_{0\leq k\leq n}|M_k| + \mathrm{E}\max_{0\leq k\leq n} Q_k + \mathrm{E}\int_0^{M_n^*} dx + \mathrm{E}\int_{M_n^*}^\infty \frac{(M_n^*)^2}{x^2} dx \\
&\leq \mathrm{E}M_n^* + \mathrm{E}Q_n^* + \mathrm{E}M_n^* + \mathrm{E}\big((M_n^*)^2 \cdot (M_n^*)^{-1}\big) = 3\mathrm{E}M_n^* + \mathrm{E}Q_n^*. \qquad \square
\end{aligned}$$

Now we introduce the Davis decomposition of the martingale. As before, we use the notation $(\Delta M)_n^* := \max_{1\leq k\leq n}|M_k - M_{k-1}|$,

$$\mathbb{1}_n := \mathbb{1}_{|\Delta M_n|\leq 2(\Delta M)_{n-1}^*}, \quad \overline{\mathbb{1}}_n := \mathbb{1} - \mathbb{1}_n.$$

Let $M = \{M_n, \mathcal{F}_n, n \geq 0\}$ be a martingale, $M_0 = 0$.

DEFINITION A2.13.– *The Davis decomposition of a martingale M is a decomposition of the form*

$$M_n = M_n' + M_n'', \qquad \text{[A2.32]}$$

where

$$\begin{aligned}
&M_0' = 0, \ \ \Delta M_n' = M_n' - M_{n-1}' = Q_n' - \mathrm{E}(Q_n'|\mathcal{F}_{n-1}), \\
&Q_n' = \Delta M_n \mathbb{1}_n, \ \ M_0'' = 0, \ \ \Delta M_n'' = M_n'' - M_{n-1}'' \qquad \text{[A2.33]} \\
&\quad = Q_n'' + \mathrm{E}(Q_n'|\mathcal{F}_{n-1}), \ \ Q_n'' = \Delta M_n \overline{\mathbb{1}}_n.
\end{aligned}$$

Note that, consequently, M_n' and M_n'' in the decomposition [A2.32] are martingales.

LEMMA A2.8.– *The components of the decomposition admit the following upper bounds:*

$$|\Delta M_n'| \leq 4(\Delta M)_{n-1}^*, \qquad \text{[A2.34]}$$

$$\sum_{k=1}^{n} |\Delta M_k^{''}| \leq \sum_{k=1}^{n} |Q_k^{''}| + \sum_{k=1}^{n} \mathrm{E}\big(|Q_k^{''}| \,\big|\, \mathcal{F}_{k-1}\big)$$

$$\leq 2(\Delta M)_n^* + \sum_{k=1}^{n} \mathrm{E}\big(|Q_k^{''}| \,\big|\, \mathcal{F}_{k-1}\big). \qquad \text{[A2.35]}$$

It follows from [A2.35] *that*

$$\mathrm{E} \sum_{k=1}^{n} |\Delta M_k^{''}| \leq 4\mathrm{E}(\Delta M)_n^*. \qquad \text{[A2.36]}$$

PROOF.– The upper bound [A2.34] immediately follows from [A2.32] and [A2.33] because $|Q_n'| \leq 2(\Delta M)_{n-1}^*$. Further, $\mathrm{E}(Q_n'|\mathcal{F}_{n-1}) = -\mathrm{E}(Q_n^{''}|\mathcal{F}_{n-1})$, whence the first inequality in [A2.35] follows. The second inequality can be checked as follows: on the set $\{|\Delta M_n| > 2(\Delta M)_{n-1}^*\}$ we have that

$$|\Delta M_n| + 2(\Delta M)_{n-1}^* \leq 2|\Delta M_n| \leq 2(\Delta M)_n^*.$$

Therefore,

$$|Q_n^{''}| = |\Delta M_n| \mathbb{1}_{\Delta M_n > 2(\Delta M)_{n-1}^*} \leq 2\big((\Delta M)_n^* - (\Delta M)_{n-1}^*\big),$$

and

$$\sum_{k=1}^{n} |Q_k^{''}| \leq 2(\Delta M)_n^*. \qquad \text{[A2.37]}$$

Inequality [A2.36] is a straightforward consequence of [A2.35] and [A2.37]. □

A2.7. Strong laws of large numbers

In this section, we formulate different conditions for the classical strong law of large numbers (SLLN), i.e. SLLN for the sequences of random variables. We consider a sequence $\{X_n, n \geq 1\}$ of random variables, $X_n \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$ and $\mathrm{E}X_n = 0$, $n \geq 1$ (centered random variables). Denote $S_n = \sum_{i=1}^n X_i$, $n \geq 1$. The first result is a standard SLLN for iid sequence; we formulate it without centering assumption.

THEOREM A2.18.– *Let* $\{X_n, n \geq 1\}$ *be an iid sequence,* $X_n \in \mathcal{L}_1(\Omega, \mathcal{F}, \mathrm{P})$, $\mathrm{E}X_n = m$, $n \geq 1$. *Then*

$$\frac{S_n}{n} \to m \;\; a.s. \; as \;\; n \to \infty.$$

The following result was proved by Kolmogorov [KOL 30].

THEOREM A2.19.– *Let* $\{X_n, n \geq 1\}$ *be a sequence of independent random variables,* $\mathrm{E}X_n = 0$*,* $X_n \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$*,* $n \geq 1$*. If* $\sum_{n=1}^{\infty} \frac{\mathrm{E}X_n^2}{n^2} < \infty$*, then* $\frac{S_n}{n} \to 0$ *a.s. as* $n \to \infty$*.*

Now, consider the Marcinkiewicz-Zygmund result ([LOÈ 78]), generalized in the paper [KOR 14].

