Statistics for Industry and Technology

Asymptotic Methods in Probability and Statistics with Applications

N. Balakrishnan
I.A. Ibragimov
V.B. Nevzorov
Editors

Springer Science+Business Media, LLC

N. Balakrishnan
Department of Mathematics
and Statistics
McMaster University
1280 Main Street West
Hamilton, Ontario L8S 4K1
Canada

I.A. Ibragimov
V.B. Nevzorov
Department of Mathematics and Mechanics
St. Petersburg State University
St. Petersburg
Russia

Library of Congress Cataloging-in-Publication Data
Asymptotic methods in probability and statistics with applications / edited by
N. Balakrishnan, I.A. Ibragimov, V.B. Nevzorov.
p. cm. (Statistics for industry and technology)
Includes bibliographical references and index.
ISBN 978-1-4612-6663-1 ISBN 978-1-4612-0209-7 (eBook)
DOI 10.1007/978-1-4612-0209-7
1. Aymptotic distribution (Probability theory) I. Balakrishnan, N., 1956–
II. Ibragimov, I.A. (Il'dar Abdulovich) III. Nevzorov, Valery B., 1946– IV. Series.
QA273.6 .A78 2001
519.2—dc21 2001025400
CIP

Printed on acid-free paper.

Originally published by Birkhäuser Boston in 2001
Softcover reprint of the hardcover 1st edition 2001

SPIN 10785725

Production managed by Louise Farkas; manufacturing supervised by Jacqui Ashri.
Typeset by the editors in LaTeX.
Cover design by Vernon Press, Boston, MA.

9 8 7 6 5 4 3 2 1

Contents

Preface

Traditions of the 150-year-old St. Petersburg School of Probability and Statistics had been developed by many prominent scientists including P. L. Chebychev, A. M. Lyapunov, A. A. Markov, S. N. Bernstein, and Yu. V. Linnik. In 1948, the Chair of Probability and Statistics was established at the Department of Mathematics and Mechanics of the St. Petersburg State University with Yu. V. Linik being its founder and also the first Chair. Nowadays, alumni of this Chair are spread around Russia, Lithuania, France, Germany, Sweden, China, the United States, and Canada.

The fiftieth anniversary of this Chair was celebrated by an International Conference, which was held in St. Petersburg from June 24–28, 1998. More than 125 probabilists and statisticians from 18 countries (Azerbaijan, Canada, Finland, France, Germany, Hungary, Israel, Italy, Lithuania, The Netherlands, Norway, Poland, Russia, Taiwan, Turkey, Ukraine, Uzbekistan, and the United States) participated in this International Conference in order to discuss the current state and perspectives of Probability and Mathematical Statistics.

The conference was organized jointly by St. Petersburg State University, St. Petersburg branch of Mathematical Institute, and the Euler Institute, and was partially sponsored by the Russian Foundation of Basic Researches.

The main theme of the Conference was chosen in the tradition of the St. Petersburg School of Probability and Statistics. The papers in this volume form a selection of invited talks presented at the conference. The papers were all refereed rigorously, and we thank all the referees who assisted us in this process. We also thank all the authors for submitting their articles for inclusion in this volume.

Thanks are also due to Mr. Wayne Yuhasz and Ms. Lauren Schultz, both at Birkhäuser (Boston), for their support and encouragement. Our final thanks go to Mrs. Debbie Iscoe for her excellent camera-ready typesetting of this entire volume.

Hamilton, Ontario, Canada *N. Balakrishnan*
St. Petersburg, Russia *I. A. Ibragimov*
St. Petersburg, Russia *V. B. Nevzorov*

Contributors

Aaronson, J.
Faculty of Exact Sciences, Tel Aviv University, Tel Aviv, Israel
e-mail: *Aaro@math.tau.ac.il*

Aleškevičienė, A.
Institute of Mathematics and Informatics, Vilnius, Lithuania
e-mail: *Stat@julius.ktl.mii.lt*

Aliev, F. A.
Department of Statistics, Ankara University, Ankara, Turkey
e-mail: *Fazil.Aliev@Science.ankara.edu.tr*

Bagdonavičius, V. B.
Department of Statistics, University of Vilnius, Vilnius, Lithuania
e-mail: *Mypass@pub.osf.lt*

Bairamov, I. G.
Department of Statistics, Ankara University, Ankara, Turkey
e-mail: *Bayramov@science.ankara.edu.tr*

Balakrishnan, N.
Department of Mathematics and Statistics, McMaster University, Hamilton, Ontario, Canada
e-mail: *bala@mcmail.cis.mcmaster.ca*

Belopolskaya, Ya.
Department of Mathematics, St. Petersburg Architecture Building University, St. Petersburg, Russia
e-mail: *Yana@Yana.abu.spb.ru*

Bühler, B.
Institute für Stochastik, Friedrich Schiller Universität, Jena, Germany
e-mail: *Buehler@minet.uni-jena.de*

Burgio, G.
University of Roma "La Sapienza," Rome, Italy
e-mail: *Burgio@scec.eco.uniromal.it*

Chebotarev, V. I.
Computer Center, Far-Eastern Branch of the Russian Academy of Sciences, Khabarovsk, Russia
e-mail: *Chebotarev@as.dv.khv.ru*

Christoph, G.
Fakultät für Mathematik, Otto–von–Guericke–Universität Magdeburg, Magdeburg, Germany
e-mail: *Gerd.Christoph@mathematik.uni-magdeburg.de*

Davydov, Yu.
Université de Sciences et Technologies de Lille, Villeneuve d'Ascq, France
e-mail: *Youri.Davydov@univ-lille1.fr*

Denker, M.
Institut für Mathematische Stochastik, Universität Göttingen, Göttingen, Germany
e-mail: *Denker@math.uni-goettingen.de*

Egorov, V.
Department of Mathematics-2, Electrotechnical University, St. Petersburg, Russia
e-mail: *Eva@probabil.usr.etu.spb.ru*

Frolov, A.
Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
e-mail: *Andrei.Frolov@pobox.spbu.ru*

Gebizlioglu, O. L.
Department of Statistics, Ankara University, Ankara, Turkey
e-mail: *Gebizli@science.ankara.edu.tr*

Götze, F.
Fakultät für Mathematik, University of Bielefeld, Bielefeld, Germany
e-mail: *Goetze@mathematik.uni-bielefeld.de*

Gushchin, A. A.
Steklov Mathematical Institute, Moscow, Russia
e-mail: *Alex@gushchin.mian.su*

Habach, L.
Université de Sciences et Technologies de Lille, Villeneuve d'Ascq, France

Haiman, G.
Université de Sciences et Technologies de Lille, Villeneuve d'Ascq, France
e-mail: *Haiman@ccr.jussieu.fr*

Hofmann, G.
Facultad de Siencias Fisicas y Matematicas, Universidad de Concepcion, Concepcion, Chile
e-mail: *Glenn@gauss.cfm.udec.cl*

Kaya, M. F.
Department of Statistics, Selcuk University, Konya, Turkey
e-mail: *Fkaya@karatay1.cc.selcuk.edu.tr*

Klebanov, L. B.
Department of Mathematics, St. Petersburg Architecture Building University, St. Petersburg, Russia
e-mail: *Lev@pdmi.ras.ru*

Kubilius, J. P.
Department of Mathematics, Vilnius University, Vilnius, Lithuania
e-mail: *Kubilius@pub.osf.lt*

Lebedev, V. A.
Laboratory of Mathematical Statistics, M. V. Lomonosov Moscow State University, Moscow, Russia
e-mail: *Lag@lms.math.msu.su*

Li, W. V.
Department of Mathematics, University of Delaware, Newark, Delaware, U.S.A.
e-mail: *Wli@math.udel.edu*

Lifshits, M.
Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
e-mail: *Lifts@mail.rcom.ru*

Linde, W.
Institute für Stochastik, Friedrich Schiller Universität, Jena, Germany
e-mail: *Lindwe@minet.uni-jena.de*

Malov, S. V.
Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
e-mail: *Sergey.Malov@pobox.spbu.ru*

Manita, A.
Department of Mathematics and Mechanics, M. V. Lomonosov Moscow State University, Moscow, Russia
e-mail: *Manita@mech.math.msu.su*

Martikainen, A.
Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
e-mail: *Alexander.Martikainen@pobox.spbu.ru*

Mason, D. M.
Department of Mathematical Statistics, University of Delaware, Newark, Delaware, U.S.A.
e-mail: *Davidm@math.udel.edu*

Mijnheer, J.
Leiden University, Leiden, The Netherlands
e-mail: *Mijnheer@rulwinw.leidenuniv.nl*

Nagaev, A. V.
Department of Mathematics, Nicholas Copernicus University, Torun, Poland
e-mail: *Nagaev@math.uni.torun.pl*

Nagaev, S. V.
Sobolev Institute of Mathematics, Novosibirsk, Russia
e-mail: *Nagaev@math.nsc.ru*

Nagaraja, H. N.
Department of Statistics, Ohio State University, Columbus, Ohio, U.S.A.
e-mail: *hnn@stat.ohio-state.edu*

Nevzorov, V. B.
Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
e-mail: *Valery.Nevzorov@pobox.spbu.ru*

Nevzorova, L. N.
Department of Economics, St. Petersburg Humanitarian University of Trade Unions, St. Petersburg, Russia
e-mail: *Ludnev@mail.ru*

Nikitin, Ya. Yu.
Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
e-mail: *Yakov.Nikitin@pobox.spbu.ru*

Nikulin, A. M.
St. Petersburg State Technical University, St. Petersburg, Russia

Nikulin, M. S.
Université Bordeaux-2, Bordeaux, France
e-mail: *Nikou@mi2s.u-bordeaux2.fr*

Padvelskis, K.
Department of Mathematics and Statistics, Kaunas Vytautas Magnus University, Kaunas, Lithuania
e-mail: *Kazimieras.Padvelskis@vdu.lt*

Petrov, V. V.
Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
e-mail: *Proba@math.spbu.ru*

Piterbarg, V.
Laboratory of Probability, M. V. Lomonosov Moscow State University, Moscow, Russia
e-mail: *Piter@julia.math.msu.su*

Prohorov, Yu.
Steklov Mathematical Institute, Moscow, Russia
e-mail: *Prohorov@mi.ras.ru*

Rémillard, B.
Département de mathématiques et d'informatique, Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada
e-mail: *bruno_remillard@uqtr.uquebec.ca*

Rukhin, A. L.
Department of Mathematics and Statistics, University of Maryland at Baltimore County, Baltimore, Maryland, U.S.A.
e-mail: *rukhin@math.umbc.edu*

Rusakov, O. V.
Department of Mathematics and Mechanics, St. Petersburg State University, St. Petersburg, Russia
e-mail: *Oleg.Rusakov@pobox.spbu.ru*

Schreiber, K.
Fakultät für Mathematik, Otto–von–Guericke–Universität Magdeburg, Magdeburg, Germany

Statulevičius, V.
Institute of Mathematics and Informatics, Vilnius, Lithuania
e-mail: *Stat@julius.ktl.mii.lt*

Stamatovich, S.
University of Montenegro, Montenegro, Yugoslavia

Steinebach, J.
Fachbereich Mathematik und Informatic, Philipps-Universität Marburg, Marburg, Germany
e-mail: *Jost@mathematik.uni-marburg.de*

Sudakov, A. V.
Steklov Mathematical Institute, St. Petersburg, Russia
e-mail: *Andrew@sudakov.pdmi.ras.ru*

Sudakov, V. N.
Steklov Mathematical Institute, St. Petersburg, Russia
e-mail: *V@sudakov.pdmi.ras.ru*

Szekely, G. J.
Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH, U.S.A.
e-mail: *Gabor@bgnet.bgsu.edu*

Theodorescu, R.
Département de mathématiques et de statistique, Université Laval, Sainte-Foy, Québec, Canada
e-mail: *radutheo@mat.ulaval.ca*

Ulyanov, V.
Moscow State University, Moscow, Russia
e-mail: *Vlad@ulyanov.mian.su*

Valkeila, E.
Department of Mathematics, University of Helsinki, Helsinki, Finland
e-mail: *Valkeila@Helsinki.fi*

Vavilov, S. A.
St. Petersburg State University, St. Petersburg, Russia

Weber, M.
Université Louis-Pasteur, Strasbourg, France
e-mail: *Weber@math.u-strasbg.fr*

Weizsäcker, H. v.
University of Kaiserslautern, Kaiserslautern, Germany
e-mail: *Weizsaecker@mathematik.uni-kl.de*

Zaitsev, A. Yu.
Stekov Mathematical Institute, St. Petersburg, Russia
e-mail: *Zaitsev@pdmi.ras.ru*

Zinger, A. A.
Department of Mathematics, Aerocosmic Academy, St. Petersburg, Russia
e-mail: *Flour@ctinet.ru*

Zolotukhin, A. Ya.
Computer Center, Far-Eastern Branch of the Russian Academy of Sciences, Khabarovsk, Russia
e-mail: *Chebotarev@as.dv.khv.ru*

PART I

Probability Distributions

1

Positive Linnik and Discrete Linnik Distributions

Gerd Christoph and Karina Schreiber

Otto–von–Guericke–Universität Magdeburg, Magdeburg, Germany

I belong to the last generation of students learning the basic knowledge of statistics by Yuri Vladimirovitsch Linnik. Only a few weeks before he died, I took my examination in Statistics. I remember this examination because Prof. Linnik welcomed me in German and he asked me (in German) about properties of "Maximum–Likelihood–Schätzungen." Never before I had heard these German terms. I was prepared to answer, as usually, in Russian. So, I learned from Prof. Linnik not only to like statistics but also during the examination that likelihood *is an old English word for probability which is used in German statistical terms too.* Gerd Christoph

Abstract: In this chapter, (continuous) positive Linnik and (nonnegative integer valued) discrete Linnik random variables are discussed. Rates of convergence and first terms of both the Edgeworth expansions and the expansions in the exponent of the distribution functions of certain sums of such random variables with nonnegative strictly stable as well as discrete stable limit laws are considered.

Keywords and phrases: Positive Linnik and discrete Linnik distributions, discrete self-decomposability, discrete stability, rates of convergence, Edgeworth expansions, expansions in the exponent with discrete stable limit law

1.1 Different Kinds of Linnik's Distributions

Linnik (1963, p. 67) proved that the functions

$$\varphi(t) = (1 + |t|^{\gamma})^{-1} \quad \text{for} \quad \gamma \in (0\,,\,2]$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

are characteristic functions of real valued symmetric random variables, which are called *symmetric Linnik distributed.* Later on, in Devroye (1990) some more parameters were considered and it was proved that

$$\varphi_\gamma(t) = (1 + c\,|t|^\gamma)^{-\beta} \quad \text{for} \quad \gamma \in (0\,,\,2],\ c > 0,\ \beta > 0$$

are characteristic functions of real valued symmetric random variables, too.

Pakes (1995, p. 294) called nonnegative random variables with Laplace–Stieltjes transforms

$$\psi_\gamma(u) = (1 + c\,u^\gamma)^{-\beta},\ u \geq 0, \quad \text{for} \quad \gamma \in (0\,,\,1],\ c > 0,\ \beta > 0 \tag{1.1}$$

positive Linnik distributed. Here, we restrict ourselves to the case $\gamma \in (0\,,\,1]$, since the corresponding function $\psi_\gamma(u)$ in case $1 < \gamma \leq 2$ is not completely monotone. Hence, $\psi_\gamma(u)$ with $1 < \gamma \leq 2$ can not be a Laplace–Stieltjes transform, see Feller (1971, p. 439).

Changing u in $\psi_\gamma(u)$ by $(1 - z)$, it was shown that

$$g_\gamma(z) = (1 + c\,(1 - z)^\gamma)^{-\beta}, \quad |z| \leq 1,\ \text{for}\ \gamma \in (0\,,\,1],\ c > 0,\ \beta > 0 \tag{1.2}$$

are probability generating functions of nonnegative integer valued random variables, which are called *discrete Linnik distributed.* See Devroye (1993) for the case $c = 1$ and Pakes (1995, p. 296).

In analogue to the generalized Pareto distribution, these random variables are redefined in Christoph and Schreiber (1998b): A random variable L_γ^λ with probability generating function

$$g_{L_\gamma^\lambda}(z) = \begin{cases} (1 + \lambda\,(1 - z)^\gamma/\beta)^{-\beta}, & \text{for}\ \ 0 < \beta < \infty\,, \\ \exp\{-\lambda\,(1 - z)^\gamma\}\,, & \text{for}\ \ \beta = \infty\,, \end{cases} \qquad |z| \leq 1\,, \tag{1.3}$$

is called *discrete Linnik distributed with characteristic exponent* $\gamma \in (0, 1]$, *scale parameter* $\lambda > 0$ *and form parameter* $\beta > 0$.

In case $\beta = \infty$, the *discrete stable* random variables denoted further by X_γ^λ and introduced in Steutel and van Harn (1979) occur in (1.3) as a natural generalization of the discrete Linnik distribution defined in (1.2) with $c = \lambda/\beta$.

If $\beta = 1$, then (1.3) gives the probability generating function of the *discrete Mittag–Leffler distribution*; see Jayakumar and Pillai (1995).

For $\gamma = 1$, well–known distributions occur:
- Poisson(λ) distribution if $\beta = \infty$,
- negative binomial distribution with probabilities

$$\mathrm{P}(L_1^\lambda = k) = \binom{-\beta}{k}\left(-\frac{\lambda}{\lambda + \beta}\right)^k \left(\frac{\beta}{\lambda + \beta}\right)^\beta, \quad k = 0, 1, \ldots\,, \tag{1.4}$$

if $\beta < \infty$ and as special case
- geometric distribution if $\beta = 1$.

If $\gamma = 1$, the probabilities (1.4) of the negative binomial random variable L_1^λ tend to the probabilities of the Poisson(λ) random variable X_1^λ as $\beta \to \infty$.

If $\gamma < 1$, the probabilities of the discrete Linnik distributions with probability generating function (1.3) may be obtained by expanding the probability generating function in a power series. Hence,

$$\mathrm{P}(L_\gamma^\lambda = k) = (-1)^k \sum_{j=0}^{\infty} \binom{\gamma j}{k} \binom{j+\beta-1}{j} \left(-\frac{\lambda}{\beta}\right)^j, \quad k = 0, 1, \ldots$$

if $\beta < \infty$, or in the case of $\beta = \infty$

$$\mathrm{P}(X_\gamma^\lambda = k) = (-1)^k \sum_{j=0}^{\infty} \binom{\gamma j}{k} \frac{(-\lambda)^j}{j!}, \quad k = 0, 1, \ldots \tag{1.5}$$

In both cases, the series are absolutely convergent, but they can not be expressed in a simple form. Again, we find

$$\mathrm{P}(L_\gamma^\lambda = k) \to \mathrm{P}(X_\gamma^\lambda = k) \quad \text{as} \quad \beta \to \infty, \quad k = 0, 1, \ldots$$

For both the discrete Linnik random variable L_γ^λ if $\beta < \infty$ and the discrete stable random variable X_γ^λ if $\beta = \infty$, asymptotic formulas for the mentioned probabilities are given in Christoph and Schreiber (1998b, 1998a). We have, as $k \to \infty$:

$$\begin{aligned} P(L_\gamma^\lambda = k) &= \frac{1}{\pi} \sum_{j=1}^{[(\gamma+1)/\gamma]} \binom{j+\beta-1}{j} \frac{(-1)^{j+1} \lambda^j \Gamma(\gamma j + 1) \sin(\gamma j \pi)}{\beta^j k^{\gamma j + 1}} \\ &\quad + O(k^{-\gamma-2}) \end{aligned}$$

if $\beta < \infty$, or in the case of $\beta = \infty$

$$P(X_\gamma^\lambda = k) = \frac{1}{\pi} \sum_{j=1}^{[(\gamma+1)/\gamma]} \frac{(-1)^{j+1} \lambda^j \Gamma(\gamma j + 1) \sin(\gamma j \pi)}{j!\, k^{\gamma j+1}} + O(k^{-\gamma-2}).$$

In analogue to (1.3), we now redefine the positive Linnik distributions using $c = \lambda/\beta$ in (1.1). A random variable W_γ^λ with Laplace–Stieltjes transform

$$\psi_{W_\gamma^\lambda}(u) = \begin{cases} (1 + \lambda u^\gamma/\beta)^{-\beta}, & \text{for } 0 < \beta < \infty, \\ \exp\{-\lambda u^\gamma\}, & \text{for } \beta = \infty, \end{cases} \qquad u \geq 0 \tag{1.6}$$

is called *positive Linnik distributed with characteristic exponent* $\gamma \in (0, 1]$, *scale parameter* $\lambda > 0$ *and form parameter* $\beta > 0$.

In case $\beta = \infty$, the nonnegative *strictly stable* random variables denoted further by S_γ^λ occur in (1.6) as a natural generalization of the positive Linnik distribution.

If $\beta = 1$, (1.6) defines probability generating function of the *Mittag–Leffler distribution* [see Pillai (1990) or Jayakumar and Pillai (1995)] with the corresponding distribution function $1 - E_\gamma(-x^\gamma)$, where $E_\gamma(x)$ is the Mittag–Leffler function. In contrast, in Bingham, Goldie and Teugels (1987, p. 329 and p. 392) Mittag–Leffler distributions as limit laws for occupation times of Markov processes are defined by the corresponding Laplace–Stieltjes transform $E_\gamma(u)$ being Mittag–Leffler functions. See also Pakes (1995, p. 294).

For $\gamma = 1$ well–known distributions occur:
– degenerate distribution at the point λ if $\beta = \infty$,
– Gamma distribution with density $(\Gamma(\beta))^{-1} (\beta/\lambda)^\beta \, \mathrm{e}^{-x\beta/\lambda} x^{\beta-1}$ as $x > 0$ if $\beta < \infty$ with the special case of
– exponential distribution with parameter $1/\lambda > 0$ if $\beta = 1$.

1.2 Self-decomposability and Discrete Self-decomposability

A real random variable W is said to be *self-decomposable* (or it belongs to the so-called *class L*) if corresponding to every $\alpha \in (0\,,\,1)$ there exists a random variable W_α such that

$$W \stackrel{\mathrm{d}}{=} \alpha W^* + W_\alpha\,, \tag{1.7}$$

where W^* and W_α are independent, $W \stackrel{\mathrm{d}}{=} W^*$ and $\stackrel{\mathrm{d}}{=}$ denotes the equality in distribution. Nondegenerate self-decomposable random variables are known to be absolutely continuous; see Fisz and Varadarajan (1963).

A real valued random variable S is called *strictly stable* if

$$S \stackrel{\mathrm{d}}{=} \alpha S^* + (1 - \alpha^\gamma)^{1/\gamma} S^{**} \quad \text{for every} \quad 0 < \alpha < 1 \tag{1.8}$$

with some $0 < \gamma \le 2$, where S^* and S^{**} are independent with the same distribution as S.

It follows from (1.7), (1.8) and $S_\alpha \stackrel{\mathrm{d}}{=} (1 - \alpha^\gamma)^{1/\gamma} S^{**}$ with some $0 < \gamma \le 2$ and $S \stackrel{\mathrm{d}}{=} S^{**}$, that strictly stable random variables are self-decomposable.

Self-decomposable random variables are important as the only possible limit laws of normalized partial sums of independent random variables. Moreover, the stable random variables occur as the only possible limit laws of normalized partial sums of independent and identically distributed (iid) random variables.

Strictly stable random variables S_γ^λ with (1.8) are nonnegative only if $\gamma \in (0\,,\,1]$. Such random variables S_γ^λ have Laplace–Stieltjes transform (1.6) with $\beta = \infty$.