THEOREM A2.20.– *Let* $\{X_n, n \geq 1\}$ *be pairwise independent identically distributed centered random variables such that* $\mathrm{E}|X_1|^p < \infty$*, for some* $1 < p < 2$*, then*

$$\frac{S_n}{n^{1/p}} \to 0 \ \textit{a.s. as} \ n \to \infty.$$

Denote Ψ the class of functions $\psi : \mathbb{R}_+ \to \mathbb{R}_+ \backslash \{0\}$, such that $\psi(x)$ is non-decreasing on $[x_0, +\infty)$ for some $x_0 > 0$, and $\sum_{n=1}^{\infty} \frac{1}{n\varphi(n)} < \infty$. The examples of $\psi \in \Psi$ are $\psi(x) = x^\delta$, $\psi(x) = (\log x)^{1+\delta}$, $\delta > 0$.

Also, recall the notion of orthogonality: we say that the random variables $\{X_n, n \geq 1\}$ are orthogonal (uncorrelated), if $X_n \in \mathcal{L}_2(\Omega, \mathcal{F}, \mathrm{P})$ and

$$\mathrm{E}\left((X_i - \mathrm{E}X_i)(X_j - \mathrm{E}X_j)\right) = 0 \text{ for } i \neq j.$$

The next two results generalize SLLN to orthogonal random variables. The first result is one of the forms of Rademacher-Menchov theorem ([RAD 22, MEN 23]).

THEOREM A2.21.– *Let* $\{X_n, n \geq 1\}$ *be a sequence of centered uncorrelated random variables,* $\{a_n, n \geq 1\}$ *be a non-decreasing sequence,* $a_n > 0$*,* $a_n \to \infty$*,* $n \to \infty$*, and*

$$\sum_{n=1}^{\infty} \frac{\mathrm{E}X_n^2}{a_n^2} \log^2 n < \infty.$$

Then $\frac{S_n}{a_n} \to 0$ *a.s. as* $n \to \infty$*.*

The second result is another form of SLLN for uncorrelated random variables. It was proved by V.V. Petrov [PET 75].

THEOREM A2.22.– *Let* $\{X_n, n \geq 1\}$ *be a sequence of centered uncorrelated random variables. If* $\sum_{i=1}^{n} \mathrm{E}X_i^2 = O\left(\frac{n^2}{\psi(n) \log^2 n}\right)$ *for some function* $\psi \in \Psi$*, then* $\frac{S_n}{n} \to 0$ *a.s. as* $n \to \infty$*.*

A2.8. Fundamental martingale related to fractional Brownian motion

Our goal is to transform fractional Brownian motion with the help of some singular kernel in order to get a martingale. For this purpose consider the kernel

$$l_H(t,s) = C'_H s^{\frac{1}{2}-H}(t-s)^{\frac{1}{2}-H}\mathbb{1}_{0<s<t},$$

where $C'_H = \left(\frac{\Gamma(3-2H)}{2H\Gamma(\frac{3}{2}-H)^3\Gamma(\frac{1}{2}+H)}\right)^{\frac{1}{2}}$. For technical simplicity, let $H \in (\frac{1}{2},1)$. Then, by using the equality

$$\int_0^1 t^{-\mu}(1-t)^{-\mu}|x-t|^{2\mu-1}dt = B(\mu, 1-\mu),$$

for any $\mu \in (0,1), x \in (0,1)$ (see, for example, [NOR 99, lemma 2.2]), and denoting $\alpha = H - \frac{1}{2}$, we obtain that for any $t > 0$

$$\begin{aligned}
&\mathrm{E}\left|\int_0^t l_H(t,s)dB_s^H\right|^2 \\
&= \left(C'_H\right)^2 2H\alpha \int_0^t\int_0^t (t-u)^{-\alpha}(t-s)^{-\alpha}u^{-\alpha}s^{-\alpha}|u-s|^{2\alpha-1}du\,ds \\
&= t^{1-2\alpha}\left(C'_H\right)^2 2H\alpha \int_0^1 u^{-\alpha}(1-u)^{-\alpha} \qquad \text{[A2.38]}\\
&\quad\left(\int_0^1 (1-s)^{-\alpha}s^{-\alpha}|u-s|^{2\alpha-1}ds\right)du \\
&= t^{1-2\alpha}\left(C'_H\right)^2 2H\alpha B(\alpha,1-\alpha)B(1-\alpha,1-\alpha) \\
&= t^{1-2\alpha}\frac{\Gamma(2-2\alpha)\Gamma(\alpha)\Gamma(1-\alpha)^3}{\Gamma(1-\alpha)^3\Gamma(\alpha)\Gamma(2-2\alpha)} = t^{1-2\alpha} < \infty.
\end{aligned}$$

Therefore, the integral $\int_0^t l_H(t,s)dB_s^H$ is well defined. Further, similarly to [A2.39], for any $0 < t < t'$, we obtain that

$$\begin{aligned}
&\mathrm{E}\int_0^t l_H(t,s)dB_s^H \int_0^{t'} l_H(t',s)dB_s^H \\
&= \left(C'_H\right)^2 2H\alpha \int_0^t (t-u)^{-\alpha}u^{-\alpha} \\
&\quad\left(\int_0^{t'} (t'-s)^{-\alpha}s^{-\alpha}|u-s|^{2\alpha-1}ds\right)du \qquad \text{[A2.39]}\\
&= \left(C'_H\right)^2 2H\alpha t^{1-2\alpha}B(\alpha,1-\alpha)B(1-\alpha,1-\alpha) = t^{1-2\alpha}.
\end{aligned}$$

Evidently, $\int_0^t l_H(t,s)dB_s^H$ is a centered Gaussian process. Moreover, from [A2.9], we obtain that for any $0 < u < t \le u' < t'$

$$\mathrm{E}\left(\int_0^{t'} l_H(t',s)dB_s^H - \int_0^{u'} l_H(u',s)dB_s^H\right)$$
$$\times\left(\int_0^{t} l_H(t,s)dB_s^H - \int_0^{u} l_H(u,s)dB_s^H\right) = 0.$$

Thus, the increments of $\int_0^t l_H(t,s)dB_s^H$ are uncorrelated, and hence independent. It follows that $M_t^H := \int_0^t l_H(t,s)dB_s^H$ is a Gaussian martingale w.r.t. its natural filtration

$$\mathcal{F}_t^H := \sigma\left\{M_u^H, 0 \le u \le t\right\}.$$

Furthermore, we have that for any $0 \le s \le t$ and $H > \frac{1}{2}$

$$\mathrm{E}\big(M_t^H - M_s^H\big)^2 = t^{2-2H} - s^{2-2H} \le (t-s)^{2-2H}$$

whence M^H is a continuous process, see remark 6.8. Its quadratic variation coincides with quadratic characteristics and both equal $\left[M^H\right]_t = t^{2-2H}$. The process M^H is called the *Molchan martingale*, or the *fundamental martingale*, since it was considered originally in the papers [MOL 69a, MOL 69b]. See also [NOR 99, MIS 08].

A2.9. Asymptotic behavior of the weighted Wiener process and fractional Brownian motion

THEOREM A2.23.– *Let* $\delta > 0$*. Then,* $\frac{W_T}{T^{1/2+\delta}} \to 0$ *a.s. as* $T \to \infty$*.*

PROOF.– Consider any sequence $\{T_n, n \ge 1\}$ such that $T_{n+1} > T_n$ and $T_n \to \infty$ as $n \to \infty$. Then it follows from the martingale property of the Wiener process that for any $\gamma > 0$

$$\mathrm{P}\left\{\sup_{T_n \le t \le T_{n+1}} \frac{|W_t|}{t^{1/2+\delta}} > \frac{1}{n^\gamma}\right\} \le \mathrm{P}\left\{\sup_{T_n \le t \le T_{n+1}} |W_t| > \frac{T_n^{1/2+\delta}}{n^\gamma}\right\}$$
$$\le \mathrm{P}\left\{\sup_{0 \le t \le T_{n+1}} |W_t| > \frac{T_n^{1/2+\delta}}{n^\gamma}\right\} \le \frac{n^{2\gamma}}{T_n^{1+2\delta}} T_{n+1}.$$

Now our goal is to choose T_n and γ so that the series $S = \sum_{n=1}^{\infty} \frac{n^{2\gamma}}{T_n^{1+2\delta}} T_{n+1}$ converges. This will be the case if we choose $\gamma = 1$, $T_n = n^{\eta}$ and $\eta > \frac{3}{2\delta}$. Then

$$\frac{n^{2\gamma}}{T_n^{2\alpha}} T_{n+1} = n^{2-(1+2\delta)\eta}(n+1)^{\eta} \leq 2^{\eta} n^{2-\delta\eta},$$

and $2 - 2\delta\eta < -1$, therefore the series S converges. And then it follows from the Borel-Cantelli lemma that for any $\omega \in \Omega'$, $\mathrm{P}\{\Omega'\} = 1$, there exists $n_0 = n_0(\omega)$ such that for any $n \geq n_0$

$$\sup_{t \geq T_n} \frac{|W_t|}{t^{1/2+\delta}} = \max_{k \geq n} \sup_{T_k \leq t \leq T_{k+1}} \frac{|W_t|}{t^{1/2+\delta}} \leq \frac{1}{n},$$

which means that $\sup_{t \geq n^{\eta}} \frac{|W_t|}{t^{\alpha}} \leq \frac{1}{n}$. The last statement implies the convergence $\frac{W_T}{T^{1/2+\delta}} \to 0, T \to \infty$ a.s. □

Now we can prove a similar result concerning the asymptotic behavior of fractional Brownian motion.

THEOREM A2.24.– *For any* $\delta > 0$ $\frac{B_t^H}{t^{H+\delta}} \to 0$ *a.s. as* $t \to \infty$.

PROOF.– According to theorem 1.10.3 from [MIS 08], for any $p > 0$, $C_p^{(H)} := \mathrm{E} \sup_{0 \leq t \leq 1} \left|B_t^H\right|^p < +\infty$. Taking this into account, together with self-similarly of B^H, we get that for any $n \geq 1$ and $\delta > 0$

$$\mathrm{P}\left\{\sup_{n \leq t \leq n+1} \frac{\left|B_t^H\right|}{t^{H+\delta}} > \frac{1}{n^{\gamma}}\right\} \leq \mathrm{P}\left\{\sup_{n \leq t \leq n+1} \left|B_t^H\right| \geq \frac{n^{H+\delta}}{n^{\gamma}}\right\}$$
$$= \mathrm{P}\left\{\sup_{\frac{n}{n+1} \leq t \leq 1} \left|B_t^H\right| \geq \frac{n^{H+\delta}}{(n+1)^H n^{\gamma}}\right\} \leq \left(1 + \frac{1}{n}\right)^{pH} \left(n^{\gamma-\delta}\right)^p \mathrm{E} \sup_{0 \leq t \leq 1} \left|B_t^H\right|^p$$
$$\leq C_p^{(H)} n^{p(\gamma-\delta)}.$$