Further, we restrict ourselves to nonnegative integer random variables, for which discrete analogues of self-decomposability and stability were introduced by Steutel and van Harn (1979). In (1.7) and (1.8), the multiplication αX of a nonnegative integer valued random variable X with a constant $\alpha \in (0,1)$ is replaced by a *dot product* $\alpha \odot X$, which is defined as a random partial sum of the first X members of a sequence of iid Bernoulli random variables $\{N_k\}_{k\geq 1}$, which are independent of X:

$$\alpha \odot X \stackrel{\mathrm{d}}{=} N_1 + \cdots + N_X \quad \text{with} \quad \alpha = \mathrm{P}(N_1 = 1) = 1 - \mathrm{P}(N_1 = 0). \tag{1.9}$$

Hence,

$$g_{\alpha\odot X}(z) = g_X(1 - \alpha + \alpha\, z) = g_X(1 \,-\, \alpha\,(1 - z))\,. \tag{1.10}$$

Then, a nonnegative integer random variable X is called *discrete self-decomposable*, if corresponding to every $\alpha \in (0\,,\,1)$ there exist independent random variables X_α and X^* with $X^* \stackrel{\mathrm{d}}{=} X$ such that

$$X \stackrel{\mathrm{d}}{=} \alpha \odot X^* \,+\, X_\alpha\,. \tag{1.11}$$

Further, a random variable X_γ is called *discrete stable* if for every $\alpha \in (0\,,\,1)$ there exist iid random variables X_γ^* and X_γ^{**} with $X_\gamma^* \stackrel{\mathrm{d}}{=} X_\gamma$ such that

$$X_\gamma \stackrel{\mathrm{d}}{=} \alpha \odot X_\gamma^* \,+\, (1 \,-\, \alpha^\gamma)^{1/\gamma} \odot X_\gamma^{**} \quad \text{for some} \quad 0 < \gamma \leq 1\,. \tag{1.12}$$

It follows from (1.11), (1.12) and $X_\alpha \stackrel{\mathrm{d}}{=} (1 \,-\, \alpha^\gamma)^{1/\gamma} \odot X_\gamma^{**}$ with some $0 < \gamma \leq 1$ and $X_\gamma \stackrel{\mathrm{d}}{=} X_\gamma^{**}$, that discrete stable random variables are discrete self-decomposable. It was also proved in Steutel and van Harn (1979) that a random variable X_γ is discrete stable iff the corresponding probability generating function has the form (1.3) in the case of $\beta = \infty$.

The relations (1.3) and (1.6) show the connection between the Laplace–Stieltjes transform of a positive Linnik random variable W_γ^λ and the probability generating function of a discrete Linnik random variable L_γ^λ. For more details, see van Harn, Steutel and Vervaat (1982) or Pakes (1995). In particular,
– degenerate distribution corresponds with Poisson law if $\beta = \infty$ and $\gamma = 1$,
– gamma distribution corresponds with negative binomial law in case of $\beta < \infty$ and $\gamma = 1$, and
– Mittag–Leffler distribution with discrete Mittag–Leffler one if $\beta = 1$.

Using the approach of Jayakumar and Pillai (1995) to prove the discrete self–decomposability of the discrete Mittag–Leffler law with probability generating function (1.3) for $\beta = 1$, we find for discrete Linnik laws with integer form parameters $\beta = m$:

$$\frac{\psi_{L_\gamma^\lambda}(z)}{\psi_{L_\gamma^\lambda}(1 - \alpha^\gamma(1 - z))} = \left(\alpha^\gamma \,+\, \frac{1 \,-\, \alpha^\gamma}{1 \,+\, \lambda(1 - z)^\gamma}\right)^m.$$

The right-hand side is the weighted average of probability generating functions of degenerate and discrete Linnik laws. Hence, the left-hand side defines a probability generating function The discrete self–decomposability of discrete Linnik laws with integer form parameters m follows now from (1.10) and (1.11). In the same manner, we obtain the self–decomposability of positive Linnik laws with integer form parameter m too.

Using representation theorems for self–decomposable or discrete self–decomposable random variables in terms of Laplace–Stieltjes transforms or probability generating functions given in Steutel and van Harn (1979), it may be proved that positive Linnik random variables are self–decomposable and discrete Linnik are discrete self–decomposable for arbitrary form parameter β. See also Christoph and Schreiber (1998b, p. 8). Hence, these random variables are infinitely divisible.

Note that the other Mittag–Leffler laws with Mittag–Leffler functions $E_\gamma(u)$ as Laplace–Stieltjes transforms are not infinitely divisible; see Bondesson, Kristiansen and Steutel (1996, Theorem 4.3).

1.3 Scaling of Positive and Discrete Linnik Laws

Let W be self-decomposable. It follows from Loéve (1977, p. 334) that if to W there correspond a real number $\alpha > 0$ and a nondegenerate random variable W_α such that (1.7) holds, then $\alpha < 1$. For $\alpha = 1$, the random variable W_1 in (1.7) is degenerate at point 0. Nevertheless, the random variable $\alpha\, W$ is a well defined for any $\alpha > 0$.

For the positive Linnik random variable W_γ^λ, we find

$$\lambda^{-1/\gamma}\, W_\gamma^\lambda \stackrel{\mathrm{d}}{=} W_\gamma^1 \quad \text{for arbitrary} \quad \lambda > 0\,. \tag{1.13}$$

Consider now the discrete Linnik random variable L_γ^λ with probability generating function (1.3), then using (1.10) with arbitrary $\alpha > 0$ we obtain

$$\lambda^{-1/\gamma} \odot L_\gamma^\lambda \stackrel{\mathrm{d}}{=} L_\gamma^1 \quad \text{for arbitrary} \quad \lambda > 0\,, \tag{1.14}$$

but the dot product in (1.9) is defined only for $\lambda^{-1/\gamma} < 1$, i.e. $\lambda > 1$. We may extend the definition of the dot product $\alpha \odot X$ for such $\alpha > 1$, whenever the functions $g_{\alpha\odot X}(z)$ defined by (1.10) is a probability generating function In dependency of X, there may exist an upper bound for such α that $g_{\alpha\odot X}(z)$ is a probability generating function. For more about scaling and the dot product, see Christoph and Schreiber (1998d).

The scaling properties (1.13) and (1.14) of the positive Linnik random variable W_γ^λ and discrete Linnik random variable L_γ^λ allow to consider only the case of $\lambda = 1$.

Let $p_\gamma(x;\lambda,\beta)$ be the density of positive Linnik random variable W_γ^λ, then by (1.13)

$$p_\gamma(x;\lambda,\beta) = \lambda^{-1/\gamma} p_\gamma(\lambda^{-1/\gamma}x;1,\beta)\,.$$

For the probabilities of discrete Linnik random variable L_γ^λ, we find by (1.14)

$$P(L_\gamma^\lambda = k) = \sum_{i=k}^{\infty} \binom{i}{k} (1-\lambda^{1/\gamma})^{i-k}\,\lambda^{k/\gamma}\,P(L_\gamma^1 = i)\,.$$

In order to avoid such difficulties, we formulate the following limit theorems for arbitrary scaling parameter.

1.4 Strictly Stable and Discrete Stable Distributions as Limit Laws

Both the positive Linnik random variable W_γ^λ with Laplace–Stieltjes transform (1.6) and the discrete Linnik random variable L_γ^λ with probability generating function (1.3) belong to the *domain of normal attraction* of the nonnegative strictly stable random variable S_γ^λ having Laplace–Stieltjes transform (1.6) with $\beta = \infty$, since

$$\psi^n_{W_\gamma^\lambda}(n^{-1/\gamma}u) \to \mathrm{e}^{-\lambda u^\gamma} \quad \text{as} \quad n \to \infty \tag{1.15}$$

and

$$\psi^n_{L_\gamma^\lambda}(n^{-1/\gamma}u) = g^n_{L_\gamma^\lambda}(\mathrm{e}^{-n^{-1/\gamma}u}) \to \mathrm{e}^{-\lambda u^\gamma} \quad \text{as} \quad n \to \infty. \tag{1.16}$$

Following Steutel and van Harn (1979), the discrete Linnik random variable L_γ^λ with probability generating function (1.3) belong to the *discrete domain of normal attraction* of the discrete stable random variable X_γ^λ having probability generating function (1.3) with $\beta = \infty$, since

$$g^n_{L_\gamma^\lambda}(1 - n^{-1/\gamma}(1-z)) \to \mathrm{e}^{-\lambda(1-z)^\gamma} \quad \text{as} \quad n \to \infty. \tag{1.17}$$

Let $W_1, W_2, \ldots, W_n$ be iid copies of W_γ^λ and $L_1, L_2, \ldots, L_n$ be iid copies of L_γ^λ. Then different limit statements follow from (1.15), (1.16) and (1.17):

$$S_n = n^{-1/\gamma}(W_1 + \ldots + W_n) \xrightarrow{\mathrm{d}} S_\gamma^\lambda \quad \text{as} \quad n \to \infty\,, \tag{1.18}$$

$$S_n^* = n^{-1/\gamma}(L_1 + \ldots + L_n) \xrightarrow{\mathrm{d}} S_\gamma^\lambda \quad \text{as} \quad n \to \infty \tag{1.19}$$

and

$$X_n = N_1 + \ldots + N_{L_1+\ldots+L_n} \xrightarrow{\mathrm{d}} X_\gamma^\lambda \quad \text{as} \quad n \to \infty, \tag{1.20}$$

where in the later case the random variables $L_1, \ldots, L_n$ are independent of the iid random variables $N_1, N_2, \ldots$ with $\mathrm{P}(N_1 = 1) = n^{-1/\gamma} = 1 - \mathrm{P}(N_1 = 0)$ and $\xrightarrow{\mathrm{d}}$ denotes the convergence in distribution.

We may interpret (1.17) also as the sum of iid random variable $L_{1\,n}, \ldots, L_{n\,n}$ in the nth series of a triangular array $\{L_{k\,n}\}$, $k = 1, 2, \ldots, n$, $n = 1, 2, \ldots$

$$X_n^* = L_{1\,n} + \ldots + L_{n\,n} \xrightarrow{\mathrm{d}} X_\gamma^\lambda \quad \text{as} \quad n \to \infty, \tag{1.21}$$

where $L_{1\,n}$ is discrete Linnik with with characteristic exponent γ, scale parameter λ/n and form parameter β. Note that $X_n \stackrel{\mathrm{d}}{=} X_n^*$.

Since the discrete Linnik random variable L_γ^λ is nonnegative integer with span 1, the random variable S_n^* is lattice with span $n^{-1/\gamma}$ but the limit random variable S_γ^λ in (1.19) is absolutely continuous for $0 < \gamma < 1$ or degenerate for $\gamma = 1$, whereas both random variables X_n and X_γ^λ in (1.20) are nonnegative integer with span 1. Relation (1.21) shows a special type of convergence to nonnegative integer valued infinitely divisible random variables.

Proposition 1.4.1 *For positive Linnik random variable W_γ^λ with Laplace–Stieltjes transform (1.6), if $\beta < \infty$ we have in (1.18)*

$$|\mathrm{P}(S_n \le x) - \mathrm{P}(S_\gamma^\lambda \le x)| = O(n^{-1}) \quad \text{as} \quad n \to \infty\,.$$

For discrete Linnik random variable L_γ^λ with probability generating function (1.3), we obtain in (1.19)

$$|\mathrm{P}(S_n^* \le x) - \mathrm{P}(S_\gamma^\lambda \le x)| = \begin{cases} O(n^{-1})\,, & \text{if } \beta < \infty\,, \\ O(n^{-1/\gamma})\,, & \text{if } \beta = \infty\,, \end{cases} \qquad \text{as } n \to \infty\,.$$

Proposition 1.4.2 *For discrete Linnik random variable L_γ^λ with probability generating function (1.3), if $\beta < \infty$ we find in (1.20)*

$$|\mathrm{P}(X_n \le x) - \mathrm{P}(X_\gamma^\lambda \le x)| = O(n^{-1}) \quad \text{as} \quad n \to \infty\,.$$

Since $X_n \stackrel{\mathrm{d}}{=} X_n^*$, the same bound holds in (1.21).

Remark 1.4.1 The results are special cases of more general ones proved in Christoph and Schreiber (1998c). They follow also from the corresponding statements about asymptotic expansions given in the next section.

Remark 1.4.2 In the case $\gamma = 1$ Proposition 1.4.1 gives a special type of rates of convergence in the weak law of large numbers for gamma distribution and negative binomial or Poisson distributions, because S_n is the arithmetic mean and the limit is the degenerate law at the point λ.

Remark 1.4.3 If $\gamma = 1$ in Proposition 1.4.2 the convergence rate for the distribution function of a random sum of Bernoulli random variables to the Poisson law is obtained. More details about Poisson approximation see Lorz and Heinrich (1991), Vellaisamy and Chaudhuri (1996) or in the monograph by Barbour, Holst and Janson (1992) and the references therein.

1.5 Asymptotic Expansions

It follows from Proposition 1.4.1 that for positive Linnik random variable W_γ^λ and discrete Linnik random variable L_γ^λ, in case of $\beta < \infty$ we obtain the same rate of convergence to the distribution of the strictly stable random variable S_γ^λ. To see the difference between the continuous and discrete random variables we consider asymptotic expansions up to such an order, where they differ.

Asymptotic expansions with a (continuous) stable limit law (corresponding to $0 < \gamma < 1$) are well investigated; see Christoph and Wolf (1993, Ch. 4) and references therein. In case of discrete stable limit laws, the Poisson limit law (which is discrete stable with $\gamma = 1$) is considered, when the underlying random variables have second or more moments; see Lorz and Heinrich (1991) or Aleskeviciene and Statulevicius (1995) and references therein. If $\gamma < 1$, then even the first moment does not exist.

Define $m = [1/\gamma]$, where $[r]$ denotes the integer part of r. Let

$$h_n(w) = \exp\left\{ w + \sum_{j=2}^{m+1} \frac{1}{j\,\beta^{j-1}\,n^{j-1}}\, w^j \right\} \tag{1.22}$$

and

$$h_n^*(w) = \mathrm{e}^w \cdot \left(1 + \sum_{j=1}^{m} P_j(w)\, n^{-j}\right) , \tag{1.23}$$

where the polynomials $P_j(w)$ are defined by the formal equation

$$\exp\left\{ \sum_{j=2}^{\infty} \frac{1}{j\,\beta^{j-1}\,n^{j-1}}\, w^j \right\} = \left(1 + \sum_{j=1}^{\infty} P_j(w)\, n^{-j}\right) ,$$

which leads to [see Christoph and Wolf (1993, p. 97)]

$$P_j(w) = \sum_{m=1}^{j} \frac{1}{m!} \sum_{s_1+\ldots+s_m=m+j} \prod_{k=1}^{m} s_k^{-1}\, w^{m+j} , \tag{1.24}$$

where the summation in the second sum of the right-hand side is carried over all integer solutions $(s_1, \ldots, s_m)$ of the equation $s_1 + \ldots + s_m = m + j$ with $s_k \geq 2$, $k = 1, \ldots, m$.

The first three polynomials are $P_1(w) = 1/(2\,\beta)\,w^2$,

$$P_2(w) = \frac{1}{\beta^2}\left(\frac{1}{8}\,w^4 + \frac{1}{3}\,w^3\right) \text{ and } P_3(w) = \frac{1}{\beta^3}\left(\frac{1}{48}\,w^6 + \frac{1}{6}\,w^5 + \frac{1}{4}\,w^4\right).$$

Remember that $\varphi_{S_\gamma^\lambda}(t)$ and $\varphi_{S_n}(t)$ are the characteristic functions of the strictly stable random variable S_γ^λ and the normalized sum S_n defined in (1.18). Put $\eta(t) = \ln \varphi_{S_\gamma^\lambda}(t)$, then

$$\eta(t) = -\,\lambda\,|t|^\gamma \exp\{-\,i\,(\pi\,\gamma\,/2)\operatorname{sgn} t\,\} = -\,\lambda(-\,i\,t)^\gamma,$$

$$\varphi_{S_\gamma^\lambda}(t) = \psi_{S_\gamma^\lambda}(-\,it) \text{ and } \varphi_{S_n}(t) = \psi^n_{W_\gamma^\lambda}(-\,it\,n^{-\,1/\gamma}) = (1\,-\,\eta(t)/\,(n\,\beta)\,)^{-\beta\,n}\,.$$

With Lemma 4.30 from Christoph and Wolf (1993), we find for $|t| < \varepsilon\,n^{1/\gamma}$ with sufficiently small $\varepsilon > 0$ that

$$|\varphi_{S_n}(t)\,-\,h_n^*(\eta(t))| \le c_1 n^{-\,m-1}(|t|^{\gamma(m+2)} + |t|^a)\,\mathrm{e}^{-\lambda\,|t|^\gamma/4}\,, \tag{1.25}$$

where $a \ge \gamma(m+2)$ and c_1 are some constants independent of $|t|$ and n.

Consider now S_n^* defined in (1.19) which is lattice with span $n^{-\,1/\gamma}$. The characteristic function of the random variable L_γ^λ is given by $\varphi_{L_\gamma^\lambda}(t) = g_{L_\gamma^\lambda}(\mathrm{e}^{it})$ with the corresponding probability generating function (1.3). Using $(1-\mathrm{e}^{it})^\gamma = (-\,it)^\gamma\,(1-\gamma(-\,it)/2) + O(|t|^{2+\gamma})$ as $|t| \to 0$, we find similar to (1.25)

$$|\varphi_{S_n^*}(t)\,-\,h_n^{**}(\eta(t))| \le c_2 n^{-\,m-1}(|t|^{\gamma(m+2)} + |t|^b)\,\mathrm{e}^{-\lambda\,|t|^\gamma/4}\,, \tag{1.26}$$

where $b \ge \gamma(m+2)$ and c_2 are some constants independent of $|t|$ and n and

$$h_n^{**}(\eta(t)) = h_n^*(\eta(t))\,-\,\gamma\,\eta(t)\,(-\,it)/2\,.$$

Let $G_n^*(x)$ and $G_n^{**}(x)$ be functions of bounded variation such that $h_n^*(\eta(t))$ and $h_n^{**}(\eta(t))$ are their Fourier-Stieltjes transforms:

$$h_n^*(\eta(t)) = \int_0^\infty \mathrm{e}^{itx}\,\mathrm{d}G_n^*(x) \quad\text{and}\quad h_n^{**}(\eta(t)) = \int_0^\infty \mathrm{e}^{itx}\,\mathrm{d}G_n^{**}(x)\,.$$

Denote the distribution function of the strictly stable random variable S_γ^λ by $G_\gamma(x;\lambda)$,

$$G_\gamma^{k,j}(x;\lambda) = \frac{\partial^{k+j}}{\partial x^k\,\partial\lambda^j}\,G_\gamma(x;\lambda)\,,\quad k=0,1 \quad\text{and}\quad j=0,1,\ldots$$

and the Fourier-Stieltjes transform of $P_j(\eta(t))\,\varphi_{S_\gamma^\lambda}(t)$ by $Q_j(x;\lambda)$. Then by (1.23) and (1.24) the functions $G_n^*(x)$ and $G_n^{**}(x)$ are linear combinations of partial derivatives of the limit law $G_\gamma(x;\lambda)$.

Proposition 1.5.1 *Let* $0 < \gamma < 1$.
i) For positive Linnik random variable W_γ^λ *with Laplace–Stieltjes transform (1.6), if* $\beta < \infty$ *we have*

$$|\mathrm{P}(S_n \le x) - G_n^*(x)| = O(n^{-[1/\gamma]-1}) \quad as \quad n \to \infty\,,$$

where

$$G_n^*(x) = G_\gamma(x;\lambda) + \sum_{j=1}^{[1/\gamma]} Q_j(x;\lambda)\, n^{-j}\,.$$

ii) For discrete Linnik random variable L_γ^λ *with probability generating function (1.3), we obtain*

$$|\mathrm{P}(S_n^* \le x) - G_n^{**}(x)| = \begin{cases} O(n^{-[1/\gamma]+1}) & \text{for } \beta < \infty \\ O(n^{-2/\gamma}) & \text{for } \beta = \infty \end{cases} \qquad as\ n \to \infty\,,$$

where, with jumps correcting function $S(x) = [x] - x + 1/2$,

$$G_n^{**}(x) = G_n^*(x) \;-\; (\gamma/2) G_\gamma^{1,1}(x;\lambda) n^{-1/\gamma} \;+\; S(x\, n^{1/\gamma})\, G_\gamma^{1,0}(x;\lambda) n^{-1/\gamma}$$

if $\beta < \infty$, *or if* $\beta = \infty$

$$G_n^{**}(x) = G_\gamma(x;\lambda) \;-\; (\gamma/2)\, G_\gamma^{1,1}(x;\lambda) n^{-1/\gamma} \;+\; S(x\, n^{1/\gamma})\, G_\gamma^{1,0}(x;\lambda) n^{-1/\gamma}\,.$$

PROOF. The first part follows from Theorem 4.11 of Christoph and Wolf (1993), which can not be used directly for the random variable L_γ^λ since it is lattice. Combining the proofs of Theorem 4.11 with that of Theorem 4.37 in Christoph and Wolf (1993) and changing the pseudomoment condition by condition (1.26) on the behavior of the characteristic functions, we obtain the second statement too. ■

Remark 1.5.1 The Edgeworth expansions of the normalized sums of positive Linnik or discrete Linnik random variables differ in a continuous term and a term considering the jumps of S_n^* both with the order $n^{-1/\gamma}$.

Consider now $P(X_n \le x) = P(X_n^* \le x)$, where X_n and X_n^* are defined in (1.20) and (1.21).

Using $(1+u)^{-\tau} = \exp\{-\tau\, \ln(1+u)\}$ and expanding $\ln(1+u)$ in series, we obtain for $|1-z| < \varepsilon$ with sufficiently small $\varepsilon > 0$

$$|g_{X_n}(z) \;-\; h_n(-\lambda\,(1-z)^\gamma)| \le c_3\, |1-z|^{(m+2)\gamma}\, n^{-m-1} \tag{1.27}$$

and expanding $\exp\{\sum\}$ in $h_n(w)$ in (1.22) in series

$$|g_{X_n}(z) \;-\; h_n^*(-\lambda\,(1-z)^\gamma)| \le c_4\, (|1-z|^{(m+2)\gamma} + |1-z|^d)\, n^{-m-1}\,, \tag{1.28}$$

where c_3, c_4 and $d \geq (m+2)\,\gamma$ are positive constants independent of $|t|$ and n.

Let $M_n(x)$ and $M_n^*(x)$ be functions of bounded variation such that

$$h_n(-\lambda\,(1-z)^\gamma) = \int_0^\infty z^x \,\mathrm{d}M_n(x) \text{ and } h_n^*(-\lambda\,(1-z)^\gamma) = \int_0^\infty z^x \,\mathrm{d}M_n^*(x)\,.$$

The function $M_n(x)$ is an *asymptotic expansion in the exponent*; see Cekanavicius (1997) and references therein which goes back to signed compound binomial approximation by LeCam (1960).

With $M_n^*(x)$, we have an *Edgeworth expansion* with discrete stable limit law which coincides for $\gamma = 1$ with the *Poisson–Charlier expansion* with the Poisson limit law due to Franken (1964). See the also the references in Aleskeviciene and Statulevicius (1995), where the approximating functions $M_n^*(x)$ with the Poisson limit law and the constants in the remainder are calculated.

In Steutel and van Harn (1979), it was proved that all nonnegative integer random variables with finite expectation are discrete normal attracted by the Poisson law. In case $0 < \gamma < 1$, we do not have even the first moment.

The probabilities of the strictly stable random variable X_γ^λ are given in (1.5). Hence, for $m = 1, 2, \ldots$, we have

$$(-\lambda\,(1-z)^\gamma)^m \,\exp\{-\lambda\,(1-z)^\gamma\} = \sum_{k=0}^\infty q_k(m) z^k \tag{1.29}$$

with the jumps

$$q_k(m) = (-1)^k \sum_{j=0}^\infty \binom{\gamma\,(j+m)}{k} \frac{(-\lambda)^{j+m}}{j!}, \quad k = 0, 1, \ldots \tag{1.30}$$

It follows from (1.29) and (1.30) that $\sum_{k=0}^\infty q_k(m) = 0$ for $m = 1, 2, \ldots$. Hence, by (1.23) and (1.23) we may calculate the approximating functions $M_n^*(x) = \sum_{0 \leq k \leq x} q_k^*$, which are functions of bounded variation with jumps q_k^*, where $\sum_{k=0}^\infty q_k^* = 1$.

Proposition 1.5.2 *For the discrete Linnik random variable L_γ with Laplace-Stieltjes transform (1.6), if $\beta < \infty$ we have*

$$|F_n(x) - M_n(x)| = O(n^{-[1/\gamma]-1}) \quad as \quad n \to \infty$$

and

$$|F_n(x) - M_n^*(x)| = O(n^{-[1/\gamma]-1}) \quad as \quad n \to \infty\,.$$

PROOF. Since both the discrete Linnik and the discrete stable random variables are nonnegative integers, we make use of an analogue of Esseen's smoothing lemma for nonnegative integer valued random variables due to Franken

(1964, Lemma 1). Let $F(x) = \sum_{0 \le k \le x} p_k$ be a lattice distribution function with jumps $p_k \ge 0$, $G(x) = \sum_{0 \le k \le x} r_k$ a function of bounded variation with jumps r_k such that $\sum_{k=0}^{\infty} r_k = 1$ and both φ_F and φ_G the corresponding Fourier-Stieltjes transforms, then

$$\sup_x |F(x) - G(x)| \le \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\varphi_F(t) - \varphi_G(t)}{\exp\{it\} - 1} \right| dt. \qquad (1.31)$$

Put $F(x) = P(X_n \le x)$ and $G(x) = M_n(x)$ or $M_n^*(x)$, then $\varphi_F(t) = \varphi_{X_n}(t)$ and $\varphi_G(t) = h_n(-\lambda(1 - e^{it})^\gamma)$ or $h_n^*(-\lambda(1 - e^{it})^\gamma)$. With (1.31) and (1.27) or (1.28) we obtain the statements of Proposition 1.5.2. ■

References

1. Aleskeviciene, A. K. and Statulevicius, V. A. (1995). Asymptotic expansions in the approximation by the Poisson law, *Lith. Math. J.* **35**, 309–329.