Choose $\gamma = \frac{\delta}{2}$ and $p > \frac{2}{\delta}$. Then

$$\sum_{n=1}^{\infty} \mathrm{P}\left\{\sup_{n \leq t \leq n+1} \frac{|B_t^H|}{t^{H+\delta}} > \frac{1}{n^{\gamma}}\right\} \leq \sum_{n=1}^{\infty} C_p^{(H)} n^{-\frac{p\delta}{2}} < \infty.$$

By the Borel-Cantelli lemma, for $\omega \in \Omega'$, $\mathrm{P}\{\Omega'\} = 1$ there exists $n_0(\omega)$ such that for $n \geq n_0(\omega)$

$$\sup_{n \leq t \leq n+1} \frac{|B_t^H|}{t^{H+\delta}} \leq \frac{1}{n^{\delta/2}}, \quad \text{or} \quad \sup_{n \leq t < \infty} \frac{|B_t^H|}{t^{H+\delta}} \leq \frac{1}{n^{\delta/2}},$$

and hence the proof follows. □

A2.10. Miscellaneous

LEMMA A2.9.– *Let* $\{\xi_n, n \geq 0\}$ *and* $\{\zeta_n, n \geq 0\}$ *be two sequences of random variables, and* $\{\xi_n\} \stackrel{d}{=} \{\zeta_n\}$. *Then* $\xi_n \to \xi_0$, $n \to \infty$ *a.s.* $\Longleftrightarrow \zeta_n \to \zeta_0$, $n \to \infty$ *a.s.*

PROOF.– We have two following equalities for the events:

$$\{\xi_n \to \xi_0\} = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left\{|\xi_k - \xi_0| < \frac{1}{m}\right\},$$

$$\{\zeta_n \to \zeta_0\} = \bigcap_{m=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left\{|\zeta_k - \zeta_0| < \frac{1}{m}\right\}.$$

Therefore,

$$\mathrm{P}\{\xi_n \to \xi_0\} = \lim_{m\to\infty} \mathrm{P}\left\{\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} \left\{|\xi_k - \xi_0| < \frac{1}{m}\right\}\right\}$$

$$= \lim_{m\to\infty} \lim_{n\to\infty} \mathrm{P}\left\{\bigcap_{k=n}^{\infty} \left\{|\xi_k - \xi_0| < \frac{1}{m}\right\}\right\}$$

$$= \lim_{m\to\infty} \lim_{n\to\infty} \lim_{N\to\infty} \mathrm{P}\left\{\bigcap_{k=n}^{N} \left\{|\xi_k - \xi_0| < \frac{1}{m}\right\}\right\}$$

$$= \lim_{m\to\infty} \lim_{n\to\infty} \lim_{N\to\infty} \mathrm{P}\left\{\bigcap_{k=n}^{N} \left\{|\zeta_k - \zeta_0| < \frac{1}{m}\right\}\right\} = \mathrm{P}\{\zeta_n \to \zeta_0\}. \qquad \square$$

LEMMA A2.10.– *Let* $\{\xi_n, n \geq 1\}$ *be a non-decreasing (non-increasing) sequence of random variables. If* $\xi_n \stackrel{\mathrm{P}}{\to} \infty$ $(\xi_n \stackrel{\mathrm{P}}{\to} 0)$ *as* $n \to \infty$, *then* $\xi_n \to \infty$ *with probability* 1 $(\xi_n \to 0$ *with probability* $1)$ *as* $n \to \infty$.

PROOF.– Consider only non-decreasing sequences, and non-increasing are considered similarly. Since $\xi_n \stackrel{\mathrm{P}}{\to} \infty$, for any $K \geq 1$ there exists $n(K) \geq 1$ such that $\mathrm{P}\left\{\xi_{n(K)} \leq K\right\} \leq 2^{-K}$. Then it follows from the Borel-Cantelli lemma that $\xi_{n(K)} \to \infty$ a.s. as $K \to \infty$, and then it follows, from the fact that ξ_n is non-decreasing, that $\xi_n \to \infty$ a.s. as $n \to \infty$. $\square$

LEMMA A2.11.– *Let* $\xi \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathrm{P})$, $\frac{1}{p} + \frac{1}{q} = 1$, $1 < p < \infty$. *Then,*

$$\sup_{\eta:\mathrm{E}|\eta|^q=1} \mathrm{E}|\xi\eta| = \left(\mathrm{E}|\xi|^p\right)^{1/p}.$$

PROOF.– Without loss of generality assume that $\mathrm{E}|\xi|^p > 0$. It follows from the Hölder inequality that for any η with $\mathrm{E}|\eta|^q = 1$ that

$$\mathrm{E}|\xi\eta| \le \big(\mathrm{E}|\xi|^p\big)^{1/p}\big(\mathrm{E}|\eta|^q\big)^{1/q} = \big(\mathrm{E}|\xi|^p\big)^{1/p}.$$

Now, substitute $\eta = \frac{|\xi|^{p-1}}{(\mathrm{E}|\xi|^p)^{1/q}}$. Taking into account the equality $(p-1)q = p$, we get that

$$\mathrm{E}|\eta|^q = \mathrm{E}\left(\frac{|\xi|^{(p-1)q}}{\mathrm{E}|\xi|^p}\right) = \frac{\mathrm{E}|\xi|^p}{\mathrm{E}|\xi|^p} = 1.$$

Further,

$$\mathrm{E}|\xi\eta| = \frac{\mathrm{E}|\xi|^p}{(\mathrm{E}|\xi|^p)^{1/q}} = \big(\mathrm{E}|\xi|^p\big)^{1/p},$$

and the proof follows. □

Bibliography

[APP 09] APPLEBAUM D., *Lévy Processes and Stochastic Calculus*, 2nd edition, Cambridge University Press, Cambridge, 2009.