2. Barbour, A. D., Holst, L. and Janson, S. (1992). *Poisson Approximation*, Oxford: Clarendon Press.

3. Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987). *Regular Variation*, Cambridge: Cambridge University Press.

4. Bondesson, L., Kristiansen, G. K. and Steutel, F. W. (1996). Infinite divisibility of random variables and their integer parts, *Statistics & Probability Letters*, **28**, 271–278.

5. Cekanavicius, V. (1997). Asymptotic expansion in the exponent: a compound Poisson approach, *Advances in Applied Probability*, **29**, 374–387.

6. Christoph, G. and Schreiber, K. (1998a). Discrete stable random variables, *Statistics & Probability Letters*, **37**, 243–247.

7. Christoph, G. and Schreiber, K. (1998b). The generalized discrete Linnik distributions, In *Advances in Stochastic Models for Reliability, Quality and Safety* (Eds., W. Kahle *et al.*), pp. 3–18, Boston: Birkhäuser.

8. Christoph, G. and Schreiber, K. (1998c). Rates of convergence to stable and discrete stable laws, In *Proceedings of 7th Vilnius Conference on Probability Theory and Mathematical Statistics and 22nd European Meeting of Statisticians* (Eds., B. Grigelionis *et al.*), Vilnius 1998, Amsterdam: VSP Publishers, Preprint Nr. 36/1998, Department of Mathematics, Otto-von-Guericke University Magdeburg (submitted).

9. Christoph, G. and Schreiber, K. (1998d). Shifted and scaled Sibuya distributions and discrete self–decomposability, Preprint Nr. 24/1998, Department of Mathematics, Otto-von-Guericke University Magdeburg (submitted).

10. Christoph, G. and Wolf, W. (1993). *Convergence Theorems with a Stable Limit Law*, Math. Research, **70**, Berlin: Akademie–Verlag.

11. Devroye, L. (1990). A note on Linnik's distribution, *Statistics & Probability Letters*, **9**, 305–306.

12. Devroye, L. (1993). A triptych of discrete distributions related to the stable law, *Statistics & Probability Letters*, **18**, 349–351.

13. Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. 2, New York: John Wiley & Sons.

14. Fish, M. and Varadarajan, V. S. (1963). A condition for absolute continuity of infinitely divisible distribution functions, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **1**, 335–339.

15. Franken, P. (1964). Approximation der Verteilungen von Summen unabhängiger nichtnegativ ganzzahliger Zufallsgrößen durch Poissonsche Verteilungen, *Math. Nachr.* **27**, 303–340.

16. van Harn, K., Steutel, F. W. and Vervaat, W. (1982). Self-Decomposable discrete distributions and branching processes, *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **61**, 97–118.

17. Jayakumar, K. and Pillai, R. N. (1995). Discrete Mittag–Leffler distribution, *Statistics & Probability Letters*, **23**, 271–274.

18. LeCam, L. (1960). An approximation theorem for Poisson binomial distribution, *Pacific Journal of Mathematics*, **10**, 1181–1197.

19. Linnik, Yu. V. (1963). Linear forms and statistical criteria I,II, *Select. Translat. Math. Statist. Prob.*, **3**, 1–90. Transl. from *Ukr. Mat. Zh.*, **5**, (1953), 207–243 and 247–290.

20. Loéve, M. (1977). *Probability Theory*, Vol. 1, Fourth edition, Berlin: Springer–Verlag.

21. Lorz, U. and Heinrich, L. (1991). Normal and Poisson approximation of infinitely divisible distribution functions, *Statistics*, **22**, 627–649.

22. Pakes, A. G. (1995). Characterization of discrete laws via mixed sums and Markov branching processes, *Stochastic Processes and their Applications*, **55**, 285–300.

23. Pillai, R. N. (1990). On Mittag–Leffler functions and related distributions, *Annals of the Institute of Statistical Mathematics*, **42**, 157–161.

24. Steutel, F. W. and van Harn, K. (1979). Discrete analogues of self–decomposability and stability, *Annals of Probability*, **7**, 893–899.

25. Vellaisamy, P. and Chaudhuri, B. (1996). Poisson and compound Poisson approximations for random sums of random variables, *Journal of Applied Probability*, **33**, 127–137.

2

On Finite–Dimensional Archimedean Copulas

S. V. Malov

St. Petersburg State University, St. Petersburg, Russia

Abstract: We investigate general finite dimensional Archimedean copulas. Some properties of generators of Archimedean copulas are under consideration. We obtain necessary and sufficient conditions for the generators of Archimedean copulas and give some properties of degenerate finite dimensional Archimedean copulas. Some examples of degenerate finite dimensional Archimedean copulas are also represented.

Keywords and phrases: Archimedean copula, complete monotone function, copula, Laplace transformation

2.1 Introduction

Let $\bar{a}, \bar{b} \in \overline{\mathbf{R}}^d$, $\bar{a} = (a_1, a_2, \ldots, a_d)$, and $\bar{b} = (b_1, b_2, \ldots, b_d)$. Introduce the partial ordering the following way. We say that $\bar{a} \le \bar{b}$ if $a_i \le b_i$ for all $i = 1, 2, \ldots, d$. A d-box is the Cartesian product of d closed intervals. For $\bar{a} \le \bar{b}$, we define the d-box $[\bar{a}, \bar{b}] = \{\bar{t} \in \overline{\mathbf{R}}^d : \bar{a} \le \bar{t} \le \bar{b}\}$ as the Cartesian product $[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_d, b_d]$ and the vertices of $[\bar{a}, \bar{b}]$ are the points $\bar{v} = (v_1, v_2, \ldots, v_d)$ such that each v_m is equal to either a_m or b_m. Introduce the following definitions.

Let $\mathfrak{A} = [\bar{a}, \bar{b}]$ be a nondegenerate n-box (i.e. $a_i < b_i$ for all $i = 1, 2, \ldots, d$). Introduce the function $\mathbf{sgn}_{\mathfrak{A}} : J \to \{-1, 0, 1\}$, where J is the set of vertices of $\mathfrak{A}$, such that

$$\mathbf{sgn}_{\mathfrak{A}}(\bar{v}) = \begin{cases} \;\;1 & \text{if } v_m = a_m \text{ for an even number of m's,} \\ -1 & \text{if } v_m = a_m \text{ for an add number of m's,} \end{cases}$$

for any vertex $v = (v_1, v_2, \ldots, v_d)$ of $\mathfrak{A}$. For degenerate n-box (i.e. $a_i = b_i$ for some $i \in \{1, 2, \ldots, n\}$), we assume that $\mathbf{sgn}_{\mathfrak{A}}(\bar{v}) = 0$ for all $\bar{v} \in J$.

N. Balakrishnan et al. (ed.), *Asymptotic Methods in Probability and Statistics with Applications*

We say that a function $F : D \to \mathbf{R}$, $D \subset \overline{\mathbf{R}}^d$, is d-increasing if for any d-box $\mathfrak{A} \subset D$

$$V_F(\mathfrak{A}) \stackrel{\text{def}}{=} \sum_{\bar{v}} \mathbf{sgn}_{\mathfrak{A}}(\bar{v})\, F(\bar{v}) \geq 0. \tag{2.1}$$

Suppose that the set D has smallest element $\bar{e} : \bar{e} \leq \bar{u}$ for all $\bar{u} \in D$. The function $F : D \to \mathbf{R}$ is said to be grounded if $F(\bar{x}) = 0$ for all $\bar{x} = (x_1, x_2, \ldots, x_d)$ such that $x_i = e_i$ for some $i \in \{1, 2, \ldots, d\}$.

It is well known that the function $F : \overline{\mathbf{R}}^d \to [0,1]$ is the distribution function of some random vector iff F is d-increasing, grounded and $F(\infty, \infty, \ldots, \infty) = 1$.

We say that a function $C : I^d \to [0,1]$, where $I^d = [0,1] \times [0,1] \times \ldots \times [0,1]$ is the d-dimensional unit cube, is a copula if it is d-increasing, grounded and

$$C(1, \ldots, 1, u_m, 1, \ldots, 1) = u_m, \quad u_m \in [0,1],$$

for all $m = 1, 2, \ldots, d$.

It is clear [see, for example, Schweizer and Sklar (1983)] that any distribution function F having marginals $F_1, F_2, \ldots, F_d$ can be represented using copula

$$F(x_1, x_2, \ldots, x_d) = C(F_1(x_1), F_2(x_2), \ldots, F_d(x_d)),\ x_1, x_2, \ldots, x_d \in \mathbf{R}. \tag{2.2}$$

A copula C is say to be Archimedean if there exist the functions $B : \mathbf{R}_+ \to [0,1]$ and $A : [0,1] \to \mathbf{R}_+$ such that B is continuous and strictly decreasing on $[0, A(0)]$, $B(0) = 1$, $\lim\limits_{u \to A(0)} B(u) = 0$, $B(u) = 0$ for all $u > A(0)$ and

$$C(u_1, u_2, \ldots, u_d) = B\Big(\sum_{j=1}^{d} A(u_j)\Big),\ u_1, u_2, \ldots u_d \in [0,1]. \tag{2.3}$$

It is not difficult to see that under continuous marginals $B(x) = A^{-1}(x)$ for all $x \in [0, A(0)]$.

Remark 2.1.1 The function F given by (2.2) is a distribution function of a random vector with marginals $F_1, F_2, \ldots, F_d$ for any marginals if this property takes place for some continuous marginals $F_1, F_2, \ldots, F_d$. Therefore, when we prove the main results, we lose no generality by the assumption that all marginal distribution functions

$$F_i(x) = U(x) = \begin{cases} 0, & x \leq 0, \\ x, & x \in (0,1], \\ 1, & x > 1 \end{cases} \tag{2.4}$$

are the distribution functions of the uniform $U(0,1)$ distribution for all $i = 1, 2, \ldots, d$.

As was given by Schweizer and Sklar (1983), based on the result of Moynihan (1978), the function $C(u_1, u_2)$ defined by (2.3) for $d = 2$ is a copula iff B is a convex function. Kimberling (1974) based on the results of Widder (1946) for Laplace transformation, proved that $C(u_1, u_2, \dots, u_d)$ in (2.3) is a copula for all $d \geq 2$ iff the function B is completely monotone, i.e. B is infinitely differentiable on $(0, \infty)$ and

$$(-1)^k B^{(k)}(t) \geq 0, \quad t \geq 0, \tag{2.5}$$

for all $k \in \mathbb{N}$.

An (n)-differentiable function satisfying (2.5) for all $k = 1, 2, \dots, n$ is said to be n-th order monotone on $[0, \infty)$ or simply n-th order monotone.

We say that the random vector having distribution function given by (2.2) and (2.3) is an Archimedean copula vector. The process having finite dimensional distributions given by (2.2) and (2.3) is said to be an Archimedean copula process.

As was mentioned by Malov (1998), any Archimedean copula sequence can be represented using the Laplace transformation via independent random variables (r.v.'s). Let $X_1, X_2, \dots, X_n$ be the Archimedean copula process having finite dimensional distributions given by (2.2) and (2.3). It is known [see, for example, Feller (1971)] that any completely monotone function $B(t)$ with $B(0) = 1$ and $\lim_{t \to \infty} B(t) = 0$ can be represented as follows:

$$B(t) = \int_0^{+\infty} e^{-st} dG(s)$$

for some distribution function G such that $G(0_+) = 0$.

Suppose that $Y_1, Y_2, \dots$ are independent and identically distributed r.v.'s having the distribution functions

$$F_i(x) = 1 - \exp(-e^x), \; x \in \mathbf{R},$$

and Y is some r.v. independent of $\{Y_k\}_{k \in \mathbb{N}}$. Introduce r.v.'s $Z = \exp(Y)$, $Z_k = \exp(Y_k)$, and $L_k = \frac{Z_k}{Z}$ for all $k \in \mathbb{N}$. Then

$$\begin{aligned}
&\mathbf{P}(L_1 > R_1(x_1), L_2 > R_2(x_2), \dots, L_n > R_n(x_n)) \\
&= \int_0^{+\infty} \mathbf{P}(Z_1 > sR_1(x_1), Z_2 > sR_2(x_2), \dots, Z_n > sR_n(x_n))\, dG(s) \\
&= \int_0^{+\infty} \exp\Big(-s \sum_{i=1}^{n} R_i(x_i)\Big)\, dG(s) = \int_0^{+\infty} \exp\Big(-s \sum_{i=1}^{n} A(F_i(x_i))\Big)\, dG(s),
\end{aligned}$$

where $G(s)$ is the distribution function of Z; $R_i(x_i) = A(F_i(x_i))$, $x \in \mathbf{R}$, for all $i \in \mathbb{N}$. The suitable choice of the distribution of Y yields us that the process $R_1^{-1}(L_1), R_2^{-1}(L_2), \dots$ is the desired Archimedean copula process.

In the special case of proportional Archimedean copula, under the condition $R_i(x) = \alpha_i R(x)$, $x \in \mathbf{R}$, for some function $R : \mathbf{R} \to [0,1]$ and some positive constants $\alpha_1, \alpha_2, \ldots, \alpha_n$, $i = 1, 2, \ldots, n$, it is convenient to use the following representation using the independent r.v.'s $Y, Y_1, Y_2, \ldots$ with the distribution functions

$$F_i(x) = 1 - \exp(-\alpha_i e^x),\ x \in \mathbf{R},$$

and the Y such that the distribution function of $Z = \exp(Y)$ is G. In this case the sequence $R^{-1}(L_1), R^{-1}(L_2), \ldots$ is the Archimedean copula process.

In Sections 2.2 and 2.3, we obtain the class of functions B such that the function C given by (2.3) is an Archimedean copula for finite $d > 2$. Also, we investigate some degenerate cases. In Section 2.4 we present some examples of Archimedean copulas.

2.2 Statements of Main Results

The following theorem gives us the class of functions which can be used to generate a finite dimensional Archimedean copula.

Theorem 2.2.1 *Let $B : [0,\infty) \to [0,1]$ be a continuous and strictly decreasing function on $[0, A(0)]$ such that $B(0)=1$, $\lim_{t\to A(0)} B(t) = 0$ and $B(t) = 0$ for all $t \geq A(0)$. In this case, the function C given by (2.3) is an Archimedean copula iff B is a (d–2)-differentiable function on $(0,\infty)$ satisfying the conditions (2.5) for $k = 1, 2, \ldots, d-2$, and $(-1)^d B^{(d-2)}$ is a convex function.*

Now we consider the following example.

Example 2.2.1 Suppose that for $d = 2$

$$B(x) = \begin{cases} 1 - x, & x \in [0,1], \\ 0, & x > 1, \end{cases}$$

and the marginal distributions are both standard uniforms $U(0,1)$, i.e. $F_1(x) = F_2(x) = U(x)$, $x \in \mathbf{R}$, where $U(x)$ is as defined in (2.4). In this case, the function $A = B^{-1}$ is given by

$$A(x) = 1 - x,\ x \in [0,1].$$

It is clear that B is a convex function. Therefore, the function

$$F(x_1, x_2) = \begin{cases} 0, & x_1 \leq 0, \text{ or } x_2 \leq 0, \text{ or } x_1 + x_2 \leq 1, \\ x_1 + x_2 - 1, & x_1 \in (0,1],\ x_2 \in (0,1],\ x_1 + x_2 > 1, \\ x_1, & x_1 \in (0,1],\ x_2 > 1, \\ x_2, & x_1 > 1,\ x_2 \in (0,1], \\ 1, & x_1 > 1,\ x_2 > 1. \end{cases}$$

calculated by (1.2) and (1.3) is a two-dimensional distribution function, but simple calculations bring us that this function is a distribution function of the vector $(X, 1-X)$, where X is the uniformly distributed on $[0,1]$ ($U(0,1)$) random variable. Therefore, this distribution is concentrated on the segment $x_1 + x_2 = 1$, $x_1 \in [0,1]$ and the distribution function $F(x_1, x_2)$ is not absolutely continuous.

Remark 2.2.1 If we choose B as in Example 2.2.1, we obtain for absolutely continuous distribution functions F_1 and F_2 that the function

$$F(x_1,x_2) = \begin{cases} 0, & \text{if } F(x_1) + F(x_2) \le 1, \\ F_1(x_1) + F_2(x_2) - 1, & \text{if } F(x_1) + F(x_2) > 1, \end{cases}$$

defined by (2.2) and (2.3) is a distribution function of the vector $(F_1^{-1}(X), F_2^{-1}(1-X))$ with X uniformly distributed on $[0,1]$. The distribution of this vector is concentrated on a one-dimensional manifold and this distribution function also is not absolutely continuous.

It is not difficult to see that the distribution function given by (2.2) and (2.3) is absolutely continuous for any absolutely continuous marginals $F_1, F_2, \dots, F_d$ if B is a monotone (d)-differentiable function satisfying the conditions (2.5) for $k = 1, 2, \dots, d$. In this case, the corresponding density function can be obtained as follows:

$$p(x_1, x_2, \dots, x_d) = B^{(d)}\Big(\sum_{j=1}^{d} A(F_j(x_j))\Big) \prod_{i=1}^{d} A'(F_i(x_i))\, p_i(x_i), \tag{2.6}$$

for all $x_1, x_2, \dots, x_d \in \mathbf{R}$, where p_i are the density functions corresponding to F_i, $i = 1, 2, \dots, d$, respectively.

Suppose that

$$\mathfrak{B}_n = \{t \in \mathbf{R}_+ : \text{ there exists } B^{(n)}(t)\}, \; n \in \mathbf{N}.$$

As was mentioned above, the function B must be $(d-2)$-differentiable and $(-1)^d B^{(d-2)}$ must be a convex function. Therefore, $\mathfrak{B}_{d-1}$ is an everywhere dense set. Consequently, for any $y \in \mathfrak{B}_{d-1}$, we can write left and right derivatives

$$B^{(d-1)}(y_-) = \lim_{\substack{t \in \mathfrak{A} \\ t \to y_-}} B^{(d-1)}(t) \quad \text{and} \quad B^{(d-1)}(y_+) = \lim_{\substack{t \in \mathfrak{A} \\ t \to y_+}} B^{(d-1)}(t).$$

The following theorem gives some properties of Archimedean copulas in the degenerate case.

Theorem 2.2.2 *Suppose that the function $F(x_1, x_2, \dots, x_d)$ is a distribution function defined by (2.2) and (2.3) of some random vector $X_1, X_2, \dots, X_d$ with absolutely continuous marginals $F_1, F_2, \dots, F_d$, respectively. Then for all $\bar{x} = (x_1, x_2, \dots, x_d) \in \mathbf{R}$ such that $\sum_{j=1}^{d} A(F_j(x_j)) \in \mathfrak{B}_d$, there exists a density function and it is given by (2.6). For $y \in \mathbf{R}_+ \backslash \mathfrak{B}_d$, we have*

$$\mathbf{P}\Big(\sum_{j=1}^{d} Z_j = y\Big) = \frac{y^d}{d!} \Delta_y B^{(d-1)},$$

where

$$\Delta_y B^{(d-1)} = | \, B^{(d-1)}(y_-) - B^{(d-1)}(y_+) \, |,$$

is the increment of $(d-1)$-th derivative of B in y, $Z_i = A(F_i(X_i))$, $i = 1, 2, \dots, d$, and for any $y \in \mathbf{R}_+ \backslash \mathfrak{B}_{d-1}$ the conditional distribution of the vector $(Z_1, Z_2, \dots, Z_{d-1})$ under the condition $Z_1 + Z_2 + \dots + Z_d = y$ is uniform in the set $\mathfrak{D} = \{t_1, t_2, \dots, t_{d-1} \geq 0 \, : \, \sum_{i=1}^{d-1} t_i < y\}$.

Remark 2.2.2 By Theorem 2.2.1, $(-1)^d B^{(d-2)}$ is a convex function. Therefore, under the condition that B is $(d-1)$-differentiable, it satisfies (2.5) for $k = 1, 2, \dots, d-1$, and $B^{(d-1)}$ is a monotone function. Under the condition that $B^{(d-1)}$ is absolutely continuous, the distribution function F defined by (2.2) and (2.3) with absolutely continuous marginals is absolutely continuous and the corresponding density function can be calculated by (2.6) for all $\bar{x}$'s, such that $\sum_{j=1}^{d} A(F_j(x_j)) \in \mathfrak{B}_d$. For other $\bar{x}$'s the density function can be defined arbitrary.

Corollary 2.2.1 *Under the conditions of Theorem 2.2.2, the distribution of the initial random vector is absolutely continuous iff B is $(d-1)$-differentiable and $B^{(d-1)}$ is an absolutely continuous function on $\mathbf{R}_+$.*

Corollary 2.2.2 *Under the conditions of Theorem 2.2.2, the distribution of the vector $(Z_1, Z_2, \dots, Z_d)$ under the condition $Z_1 + Z_2 + \dots + Z_d = y$ for $y \in \mathbf{R}_+ \backslash \mathfrak{B}_{d-1}$ is uniform in the set $\mathfrak{D}^* = \{t_1, t_2, \dots, t_d \geq 0 \, : \, \sum_{i=1}^{d} t_i = y\}$.*

Usually in survival analysis, distributions are given by survival functions. Suppose that $X_1, X_2, \dots, X_d$ is an Archimedean copula vector having the distribution function given by (2.2) and (2.3) with absolutely continuous marginals. Introduce the r.v.'s $Y_i : Y_i = -X_i$, $i \in \mathbb{N}$. Then as was mentioned by Bagdonavicius, Malov and Nikulin (1997), the random vector $(Y_1, Y_2, \dots, Y_d) : Y_i = -X_i$, $i = 1, 2, \dots, d$, has the survival function

$$\begin{aligned} S(x_1, x_2, \ldots, x_d) &= \mathbf{P}(Y_1 > x_1, Y_2 > x_2, \ldots, Y_d > x_d) \\ &= \mathbf{P}(X_1 \le -x_1, X_2 \le -x_2, \ldots, X_d \le -x_d) \\ &= B\Big(\sum_{i=1}^{d} A(F_i(-x_i))\Big) = B\Big(\sum_{i=1}^{d} A(S_i(x_i))\Big), \end{aligned}$$

where $S_1, S_2, \ldots, S_d$ are the survival functions of $Y_1, Y_2, \ldots, Y_d$. In this case, $(Y_1, Y_2, \ldots, Y_d)$ is said to be a survival Archimedean copula vector. Theorems 2.2.1 and 2.2.2 can be also formulated in terms of survival Archimedean copulas.

2.3 Proofs

When we prove Theorem 2.2.1, we lose no generality by assuming that the marginal distributions are all the standard uniform $U(0,1)$.

PROOF OF THEOREM 2.2.1. Let $(X_1, X_2, \ldots, X_d)$ be a random vector having distribution function F defined by (2.2) and (2.3) with the standard uniform marginals. Introduce the following notation

$$R(x) = A(U(x)), \quad x \in \mathbf{R},$$

where $U(x)$ is given by (2.4). Then for any $n \in \{2, 3, \ldots, d\}$, the vectors $(X_{i_1}, X_{i_2}, \ldots, X_{i_n})$ are Archimedean copula vectors with distribution functions

$$F_n(x_1, x_2, \ldots, x_n) = B\Big(\sum_{j=1}^{n} R(x_i)\Big), \quad x_1, x_2, \ldots, x_n \in \mathbf{R} \tag{2.7}$$

for all populations of indexes $\{i_1, i_2, \ldots, i_n\} \subseteq \{1, 2, \ldots, d\}$. We use induction based on the case $n = 2$ which was given by Schweizer and Sklar (1983).

Suppose that we have proved that the function B is $(n-2)$-differentiable function such that the conditions (2.5) take place for all $k = 1, 2, \ldots, n-2$, and $(-1)^{(n)} B^{(n-2)}$ is a convex monotone function for some $n \in \{2, 3, \ldots, d-1\}$. Then the distribution function of the vector $X_1, X_2, \ldots, X_{n+1}$ can be represented as follows:

$$\begin{aligned} B\Big(\sum_{j=1}^{n+1} R(x_i)\Big) &= \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_{n-2}} B^{(n-2)}\Big(\sum_{i=1}^{n-2} R(t_i) \\ &\quad + \sum_{j=n-1}^{n+1} R(x_j)\Big)\, dR(x_1)\, dR(x_2) \ldots dR(x_{n-2}). \end{aligned}$$

Using the convex property of $B^{(n-2)}$, the function under the integral can be represented for any fixed $t_1, t_2, \ldots, t_{n-2}$ in the following way:

$$B^{(n-2)}\Big(\sum_{i=1}^{n-2} R(t_i) + \sum_{j=n-1}^{n+1} R(x_j)\Big) = \int\limits_{\substack{(-\infty,x_{n-1}]\\ \Sigma\in\mathfrak{B}_{n-1}}} B^{(n-1)}\Big(\sum_{i=1}^{n-1} R(t_i) + R(x_n) + R(x_{n+1})\Big)\, dt_{n-1},$$

where $\Sigma = \sum_{i=1}^{n-1} R(t_i) + R(x_n) + R(x_{n+1})$. Therefore, it is possible to define the conditional distribution function a.s. on $t_1, t_2, \dots, t_{n-1}$:

$$\begin{aligned}\tilde{F}_{\bar{t}}(x_1,x_2) &= \mathbf{P}(X_{n+1} < x_1, X_n < x_2 \mid X_1 = t_1, X_2 = t_2, \dots, X_{n-1} = t_{n-1})\\ &= \frac{B^{(n-1)}\Big(\sum_{i=1}^{n-1} R(t_i) + R(x_n) + R(x_{n+1})\Big)}{B^{(n-1)}\Big(\sum_{i=1}^{n-1} R(t_i)\Big)}, \qquad x_{k+1}, x_k \in \mathbf{R}.\end{aligned}$$

In this case

$$\begin{aligned}&\mathbf{P}(X_{n+1} \in (x,y), X_n \in (x,y) \mid X_1 = t_1, X_2 = t_2, \dots, X_{n-1} = t_{n-1})\\ &\quad= \Big[B^{(n-1)}\Big(\sum_{i=1}^{n-1} R(t_i)+R(y)+R(y)\Big)+B^{(n-1)}\Big(\sum_{i=1}^{n-1} R(t_i)+R(x)+R(x)\Big)\\ &\qquad - 2\,B^{(n-1)}\Big(\sum_{i=1}^{n-1} R(t_i) + R(x) + R(y)\Big)\Big] \Big/ B^{(n-1)}\Big(\sum_{i=1}^{n-1} R(t_i)\Big). \qquad (2.8)\end{aligned}$$

By the properties of distribution functions,

$$\mathbf{P}(X_{k+1} \in (x,y), X_k \in (x,y) \mid X_1 = t_1, X_2 = t_2, \dots, X_{n-1} = t_{n-1}) \ge 0. \quad (2.9)$$

The left side of this inequality is given by (2.8) a.s. for $(t_1, t_2, \dots, t_{n-1}) \in \mathbf{R}^{n-1}$. By letting $t_i \to \infty$, $i = 1, 2, \dots, n-1$, we obtain by (2.8) and (2.9) that

$$\frac{B^{(n-1)}(2\,R(y)) + B^{(n-1)}(2\,R(x)) - 2\,B^{(n-1)}(R(x) + R(y))}{B^{(n-1)}\Big(\sum_{i=1}^{n-1} R(t_i)\Big)} \ge 0$$

or

$$(-1)^{(n-1)}\Big(B^{(n-1)}(2\,R(y)) + B^{(n-1)}(2\,R(x)) - 2\,B^{(n-1)}(R(x) + R(y))\Big) \ge 0$$

a.s. for all $x, y \in \mathbf{R}$. Therefore, the function $(-1)^{n-1}\,B^{(n-1)}$ is a convex function. Then $B^{(n-1)}$ is continuous and satisfy the condition (2.5) for $k =$

$n-1$. By the condition $\lim_{t\to\infty} B(t) = 0$, we obtain that $\lim_{t\to\infty} B^{(n-1)}(t) = 0$ and $(-1)^{n-1} B^{(n-1)}$ is a monotone function.

Therefore, B is a $(d-2)$-differentiable monotone function satisfying the condition (2.5) for all $k = 1, 2, \ldots, d-2$, and $(-1)^d B^{(d-2)}$ is a convex function.

Conversely, let $B : [0, \infty) \to [0, 1]$ be a continuous function, strictly decreasing on $[0, A(0)]$ and such that $B(0) = 1$, $\lim_{t\to A(0)} B(t) = 0$ and $B(t) = 0$ for all $t \geq A(0)$. Also we assume that it is $(d-2)$-differentiable and satisfy the condition (2.5) for all $k = 1, 2, \ldots, d-2$ and $(-1)^d B^{(d-2)}$ is a convex function. It is easy to see that the function $F(x_1, x_2, \ldots, x_n)$ defined by (2.2) and (2.3) is continuous and satisfies the following conditions:

$$\lim_{x_i \to -\infty} F(x_1, x_2, \ldots, x_d) = 0 \text{ for all } i = 1, 2, \ldots, d.$$

and

$$\lim_{\substack{x_i \to \infty \\ i=1,2,\ldots,d}} F(x_1, x_2, \ldots, x_d) = 1.$$

Now we need to prove that the condition (2.1) holds for an arbitrary chosen set $\mathfrak{A} = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_d, b_d] \in \mathbf{R}^d$ such that $a_i \leq b_i$ for all $i = 1, 2, \ldots, d$.

Introduce, for any fixed $c \geq 0$, the function $\tilde{B}_c$ such that

$$\tilde{B}_c(t) = (-1)^d B^{(d-2)}(c + t), \quad t \in \mathbf{R}_+.$$

Then $\tilde{B}_c(t)$ is a convex function for any $c \geq 0$. Following the proof for $d = 2$ given in Schweizer and Sklar (1983), we obtain that for all $[x_1, y_1] \times [x_2, y_2] \subseteq \mathbf{R}^2$ the condition

$$\begin{aligned}\tilde{B}_c(R(y_1) + R(y_2)) &+ \tilde{B}_c(R(x_1) + R(x_2)) \\ &- \tilde{B}_c(R(x_1) + R(y_2)) - \tilde{B}_c(R(y_1) + R(x_2)) \geq 0 \quad (2.10)\end{aligned}$$

holds for any $c \geq 0$.

The function $F(x_1, x_2, \ldots, x_d)$ can be represented as

$$\begin{aligned}&F(x_1, x_2, \ldots, x_d) \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_{d-2}} B^{(d-2)}\Big(\sum_{i=1}^{d-2} R(t_i) + R(x_{d-1}) + R(x_d)\Big) \\ &\qquad\qquad dR(t_1) dR(t_2) \ldots dR(t_{d-2}).\end{aligned}$$

Then for any $\mathfrak{A} = [a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_d, b_d] \in \mathbf{R}^d$ such that $a_i \leq b_i$ for all $i = 1, 2, \ldots, d$, the left hand side of (2.1) can be rewritten as follows:

$$\sum_{\bar v} \mathbf{sgn}_{\mathfrak{A}}(\bar v)\,F(\bar v) = \int_{a_1}^{b_1}\int_{a_2}^{b_2}\dots\int_{a_{d-2}}^{b_{d-2}} \Big[B^{(d-2)}\Big(\sum_{i=1}^{d-2} R(t_i) + R(b_{d-1}) + R(b_d)\Big)$$
$$+ B^{(d-2)}\Big(\sum_{i=1}^{d-2} R(t_i) + R(a_{d-1}) + R(a_d)\Big)$$
$$- B^{(d-2)}\Big(\sum_{i=1}^{d-2} R(t_i) + R(b_{d-1}) + R(a_d)\Big)$$
$$- B^{(d-2)}\Big(\sum_{i=1}^{d-2} R(t_i) + R(a_{d-1}) + R(b_d)\Big)\Big]$$
$$dR(t_1)\,dR(t_2)\dots dR(t_{d-2}).$$

By choosing $c = \sum_{i=1}^{d-2} R(t_i)$ in (2.10) we obtain that the function under this integral is nonnegative for odd d and nonpositive for even d. Therefore, the last integral is nonnegative and the condition (2.1) holds. Theorem is proved. ∎

Proof of Theorem 2.2.2. Let $(X_1, X_2, \dots, X_d)$ be a random vector with a distribution function F given by (2.2) and (2.3). It is clear that under the condition that B is a (d)-differentiable function at some $y \in \mathbf{R}_+$, the function $F(x_1, x_2, \dots, x_d)$ defined by (2.2) and (2.3) has a d-th partial derivative with respect to $x_1, x_2, \dots, x_d$ for all $\bar a = (a_1, a_2, \dots, a_d)$ such that $\sum_{i=1}^{d} A(F_i(a_i)) = y$, which can be calculated in the following way:

$$\frac{\partial F(x_1, x_2, \dots, x_d)}{\partial x_1\,\partial x_2, \dots, \partial x_d}\bigg|_{\bar a} = B^{(d)}\Big(\sum_{i=1}^{d} A(F_i(a_i))\Big) \prod_{i=1}^{d} A'(F_i(a_i))\,\frac{dF_i(x)}{dx}\bigg|_{a_i}.$$

Therefore, the density function for any $\bar x \in \mathfrak{B}_d$ can be calculated by (2.6).

Further, without loss of generality we assume that the random vector $(X_1, X_2, \dots, X_d)$ has the standard uniform marginal distributions [i.e. the distribution function of this vector is defined by (2.7)]. By Theorem 2.2.1, the function $B^{(d-2)}$ is convex. Therefore, $B^{(d-2)}$ is absolutely continuous and

$$\mathbf{P}\Big(\sum_{i=1}^{d-1} R(X_i) \in \mathfrak{B}_{d-1}\Big) = 0.$$

Introduce for all $y \in \mathfrak{B}_{d-1}$ and for all $x_1, x_2, \dots, x_{d-1}$ a.s. the conditional distribution function of the sum

$$\mathbf{P}\Big(\sum_{i=1}^{d} R(X_i)<y \mid X_1=x_1, X_2=x_2, \ldots, X_{d-1}=x_{d-1}\Big)$$
$$= \begin{cases} \dfrac{B^{(d-1)}(y)}{B^{(d-1)}\Big(\sum\limits_{i=1}^{d-1} R(x_i)\Big)}, & \sum\limits_{i=1}^{d-1} R(x_i)<y \\ 0, & \text{in other cases.} \end{cases}$$

Therefore, for any $y \in \mathfrak{B}_{d-1}$

$$\begin{aligned} &\mathbf{P}\Big(\sum_{i=1}^{d} R(X_i) < y\Big) \\ &= \int\int\cdots\int_{R(t_1)+R(t_2)+\ldots+R(t_{d-1})<y} B^{(d-1)}(y)\, dR(t_1)dR(t_2)\ldots dR(t_{d-1}) \\ &= B^{(d-1)}(y) \int\int\cdots\int_{R(t_1)+R(t_2)+\ldots+R(t_{d-1})<y} dR(t_1)dR(t_2)\ldots dR(t_{d-1}) \\ &= \frac{y^{d-1}}{(d-1)!} B^{(d-1)}(y). \end{aligned}$$

As was mentioned above, the set $\mathfrak{B}_{d-1}$ is everywhere dense in $\mathbf{R}_+$. Thus, we can find two sequences $\{y_i\}$ and $\{z_i\}$ such that $y_i \xrightarrow[i\to\infty]{} y_-$, $z_i \xrightarrow[i\to\infty]{} y_+$ and $y_i, z_i \in \mathfrak{B}_{n-1}$ for all $i \in \mathbf{N}$. Consequently,

$$\begin{aligned} &\mathbf{P}(Z_1 + Z_2 + \ldots + Z_d=y) \\ &= \lim_{\substack{y_i\to y_- \\ z_i\to y_+}} \left| \frac{y_i^{d-1}}{(d-1)!} B^{(d-1)}(y_i) - \frac{z_i^{d-1}}{(d-1)!} B^{(d-1)}(z_i) \right| \\ &= \frac{y^{d-1}}{(d-1)!} \left| B^{(d-1)}(y_+) - B^{(d-1)}(y_-) \right| \\ &= \frac{y^{d-1}}{(d-1)!} \Delta_y B^{(d-1)}. \end{aligned}$$

Now we assume that $y \in \mathbf{R}\backslash\mathfrak{B}_{d-1}$. In this case, $\mathbf{P}(\sum_{j=1}^{d} Z_j = y) > 0$, where Z's were defined above, and we can write the conditional distribution function $\hat{F}(X_1, X_2, \ldots, X_{d-1})$ under the condition $\sum_{j=1}^{d} Z_j = y$ as follows:

$$\begin{aligned} &\hat{F}(x_1, x_2, \ldots, x_{d-1}) \\ &= \mathbf{P}\Big(Z_1 < x_1, Z_2 < x_2, \ldots, Z_{d-1} < x_{d-1}, \sum_{j=1}^{d} Z_j = y\Big) \Big/ \mathbf{P}\Big(\sum_{j=1}^{d} Z_j=y\Big) \end{aligned}$$

for all $x_1, x_2, \ldots, x_{d-1} \in \mathbf{R}$.

For any $t \in \mathfrak{B}_{d-1}$, it is possible to calculate that

$$\mathbf{P}\Big(Z_1 < x_1, Z_2 < x_2, \dots, Z_{d-1} < x_{d-1}, \sum_{j=1}^{d} Z_j < t\Big)$$
$$= \underset{\substack{t_i \in [0,x_i), i=1,2,\dots,d-1 \\ t_1+t_2+\dots+t_{d-1}<t}}{\int\int\dots\int} B^{(d-1)}(t)\, dt_1 dt_2 \dots dt_{d-1}$$
$$= B^{(d-1)}(t) \underset{\substack{t_i \in [0,x_i), i=1,2,\dots,d-1 \\ t_1+t_2+\dots+t_{d-1}<t}}{\int\int\dots\int} dt_1 dt_2 \dots dt_{d-1}.$$

Then,

$$\mathbf{P}\Big(Z_1 < x_1, Z_2 < x_2, \dots, Z_{d-1} < x_{d-1}, \sum_{j=1}^{d} Z_j = y\Big)$$
$$= \mathbf{P}\Big(Z_1 < x_1, Z_2 < x_2, \dots, Z_{d-1} < x_{d-1}, \sum_{j=1}^{d} Z_j < y_+\Big)$$
$$-\mathbf{P}\Big(Z_1 < x_1, Z_2 < x_2, \dots, Z_{d-1} < x_{d-1}, \sum_{j=1}^{d} Z_j < y_-\Big)$$
$$= \Delta_y B^{(d-1)} \underset{\substack{t_i \in [0,x_i), i=1,2,\dots,d-1 \\ t_1+t_2+\dots+t_{d-1}<t}}{\int\int\dots\int} dt_1 dt_2 \dots dt_{d-1}$$

and

$$\mathbf{P}\Big(Z_1<x_1, Z_2<x_2, \dots, Z_{d-1}<x_{d-1} \;\Big|\; \sum_{j=1}^{d} Z_j = y\Big)$$
$$= \frac{(d-1)!}{y^{d-1}} \underset{\substack{t_i \in [0,x_i), i=1,2,\dots,d-1 \\ t_1+t_2+\dots+t_{d-1}<t}}{\int\int\dots\int} dx_1 dx_2 \dots dx_{d-1}.$$

Consequently, the conditional distribution of $(Z_1, Z_2, \dots, Z_{d-1})$ under the condition $\sum_{j=1}^{d} Z_j = y$ is uniform in $\mathfrak{D} = \{t_1, t_2, \dots, t_{d-1} \geq 0 : \sum_{i=1}^{d} t_i \leq y\}$. Theorem is proved. ■

2.4 Some Examples

Now we represent some examples of finite dimensional Archimedean copulas.

Example 2.4.1 Let the function $B : [0, \infty) \to [0, 1]$ be such that

$$B(t) = \begin{cases} (t-1)^2, & t \in [0,1], \\ 0, & \text{in other cases.} \end{cases}$$

By Theorem 2.2.1, this function can generate Archimedean copulas only for $d = 2$ and $d = 3$. For $d = 2$, the direct calculation (2.3) gives us the function $C : [0, 1]^2 \to [0, 1]$ such that

$$C(x_1, x_2) = \begin{cases} (\sqrt{x_1} + \sqrt{x_2} - 1)^2, & \sqrt{x_1} + \sqrt{x_2} > 1, \\ 0, & \text{in other cases.} \end{cases}$$

By Corollary 2.2.1, the distribution function F defined by (2.7), viz.

$$F(x_1, x_2) = \begin{cases} 0, & x_1 \le 0 \text{ or } x_2 \le 0, \\ 0, & x_1 > 0,\ x_2 > 0,\ \sqrt{x_1} + \sqrt{x_2} \le 1, \\ (\sqrt{x_1} + \sqrt{x_2} - 1)^2, & x_1 \le 0,\ x_2 \le 0,\ \sqrt{x_1} + \sqrt{x_2} > 1, \\ x_1, & 0 < x_1 \le 1,\ x_2 > 1, \\ x_2, & x_1 > 1,\ 0 < x_2 \le 1, \\ 1, & \text{in other cases,} \end{cases}$$

of an Archimedean copula vector with $U(0, 1)$ marginals is absolutely continuous and the density function of this vector has the following form:

$$p(x_1, x_2) = \begin{cases} \dfrac{1}{2\sqrt{x_1 x_2}}, & 0 < x_1 \le 1,\ 0 < x_2 \le 1,\ \sqrt{x_1} + \sqrt{x_2} > 1, \\ 0, & \text{in other cases.} \end{cases}$$

In the case $d = 3$, we obtain that

$$C(x_1, x_2, x_3) = \begin{cases} (\sqrt{x_1} + \sqrt{x_2} + \sqrt{x_3} - 2)^2, & \sqrt{x_1} + \sqrt{x_2} + \sqrt{x_3} > 2, \\ 0, & \text{in other cases.} \end{cases}$$

Then the distribution function F defined in (2.7) is

$$F(x_1, x_2, x_3) = \begin{cases} (\sqrt{x_1} + \sqrt{x_2} + \sqrt{x_3} - 2)^2, & x_1 \in (0,1],\ x_2 \in (0,1],\ x_3 \in (0,1] \\ & \sqrt{x_1} + \sqrt{x_2} + \sqrt{x_3} > 2, \\ (\sqrt{x_1} + \sqrt{x_2} - 1)^2, & x_1 \in (0,1],\ x_2 \in (0,1],\ x_3 > 1, \\ & \sqrt{x_1} + \sqrt{x_2} > 1, \\ (\sqrt{x_1} + \sqrt{x_3} - 1)^2, & x_1 \in (0,1],\ x_2 > 1,\ x_3 \in (0,1], \\ & \sqrt{x_1} + \sqrt{x_3} > 1, \\ (\sqrt{x_2} + \sqrt{x_3} - 1)^2, & x_1 > 1,\ x_2 \in (0,1],\ x_3 \in (0,1], \\ & \sqrt{x_2} + \sqrt{x_3} > 1, \\ x_1, & x_1 \in (0,1],\ x_2 > 1,\ x_3 > 1, \\ x_2, & x_1 > 1,\ x_2 \in (0,1],\ x_3 > 1, \\ x_3, & x_1 > 1,\ x_2 > 1,\ x_3 \in (0,1], \\ 1, & x_1 > 1,\ x_2 > 1,\ x_3 > 1, \\ 0, & \text{in other cases.} \end{cases}$$

It is easy to see that the absolutely continuous component of a vector (X_1, X_2, X_3) having distribution function F is 0, and by Theorem 2.2.2 we obtain that $\sqrt{X_1} + \sqrt{X_2} + \sqrt{X_3} = 2$.

In the following example, the function A from Example 2.4.1 is taken as a generator of an Archimedean copula.

Example 2.4.2 Suppose that the function $B : [0, \infty) \to [0, 1]$ is defined by the following relation

$$B(t) = \begin{cases} 1 - \sqrt{t}, & t \in [0, 1], \\ 0, & \text{in other cases.} \end{cases}$$

It is clear that the function B is convex but it is not differentiable for $t = 1$ $[B'(1_-) = -1/2$ and $B'(1_+) = 0]$. Therefore, the function B can generate an Archimedean copula only for $d = 2$. This copula has the following form:

$$C(x_1, x_2) = \begin{cases} 1 - \sqrt{(1-x_1)^2 + (1-x_2)^2}, & (1-x_1)^2 + (1-x_2)^2 \le 1, \\ 0, & \text{in other cases.} \end{cases}$$

Calculate the corresponding distribution function with $U(0, 1)$ marginals

$$F(x_1, x_2) = \begin{cases} 0, & x_1 \le 0 \text{ or } x_2 \le 0, \\ 0, & x_1 \in (0, 1],\ x_2 \in (0, 1], \\ & (1-x_1)^2 + (1-x_2)^2 > 1, \\ 1 - \sqrt{(1-x_1)^2 + (1-x_2)^2}, & x_1 \in (0, 1],\ x_2 \in (0, 1], \\ & (1-x_1)^2 + (1-x_2)^2 \le 1, \\ x_1, & x_1 \in (0, 1],\ x_2 > 1, \\ x_2, & x_1 > 1,\ x_2 \in (0, 1], \\ 1, & \text{in other cases.} \end{cases}$$

Then,

$$\frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = \begin{cases} \dfrac{(x_1 - 1)(x_2 - 1)}{\left((x_1-1)^2 + (x_2-1)^2\right)^{3/2}}, & x_1 \in (0, 1], x_2 \in (0, 1],\ (1-x_1)^2 + (1-x_2)^2 < 1, \\ 0, & x_1 \notin (0, 1] \text{ or } x_2 \notin (0, 1] \\ & \text{or } (1-x_1)^2 + (1-x_2)^2 > 1. \end{cases}$$

Let (X_1, X_2) be the random vector with distribution function F. By Theorem 2.2.2, we obtain that $\mathbf{P}((1 - X_1)^2 + (1 - X_2)^2 = 1) = 1/2$. This fact is easy to obtain by direct calculations:

$$\int\limits_{\substack{x_1\in(0,1],x_2\in(0,1]\\(1-x_1)^2+(1-x_2)^2<1}}\int \frac{(x_1-1)(x_2-1)\,dx_1dx_2}{\left((x_1-1)^2+(x_2-1)^2\right)^{3/2}}$$

$$=\int_0^1 dr\int_0^{\pi/2}\sin\phi\cos\phi\,d\phi=1/2=1-1/2.$$

In the following example we consider the copula generated by a function which is not differentiable at two points.