[BAC 95] BACHELIER L., *Théorie de la spéculation. Théorie mathématique du jeu.*, Les Grands Classiques Gauthier-Villars, Éditions Jacques Gabay, Sceaux, 1995.

[BAR 01] BARNDORFF-NIELSEN O.E., MIKOSCH T., RESNICK S.I. (eds), *Lévy Processes. Theory and Applications.*, Birkhäuser, Boston, 2001.

[BIA 08] BIAGINI F., HU Y., ØKSENDAL B. *et al.*, *Stochastic Calculus for Fractional Brownian Motion and Applications*, Springer-Verlag London, Ltd., London, 2008.

[BIL 95] BILLINGSLEY P., *Probability and Measure*, 3rd edition, John Wiley & Sons, Inc., New York, 1995.

[BIL 99] BILLINGSLEY P., *Convergence of Probability Measures*, 2nd edition, John Wiley & Sons, Inc., New York, 1999.

[BJÖ 04] BJÖRK T., *Arbitrage Theory in Continuous Time*, 2nd edition, Oxford University Press, Oxford, 2004.

[BOJ 04] BOJDECKI T., GOROSTIZA L.G., TALARCZYK A., "Sub-fractional Brownian motion and its relation to occupation times", *Statistics & Probability Letters*, vol. 69, no. 4, pp. 405–419, 2004.

[BOR 17] BORTKIEWICZ L., *Die Iterationen, ein Beitrag zur Wahrscheinlichkeitstheorie*, Springer, Berlin, 1917.

[BOR 02] BORODIN A.N., SALMINEN P., *Handbook of Brownian Motion – Facts and Formulae*, 2nd edition, Birkhäuser Verlag, Basel, 2002.

[BRO 06] BROCKWELL P.J., DAVIS R.A., *Time Series: Theory and Methods*, Springer, New York, 2006.

[BRZ 99] BRZEŹNIAK Z., ZASTAWNIAK T., *Basic Stochastic Processes. A Course through Exercises*, Springer-Verlag London, Ltd., London, 1999.

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

[BUL 05] BULINSKIY A.V., SHIRYAEV A.N., *Theory of Stochastic Processes*, Fizmatlit, Moscow, 2005.

[CHE 01] CHERNY A.S., "On the uniqueness in law and the pathwise uniqueness for stochastic differential equations", *Theory of Probability and Its Applications*, vol. 46, no. 3, pp. 406–419, 2001.

[CHU 79] CHUNG K.L., *Elementary Probability theory with Stochastic Processes*, 3rd edition, Springer-Verlag, New York-Heidelberg, 1979.

[CHU 90] CHUNG K.L., WILLIAMS R.J., *Introduction to Stochastic Integration*, 2nd edition, Birkhäuser Boston, Inc., Boston, 1990.

[COH 15] COHEN S.N., ELLIOTT R.J., *Stochastic Calculus and Applications*, 2nd edition, Birkhäuser/Springer, New York, 2015.

[DEL 78] DELLACHERIE C., MEYER P.-A., *Probabilities and Potential*, North-Holland Publishing Co., Amsterdam-New York, 1978.

[DEL 17] DEL MORAL P., PENEV S., *Stochastic Processes. From Applications to Theory*, CRC Press, Boca Raton, 2017.

[DOO 53] DOOB J.L., *Stochastic Processes*, John Wiley & Sons, Inc., New York, 1953.

[DVO 61] DVORETZKY A., ERDÖS P., KAKUTANI S., "Nonincrease everywhere of the Brownian motion process", *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. II, pp. 103–116, 1961.

[DOZ 15] DOZZI M., MISHURA Y., SHEVCHENKO G., "Asymptotic behavior of mixed power variations and statistical estimation in mixed models", *Stat. Inference Stoch. Process.* vol. 18, no. 2, pp. 151–175, 2015

[EIN 05] EINSTEIN A., "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen", *Annalen der Physik*, vol. 322, no. 8, pp. 549–560, 1905.

[FEL 49] FELLER W., "On the theory of stochastic processes, with particular reference to applications", *Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, 1945, 1946*, pp. 403–432, 1949.

[GIH 72] GIHMAN I.I., SKOROHOD A.V., *Stochastic Differential Equations*, Springer-Verlag, New York-Heidelberg, 1972.

[GIK 04a] GIKHMAN I.I., SKOROKHOD A.V., *The Theory of Stochastic Processes. I*, Springer-Verlag, Berlin, 2004.

[GIK 04b] GIKHMAN I.I., SKOROKHOD A.V., *The Theory of Stochastic Processes. II*, Springer-Verlag, Berlin, 2004.

[GIK 07] GIKHMAN I.I., SKOROKHOD A.V., *The Theory of Stochastic Processes. III*, Springer, Berlin, 2007.

[GUS 10] GUSAK D., KUKUSH A., KULIK A. *et al.*, *Theory of Stochastic Processes. With Applications to Financial Mathematics and Risk Theory*, Springer, New York, 2010.

[GUS 15] GUSHCHIN A.A., *Stochastic Calculus for Quantitative Finance*, ISTE Press Elsevier Ltd, Oxford, 2015.

[HAC 07] HACCOU P., JAGERS P., VATUTIN V.A., *Branching Processes: Variation, Growth, and Extinction of Populations*, Cambridge University Press, Cambridge, 2007.

[HEY 97] HEYDE C.C., *Quasi-likelihood and Its Application*, Springer-Verlag, New York, 1997.