Example 2.4.3 Let the function $B:[0,\infty)\to[0,1]$ be defined as follows:

$$B(t)=\begin{cases}(2t-1)^2, & t\in[0,1/4],\\ \dfrac{1-\sqrt{t}}{2}, & t\in(1/4,1].\end{cases}$$

The function B is convex, but it has no derivative at the points $t=1/4$ and $t=1$. The copula generated by B is

$$C(x_1,x_2)$$
$$=\begin{cases}\dfrac{1-\sqrt{(1-2x_1)^2+(1-2x_2)^2}}{2}, & \begin{matrix}x_1\in(0,1/4],\ x_2\in(0,1/4],\\(1-2x_1)^2+(1-2x_2)^2\le 1,\end{matrix}\\ 1-\sqrt{(1-2x_1)^2+(1-\sqrt{x_2})/2}, & \begin{matrix}x_1\in(0,1/4],\ x_2\in(1/4,1],\\(1-2x_1)^2+(1-\sqrt{x_2})/2\le 1,\end{matrix}\\ 1-\sqrt{(1-\sqrt{x_2})/2+(1-2x_1)^2}, & \begin{matrix}x_1\in(1/4,1],\ x_2\in(0,1/4],\\(1-2x_2)+(1-\sqrt{x_1})/2\le 1,\end{matrix}\\ \left(1-\sqrt{1-(\sqrt{x_2}+\sqrt{x_1})/2}\right)\Big/2, & \begin{matrix}x_1\in(1/4,1],\ x_2\in(1/4,1],\\\sqrt{x_1}+\sqrt{x_2}\le 3/2,\end{matrix}\\ (1-(\sqrt{x_1}+\sqrt{x_2}))^2, & \begin{matrix}x_1\in(1/4,1],\ x_2\in(1/4,1],\\\sqrt{x_1}+\sqrt{x_2}>3/2,\end{matrix}\\ 1, & x_1>1 \text{ or } x_2>1 \text{ or } x_3>1.\\ 0, & \text{in other cases.}\end{cases}$$

Then the distribution function given by (2.7) with $U(0,1)$ marginals is

$$F(x_1,x_2)=\begin{cases}0, & x_1\le 0 \text{ or } x_1\le 0,\\ C(x_1,x_2), & x_1\in(0,1],\ x_2\in(0,1],\\ x_1, & x_1\in(0,1],\ x_2>1,\\ x_2, & x_1>1,\ x_2\in(0,1],\\ 1, & x_1>1,\ x_2>1.\end{cases}$$

Let (X_1,X_2) be the Archimedean copula vector having the distribution function F. In this example, $A(t)=B(t),\ t\in(0,1)$. Also, it is important to mention that $\Delta_{1/4}B = 3/2$ and $\Delta_1 B = 1/4$. Therefore, by Theorem 2.2.2, $\mathbf{P}(A(X_1) + A(X_2) = 1/4) = 3/8$ and $\mathbf{P}(\sqrt{X_1} + \sqrt{X_2} = 3/2) = 1/4$. Thus, 3/8 of unit measure is concentrated on the manifold $A(x_1) + A(x_2) = 1/4$, $x_1 \in [0,1]$, and 1/4 of unit measure is concentrated on the manifold $\sqrt{x_1} + \sqrt{x_2} = 3/2$, $x_1 \in [1/4, 1]$. For any other point $(x_1, x_2) \in \mathbf{R}^2$, the derivative of $F(x_1, x_2)$ exists:

$$\frac{\partial^2 F(x_1,x_2)}{\partial x_1 \partial x_2} = \begin{cases} \dfrac{2(1-2x_1)(1-2x_2)}{\left((1-2x_1)^2 + (1-2x_2)^2\right)^{3/2}}, & x_1 \in (0,1/4],\ x_2 \in (0,1/4],\ (1-2x_1)^2 + (1-2x_2)^2 \le 1/4, \\ \dfrac{1-2x_1}{4\sqrt{x_2((1-2x_1)^2 + (1-\sqrt{x_2})/2)^3}}, & x_1 \in (0,1/4],\ x_2 \in (1/4,1],\ (1-2x_1)^2 \le (1-\sqrt{x_2})/2, \\ \dfrac{1-2x_2}{4\sqrt{x_1((1-\sqrt{x_1})/2 + (1-2x_2)^2)^3}}, & x_1 \in (1/4,1],\ x_2 \in (0,1/4],\ (1-2x_2)^2 \le (1-\sqrt{x_1})/2, \\ \dfrac{1}{128\sqrt{x_1 x_2 (1-(\sqrt{x_1}+\sqrt{x_2})/2)^3}}, & x_1 \in (1/4,1],\ x_2 \in (1/4,1],\ \sqrt{x_1}+\sqrt{x_2} \le 3/2, \\ 1/(2\sqrt{x_1 x_2}), & x_1 \in (1/4,1],\ x_2 \in (1/4,1],\ \sqrt{x_1}+\sqrt{x_2} > 3/2, \\ 0, & \text{in other cases} \end{cases}$$

and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\partial^2 F(x_1,x_2)}{\partial x_1 \partial x_2}\, dx_1 dx_2 = 3/8.$$

Acknowledgement. This research was partially supported by the Russian Fund of Fundamental Investigations (grant N 96-01-00549).

References

1. Bagdonavičius V., Malov, S. and Nikulin, M. (1997). On characterizations and semiparametric regression estimation in Archimedean copula, *Technical Report 97007 of MAB*, University of Bordeaux 1, Bordeaux.

2. Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Second edition, New York: John Wiley & Sons.

3. Malov, S. V. (1998). Random variables generated by ranks in dependent schemes, *Submitted for publication.*

4. Moynihan, R. (1978). On τ_T-semigroups of probability distribution functions, II, *Aequationes Mathematicae*, **17**, 19–40.

5. Kimberling, C. H. (1973). Exchangeable events and completely monotonic sequences, *Rocky Mountain Journal of Mathematics*, **3**, 565–574.

6. Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity, Aequationes Mathematicae, **10**, 152–164.

7. Schweizer, B. and Sklar, A. (1983). *Probabilistic Metric Spaces*, Amsterdam: North-Holland.

8. Widder, D. V. (1946). *The Laplace Transform*, Princeton, New Jersey: Princeton University Press.

PART II

Characterizations of Distributions

3

Characterization and Stability Problems for Finite Quadratic Forms

G. Christoph, Yu. Prohorov, and V. Ulyanov

Otto–von–Guericke–Universität Magdeburg, Magdeburg, Germany
Steklov Mathematical Institute, Moscow, Russia
Moscow State University, Moscow, Russia

Abstract: Sufficient conditions are given under which the distribution of a finite quadratic form in independent identically distributed symmetric random variables defines uniquely the underlying distribution. Moreover, a stability theorem for quadratic forms is proved.

Keywords and phrases: Quadratic forms, characterization problem, stability problem

3.1 Introduction

Let $Z_1, \ldots, Z_n$ be independent identically distributed (i.i.d.) standard normal random variables and $a_1, \ldots, a_n$ be real numbers with $a_1^2 + \ldots + a_n^2 \neq 0$. Suppose that $X_1, \ldots, X_n$ are i.i.d. random variables such that

$$a_1 Z_1 + \ldots + a_n Z_n \stackrel{\mathrm{d}}{=} a_1 X_1 + \ldots + a_n X_n,$$

where $\stackrel{\mathrm{d}}{=}$ denotes the equality in distribution. Then, by Cramér's decomposition theorem for the normal law [see Linnik and Ostrovski (1972, Theorem 3.1.4)], the X_i are standard normal too.

Lukacs and Laha (1964, Theorem 9.1.1) considered a more general problem. Namely, let $X_1, \ldots, X_n$ be i.i.d. random variables such that their linear combination $L = a_1 X_1 + \ldots + a_n X_n$ has analytic characteristic function and

$$a_1^s + \ldots + a_n^s \neq 0 \quad \text{for all} \quad s = 1, 2, \ldots$$

Then the distribution of X_1 is uniquely determined by that of L.

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

The aim of this Chapter is to obtain a similar characterization property for quadratic forms in i.i.d. random variables $Z_1, \ldots, Z_n$. Furthermore, we state a stability property of such quadratic forms.

3.2 Notations and Main Results

Consider a symmetric matrix $A = (a_{ij})_{i,j=1}^{n}$. Let

$$Q(x_1, \ldots, x_n) = \sum_{i,j=1}^{n} a_{ij} x_i x_j$$

be a quadratic form in variables $x_1, \ldots, x_n$. Assume that Q is non-degenerate in the sense that A is not a zero matrix. Suppose $Z_1, \ldots, Z_n$ are i.i.d. random variables with a symmetric distribution F.

We say that a pair (Q, F) has a *characterization property* (CP) iff for a sequence of i.i.d. symmetric random variables $X_1, \ldots, X_n$, the equality

$$Q(Z_1, \ldots, Z_n) \stackrel{\mathrm{d}}{=} Q(X_1, \ldots, X_n) \tag{3.1}$$

implies

$$Z_1 \stackrel{\mathrm{d}}{=} X_1.$$

Remark 3.2.1 We require in the definition of CP that the random variables $X_1, \ldots, X_n$ are symmetric. Otherwise the problem does not have solution even in the case $n = 1$ and $Q(x_1) = x_1^2$. Equation (3.1) holds for $X_1 = Z_1$ as well as for $X_1 = |Z_1|$.

Remark 3.2.2 With a symmetric distribution F an answer is trivial in the one dimensional case, i.e. any pair (Q, F) has CP. Therefore we assume that $n \geq 2$ everywhere below.

In this Chapter, sufficient conditions are given under which the pair (Q, F) has CP. The solution of the problem depends also on the coefficients of the matrix A, where the following possibilities occur:

1. $a_{ii} = 0$ for all $i = 1, 2, \ldots, n$.

2. $a_{ii} \neq 0$ for some $i = 1, 2, \ldots, n$.

   2.1. $a_{11}^{2k+1} + a_{22}^{2k+1} + \ldots + a_{nn}^{2k+1} \neq 0$ for all $k = 0, 1, 2, \ldots$

   2.2. $a_{11} + a_{22} + \ldots + a_{nn} = 0$.

      2.2.1. $a_{ij} = 0$ for all $i \neq j$.

2.2.2. $a_{ij} \neq 0$ for some $i \neq j$.

2.3. $a_{11}^{2k+1} + a_{22}^{2k+1} + \ldots + a_{nn}^{2k+1} = 0$ for some $k = 1, 2, \ldots$

Here, we consider cases 1, 2.1 and 2.2.1.

Define now a class $\mathcal{F}$ of probability distributions so that $F \in \mathcal{F}$ iff the following two conditions are satisfied:

(C1) F has moments $\alpha_k = \int_{-\infty}^{\infty} x^k dF(x)$ of all orders k.

(C2) F is uniquely specified by $\alpha_1, \alpha_2, \ldots$

The following examples demonstrate when probability distribution $F \in \mathcal{F}$.

Example 3.2.1 If F has an analytic characteristic function, then $F \in \mathcal{F}$.

Remember [see Lukacs (1970, §7.2)] that a characteristic function is analytic iff

(i) the condition (C1) is satisfied and

(ii) $\limsup_{n\to\infty} \alpha_{2n}^{1/(2n)}/(2n) < \infty$.

The latter condition leads to (C2); see Lukacs and Laha (1964, Ch. 9).

We say that a probability distribution F satisfies *Cramér's condition* CC iff

$$\int_{-\infty}^{\infty} \exp\{h|x|\} dF(x) < \infty \quad \text{for some} \quad h > 0.$$

Example 3.2.2 Let F satisfies CC, then $F \in \mathcal{F}$.

It follows from the fact that F satisfies CC iff its characteristic function is analytic [see Lukacs (1970, §7.2)].

Example 3.2.3 If the moments $\{\alpha_k\}$ of F satisfy Carleman condition, i.e.

$$\sum_{n=1}^{\infty} \alpha_{2n}^{-1/(2n)} = \infty\,, \tag{3.2}$$

then $F \in \mathcal{F}$.

In fact, the condition (3.2) yields the uniqueness of the moment problem for $\mathcal{F}$; see, for example, Shohat and Tamarkin (1970, Theorem 1.10).

Note that Carleman condition is weaker than CC. Other examples of probability distributions belonging to $\mathcal{F}$ as well as detailed discussion concerning the moment problem and other related topics; see Akhiezer (1965), Feller (1971, Sec. VII.3) and Stoyanov (1987, Sec. 8.12 and 11).

Theorem 3.2.1 *Let $F \in \mathcal{F}$ and the matrix A be such that $a_{ii} = 0$ for all $i = 1, 2, \ldots, n$. Then, (Q, F) has CP.*

Example 3.2.4 Let Z_1, Z_2, Z_3 be i.i.d. standard normal random variables and X_1, X_2, X_3 be i.i.d. symmetric random variables such that

$$Z_1 Z_2 - Z_2 Z_3 \stackrel{\mathrm{d}}{=} X_1 X_2 - X_2 X_3 \,,$$

then by Theorem 3.2.1 the random variables X_1, X_2, X_3 are standard normal.

Theorem 3.2.2 *Let $F \in \mathcal{F}$ and the matrix A be such that for all $k = 0, 1, 2, \dots$ $a_{11}^{2k+1} + a_{22}^{2k+1} + \dots + a_{nn}^{2k+1} \neq 0$. Then, (Q, F) has CP.*

Example 3.2.5 Let Z_1, Z_2 be i.i.d. random variables with distribution F and density function

$$p(x) = (1/4)\,\exp\{-\,|x|^{1/2}\}, \quad x \in (-\,\infty\,,\,\infty) \tag{3.3}$$

Then, $F \in \mathcal{F}$; see Stoyanov (1987, p. 98).

Let X_1, X_2 be i.i.d. symmetric random variables such that

$$2\,Z_1^2 \,+\, 4\,Z_1\,Z_2 \,-\, Z_2^2 \stackrel{\mathrm{d}}{=} 2\,X_1^2 \,+\, 4\,X_1\,X_2 \,-\, X_2^2\,.$$

Then by Theorem 3.2.2, the random variables X_1 and X_2 have the density function defined in (3.3) too.

Theorem 3.2.3 *Let $a_{ii} \neq 0$ for some $i = 1, 2, \dots, n$, but $a_{11} + a_{22} + \dots + a_{nn} = 0$ and $a_{ij} = 0$ for all $i \neq j$. Then for any F, the pair (Q, F) does not have CP.*

Example 3.2.6 Let Z be a random variable with symmetric distribution F independent of the random variable ζ with $\mathrm{P}(\zeta = 1) = \mathrm{P}(\zeta = -\,1) = 1/2$ and let $c > 0$ be a real constant. Put

$$X = \zeta\,(Z^2 \,+\, c)^{1/2}\,.$$

Suppose now that both $Z, Z_1, Z_2, \dots, Z_n$ are i.i.d. and $X, X_1, X_2, \dots, X_n$ are i.i.d. too. Under the conditions of Theorem 3.2.3 varying the constant c, we find a family of symmetric distributions of X_1 such that (3.1) holds. In particular, if

$$Z_1^2 - Z_2^2 \stackrel{\mathrm{d}}{=} X_1^2 - X_2^2\,,$$

then the distributions of X_1 and Z_1 may differ.

Example 3.2.6 proves Theorem 3.2.3. The proofs of Theorems 3.2.1 and 3.2.2 are given in Section 3.4. They are based on the following:

a) If $F \in \mathcal{F}$, then X_1 has moments $\mathrm{E}X_1^k$ of all orders k.

b) Under the given conditions, we have

$$\mathrm{E}\,X_1^k = \mathrm{E}\,Z_1^k \quad \text{for all} \quad k = 1, 2, \dots$$

Moreover, we also prove also a stability theorem.

Theorem 3.2.4 *Suppose that the pair (Q, F) has CP. Let $X_{N,1}, \ldots, X_{N,n}$ for $N = 1, 2, \ldots$ be a series of i.i.d. symmetric random variables and*

$$Q(X_{N,1}, \ldots, X_{N,n}) \xrightarrow{d} Q(Z_1, \ldots, Z_n) \quad as \quad N \to \infty,$$

where $\xrightarrow{\mathrm{d}}$ denotes the convergence in distribution. Then,

$$X_{N,1} \xrightarrow{d} Z_1 \quad as \quad N \to \infty.$$

Theorem 3.2.4 will be proved in Section 3.4 using the tightness of the converging sequences of quadratic forms.

3.3 Auxiliary Results

At first, we give some simple relations for a quadratic form which enable us to remove undesirable elements to get inequalities between tail probabilities of X_1 and $Q(X_1, \ldots, X_n)$. Denote

$$\mathrm{tr}A = a_{11} + \ldots + a_{nn} \quad \text{and} \quad M = \max_{i,j} |a_{i,j}|.$$

Lemma 3.3.1 *We have*

$$a_{11}x_1^2 + \ldots + a_{nn}x_n^2 = 2^{-n} \sum_{\varepsilon(1,n)} Q(\varepsilon_1 x_1, \ldots, \varepsilon_n x_n) \tag{3.4}$$

and

$$a_{11}x_1^2 + \ldots + a_{nn}x_n^2 + 2a_{12}x_1x_2 = 2^{2-n} \sum_{\varepsilon(3,n)} Q(x_1, x_2, \varepsilon_3 x_3, \ldots, \varepsilon_n x_n), \tag{3.5}$$

where $\sum_{\varepsilon(i,n)}$ for $i \le n$ denotes the summation over all vectors $\varepsilon(i,n) = (\varepsilon_i, \ldots, \varepsilon_n)$ with $\varepsilon_j \in \{-1, 1\}$.

Lemma 3.3.2 *Assume that* $\mathrm{tr}A = 0$ *and put*

$$Q^*(x_1, \ldots, x_n) = a_{11}\,x_1^2 + \ldots + a_{nn}\,x_n^2 + 2\,a_{12}\,x_1\,x_2.$$

Then

$$\begin{aligned} &2\,a_{12}\,(x_1x_2 + x_nx_1 + \ldots + x_2x_3) \\ &= Q^*(x_1, x_2 \ldots, x_n) + Q^*(x_n, x_1 \ldots, x_{n-1}) + \ldots + Q^*(x_2, x_3 \ldots, x_n, x_1). \end{aligned}$$

Lemma 3.3.3 *Let $X_1, \dots, X_n$ be i.i.d. symmetric random variables. Then for any permutation $(i_1, \dots, i_n)$ of indices $(1, \dots, n)$ and any vector $(\varepsilon_1, \dots, \varepsilon_n)$ with $\varepsilon_j \in \{-1, 1\}$, we have*

$$Q(X_1, \dots, X_n) \stackrel{\mathrm{d}}{=} Q(\varepsilon_1 X_{i_1}, \dots, \varepsilon_n X_{i_n}).$$

Proofs of Lemmas 3.3.1–3.3.3 are obvious.

Lemma 3.3.4 *If $a_{11} \neq 0$, then*

$$|Q(x_1, \dots, x_n)| \geq 0.75\, |a_{11}| x_1^2 - c_1(A)(x_2^2 + \dots + x_n^2),$$

with $c_1(A) = \frac{4}{|a_{11}|} \sum_{j=2}^{n} a_{1j}^2 + \max_{2 \leq i \leq n} \left\{ \sum_{j=2}^{n} |a_{ij}| \right\}$.

PROOF. Since

$$Q(x_1, \dots, x_n) = a_{11} x_1^2 + 2x_1(a_{12}x_2 + \dots + a_{1n}x_n) + Q(0, x_2, \dots, x_n),$$

we find

$$|Q(x_1, \dots, x_n)| \geq |a_{11}| x_1^2 - |2x_1(a_{12}x_2 + \dots + a_{1n}x_n)| - |Q(0, x_2, \dots, x_n)|.$$

Using in the second term of the right hand side

$$2|ab| \leq a^2 + b^2 \text{ with } a = \frac{1}{2}\sqrt{|a_{11}|}\, x_1 \text{ and } b = \frac{2}{\sqrt{|a_{11}|}}(a_{12}x_2 + \dots + a_{1n}x_n),$$

and $2|a_{1i}a_{1j}x_i x_j| \leq a_{1i}^2 x_j^2 + a_{1j}^2 x_i^2$, we obtain

$$|2x_1(a_{12}x_2 + \dots + a_{1n}x_n)| \leq \frac{1}{4}|a_{11}| x_1^2 - \frac{4}{|a_{11}|} \sum_{j=2}^{n} a_{1j}^2 (x_2^2 + \dots + x_n^2).$$

The inequality $2|x_i x_j| \leq x_i^2 + x_j^2$ leads to

$$|Q(0, x_2, \dots, x_n)| \leq \max_{2 \leq i \leq n} \left\{ \sum_{j=2}^{n} |a_{ij}| \right\} (x_2^2 + \dots + x_n^2)$$

which completes the proof of Lemma 3.3.4. ∎

We now prove inequalities between tail probabilities of both X_1 and $Q_X = Q(X_1, \dots, X_n)$.

Lemma 3.3.5 *Let $X_1, \dots, X_n$ be i.i.d. symmetric random variables. For any positive u, we have*

$$\mathrm{P}\{|Q_X| \geq u\} \leq n\, \mathrm{P}\{X_1^2 \geq u/(M n^2)\}.$$

PROOF. The obvious inequality

$$|Q(x_1, x_2, \ldots, x_n)| \le n\, M\, (x_1^2 + \ldots + x_n^2) \tag{3.6}$$

leads to the statement. ■

Denote by $m = \mathrm{med} X_1^2$ a median of X_1^2, i.e.

$$\mathrm{P}\{X_1^2 \ge m\} \ge 1/2 \quad \text{and} \quad \mathrm{P}\{X_1^2 \le m\} \ge 1/2\,.$$

Lemma 3.3.6 *There are positive constants c_1 and c_2 depending only on the elements of matrix A such that*

(a) if $a_{ii} = 0$ for all $i = 1, 2, \ldots, n$, then for any $u \ge 0$ we have

$$\mathrm{P}^2\{|X_1| \ge \sqrt{u}\} \le 2^{n-2}\mathrm{P}\{|Q_X| \ge c_1\, u\}\,; \tag{3.7}$$

(b) if $a_{ii} \ne 0$ for some $i = 1, 2, \ldots, n$, then for any $u \ge 0$ we have

$$\mathrm{P}\{X_1^2 \ge u + c_2\,(n-1)\,m\} \le 2^{n-1}\,\mathrm{P}\{|Q_X| \ge c_1\, u\}\,. \tag{3.8}$$

PROOF. *Case a:* Since A is not a zero matrix, there exists $a_{ij} \ne 0$ with $i \ne j$. Without lose of generality, we may assume $a_{12} \ne 0$. Then using (3.5), we get

$$2|a_{12}X_1X_2| \le 2^{2-n} \sum_{\varepsilon(3,n)} |Q(X_1, X_2, \varepsilon_3 X_3, \ldots, \varepsilon_n X_n)|\,.$$

Therefore, by Lemma 3.3.3, we have for any positive u

$$\mathrm{P}^2\{|X_1| \ge \sqrt{u}\} \le \mathrm{P}\{2|a_{12}X_1X_2| \ge 2|a_{12}|u\} \le 2^{n-2}\mathrm{P}\{|Q_X| \ge 2|a_{12}|u\}.$$

Case b: Without loss of generality, we assume $a_{11} \ne 0$. Put

$$\alpha = 4\, c_1(A)(n-1)\, m\, /(3\, |a_{11}|)\,,$$

where $c_1(A)$ is defined in Lemma 3.3.4 and m is the median of X_1^2.
For any $u \ge 0$, we find now

$$\begin{aligned} \mathrm{P}\{X_1^2 \ge u + \alpha\} &\le 2^{n-1}\,\mathrm{P}\{X_1^2 \ge u + \alpha\,, X_2^2 \le m\,, \ldots, X_n^2 \le m\} \\ &\le 2^{n-1}\,\mathrm{P}\{|Q_X| \ge 0.75\,|a_{11}|\,(u+\alpha) - c_1(A)\,(n-1)\,m\} \\ &\le 2^{n-1}\,\mathrm{P}\{|Q_X| \ge 0.75\,|a_{11}|\,u\} \end{aligned}$$

and Lemma 3.3.6 is proved. ■

Using the last two lemmas, we find the following statement which is of its own interest.

Lemma 3.3.7 *Random variables X_1 and $|Q_X|^{1/2}$ satisfy or do not satisfy CC simultaneously.*

PROOF. It follows from Lemma 3.3.5 and the equality for any $h > 0$

$$\mathrm{E}\exp\{h|X|\} = 1 + h\int_0^\infty \exp\{hu\}\,\mathrm{P}\{|X| \geq u\}du \tag{3.9}$$

that $|Q_X|^{1/2}$ satisfies CC if X_1 satisfies CC.

Suppose now $\mathrm{E}\exp\{h_0|Q_X|^{1/2}\} < \infty$ and $a_{ii} = 0$ for all $i = 1, 2, \ldots, n$. By (3.9), (3.7) and the Markov inequality, we find

$$\begin{aligned}\mathrm{E}\exp\{hX_1\} &\leq 1 + 2^{(n-2)/2}h\int_0^\infty \exp\{hu\}\left(\mathrm{P}\{|Q_X|^{1/2} \geq c_1^{1/2}u\}\right)^{1/2}du \\ &\leq 1 + 2^{(n-2)/2}h\int_0^\infty \exp\{hu\}\left(\frac{\mathrm{E}\exp\{h_0|Q_X|^{1/2}\}}{\exp\{h_0c_1^{1/2}u\}}\right)^{1/2}du < \infty.\end{aligned}$$

Hence, CC holds with some $0 < h < h_0\,c_1^{1/2}$.