[IKE 89] IKEDA N., WATANABE S., *Stochastic Differential Equations and Diffusion Processes*, 2nd edition, North-Holland Publishing Co., Amsterdam, 1989.

[ITÔ 44] ITÔ K., "Stochastic integral", *Proceedings of the Imperial Academy Tokyo*, vol. 20, pp. 519–524, 1944.

[ITÔ 74] ITÔ K., MCKEAN H.P., *Diffusion Processes and their Sample Paths*, 2nd edition, Springer-Verlag, Berlin-New York, 1974.

[JAC 03] JACOD J., SHIRYAEV A.N., *Limit Theorems for Stochastic Processes*, 2nd edition, Springer-Verlag, Berlin, 2003.

[JEA 09] JEANBLANC M., YOR M., CHESNEY M., *Mathematical Methods for Financial Markets*, Springer Finance, Springer, London, 2009.

[JOS 06] JOST C., "Transformation formulas for fractional Brownian motion", *Stochastic Processes and their Applications*, vol. 116, no. 10, pp. 1341–1357, 2006.

[KAL 02] KALLENBERG O., *Foundations of Modern Probability*, 2nd edition, Springer-Verlag, New York, 2002.

[KAR 91] KARATZAS I., SHREVE S.E., *Brownian Motion and Stochastic Calculus*, 2nd edition, Springer-Verlag, New York, 1991.

[KHI 34] KHINTCHINE A., "Korrelationstheorie der stationären stochastischen Prozesse", *Mathematische Annalen*, vol. 109, no. 1, pp. 604–615, 1934.

[KOL 30] KOLMOGOROV A., "Sur la loi forte des grands nombres", *Comptes Rendus de l'Académie des Sciences*, vol. 191, pp. 910–912, 1930.

[KOL 31] KOLMOGOROV A., "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung", *Mathematische Annalen*, vol. 104, no. 1, pp. 415–458, 1931.

[KOL 77] KOLMOGOROV A., *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer-Verlag, Berlin, New York, 1977, Reprint of the 1933 original.

[KOR 07] KORALOV L.B., SINAI Y.G., *Theory of Probability and Random Processes*, 2nd edition, Springer, Berlin, 2007.

[KOR 14] KORCHEVSKY V., "Marcinkiewicz–Zygmund Strong Law of Large Numbers for Pairwise iid Random Variables", arXiv preprint arXiv:1404.7454, 2014.

[KUT 04] KUTOYANTS Y.A., *Statistical Inference for Ergodic Diffusion Processes*, Springer-Verlag London, Ltd, London, 2004.

[KUZ 84] KUZNETSOV S.E., "Inhomogeneous Markov processes", *Journal of Soviet Mathematics*, vol. 25, no. 5, pp. 1380–1498, 1984.

[LAW 06] LAWLER G.F., *Introduction to Stochastic Processes*, 2nd edition, Chapman & Hall/CRC, Boca Raton, 2006.

[LEG 16] LE GALL J.-F., *Brownian motion, martingales, and stochastic calculus*, Springer, Cham, 2016.

[LÉV 65] LÉVY P., *Processus stochastiques et mouvement brownien. Suivi d'une note de M. Loève*, 2ème édition revue et augmentée, Gauthier-Villars & Cie, Paris, 1965.

[LIP 89] LIPTSER R.S., SHIRYAYEV A.N., *Theory of Martingales*, Kluwer Academic Publishers, Dordrecht, 1989.

[LIP 01] LIPTSER R.S., SHIRYAEV A.N., *Statistics of Random Processes I. General Theory*, Springer-Verlag, Berlin, 2001.

[LOÈ 77] LOÈVE M., *Probability Theory. I*, 4th edition, Springer-Verlag, New York-Heidelberg, 1977.

[LOÈ 78] LOÈVE M., *Probability Theory. II*, 4th edition, Springer-Verlag, New York-Heidelberg, 1978.

[MAN 68] MANDELBROT B.B., VAN NESS J.W., "Fractional Brownian motions, fractional noises and applications", *SIAM Review*, vol. 10, pp. 422–437, 1968.

[MCK 69] MCKEAN JR H.P., *Stochastic Integrals*, Academic Press, New York-London, 1969.

[MEN 23] MENCHOFF D., "Sur les séries de fonctions orthogonales. (Première Partie. La convergence.)", *Fundamenta Mathematicae*, vol. 4, pp. 82–105, 1923.

[MIS 08] MISHURA Y., *Stochastic calculus for fractional Brownian motion and related processes*, Springer-Verlag, Berlin, 2008.

[MIS 14] MISHURA Y., "Standard maximum likelihood drift parameter estimator in the homogeneous diffusion model is always strongly consistent", *Statistics & Probability Letters*, vol. 86, pp. 24–29, 2014.

[MOL 69a] MOLCHAN G.M., "Gaussian processes with spectra which are asymptotically equivalent to a power of λ", *Theory of Probability and Its Applications*, vol. 14, no. 3, pp. 530–532, 1969.

[MOL 69b] MOLCHAN G.M., GOLOSOV J.I., "Gaussian stationary processes with asymptotic power spectrum", *Soviet Mathematics. Doklady*, vol. 10, pp. 134–137, 1969.

[NOR 99] NORROS I., VALKEILA E., VIRTAMO J., "An elementary approach to a Girsanov formula and other analytical results on fractional Brownian motions", *Bernoulli*, vol. 5, no. 4, pp. 571–587, 1999.

[NOU 12] NOURDIN I., *Selected Aspects of Fractional Brownian Motion*, Springer, Milan; Bocconi University Press, Milan, 2012.

[NUA 06] NUALART D., *The Malliavin Calculus and Related Topics*, 2nd edition, Springer-Verlag, Berlin, 2006.

[ØKS 03] ØKSENDAL B., *Stochastic Differential Equations. An Introduction with Applications*, 6th edition, Springer, Berlin, 2003.

[PAL 33] PALEY R.E.A.C., WIENER N., ZYGMUND A., "Notes on random functions", *Mathematische Zeitschrift*, vol. 37, no. 1, pp. 647–668, 1933.

[PET 75] PETROV V.V., *Sums of Independent Random Variables*, Springer-Verlag, New York-Heidelberg, 1975.

[PRO 04] PROTTER P.E., *Stochastic Integration and Differential Equations*, 2nd edition, Springer, Berlin, 2004.

[RAD 22] RADEMACHER H., “Einige Sätze über Reihen von allgemeinen Orthogonalfunktionen”, *Mathematische Annalen*, vol. 87, nos. 1–2, pp. 112–138, 1922.

[RES 92] RESNICK S., *Adventures in Stochastic Processes*, Birkhäuser Boston, Inc., Boston, 1992.

[REV 99] REVUZ D., YOR M., *Continuous Martingales and Brownian Motion*, 3rd edition, Springer-Verlag, Berlin, 1999.

[ROG 00a] ROGERS L.C.G., WILLIAMS D., *Diffusions, Markov Processes, and Martingales, vol. 1, Foundations*, Cambridge University Press, Cambridge, 2000.

[ROG 00b] ROGERS L.C.G., WILLIAMS D., *Diffusions, Markov Processes, and Martingales, vol. 2, Itô calculus*, Cambridge University Press, Cambridge, 2000.

[ROS 96] ROSS S.M., *Stochastic Processes*, John Wiley & Sons, Inc., New York, 1996.

[RUS 06] RUSSO F., TUDOR C.A., “On bifractional Brownian motion”, *Stochastic Processes and their Applications*, vol. 116, no. 5, pp. 830–856, 2006.

[SAM 93] SAMKO S.G., KILBAS A.A., MARICHEV O.I., *Fractional Integrals and Derivatives. Theory and Applications*, Gordon and Breach Science Publishers, Yverdon, 1993.

[SAT 13] SATO K., *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, Cambridge, 2013.

[SCH 03] SCHOUTENS W., *Lévy Processes in Finance: Pricing Financial Derivatives*, Wiley, New York, 2003.

[SCH 14] SCHILLING R.L., PARTZSCH L., *Brownian Motion. An Introduction to Stochastic Processes*, 2nd edition, De Gruyter, Berlin, 2014.

[SEB 03] SEBER G.A.F., LEE A.J., *Linear Regression Analysis*, 2nd edition, John Wiley & Sons, Hoboken, 2003.

[SHE 15] SHEVCHENKO G., “Fractional Brownian motion in a nutshell”, *Analysis of Fractional Stochastic Processes*, *International Journal of Modern Physics: Conference Series*, vol. 36, pp. 1560002 (16 pages), 2015.

[SHI 99] SHIRYAEV A.N., *Essentials of Stochastic Finance. Facts, Models, Theory*, World Scientific Publishing Co., Inc., River Edge, 1999.

[SHR 04] SHREVE S.E., *Stochastic Calculus for Finance. II. Continuous-Time Models*, Springer-Verlag, New York, 2004.

[SKO 65] SKOROKHOD A.V., *Studies in the Theory of Random Processes*, Addison-Wesley Publishing Co., Inc., Reading, 1965.

[STR 06] STROOCK D.W., VARADHAN S. R.S., *Multidimensional Diffusion Processes*, Springer-Verlag, Berlin, 2006.

[VON 06] VON SMOLUCHOWSKI M., "Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen", *Annalen der Physik*, vol. 326, no. 14, pp. 756–780, Wiley Online Library, 1906.

[WIL 91] WILLIAMS D., *Probability with Martingales*, Cambridge University Press, Cambridge, 1991.

[ZAB 04] ZABCZYK J., *Topics in Stochastic Processes*, Scuola Normale Superiore, Pisa, 2004.

Index

Theory and Statistical Applications of Stochastic Processes,
First Edition. Yuliya Mishura and Georgiy Shevchenko.

F, G

H, I, K

L, M, N

O, P

Q, R, S

T, U, W

Other titles from

iSTE

in

Mathematics and Statistics

2017

CARONI Chysseis
First Hitting Time Regression Models: Lifetime Data Analysis Based on Underlying Stochastic Processes
(Mathematical Models and Methods in Reliability Set – Volume 4)

CELANT Giorgio, BRONIATOWSKI Michel
Interpolation and Extrapolation Optimal Designs 2: Finite Dimensional General Models

CONSOLE Rodolfo, MURRU Maura, FALCONE Giuseppe
Earthquake Occurrence: Short- and Long-term Models and their Validation
(Statistical Methods for Earthquakes Set – Volume 1)

D'AMICO Guglielmo, DI BIASE Giuseppe, JANSSEN Jacques, MANCA Raimondo
Semi-Markov Migration Models for Credit Risk
(Stochastic Models for Insurance Set – Volume 1)

HARLAMOV Boris
Stochastic Analysis of Risk and Management
(Stochastic Models in Survival Analysis and Reliability Set – Volume 2)

KERSTING Götz, VATUTIN Vladimir
Discrete Time Branching Processes in Random Environment
(Branching Processes, Branching Random Walks and Branching Particle Set – Volume 1)

NIKULIN Mikhail, CHIMITOVA Ekaterina
Chi-squared Goodness-of-fit Tests for Censored Data
(Stochastic Models in Survival Analysis and Reliability Set – Volume 3)