Let now $a_{ii} \neq 0$ for some $i = 1, 2, \ldots, n$. Then by (3.9), (3.8) and $\mathrm{P}\{|X_1| \geq u + c_2(n-1)m\} \leq \mathrm{P}\{|X_1|^2 \geq u^2 + c_2\,(n-1)m\}$, we find

$$\mathrm{E}\exp\{h|X_1|\} \leq c_3 + c_4\int_0^\infty e^{hu}\,\mathrm{P}\{|Q_X|^{1/2} \geq c_1^{1/2}u\}du$$

with some finite constants c_3 and c_4. Hence, X_1 satisfies CC. ■

Lemma 3.3.8 *Random variables X_1 and Q_X have moments of all orders simultaneously.*

PROOF. Let X_1 have moments of all orders, then by (3.6) $E|Q_X|^k < \infty$ for $k = 1, 2, \ldots$, too. If Q_X has moments of all orders, then the existence of the absolute moments of all orders of X_1 follows now from the equality

$$\mathrm{E}|X_1|^k = k\int_0^\infty u^{k-1}\mathrm{P}\{|X_1| \geq u\}du \quad \text{for any integer} \quad k \geq 1,$$

Lemma 3.3.6 and Markov inequality in the same way as in the second part of the proof of Lemma 3.3.7. ■

Lemma 3.3.9 *Let $a_{ii} = 0$ for all $i = 1, 2, \ldots, n$. Then, $\mathrm{E}\,Q_X^{2k}$ is an increasing function of $\beta_{2k} = \mathrm{E}\,X_1^{2k}$ for all $k = 1, 2, \ldots$*

PROOF. We have

$$\mathrm{E}\,Q_X^{2k} = B\beta_{2k}^2 + C\beta_{2k} + D \tag{3.10}$$

for all $k = 1, 2, \ldots$, where B, C and D depend on the matrix A and β_{2k-2l} with $l = 1, 2, \ldots, k-1$. It is enough to prove that $B > 0$ and $C \geq 0$.

We obtain

$$\mathrm{E}\,Q_X^{2k} = 2^{2k}\mathrm{E}\sum_{i'<j'} a_{i_1j_1}X_{i_1}X_{j_1}\ldots a_{i_{2k}j_{2k}}X_{i_{2k}}X_{j_{2k}}, \tag{3.11}$$

where $\sum\limits_{i'<j'}$ denotes the summation over all $2k$ pairs:

$$1 \le i_1 < j_1 \le n, \ldots, 1 \le i_{2k} < j_{2k} \le n\,.$$

It is clear from (3.11) that B in the representation (3.10) equals

$$B = 2^{2k} \sum_{1 \le i < j \le n} a_{ij}^{2k} > 0\,.$$

In order to prove that $C \ge 0$, we introduce notation for a finite set $M = \{m_1, \ldots, m_l\}$ of integers $m_1, \ldots, m_l$. Let $\#(M)$ be the number of elements in M and $\#^*(M)$ be the number of different elements in M. For example, if $M = \{3, 2, 2, 1\}$, then $\#(M) = 4$ and $\#^*(M) = 3$.

The coefficient C in (3.10) up to factors β_{2k-2l} with $l = 1, 2, \ldots, k-1$ is a sum of products $2^{2k}\, a_{i_1 j_1} \ldots a_{i_{2k} j_{2k}}$ [see (3.11)] such that the set of their indices $E = \{i_1, j_1, \ldots, i_{2k}, j_{2k}\}$ satisfies the following three conditions:

a) There is a subset $E_1 \subset E$ with $\#(E_1) = 2k$ and $\#^*(E_1) = 1$. This yields that a corresponding summand in (3.11) has a factor β_{2k}.

b) $\#^*(E \setminus E_1) \ge 2$. It implies that we consider a summand with factor β_{2k}, but not β_{2k}^2. Note that $\#^*(E) = \#^*(E \setminus E_1) + \#^*(E_1)$.

c) Each value from the set $E \setminus E_1$ is taken by even number of elements from $E \setminus E_1$. Otherwise, the corresponding expectation equals to zero since the random variables X_j, $j = 1, 2, \ldots, n$, are symmetric and independent.

It follows from the above three conditions that $C \ge 0$. Thus, Lemma 3.3.9 is proved. ■

A similar idea of monotony was used by Khakhubiya (1965).

3.4 Proofs of Theorems

Proof of Theorem 3.2.1. We get from (3.1) and Lemma 3.3.8 that X_1 has moments of all orders. Obviously, $\mathrm{E}X_1^{2k+1} = \mathrm{E}Z_1^{2k+1} = 0$ for all $k = 0, 1, 2, \ldots$ We now show that

$$\mathrm{E}\,X_1^{2k} = \mathrm{E}\,Z_1^{2k} \quad \text{for all} \quad k = 1, 2, \ldots \tag{3.12}$$

Comparing moments of $Q_X = Q(X_1, \ldots, X_n)$ and $Q_Z = Q(Z_1, \ldots, Z_n)$, we get (3.12). In fact, it follows from (3.1) that

$$\mathrm{E}\,Q_X^{2k} = \mathrm{E}\,Q_Z^{2k} \quad \text{for all} \quad k = 1, 2, \ldots \tag{3.13}$$

Since $a_{ii} = 0$ for all $i = 1, 2, \ldots, n$, in $\mathrm{E}\, Q_X^{2k}$ there occur only moments $\mathrm{E}\, X_1^j$ up to order $2k$. Taking $k = 1$ in (3.13), we get

$$(\mathrm{E}\, X_1^2)^2 \ \mathrm{tr}(A^2) = (\mathrm{E}\, Z_1^2)^2 \ \mathrm{tr}(A^2) .$$

Therefore, we obtain (3.12) for $k = 1$.

Then taking $k \geq 2$ in (3.13) and using Lemma 3.3.9, we get (3.12) for $k \geq 2$ by induction.

Since $F \in \mathcal{F}$, it is uniquely specified by its moments. We proved that all moments of X_1 and Z_1 coincide, respectively. Hence, the distribution of X_1 is uniquely defined by its moments too, and Theorem 3.2.1 is proved. ■

PROOF OF THEOREM 3.2.2. Similar to the proof of Theorem 3.2.1, it is enough to show that (3.12) holds. With (3.1), we find now

$$\mathrm{E}\, Q_X^k = \mathrm{E}\, Q_Z^k \quad \text{for all} \quad k = 1, 2, \ldots \tag{3.14}$$

Taking $k = 1$ in (3.14), we obtain

$$\mathrm{E}\, X_1^2 \ \mathrm{tr} A = \mathrm{E}\, Z_1^2 \ \mathrm{tr} A ,$$

i.e. we get (3.12) for $k = 1$ since $\mathrm{tr} A \neq 0$.

The proof of (3.12) for $k \geq 2$ can be done by induction using (3.14) and the conditions on the elements of matrix A. ■

PROOF OF THEOREM 3.2.4. Put

$$Q_{X,N} = Q(X_{N,1}, \ldots, X_{N,n}) \quad \text{and} \quad Q_Z = Q(Z_1, \ldots, Z_n) .$$

Since $Q_{X,N} \xrightarrow{d} Q_Z$ as $N \to \infty$, the sequence $\{Q_{X,N}\}$ is relatively compact. It is known [see Prohorov (1956)] that $\{Q_{X,N}\}$ is relatively compact if and only if $\{Q_{X,N}\}$ is tight, i.e.

$$\sup_N \mathrm{P}\{|Q_{X,N}| > v\} \to 0 \quad \text{as} \quad v \to +\infty . \tag{3.15}$$

In order to prove Theorem 3.2.4, it is enough to show that $\{X_{N,1}\}$ is also tight. In fact, in this case for any infinite subset of $\{X_{N,1}\}$ there exists a subsequence $\{X_{N_k,1}\}$ which converges in distribution to some symmetric random variable V_1. Since Q is continuous in each argument, we have

$$Q(X_{N_k,1}, \ldots, X_{N_k,n}) \xrightarrow{d} Q(V_1, \ldots, V_n) \quad \text{as} \quad N_k \to \infty,$$

where $V_1, \ldots, V_n$ are i.i.d. symmetric random variables. From the assumption of Theorem 3.2.4, we get

$$Q(V_1, \ldots, V_n) \stackrel{d}{=} Q(Z_1, \ldots, Z_n).$$

It yields $V_1 \stackrel{d}{=} Z_1$. Therefore, any limit point of $\{X_{N,1}\}$ has the same distribution as Z_1, which proves the statement of Theorem 3.2.4.

In order to prove the tightness of $\{X_{N,1}\}$, we consider two cases with respect to diagonal elements of A.

Case 1 if $a_{ii} = 0$ for all $i = 1, 2, \ldots, n$. It follows from *Case a* of Lemma 3.3.6 that $\{X_{N,1}\}$ is tight when (3.15) holds.

Case 2 if $a_{ii} \neq 0$ for some $i = 1, 2, \ldots, n$. It follows from (3.8) that the sequence $\{X_{N,1}^2 - c_2(n-1)\, m_N\}$ with $m_N = \operatorname{med}(X_{N,1}^2)$ is also tight when (3.15) holds. Therefore, it is enough to show that

$$\sup_N m_N \leq c < \infty \tag{3.16}$$

with some absolute constant c.

Put

$$\widehat{Q}(x_1, \ldots, x_n) = \sum_{i=1}^{n} (|a_{ii}| - a_{ii}) x_i^2 + Q(x_1, \ldots, x_n).$$

The quadratic form $\widehat{Q}$ differs from Q only by the diagonal elements of the matrix A, which are $|a_{ii}|$ in $\widehat{Q}$ instead of a_{ii} in $Q, i = 1, 2, \ldots, n$. Using (3.4) of Lemma 3.3.1 and Lemma 3.3.3, we get

$$\begin{aligned}
2^{-n} &\leq \mathrm{P}\{a_{11} X_{N,1}^2 \geq |a_{11}| m_N, \ldots, |a_{nn}| X_{N,n}^2 \geq |a_{nn}| m_N\} \\
&\leq \mathrm{P}\left\{\sum_{i=1}^{n} |a_{ii}| X_{N,i}^2 \geq m_N \sum_{i=1}^{n} |a_{ii}|\right\} \\
&\leq \mathrm{P}\left\{2^{-n} \sum_{\varepsilon(1,n)} |\widehat{Q}(\varepsilon_1 X_{N,1}, \ldots, \varepsilon_n X_{N,n})| \geq m_N \sum_{i=1}^{n} |a_{ii}|\right\} \\
&\leq 2^n \mathrm{P}\{|\widehat{Q}_{X,N}| \geq m_N \sum_{i=1}^{n} |a_{ii}|\}.
\end{aligned}$$

Comparing the last inequality with (3.15), we find (3.16). It proves the tightness of $\{X_{N,1}\}$ in this case too. Thus, Theorem 3.2.4 is proved. ■

Acknowledgements. This research was carried out with partial support from RFBR and INTAS under grants RFBR 96-01-01919 and INTAS-RFBR 95-0099.

References

1. Akhizer, N. I (1961). *Classical Problem of Moments*, Moscow: FML.

2. Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. 2, New York: John Wiley & Sons.

3. Khakhubiya, T. G. (1965). A lemma on random determinants and its application to the characterization of multivariate distributions, *Theory of Probability and Its Applications*, **10**, 685–689.

4. Linnik, Yu. V. and Ostrovski, I. V. (1972). *Decomposition of Random Variables and Vectors*, Moscow: Nauka. English Translation in *American Mathematical Society Translations*, **48** (1977), Providence, Rhode Island: American Mathematical Soceity.

5. Lukacs, E. (1970). *Characteristic Functions*, Second edition, London: Griffin.

6. Lukacs, E. and Laha, R. G. (1964). *Applications of Characteristic Functions*, London: Griffin.

7. Prohorov, Yu. (1956). Convergence of random processes and limit theorems in probability theory, *Theory of Probability and Its Applications*, **1**, 157–214.

8. Shohat, J. A. and Tamarkin, J. D. (1970). *The Problem of Moments*, Providence, Rhode Island: American Mathematical Society.

9. Stoyanov, J. M. (1987). *Counterexamples in Probability*, New York: John Wiley & Sons.

4

A Characterization of Gaussian Distributions by Signs of Even Cumulants

L. B. Klebanov and G. J. Szekely

St. Petersburg Architecture Building University, St. Petersburg, Russia
Bowling Green State University, Bowling Green, Ohio

Abstract: Let $f(t)$ be a characteristic function, analytic in some neighborhood of the origin, and let $\{\kappa_{2j}\}_{j=1}^{\infty}$ be a sequence of its even cumulants. According to a classical result of Marcinkiewitz, if all but finitely many cumulants are 0, then f is Gaussian. In this chapter, we prove the following generalization. Denote by λ_n the sequence of sign changes in the sequence $\{(-1)^j \kappa_{2j}\}_{j=1}^{\infty}$. If f has no zeros on the real line and $\sum_{n=1}^{\infty} 1/\lambda_n < \infty$, then f is Gaussian. We conjecture that for non-Gaussian characteristic functions f without zeros on the real line, there is a fixed j_0 such that $\kappa_{2j} > 0$ for all $j > j_0$.

Keywords and phrases: Cumulants, sign changes, Gaussian distribution

4.1 A Conjecture and Main Theorem

Let $f(t)$ be a characteristic function, analytic in some neighborhood of the origin and let $\{\kappa_{2j}\}_{j=1}^{\infty}$ be a sequence of its even cumulants (by definition, $\kappa_j = (-i)^j \frac{d^j \log f(t)}{dt^j}|_{t=0}$). By a classical result of Marcinkiewitz (1938), if $\kappa_{2j} = 0$ for all $j > j_0$, then the corresponding distribution is Gaussian. In this Chapter, we prove the following generalization. Denote by λ_n the sequence of sign changes in the sequence $\{\kappa_{2j}^*\}_{j=1}^{\infty}$, where $\kappa_{2j}^* = (-1)^j \kappa_{2j}$, that is, $\lambda_1 = 1$, $\lambda_2 = m$ if $\kappa_{2m}^* < 0$, and the inequality $\kappa_{2j}^* < 0$ does not hold for $j < m$, $\lambda_3 = n$ if $\kappa_{2n}^* > 0$, and the inequality $\kappa_{2j}^* > 0$ does not hold for $m < j < n$, etc.

Suppose that an entire characteristic function f has no zeros. If j is big enough, then κ_{2j} is strictly positive except the Gaussian case when $\kappa_{2j} = 0$ for $j = 1, 2, \ldots$.

For infinitely divisible distributions, this conjecture is a simple consequence of Ramachandran (1969). What we can prove here is the following theorem.

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

Theorem *Let $f(t)$ be a characteristic function, analytic in some neighborhood of the origin, and let λ_n be the sequence of sign changes of the sequence $\{\kappa^*_{2j}\}_{j=1}^{\infty}$. If f has no zeros on the real line and $\sum_{n=1}^{\infty} 1/\lambda_n < \infty$ then f is Gaussian.*

PROOF. First we suppose that f is a symmetric characteristic function.

We are going to show that $f(t)$ is an entire function. Suppose indirectly that this is not true, and let $\{t : |t| < R\}$ $(0 < R < \infty)$ be the maximal circle in which the function $\varphi(t) = \log f(t)$ is analytic. According to Fabri's theorem [see, for example, Bieberbach (1955)], if an analytic function has real Taylor coefficients and for the corresponding sequence λ_n of sign changes $1/\lambda_n \to 0$ as $n \to \infty$, then $t = R$ is a point of singularity of φ . But $f(t)$ is analytic in a circle centered at the origin, and (being a characteristic function) analytic in a strip containing the real line. Since f has no real zeros, the point $t = R$ cannot be a singularity of φ and, therefore, both φ and f are entire functions. Thus, f cannot have any complex zeros.

Since f is an entire function,

$$M(r) = M(r; f) = \max_{|t|\le r} |f(t)| = \max(|f(ir)|, |f(-ir)|).$$

Using the notation $T = \{z : |\arg z| \le \pi\sigma\}$ where $\sigma > 0$ is arbitrary, we have

$$\max_{|t|\le r,\ t\in T} |f(t)| \le M(r \sin(\pi\sigma)),$$

and therefore

$$\max_{|z|\le r,\ z\in T} Re\varphi(z) \le \log(M(r \sin \pi\sigma)).$$

According to Carathéodory's theorem [see Pólya and Szegő (1964)].

$$\max_{|z|\le r,\ z\in T} |\varphi(z)| \le C \log(M(r \sin \pi\sigma)).$$

Sheremeta (1975) obtained the following result. *Let an entire function f have real Taylor coefficients and suppose $\sum_{n=1}^{\infty} 1/\lambda_n < \infty$. Then,*

$$\lim_{x\to\infty} \frac{\log|\varphi(x)|}{\log M(x; \varphi)} = 1.$$

In our case, this implies

$$\log M(r; \varphi) \le C_1 \log(M(r \sin \pi\sigma))$$

and consequently

$$M(r) \le (M(r \sin \pi\sigma))^{C_1}.$$

But $\sigma > 0$ can be arbitrarily small and thus from the previous inequality, we see that f has finite order. Hence, f is an entire characteristic function of finite order and has no zeros. Thus by Marcinkiewicz's theorem, it is Gaussian.

To complete the proof, observe that if f is not necessarily real valued, then we can apply the proof above for the real valued $f(t)f(-t)$, and finally apply Cramér's classical result: all components of Gaussian distributions are Gaussian. ■

4.2 An Example

Let us now give an example showing that the condition of our Theorem "f has no zeros on the real line" is essential.

Consider the function $(1-t^2)e^{-t^2/2}$. It is easy to see that the function is a characteristic function of the distribution with the density $cx^2e^{-x^2/2}$, where c is a normalizing constant. The function $1/(1+t^2)$ is the characteristic function of Laplacian distribution. Therefore, it is clear that the function

$$f(t) = \frac{1-t^2}{1+t^2}e^{-t^2/2}$$

is a characteristic function of the corresponding convolution. We have

$$f(t) = \exp\left(\sum_{k=0}^{\infty}(\frac{1}{k+1} - (-1)^k\frac{1}{k+1})t^{2k+2} - t^2/2\right).$$

From here, we see that $\kappa_{2k}^* \geq 0$ for all $k = 1, 2, \ldots$. Now we see that the condition on the zeros in the Theorem is essential.

Acknowledgements. L. B. Klebanov's research was partially supported by RFFI Grants 95-01-01260, 96-01-00852. G. J. Szekely's research was supported by Hungarian NSF, OTKA T016237, and BGSU, OH FRC Grant.

References

1. Bieberbach, L. (1955). *Analytische Forsetzung*, Berlin: Springer-Verlag.

2. Marcinkiewitz, J. (1938). Sur un propriété de la loi de Gauss, *Mathematisch Zeitschrift*, **44**, 612–618.

3. Ramachandran, B. (1969). On characteristic functions and moments, *Sankhyā, Series A*, **31**, 1–12.

4. Sheremeta, M. N. (1975). A property of entire functions with real Taylor coefficients, *Mathematical Notes of the Sciences of the USSR*, **18**, 823–827.

5. Pólya, G. and Szegő, G. (1964). *Aufgaben and Lehrsätze aus der Analysis*, **1**, Berlin: Springer-Verlag.

5

On a Class of Pseudo-Isotropic Distributions

A. A. Zinger

Aerocosmic Academy, St. Petersburg, Russia

Abstract: A class of distributions of random vectors in $\mathbb{R}^n$, such that distribution of any linear statistics belongs to the same multiplicative type, is considered. Results are then developed for the description of translated moments of linear statistics.

Keywords and phrases: Linear statistics, pseudo-isotropic distribution, scale function, negative definite function, translated moments

5.1 Introduction

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random vector in $\mathbb{R}^n$ $(n \geq 2)$ and $L = t_1 X_1 + \cdots + t_n X_n$ be a linear statistic with coefficients vector $\boldsymbol{t} = (t_1, \ldots, t_n)$. We will be interested here in distributions of vectors $\boldsymbol{X}$, for which the distributions of linear statistics L belong to the same multiplicative type for any $\boldsymbol{t} \in \mathbb{R}^n$, i.e.,

$$P\{t_1 X_1 + \cdots + t_n X_n < x\} = P\{\boldsymbol{c}(t_1, \ldots, t_n)\xi < x\}, \; x \in \mathbb{R}, \tag{5.1}$$

for some $\boldsymbol{c} \in \mathbb{R}^n$ and random variable ξ. Distribution of $\boldsymbol{X}$, satisfying (5.1), will be called *pseudo-isotropic distribution*, and the corresponding function $\boldsymbol{c}$ will be called the *scale function*. Let us also denote $(\boldsymbol{x}, \boldsymbol{y}) = x_1 y_1 + \cdots + x_n y_n$, $|\boldsymbol{x}| = (\boldsymbol{x}, \boldsymbol{x})^{1/2}$ for any $\boldsymbol{x} = (x_1, \ldots, x_n)$, $\boldsymbol{y} = (y_1, \ldots, y_n) \in \mathbb{R}^n$. A simple example of the pseudo-isotropic distribution is spherically symmetric distributions, corresponding to the scale function $\boldsymbol{c}(\boldsymbol{t}) = |\boldsymbol{t}|$. These distributions were described by Shoenberg (1938). Similar results were obtained in Cambanis, Kenner and Simons (1983) for the scale function $\boldsymbol{c}(\boldsymbol{t}) = \sum_{j=1}^n |t_j|$. Further investigations were mainly devoted to the case when the scale function $\boldsymbol{c}(\boldsymbol{t}) = \left(\sum_{j=1}^n |t_j|^\alpha\right)^{1/\alpha}$; see, for example, Kuritsin (1989), Koldobsky (1992), and Gnating (1998). Special interest in this case is probably due to the connection with l_p-norms. In the general case, however, there is almost nothing

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

known until now. The main goal of this chapter is to construct a sufficiently wide class of pseudo-isotropic distributions corresponding to scale functions of a general type. The results obtained here turn out to be useful for some problems, connected with linear statistics. One result in this direction will also be presented. All investigations on this subject till now make an assumption of existence of finite absolute power moment of some positive order $\alpha_0 \leq 2$. The case $\alpha_0 = 2$ gives a scale function that is a positive definite quadratic form and elliptically countered distributions as the corresponding pseudo-isotropic distributions. So, we will assume that for some $0 < \alpha_0 < 2$

$$E|X|^{\alpha_o} < \infty. \tag{5.2}$$

It is known [see Kuritsin (1989)] that by assumption (5.2) the scale function $\boldsymbol{c}(\boldsymbol{t})$ should be even, positive, continuous, and homogeneous of the first order, admitting representation

$$\boldsymbol{c}(\boldsymbol{t}) = \left(\int_{S^{n-1}} |(\boldsymbol{t},\boldsymbol{e})|^{\alpha_0}\sigma(d\boldsymbol{e})\right)^{1/\alpha_0}, \qquad \boldsymbol{t} \in \mathbb{R}^n, \tag{5.3}$$

where σ is a finite measure on Borel subsets of the unit sphere S^{n-1}. It is evident that the characteristic function of the pseudo-isotropic distribution $\varphi(\boldsymbol{t})$ should have a form

$$\varphi(\boldsymbol{t}) = h(\boldsymbol{c}(\boldsymbol{t})), \qquad \boldsymbol{t} \in \mathbb{R}^n.$$

It is also known [see, for example, Kuritsin (1989)] that the class of pseudo-isotropic distributions corresponding to the scale function $\boldsymbol{c}(\boldsymbol{t})$, given by (5.3) is certainly non-empty, because the Levy-Feldhaim distribution with characteristic function $\varphi_0(\boldsymbol{t})$ given by

$$\varphi_0(\boldsymbol{t}) = \exp(q(\boldsymbol{c}(\boldsymbol{t}))^{\alpha_0})$$

for any coefficient $q > 0$, belongs to this class of distributions.

5.2 The Main Results

Before we go over to the main results of this paper, we should mention some known facts. An even function $\omega : \mathbb{R}^n \to \mathbb{R}$ is negative definite, if for any integer $k \geq 1$ and for arbitrary $\boldsymbol{t}^1, \ldots, \boldsymbol{t}^k \in \mathbb{R}^n$, $u_1, \ldots, u_k \in \mathbb{R}$, $u_1 + \cdots + u_k = 0$, the inequality

$$\sum_{\mu=1}^{k} \sum_{\nu=1}^{k} \omega(t^\nu - t^\mu) \leq 0$$

holds. Negative definite continuous functions satisfy the known Levy representation, which gives in case of even functions

$$\omega(\boldsymbol{t}) = \int_{\mathbb{R}^n\setminus\{0\}} (\cos((\boldsymbol{t}, \boldsymbol{x})) - 1) s(d\boldsymbol{x}),$$

where s is a σ-finite measure on the class of Borel subsets of $\mathbb{R}^n$ satisfying the condition

$$\int_{\mathbb{R}^n\setminus\{0\}} \frac{|\boldsymbol{x}|^2}{1+|\boldsymbol{x}|^2} s(d\boldsymbol{x}). \tag{5.4}$$

We will also use the formula [see Gradshtein and Ryzhik (1994)]

$$c(z) = \int_0^\infty \frac{1-\cos a\rho}{\rho^{1+z}} d\rho = \frac{\pi}{2} \frac{|a|^2}{\Gamma(1+z)\sin\frac{\pi}{2}z}, \quad 0 < \mathrm{Re} z < 2. \tag{5.5}$$

In this chapter, we will the prove the following results.

Theorem 5.2.1 *Let $\omega : \mathbb{R}^n \to \mathbb{R}_+$ be even, continuous, homogeneous of order $0 < \alpha_0 < 2$. Then*

(i) $\boldsymbol{c}(\boldsymbol{t}) = (\omega(\boldsymbol{t}))^{1/\alpha_0}$ *is a scale function of some nondegenerate pseudo-isotropic distribution if ω is negative definite, and the same is true for $(\omega(t))^{\alpha/\alpha_0}$*

(ii) for any complex z, satisfying the condition $0 < \mathrm{Re} z \leq \alpha_0$, the representation

$$\int_{S^{n-1}} |(\boldsymbol{t}, \boldsymbol{e})|^z \sigma(d\boldsymbol{e}) = (\omega(\boldsymbol{t}))^{z/\alpha_0}, \qquad \boldsymbol{t} \in \mathbb{R}^n \tag{5.6}$$

where σ is a complex-valued finite measure on the class of Borel subsets of the unit sphere S^{n-1}, holds.

Theorem 5.2.2 *Let ω be a function, satisfying the conditions of Theorem 5.2.1 with $0 < \alpha_0 < 2$ its homogeneity order. For any integer $N \geq 1$, let us take complex $\kappa_1, \ldots, \kappa_N$ and $z_1, \ldots, z_N$ under the condition $\mathrm{Re} z_j = \alpha \in (0, \alpha_0]$, $j = 1, \ldots, N$, and construct the function*

$$\varphi(\boldsymbol{t}) = \exp\{-\mathrm{Re} \sum_{j=1}^{N} \kappa_j (\omega(\boldsymbol{t}))^{z_j/\alpha_0}\}. \tag{5.7}$$

Then, $\varphi(\boldsymbol{t})$ is a characteristic function of the pseudo-isotropic distribution with scale function $(\omega(t))^{1/\alpha_o}$, if

$$\sigma = \mathrm{Re} \sum_{j=1}^{N} \kappa_j \rho^{z_j} c(z_j)^{-1} \sigma_{z_j} \tag{5.8}$$

is a finite measure on Borel subsets of S^{n-1} for any $\rho > 0$, and σ_{z_j} is the solution of (5.6) for $z = z_j$, $j = 1, \ldots, N$.

Note. The function $\varphi(\boldsymbol{t})$ in this theorem turns out to be an infinitely divisible characteristic function. In this class of characteristic functions, the condition of Theorem 5.2.2 is also necessary.

The stated results are closely connected with the problem of the reconstruction of reconstructing from means of translated moments of linear statistics, considered earlier by Zinger (1997) [see also Kakosyan, Klebanov and Zinger (1989)]. An application of the concept of negative definiteness is also productive here as it allows us to give a new simple condition for a function to be a translated moment of linear statistics. We give here one possible result in this direction.

Theorem 5.2.3 *Let $n \geq 1$ be an integer and $0 < \alpha < 2$. Then, some even function $\psi : \mathbb{R}^{n+1} \to \mathbb{R}_+$ may be presented in the form*

$$\psi(t_1, \ldots, t_n, \tau) = E|(\boldsymbol{t}, \boldsymbol{X}) - \tau|^\alpha, \qquad \boldsymbol{t} \in \mathbb{R}^n,\ \tau \in \mathbb{R} \tag{5.9}$$

for some random vector $\boldsymbol{X}$ in $\mathbb{R}^n$, iff ψ is continuous, homogeneous of 00 order α and negative definite.

5.3 Proofs

In this section, we provide proofs of the theorems stated in the last section.

Proof of Theorem 5.2.1. Necessity of (i) follows immediately from (5.3). If $(\omega(\boldsymbol{t}))^{1/\alpha_0}$ is a scale function, then

$$\omega(\boldsymbol{t}) = \int_{S^{n-1}} |(\boldsymbol{t}, \boldsymbol{e})|^{\alpha_0} \sigma_{\alpha_0}(d\boldsymbol{e}), \tag{5.10}$$

where σ_{α_0} is a finite measure on Borel subsets of S^{n-1}. Multiplying both sides of (5.10) by $\boldsymbol{c}(\alpha_0)$, we will have

$$\boldsymbol{c}(\alpha_0)\omega(\boldsymbol{t}) = \int_{\mathbb{R}^n} \{1 - \cos((\boldsymbol{t}, \boldsymbol{x}))\} S_{\alpha_0}(d\boldsymbol{x}), \tag{5.11}$$

where

$$S_{\alpha_0}(d\boldsymbol{x}) = \frac{d\rho}{\rho^{1+\alpha_0}} \sigma_{\alpha_0}(d\boldsymbol{e}), \qquad x = \rho \boldsymbol{e},\ \rho \in \mathbb{R}_+,\ \boldsymbol{e} \in S^{n-1}.$$

It follows from (5.11) that $-\omega$ is negative definite ($\boldsymbol{c}_{\alpha_0} > 0$). For checking the sufficiency of (i), let us consider the representation

$$\omega(\boldsymbol{t}) = \int_{\mathbb{R}^{n-1}} \{1 - \cos((\boldsymbol{t}, \boldsymbol{x}))\} S(d\boldsymbol{x}), \tag{5.12}$$

where the measure S satisfies (5.4), and deduce from (5.12) due to the homogeneity of ω (for $\lambda > 0$)

$$\lambda^{\alpha_0}\omega(\boldsymbol{t}) = \int \{1 - \cos((\boldsymbol{t}, \boldsymbol{x}/\lambda))\} S(d\boldsymbol{x}). \tag{5.13}$$

In the integral on the right hand side of (5.13), using the coordinates transformation

$$\boldsymbol{x} = \lambda\rho\boldsymbol{e}, \ \rho \in \mathbb{R}_+, \qquad \boldsymbol{e} \in S^{n-1}$$

and then comparing the result with (5.12), we obtain

$$S(d\boldsymbol{x}) = \frac{d\rho}{\rho^{1+\alpha_0}}\sigma_{\alpha_0}(d\boldsymbol{e}),$$

where σ_{α_0} is some finite measure on Borel subsets of S^{n-1}. This means that

$$\omega(\boldsymbol{t}) = \int_{S^{n-1}} |(\boldsymbol{t}, \boldsymbol{e})|^{\alpha_0}\sigma_{\alpha_0}(d\boldsymbol{e})$$

and $\boldsymbol{c}(\boldsymbol{t}) = (\omega(\boldsymbol{t}))^{1/\alpha_0}$ is a scale function of a pseudo-isotropic distribution in which case, one can take the Levy–Feldhaim distribution with characteristic function

$$\varphi_0(\boldsymbol{t}) = \exp\{-\omega(\boldsymbol{t})\}. \tag{5.14}$$

Next, let us consider (ii). Because of the positiveness and homogeneity of the function $\omega(\boldsymbol{t})$ we have

$$m|\boldsymbol{t}|^{\alpha_0} \leq \omega(\boldsymbol{t}) \leq M|\boldsymbol{t}|^{\alpha_0}$$

for some appropriate positive constants m and M. So, the distribution with characteristic function $\varphi_0(\boldsymbol{t})$ defined in (5.14) possesses an infinitely differentiable density function

$$p_0(\boldsymbol{x}) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \exp\{-i(\boldsymbol{t}, \boldsymbol{x}) - \omega(\boldsymbol{t})\} d\boldsymbol{t}.$$

This distribution, in addition, has finite absolute power moments of order $0 < \alpha < \alpha_0$. We can rewrite the definition of pseudo-isotropic distribution with scale function $\boldsymbol{c}(\boldsymbol{t})$ as follows: a random vector X has a pseudo-isotropic distribution if

$$(\boldsymbol{t}, \boldsymbol{X}) \stackrel{d}{=} \boldsymbol{c}(\boldsymbol{t})\xi, \qquad \boldsymbol{t} \in \mathbb{R}^n \tag{5.15}$$

for some random variable ξ. Here, the symbol $\stackrel{d}{=}$ denotes equality in distribution. We can take in (5.15) the random vector $\boldsymbol{X}_0$ (for $\boldsymbol{X}$) with characteristic function (5.14) and corresponding ξ_0 (for ξ) random variable having symmetric stable distribution with parameter α_0. In this case, we obtain from (5.15)

$$E|(\boldsymbol{t}, \boldsymbol{X}_0)|^z = (\omega(\boldsymbol{t}))^{\frac{z}{\alpha_0}} E|\xi_0|^z \tag{5.16}$$

for complex z, satisfying condition $0 < \mathrm{Re} z = \alpha < \alpha_o$. It is known [see Zolotarev (1983)] that for such z

$$E|\xi_0|^z = 2\Gamma(1+z)\Gamma(1-\frac{z}{\alpha_0})\frac{1}{z}\sin\frac{\pi}{2}z \neq 0$$

and so we have from (5.16)

$$\frac{1}{E|\xi_0|^z}\int_{\mathbb{R}^n}|(\boldsymbol{t},\boldsymbol{x})|^z p_0(x)dx = (\omega(\boldsymbol{t}))^{z/\alpha_0} \tag{5.17}$$

We obtain (5.6) from (5.17) immediately with

$$\sigma(A) = \frac{1}{E|\xi_0|^z}\int_{\mathbb{R}_+\times A} p_0(\rho\boldsymbol{e})\rho^{n-1+z}d\rho\gamma(d\boldsymbol{e}), \tag{5.18}$$

where γ denotes the uniform measure on S^{n-1}. It is obvious, that for real z, (5.18) gives a measure. Now we will extend this result to the case $\mathrm{Re} z = \alpha_0$. We can proceed here in the same way as Kuritsin (1989), in spite of the fact, that we have complex exponents. We fix $z = \alpha_o + is$ and consider a sequence $\{z_k : k = 1, 2, \ldots, z_k = \alpha_k + is, \alpha_k < \alpha_o, k \to \infty\}$, from which may be chosen the subsequence $\{z_{k_\nu} : \nu = 1, 2, \ldots\}$, such that there exists a weak limit σ_{α_0+is} for σ_{k_ν} by $\nu \to \infty$, where σ_{k_ν} satisfies (5.6), and this limit is finite complex-valued measure on S^{n-1}, also satisfying (5.6). Concluding the proof of Theorem 5.2.1, we should mention that in the class of solutions of (5.6) the solutions are unique, which may be proved in the same way as Zinger (1997) [see also Kakosyan, Klebanov and Zinger (1989)]. ∎

Proof of Theorem 5.2.2. We can deduce from (5.6), similar to (5.11), that

$$(\omega(\boldsymbol{t}))^{z/\alpha_0} = \frac{1}{c(z)}\int_{\mathbb{R}^n\setminus\{0\}}(1-\cos((\boldsymbol{t},\boldsymbol{x})))s_z(d\boldsymbol{x}),\ \boldsymbol{t}\in\mathbb{R}^n\ 0<\mathrm{Re} z\le\alpha_0, \tag{5.19}$$

where $c(z)$ defined by (5.5) is nonzero, and

$$S_z(d\boldsymbol{x}) = \frac{d\rho}{\rho^{1+z}}\sigma_z(d\boldsymbol{e}), \qquad x = \rho\boldsymbol{e}$$

with σ_z as in (5.18). Using (5.19), we have

$$\mathrm{Re}\sum_{j=1}^{N}\kappa_j(\omega(\boldsymbol{t}))^{z/\alpha_0} = \int_{\mathbb{R}^n\setminus\{0\}}\{1-\cos(\boldsymbol{t},\boldsymbol{x})\}\tilde{s}(d\boldsymbol{x}),$$

where

$$\tilde{s}(d\boldsymbol{x}) = \mathrm{Re}\sum_{j=1}^{N}\kappa_j c^{-1}(z)\frac{\sigma_{z_j}(d\boldsymbol{e})}{\rho^{1+z_j}}\,d\rho.$$

Condition (5.8) provides $\tilde{s}$ to be a measure, satisfying (5.4). This means that (5.7) is a characteristic function of pseudo-isotropic distribution and this distribution is infinitely divisible. It is easy to see that in this class of distributions, (5.8) is also necessary.

Verification of (5.8) to be a measure becomes very simple, when we choose $z_j = \alpha + id_j$ for some $d_j > 0$, $j = 1, \ldots, N$. In this case, we need to deal with positivity condition for trigonometric polynomials, which is well worked out.

Using characteristic function (5.7), we can essentially extend the variety of pseudo-isotropic distributions. Multiplicative convolutions of pseudo-isotropic distributions with arbitrary random coefficients once again, give again a pseudo-isotropic distribution. More precisely, if $\varphi(\boldsymbol{t})$ is a characteristic function of a pseudo-isotropic distribution, then

$$\varphi_\eta(\boldsymbol{t}) = E\varphi(\eta \boldsymbol{t})$$

(for any random variable η) is also a characteristic function of a pseudo-isotropic distribution, with the same scale function. ■

Proof of Theorem 5.2.3. We note that the situation is quite similar to one (i) of Theorem 5.2.1. So, $\psi(t_1, \ldots, t_n, \tau)$ belongs to the convex hull of $|(\boldsymbol{t}, \boldsymbol{x}) - \tau|^\alpha$, $\boldsymbol{t} = (t_1, \ldots, t_n)$ and $\boldsymbol{x} = (x_1, \ldots, x_n)$. For checking sufficiency one can notice that the conditions of (i) of Theorem 5.2.1 are fulfilled, and the following the presentation for ψ [see Kuritsin (1989)]

$$\psi(t_1, \ldots, t_n, \tau) = \int_{S^n} |(t_1 e_1 + \cdots + t_n e_n + \tau e_{n+1})|^\alpha \sigma(d\boldsymbol{e}) \tag{5.20}$$

is true. (5.9) follows from (5.20) after appropriate change of variables on the right hand side of (5.20). In conclusion, we may mention that, in the same manner one can also treat odd translated moments and probabilities to get into half-spaces. But, we intend to present these results in our next paper. ■

Acknowledgement. The author is grateful to L. B. Klebanov for useful discussions and help in preparing this paper. This research was partially supported by RFFI Grants 96-01-00852 and 96-15-96199.

References

1. Cambanis, S., Kenner, R. and Simons, G. (1983). On α-symmetric multivariate distributions, *Journal of Multivariate Analysis*, **13**, 213–233.

2. Gradshtein, I. S. and Ryzhik, I. M. (1994). *Table of Integrals, Series and Products*, Fifth edition, Boston: Academic Press.

3. Gnating, T. (1998). On α-symmetric multivariate characteristic functions, *Journal of Multivariate Analysis*, **64**, 131–147.

4. Kakosyan, A. V., Klebanov, L. B. and Zinger, A. A. (1989). Characterization of distributions by average of statistics and some probability metrics, In *Stability Problems for Stochastic Models. Proceedings of the Seminar, The Institute for Systems Studies*, pp. 47–54.

5. Koldobsky, A. (1992). Shoenberg's problem on positive definite functions, *Report*, University of Missouri at Columbia, Department of Mathematics, 1–12.

6. Kuritsin, Yu. G. (1989). Multidimensional versions and two Shoenberg problems, *Stability Problems for Stochastic Models. Proceeding of the Seminar, The Institute for Systems Studies*, pp. 72–78.

7. Shoenberg, I. J. (1938). Metric spaces and completely monotone functions, *Annals of Mathematics*, **39**, 811–841.

8. Zinger, A. A. (1997). On reconstruction of distributions by the translated moments of linear statistics, *Journal of Mathematical Sciences*, **83**, 109–115.

9. Zolotarev, V. M. (1983). *One-Dimensional Stable Distributions*, Moscow: Nauka.

PART III

Probabilities and Measures in High-Dimensional Structures

6

Time Reversal of Diffusion Processes in Hilbert Spaces and Manifolds

Ya. Belopolskaya

St. Petersburg Architecture Building University, St. Petersburg, Russia

Abstract: We describe some results of the theory of diffusion processes in infinite dimensional Hilbert spaces and manifolds and apply them to investigation of invariant measures and time reversal of diffusion processes.

Keywords and phrases: Hilbert space and manifold, diffusion process, invariant measure, time reversal

6.1 Diffusion in Hilbert Space

The development of the theory of infinite dimensional diffusion processes was started by Gross and Dalecky and intensively developed during the last decades. Nevertheless, there are still many open problems in the field both in the framework of linear spaces and smooth manifolds. In this chapter, we follow the line exposed in Belopolskaya and Dalecky (1990) and Dalecky and Fomin (1991) and discuss problems concerning invariant measures of infinite dimensional diffusion processes and description of their time reversal. We extend here the results due to Dalecky and Steblovskaya (1996) concerning invariant measures of diffusion processes using previous results from Belopolskaya (1998) as well. In the construction of time reversal of a diffusion process, we use the approach developed by Nagasawa (1961) in a finite dimensional framework. Notice that our results are close as well to the results received by Follmer and Wakolbinger (1986).

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, $H_+ \subset H \subset H_-$ be a Gelfand triple of Hilbert spaces with dense Hilbert-Schmidt imbedding, $w(t) \in H_-$ be a standard Wiener process in H, and $\mathcal{F}_t \subset \mathcal{F}$ be a set of σ-fields adopted to $w(t)$. Given nonrandom vector field $a(t,x) \in H$ and Hilbert-Schmidt operator

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

field $A(t,x)$, $\quad x \in H, t \in [0,T]$ consider a Cauchy problem for SDE

$$d\xi = a(t,\xi(t))dt + A(t,\xi(t))dw, \quad \xi(0) = \xi_0 \in H. \tag{6.1}$$

If $a(t,x)$ and $A(t,x)$ are smooth enough functions with sublinear growth in x, then there exists the unique Markov solution $\xi(t)$ of (6.1) if $\xi(0) = \xi_0 \in H$ is $\mathcal{F}_0$-measurable [see Belopolskaya and Dalecky (1990)].

Let $B(H)$ and $\mathcal{M}(H)$ be the space of real valued bounded measurable functions and Borel measures defined on H, and $C^k(H)$ and $\mathcal{M}^k(H)$ denote spaces of C^k-smooth functions and measures correspondingly. Denote by $P(s,x,t,G) = P\{\xi(t) \in G|\xi(s) = x\}$, $0 \le s \le t \le T, x \in H, G \in \mathcal{B}_H$, the transition probability of the Markov process $\xi(t) \in H$, and consider evolution families $V(t,s)$ and $V^*(s,t)$

$$V(t,s)f(x) = Ef(\xi(t)) \;=\; \int_H f(y)P(s,x,t,dy), \quad f \in B(H), \tag{6.2}$$

$$V^*(s,t)\nu(dy) \;=\; \int_H \nu(dx)P(s,x,t,dy), \quad \nu \in \mathcal{M}(H) \tag{6.3}$$

dual in pairing $\langle f,\nu\rangle = \int_H f(y)\nu(dy)$.

A measure $\nu \in \mathcal{M}^2(H)$ is said to be an invariant measure of the diffusion process $\xi(t)$ if $V^*(s,t)\nu = \nu$. It is known [see Dalecky and Steblovskaya (1996) and Belopolskaya (1998)] that a measure ν is an invariant measure of $\xi(t)$ if

$$m = div\,[\frac{1}{2}\sum_{k=1}^{\infty} \nabla_{A^k}(A^k\nu) - a\nu] = 0. \tag{6.4}$$

Let $L_2(H,\nu)$ be a Hilbert space of square integrable functions on H (with respect to ν) with the inner product $\langle\langle g,f\rangle\rangle$. Assume that $A(t,x) = A(x)$, $a(t,x) = a(x)$ and ν is an invariant measure of the solution $\xi(t)$ to (6.1). Define $V(t)$ and $V^+(t)$ in $L_2(H,\nu)$ by

$$\int_H g(x)\int_H f(y)P(t,x,dy)\nu(dx) = \langle g, V(t)f\rangle = \langle V(t)^+g, f\rangle \tag{6.5}$$

for any measurable bounded functions f and g.

Semigroups V_t and V_t^+ are called dual with respect to the invariant measure ν if (6.5) holds.

We show that the evolution family $V^+(t)$ coincides with the evolution family $\hat{V}(t)$ generated by the time reversal of $\xi(t)$ and derive the stochastic equation for the time reversal process.

Denote by $\hat{\xi}(t)$ the time reversal of $\xi(t)$ given by $\hat{\xi}(t) = \xi(T-t), \quad t \in [0,T]$. The process $\hat{\xi}(t)$ is a diffusion process as well. To check it, consider a partition $0 = t_0 < t_1 < \ldots < t_n = T$, denote by $\Delta_k(t) = t_k - t_{k-1}$, and notice that

$$\begin{aligned}
&Ef(\hat{\xi}(t_0),\ldots,\hat{\xi}(t_n))\\
&= Ef(\xi(1-t_0),\ldots,\xi(1-t_n))\\
&= \int_H f(x_0,\ldots,x_n)\nu(dx_0)\frac{P(t,x_1,dx_0)}{\nu(dx_0)}\frac{\nu(dx_1)P(\Delta_1(t),x_2,dx_1)}{\nu(dx_1)}\\
&\quad\ldots\frac{P(\Delta_n(t),x_n,dx_{n-1})}{\nu(dx_{n-1})}\nu(dx_n).
\end{aligned}\tag{6.6}$$

Consider $\hat{P}(t,x,dy) = \dfrac{P(t,y,dx)\nu(dy)}{\nu(dx)}$. It follows from (6.6) that

$$\begin{aligned}
&Ef(\hat{\xi}(t_0),\ldots,\hat{\xi}(t_n))\\
&= \int_H f(x_0,\ldots,x_n)\nu(dx_0)\hat{P}(t_1,x_0,dx_1)\ldots\hat{(}\Delta_n(t),x_{n-1},dx_n)f(x_0,\ldots,x_n)
\end{aligned}$$

and hence $\hat{P}(t,x,dy)$ is the transition probability of the diffusion process $\hat{\xi}(t)$. Finally,

$$\begin{aligned}
\langle g, V_t f\rangle &= \int_H g(x)\int_H f(y)P(t,x,dy)\nu(dx)\\
&= \int_H\int_H g(x)[\frac{P(t,x,dy)}{\nu(dy)}\nu(dy)]\nu(dx)\\
&= \int_H \hat{V}_t^+ g(y)f(y)\nu(dy) = \langle V_t^+ g, f\rangle.
\end{aligned}$$

Thus, the following assertion is proved.

Theorem 6.1.1 *Let $\nu(dy)$ be an invariant measure of the process $\xi(t)$. Let V_t and $\hat{V}_t$ be evolution families in $L_2(H,\nu)$ generated by the processes $\xi(t)$ and $\hat{\xi}(t)$ respectively. Then $\hat{V}_t = V_t^+$, where V_t^+ is given by (6.5).*

Consider a pair of diffusion processes $\xi(t)$ and $\eta(t)$ in H such that

$$d\xi = a(\xi(t))dt + A(\xi(t))dw, \quad \xi(0) = \xi_0 \tag{6.7}$$

$$d\eta = \hat{a}(\eta(t))dt + A(\eta(t))dw, \quad \eta(0) = \xi_0 \tag{6.8}$$

and let the distribution ν of $\xi_0 \in H$ be a smooth measure with vector logarithmic derivative λ.