SIMON Jacques
Banach, Fréchet, Hilbert and Neumann Spaces
(Analysis for PDEs Set – Volume 1)

2016

CELANT Giorgio, BRONIATOWSKI Michel
Interpolation and Extrapolation Optimal Designs 1: Polynomial Regression and Approximation Theory

CHIASSERINI Carla Fabiana, GRIBAUDO Marco, MANINI Daniele
Analytical Modeling of Wireless Communication Systems
(Stochastic Models in Computer Science and Telecommunication Networks Set – Volume 1)

GOUDON Thierry
Mathematics for Modeling and Scientific Computing

KAHLE Waltraud, MERCIER Sophie, PAROISSIN Christian
Degradation Processes in Reliability
(Mathematial Models and Methods in Reliability Set – Volume 3)

KERN Michel
Numerical Methods for Inverse Problems

RYKOV Vladimir
Reliability of Engineering Systems and Technological Risks
(Stochastic Models in Survival Analysis and Reliability Set – Volume 1)

2015

DE SAPORTA Benoîte, DUFOUR François, ZHANG Huilong
Numerical Methods for Simulation and Optimization of Piecewise Deterministic Markov Processes

DEVOLDER Pierre, JANSSEN Jacques, MANCA Raimondo
Basic Stochastic Processes

LE GAT Yves
Recurrent Event Modeling Based on the Yule Process
(Mathematical Models and Methods in Reliability Set – Volume 2)

2014

COOKE Roger M., NIEBOER Daan, MISIEWICZ Jolanta
Fat-tailed Distributions: Data, Diagnostics and Dependence
(Mathematical Models and Methods in Reliability Set – Volume 1)

MACKEVIČIUS Vigirdas
Integral and Measure: From Rather Simple to Rather Complex

PASCHOS Vangelis Th
Combinatorial Optimization – 3-volume series – 2nd edition
Concepts of Combinatorial Optimization / Concepts and Fundamentals – volume 1
Paradigms of Combinatorial Optimization – volume 2
Applications of Combinatorial Optimization – volume 3

2013

COUALLIER Vincent, GERVILLE-RÉACHE Léo, HUBER Catherine, LIMNIOS Nikolaos, MESBAH Mounir
Statistical Models and Methods for Reliability and Survival Analysis

JANSSEN Jacques, MANCA Oronzio, MANCA Raimondo
Applied Diffusion Processes from Engineering to Finance

SERICOLA Bruno
Markov Chains: Theory, Algorithms and Applications

2012

BOSQ Denis
Mathematical Statistics and Stochastic Processes

CHRISTENSEN Karl Bang, KREINER Svend, MESBAH Mounir
Rasch Models in Health

DEVOLDER Pierre, JANSSEN Jacques, MANCA Raimondo
Stochastic Methods for Pension Funds

2011

MACKEVIČIUS Vigirdas
Introduction to Stochastic Analysis: Integrals and Differential Equations

MAHJOUB Ridha
Recent Progress in Combinatorial Optimization – ISCO2010

RAYNAUD Hervé, ARROW Kenneth
Managerial Logic

2010

BAGDONAVIČIUS Vilijandas, KRUOPIS Julius, NIKULIN Mikhail
Nonparametric Tests for Censored Data

BAGDONAVIČIUS Vilijandas, KRUOPIS Julius, NIKULIN Mikhail
Nonparametric Tests for Complete Data

IOSIFESCU Marius *et al.*
Introduction to Stochastic Models

VASSILIOU PCG
Discrete-time Asset Pricing Models in Applied Stochastic Finance

2008

ANISIMOV Vladimir
Switching Processes in Queuing Models

FICHE Georges, HÉBUTERNE Gérard
Mathematics for Engineers

HUBER Catherine, LIMNIOS Nikolaos *et al.*
Mathematical Methods in Survival Analysis, Reliability and Quality of Life

JANSSEN Jacques, MANCA Raimondo, VOLPE Ernesto
Mathematical Finance

2007

HARLAMOV Boris
Continuous Semi-Markov Processes

2006

CLERC Maurice
Particle Swarm Optimization

This book is concerned with the theory of stochastic processes and the theoretical aspects of statistics for stochastic processes. It combines classic topics such as construction of stochastic processes, associated filtrations, processes with independent increments, Gaussian processes, martingales, Markov properties, continuity and related properties of trajectories with contemporary subjects: integration with respect to Gaussian processes, Itô integration, stochastic analysis, stochastic differential equations, fractional Brownian motion and parameter estimation in diffusion models.

The presentation is made as self-contained as possible, with complete proofs of the facts which are often either omitted from textbooks or are replaced by informal or heuristic arguments. Some auxiliary material, related mainly to different subjects of real analysis and probability theory, is included in the comprehensive appendix. The book is targeted at the widest audience: students of mathematical and related programs, postgraduate students, postdocs, lecturers, researchers and practitioners in any field concerned with the application of stochastic processes will find this book to be a valuable resource.

Yuliya Mishura is Full Professor and Head of the Department of Probability Theory, Statistics and Actuarial Mathematics at Taras Shevchenko National University of Kyiv. Her research interests include martingale theory, fractional processes, stochastic differential equations, the statistics of stochastic processes and financial mathematics.

Georgiy Shevchenko is Full Professor in the Department of Probability Theory, Statistics and Actuarial Mathematics at Taras Shevchenko National University of Kyiv. His research interests include stochastic analysis, stochastic differential equations, models with long memory, the statistics of stochastic processes and financial mathematics.

www.iste.co.uk

9 781786 300508