Theorem 6.1.2 *Let $\nu \in \mathcal{M}^2(H)$. Then ν is an invariant measure for a pair of diffusion processes $\xi(t)$ and $\eta(t)$ satisfying (6.7) and (6.8) if and only if the drift coefficients $a(x)$ and $\hat{a}(x)$ satisfy*

$$a(x) + \hat{a}(x) = B(x)\lambda(x) + \nabla_{A^k(x)}A^k(x), \quad B(x) = A^*(x)A(x) \tag{6.9}$$

and

$$div(a(x) - \hat{a}(x)) + (a(x) - \hat{a}(x), \lambda(x)) = 0. \tag{6.10}$$

PROOF. Consider generators of semigroups V_t and $\hat{V}_t$ having the form

$$\mathcal{A}f(x) = \frac{1}{2} f''(x)(A^k(x), A^k(x)) + \nabla_{a(x)} f(x)$$

and

$$\hat{\mathcal{A}}g(x) = \frac{1}{2} g''(x)(A^k(x), A^k(x)) + \nabla_{\hat{a}(x)} g(x).$$

Here and below, we assume summing over all repeating indices. If V_t and $\hat{V}_t$ are dual then the relation $\langle \hat{\mathcal{A}}g, f\rangle_\mu = \langle g, \mathcal{A}f\rangle_\mu$ should hold. Using integration by part formula, one can check that if

$$\begin{aligned}\frac{1}{2} div\, (\nabla_{A^k(x)} A^k(x)) - div(\hat{a}(x)) \;&+\; \frac{1}{2}((\nabla A^k(x)\lambda, A^k(x)) + (A^k(x)\lambda, A^k(x)\lambda) \\ &+\; (\nabla_{A^k(x)} A^k(x), \lambda)) - (\hat{a}, \lambda) = 0,\end{aligned}$$

$$(\nabla_{A^k(x)} f, A^k(x)\lambda) + \nabla_{\nabla_{A^k(x)} A^k(x)} f) = (\hat{a}(x) + a(x), \nabla f)$$

hold, then

$$\begin{aligned}&\int_H [\frac{1}{2} g''(x)(A^k(x), A^k(x)) + \nabla_{\hat{a}(x)} g(x)] f(x) \nu(dx) \\ &\quad = \int_H g[\frac{1}{2} f''(x)(A^k(x), A^k(x)) + \nabla_{a(x)} f(x)] \nu(dx).\end{aligned}$$

Notice that (6.4) and the invariance of ν with respect to V_t^+ yield (6.9). As far as ν is invariant with respect to V_t as well, we have

$$\begin{aligned} m(y) &= \frac{1}{2} B_{ij}[\nabla_j \lambda_i + \lambda_i \lambda_j] + (\nabla_i B_{ij} - \hat{a}_j)\lambda_j + \frac{1}{2}\nabla_i \nabla_j B_{ij} - div\, \hat{a} \\ &= \frac{1}{2} B_{ij}[\nabla_j \lambda_i + \lambda_i \lambda_j] + (\nabla_i B_{ij} - a_j)\lambda_j + \frac{1}{2}\nabla_i \nabla_j B_{ij} - div\, a = 0 \end{aligned}$$

that yields (6.10). ■

Remark 6.1.1 Given $\xi(t)$ satisfying (6.7), the time reversal process $\hat{\xi}(t)$ is governed by (6.8). Denote by S and U two smooth scalar (potential) functions such that $\nabla U = \frac{1}{2}\lambda$ and $\nabla S = \frac{1}{2}(a - \hat{a})$. Then both a and $\hat{a}$ could be represented in terns of S and U as $a = B\nabla(U+S)$ and $\hat{a} = B\nabla(U-S)$. This representation is very important in Nelson mechanics.

It is known due to Nagasawa (1961) that if an arbitrary measure ν is choosen (omitting the requirement that it should be invariant with respect to $\xi(t)$) to define the time reversal process $\hat{\xi}(t)$, one has to to consider the random time $L(\omega)$ called co-optional time instead of the constant time T to get a homogenous time reversal process. Using the co-optional time, it is possible as well to extend the above consideration to time dependent drifts .

By definition, $L(\omega)$ is a co-optional time if $\{\omega : s < L - t\} = \{\omega : s < L \circ \Sigma_t, t, s \geq 0\}$, where Σ_t is the shift operator in the trajectory space $\Sigma_t w(s) = w(s+t)$. This property is equivalent to $L \circ \Sigma_t = (L-t)^+, t \geq 0$.

The following statements are proved similar to Nagasawa (1961).

Theorem 6.1.3 *Let $\xi(t)$ be a solution to (6.7) and the distribution ν of ξ_0 has smooth vector logarithmic derivative λ. Then the time reversal process $\hat{\xi}(t)$ from a cooptional time $L(\omega)$ with $\hat{\xi}(0) = \xi_0$ is a diffusion process and the corresponding semigroup $\hat{V}_t$ is dual to V_t, $\langle g, V_t f\rangle_m = \langle \hat{V}_t g, f\rangle_m$ with respect to a measure m defined by $m(G) = E[\int_0^\infty \chi_G(\xi(t))dt]$.*

6.1.1 Duality of time inhomogenous diffusion processes

Consider SDE with time inhomogenous diffusion coefficients

$$d\gamma = a(t,\gamma(t))dt + b(t,\gamma(t))dt + A(t,\gamma(t))dw. \tag{6.11}$$

Notice that we need two components in the drift coefficient since they are responsible for different phenomena.

Let $\gamma(t)$ be a time inhomogenous diffusion process defined on the interval $[0,1]$ which satisfies (6.11) and at the moment $s < t$ its value $\gamma(s)$ be a random variable with given distribution μ_s. Denote by $P(s,x,t,G) = P\{\gamma(t) \in G | \gamma(s) = x\}$ its transition probability. The process $\gamma(t)$ gives rise to evolution families

$$(Z_t^s f)(x) = \int_H f(t,y)P(s,x,t,dy), \quad [(Z_t^s)^*\mu]_t(dy) = \int_H \mu_s(dx)P(s,x,t,dy).$$

Recall that a time inhomogenous diffusion process $\gamma(t)$ could be considered as a component of the time homogenous process $\kappa(t) = (t,\gamma(t))$ and hence we could apply the above considerations to this new process choosing $\mu_0(dy) = \nu(dy)$, $\mu_t(dy) = ((Z_t^0)^*\nu)(dy)$ and

$$m(dt,dy) = E\int_0^1 d\tau \chi_{\tau,\gamma(\tau)}(dt,dy) = \mu_t(dy)dt. \tag{6.12}$$

Denote by $(\Gamma_s u)(x) = \int_0^1 \int_H u(\tau,y)P(s,x,\tau,dy)d\tau$ and use the time homogenous duality relation $\langle g, \Gamma_t f\rangle_m = \langle \Gamma_t^+ g, f\rangle_m$ to derive

$$\langle g, Z_t^s f\rangle_{\mu_s} = \langle \hat{Z}_t^s g, f\rangle_{\mu_t} = \langle (Z_t^s)^+ g, f\rangle_{\mu_t}, \quad s < t. \tag{6.13}$$

Here, $\hat{Z}_t^s g(t,y) = \int_H g(s,x)\hat{P}(t,y,s,dx)$ and

$$\hat{P}(t,y,s,dx) = \frac{P(s,x,t,dy)}{\mu_t(dy)}\mu_s(dx). \tag{6.14}$$

Given the space-time process $\kappa(t) = (t, \gamma(t))$ and its time reversal, define $M_r f(s,x) = Z^s_{s+r} f(s,x)$, $r \geq 0$, and $\hat{M}_r g(t,y) = \hat{Z}^{t-r}_t g(t,y)$, $r \geq 0$. Then

$$\begin{aligned}\langle g, M_r f\rangle_m &= \int_0^{1-r} \langle g, Z^s_{s+r} f\rangle_{\mu_s} ds \\ &= \int_r^1 \langle g, Z^{t-r}_t f\rangle_{\mu_{t-r}} dt \\ &= \int_r^1 \langle (\hat{Z}^{t-r}_t) g, f\rangle_{\mu_t} dt \\ &= \langle \hat{M}_r g, f\rangle_m. \end{aligned} \tag{6.15}$$

As a result, the duality relation

$$\langle g, M_r f\rangle_m = \langle \hat{M}_r g, f\rangle_m \tag{6.16}$$

holds.

Let us compute the difference

$$\begin{aligned}&\int_0^T \int_H [\mathcal{B}(t) f(x) g(x) - f(x)\hat{\mathcal{B}}(t) g(x)]\mu_t(dx)dt \\ &= \int_0^T \int_H f(x)g(x)(\mathcal{B}^*(t)\mu)_t(dx)dt - \int_0^T \int_H f(x)([a(t,x) + \hat{a}(t,x) \\ &\quad - A^k(t,x)A^k(t,x)\lambda(t,x) - \nabla_{A^k} A^k], \nabla g(x))\mu_t(dx)dt. \end{aligned} \tag{6.17}$$

Here

$$\mathcal{B}(t) f = \frac{\partial f}{\partial t} + \frac{1}{2} f''(A^k, A^k) + (b + a, f') \tag{6.18}$$

and

$$\mathcal{B}^* \mu = -\frac{\partial \mu}{\partial t} + \frac{1}{2}\mu''(A^k, A^k) + div[(b+a)\mu] + div(\nabla_{A^k} A^k \mu), \tag{6.19}$$

while the corresponding adjoint (in $L_2(H,m)$) operators have the form

$$\hat{\mathcal{B}}(t) g = -\frac{\partial g}{\partial t} - \frac{1}{2} g''(A^k, A^k) + ((\hat{a} - b), \nabla g), \tag{6.20}$$

$$(\hat{\mathcal{B}}^*(t))\mu = \frac{\partial \mu}{\partial t} - \frac{1}{2}\mu''(A^k, A^k) - div\,(\nabla_{A^k} A^k \mu) + div((\hat{a} - b)\mu). \tag{6.21}$$

Let us prove that the duality relations for a pair of space-time diffusion processes have the form

$$a + \hat{a} = B\lambda + \nabla_{A^k} A^k, \quad \frac{\partial \mu}{\partial t} + div(b + \frac{a - \hat{a}}{2})\mu = 0. \tag{6.22}$$

To derive (6.22), consider the diffusion process $\gamma(t)$ and its dual in the sense of (6.16) with respect to $m(dt,dy) = \mu_t(dy)dt$. Since $\mathcal{B}^*(t)\mu = 0$ and $\hat{\mathcal{B}}^*(t)\mu = 0$, we deduce from (6.17) the first relation in (6.22). To derive the second one, notice that

$$\begin{aligned}&\int_0^T \int_H [\mathcal{B}(t)f(x)g(x) - f(x)\hat{\mathcal{B}}(t)g(x)]\mu_t(dx)dt \\ &= \int_0^T \int_H f(x)g(x)(\hat{\mathcal{B}}^*(t)\mu)_t(dx)dt - \int_0^T \int_H g(x)([a(t,x)+\hat{a}(t,x) \\ &\quad - A^k(t,x)A^k(t,x)\lambda(t,x) - \nabla_{A^k(t,x)}A^k(t,x)], \nabla f(x))\mu_t(dx)dt.\end{aligned} \tag{6.23}$$

Hence to ensure (6.16), we need

$$(\mathcal{B}^+)^*(t)\mu = 0. \tag{6.24}$$

Finally, taking into account (6.23) and (6.24), we get (6.22). Notice that on the contrary, if (6.22) holds, then we can easily check with the help of (6.17) and (6.23) that (6.16) is valid.

In applications, the case $a(t,x) - \hat{a}(t,x) = 2B(t,x)\nabla S(t,x)$ with a potential function $S(t,x)$ is rather important.

Consider next symmetric semigroups generated by diffusion processes. Recall that M_r is called symmetric with respect to $m(dx,dt) = m_t(dx)dt$ if

$$\int_0^T \int_H M_r f(t,x)g(t,x)\mu_t(dx)dt = \int_0^T \int_H f(t,x)M_r g(t,x)\mu_t(dx)dt.$$

In other words, the semigroup M_r is symmetric if $M_r = M_r^+$. To derive the conditions on coefficients of (6.11) and the measure μ_t to ensure $M_r = M_r^+$ consider the difference

$$\int_0^T \int_H \mathcal{B}f(t,x)g(t,x)\mu_t(dx)dt - \int_0^T \int_H f(t,x)\mathcal{B}^+g(t,x)\mu_t(dx)dt,$$

where $\mathcal{B}$ and $\mathcal{B}^+$ are given by (6.18) and (6.20), respectively.

It is easy to check by differentiation by part formula that

$$\begin{aligned}\int_0^T \int_H \mathcal{B}f(t,x)g(t,x)\mu_t(dx)dt &= \int_0^T \int_H \mathcal{B}^+g(t,x)f(t,x)\mu_t(dx)dt \\ &+ \int_0^T \int_H f(t,x)g(t,x)M_r^*\mu_t(dx)dt \\ &+ \int_0^T \int_H f\nabla_{A^k}g\nabla_{A^k}\mu_t(dx)dt \\ &+ \int_0^T \int_H fg div[(b+a)\mu_t(dx)]dt.\end{aligned}$$

Consider in addition backward Kolmogorov equation $\hat{\mathcal{B}}(t)g = 0$ for time reversal process and forward Kolmogorov equation $(\hat{\mathcal{B}}^*(t))\hat{\mu} = 0$, where $\hat{\mu}_s(dx) = \hat{\mu}_t(dx)q(s,x,t,y)$, $\quad q(s,x,t,y) = \frac{P(s,x,t,dy)}{\mu_t(dy)}$. Recall that

$$Q_s^t f(s,x) = \int_H f(t,y)P(s,x,t,dy), \quad \hat{Q}_s^t g(t,y) = \int_H g(s,x)\hat{P}(t,y,s,dx)$$

and $\hat{P}(t,y,s,dx) = \mu_s(dx)\frac{P(s,x,t,dy)}{\mu_t(dy)}$. Finally

$$\int_H (\hat{Q}_s^t g)(t,y)\mu_t(dy) = \int_H g(s,x)(\hat{Q}_s^t)^*\mu_t(dx) = \int_H g(s,x)q(s,x,t,y)\mu_s(dx)$$

and as a result we prove that $\hat{\mu}_t(dx)$ solves (6.21).

The invariance of $m(dt,dy) = \mu_t(dy)dt$ and the symmetry of M_r yield $\langle M_r f, g\rangle_m = \langle f, \hat{M}_r g\rangle_m$ and hence $\mathcal{B}^* = \hat{\mathcal{B}}^*$.

In particular, consider the duality relations

$$\frac{\partial \mu}{\partial s} + div(b + \frac{a - \hat{a}}{2}\mu) = 0,$$

and

$$a(t,x) + \hat{a}(t,x) = B\lambda(t,x)$$

deduced from (6.22) for $A(t,x) = A$ and $T = 1$. Choose $b(t,x) \equiv 0$ and

$$a(t,x) = \frac{1}{2}B\lambda(t,x) = -\frac{x}{t}, \quad \hat{a}(t,x) = -\frac{x}{1-t};$$

then $\xi(t)$ and $\hat{\xi}(t)$ solve respectively SDE

$$d\xi = Adw - \frac{\xi}{t}dt, \quad d\hat{\xi} = Adw - \frac{\hat{\xi}}{1-t}dt.$$

Stochastic process $\hat{\xi}(t)$ with $\hat{\xi}(0) = 0$ is called a Brownian bridge and the above result for this process was derived in Follmer and Wakolbinger (1986).

6.2 Diffusion on Hilbert Manifold

Let M be a Hilbert manifold, B be its model space. We say that M is equipped with a Hilbert-Schmidt structure if its model space has a structure of a rigged Hilbert space $H_+ \subset H \subset H_- = B$.

Denote by $exp : TM \to M$ an exponential mapping on M corresponding to a fixed connection on M and assume that it is the Levi-Chivitta connection generated by a metric G on M.

Assume in addition that ∇ denotes the covariant derivative corresponding to this connection and moreover that the tangent bundle TM has a Hilbert-Schmidt structure. Recall [see Belopolskaya and Dalecky (1990)] that it means the following. Given a local trivialization of the Riemannian bundle over a neighborhood U_y of a point $y \in M$, one may choose a set of local sections $\{e_k(x)\}_{k=1}^{\infty}$ that make an orthonormal basis in each $\mathcal{H}_x = \gamma^{-1}(x), x \in U_y$, and such that $(e_k(x), e_k(x))_{\mathcal{H}_x} = \langle e_k(x), G(x)e_k(x)\rangle_x = \delta_{ik}$ (the Kronecker symbol). Here $\langle \cdot, \cdot \rangle_x$ is a natural pairing between $\mathcal{H}_x$ and $\mathcal{H}_x^*$, $(\cdot, \cdot)$ is an inner product in $\mathcal{H}_x$.

Let Φ be a section of $L(\gamma, \gamma^*)$. Given an orthonormal basis $e_k(x)$, define

$$\begin{aligned} Tr_G\Phi(x) &= \sum_{k=1}^{\infty}\langle e_k(x), \\ \Phi(x)e_k(x)\rangle_x &= \sum_{k=1}^{\infty}(e_k(x), \\ G^{-1}(x)\Phi(x)e_k(x))_{\mathcal{H}_x} &= Tr_{\mathcal{H}_x}G^{-1}(x)\Phi(x), \end{aligned}$$

assuming that $G^{-1}(x)\Phi(x)$ is a nuclear operator in $\mathcal{H}_x$. In the sequel, we omit subscripts in notations $(\cdot, \cdot)_{\mathcal{H}_x}$ and $\langle \cdot, \cdot \rangle_x$ if it will not lead to any confusion.

A manifold M is said to be equipped with a Hilbert-Schmidt structure (τ_x, γ_x, i) [see Belopolskaya and Dalecky (1990)] if given the Hilbert bundle γ one may define a bundle embedding $i : \gamma \to \tau$ with $\tau_x \circ i = \gamma$ possessing the following property: for each $x \in M$, the map $i_x : \mathcal{H}_x = \gamma^{-1}(x) \to T_xM$ belongs to $L_{12}(\mathcal{H}_x, T_xM)$ and $i_x\mathcal{H}_x$ is a dense subset of T_xM. This structure is called nuclear if the map i_x belongs to $L_{11}(\mathcal{H}_x, T_xM)$. In this case, given a Riemannian bundle γ with the inner product

$$(\xi_x, \eta_x)_{\mathcal{H}_x} = \langle \xi_x, G(x)\eta_x\rangle_x,$$

we say that the HS-structure is Riemannian. The affine connection on M is called Hilbert-Schmidt affine connection if the local connection coefficient possesses the property

$$\Gamma_x^{\gamma} = \Gamma_x^{\tau}|_{B\times H} : B \times H \to H, x \in U.$$

Introduce $\Gamma_x^{\gamma^*} : B \times H^* \to H^*$ by

$$\langle z, \Gamma_x^{\gamma^*}(y, v)\rangle = -\langle \Gamma_x^{\gamma}(y, z), v\rangle.$$

Denote by $\sigma_k(\gamma)$ the class of vector fields belonging to $\sigma_k(\tau)$ and valued in $\mathcal{H}_x \subset T_xM$ for each $x \in M$. Given $\eta \in \sigma_k(\gamma)$ and $\xi \in \sigma_k(\tau)$, put

$$\nabla_\xi^{\gamma}\eta = \nabla_\xi^{\tau}\eta, \quad \nabla_\xi^{\gamma}\eta(x) = \eta_x'\xi_x + \Gamma_x^{\gamma}(\xi_x, \eta_x).$$

Hence, $\nabla_\xi^{\gamma}\eta \in \sigma_k(\tau)$ if Γ_x^{γ} is smooth enough.

Let M be a manifold equipped with a Riemannian structure (γ, τ, i). An affine connection on the manifold is said to be compatible with this structure if $\nabla_z(\phi, \psi) = (\nabla_z^\gamma \phi, \psi) + (\phi, \nabla_z^\gamma \psi)$ holds for all $\phi, \psi \in \sigma_1(\gamma), z \in \sigma_1(\tau)$.

Consider $i : \gamma \to \tau$ and $i^* : \tau^* \to \gamma^*$. For each point $x \in X$, we equip T_xM with an HS-structure

$$T_x^*M \overset{j_x = G_x^{-1} i_x^*}{\longrightarrow} G^{-1}\mathcal{H}_x^* = \mathcal{H}_x \overset{i_x}{\longrightarrow} T_xM.$$

As a result, we get a rigged Hilbert space $\mathcal{H}_x^+ = G^{-1}T_x^M \subset \mathcal{H}_x \subset T_xM = \mathcal{H}_x^-$ with the pairing $(g, h)_{\mathcal{H}_x} = \langle g, G(x)h\rangle_x$. Denote by $\sigma_k^*(\gamma)$ a class of vector fields belonging to $\sigma_k(\tau)$ which satisfy the condition $G(x)z_x \in i^*T_x^*M$.

We say that the divergence of the vector field $\eta \in \sigma_k^*(\gamma)$ exists if $\nabla^\gamma \eta|_H$ is a nuclear operator and define it by $div_G \eta(x) = Tr_G G\nabla^\gamma \eta$.

Notice that if both the Riemannian metrics and connection are nuclear, then the vector field $\eta \in \sigma_k^*(\gamma)$ possesses a finite divergence and $div_G\eta(x) = (\nabla, \eta(x))_{\mathcal{H}_x}$, where $\nabla = D + \Gamma$ and Γ is defined by $(\gamma_x, \eta_x)_{\mathcal{H}_x} = Tr\, G(x)\hat{\Gamma}_x^\gamma(\eta_x)$ and

$$\hat{\Gamma}_x^\gamma = -G^{-1}(x) \sum_{k=1}^{\infty} \Gamma_x^{\gamma^*}(e_k(x), G(x)e_k(x)).$$

Morover, if $R(\eta_1, e_k, \eta_2)$ is a curvature tensor of the given connection, then the Ricci tensor $R(\eta_1, \eta_2) = \sum_k \langle R(\eta_1, e_k, \eta_2), Ge_k\rangle_x$ is finite and it holds

$$\kappa(z_1, z_2) = \nabla_{z_1} div_G z_2 - div_G \nabla_{z_1} z_2 = R(z_1, z_2) - Tr_G \nabla z_2 \nabla z_1. \tag{6.25}$$

Consider a diffusion process $\xi(t) \in M$ satisfying

$$d\xi = exp^M_{\xi(t)}(a(t, \xi(t))dt + A(t, \xi(t))dw), \quad \xi(0) = \xi_0 \tag{6.26}$$

or in local chart

$$d\xi = a(t, \xi(t))dt + A(t, \xi(t))dw - \frac{1}{2}\Gamma(\xi(t))(A^k(\xi(t)), A^k(\xi(t)))dt, \xi(0) = \xi_0. \tag{6.27}$$

Here and below, we omit notations connected with a chart if it will not lead to any confusion.

Let $L_{12}(\gamma, \tau)$ be the bundle of Hilbert-Schmidt operators acting from a Hilbert subbundle $\gamma : \mathcal{K} \to M$ of the tangent bundle $\tau : TM \to M$. We assume that given sections $a(t, x)$ and $A(t, x)$ of TM and $L_{12}(\gamma, \tau)$, respectively, are nonrandom, smooth and bounded. It is known that under these assumptions there exists a unique solution $\xi(t) \in M$ to (6.26) possessing the Markov property. Denote by

$$P(s, x, t, K) = P\{\xi(t) \in K | \xi(0) = x\}, 0 \le s \le t \le T, x \in M, G \in \mathcal{B}_M$$

the transition probability of the process $\xi(t)$.

Denote by $B(M)$ the space of measurable bounded scalar functions on M and by $\mathcal{M}(M)$ the space of measures defined on a σ-algebra $\mathcal{B}_M$ of Borel sets of M. It is well known that

$$V(t,s)f(x) = \int_M f(y)P(s,x,t,dy) = Ef(\xi(t)) \tag{6.28}$$

and

$$V^*(s,t)\nu(dy) = \int_M \nu(dx)P(s,x,t,dy) \tag{6.29}$$

is a pair of evolution families acting in $B(M)$ and $\mathcal{M}(M)$, respectively. The infinitesimal operator of $V(t,s)$ has the form

$$\mathcal{A}(s)f(x) = \frac{1}{2}[\nabla_{A^k(s,x)}\nabla_{A^k(s,x)} - \nabla_{\nabla_{A^k(s,x)}A^k(s,x)}]f(x) + \nabla_{a(s,x)}f(x), \tag{6.30}$$

or in local chart

$$\mathcal{A}f(x) = \frac{1}{2}[f''(x)(A^k, A^k) - (\Gamma(x)(A^k, A^k), f'(x))] + (a, f'(x)). \tag{6.31}$$

To derive the expression for the infinitesimal operator of $V^*(s,t)$, notice that $V(t,s)$ and $V^*(s,t)$ are dual in the natural pairing between $\mathcal{M}(M)$ and $B(M)$ given by $\langle f, \nu\rangle = \int_M f(y)\nu(dy)$ and hence $\mathcal{A}^*(t)\nu$ could be computed as an adjoint operator to $\mathcal{A}(s)$ using differentiation by part formulas.

Given a measure μ on $B = H_-$, denote by $\nabla_z^*\mu(dx) = D_z\mu(dx) = [(\lambda, z)_H + divz]\mu(dx)$.

To make both terms in square brackets of the last relation invariant, rewrite $\nabla_z^*\mu(dx)$ in the form $\nabla_z^*\mu(dx) = (\Lambda, z)_{\mathcal{H}_x} + div_G z$, where $\Lambda = \lambda - \Gamma_i^i$. Applying integration by parts formula, we derive that the generator $\mathcal{A}^*$ of $V^*(s,t)$ has the form

$$\begin{aligned}
\mathcal{A}^*\mu &= \frac{1}{2}[\nabla_{A^k}^*\nabla_{A^k}^*\mu + \nabla_{\nabla_{A^k}A^k}^*\mu + div_G A^k[\nabla_{A^k}^*\mu + div_G A^k\mu] \\
&\quad + \nabla_{A^k}^*(div_G A^k\mu) + 2div_G A^k\nabla_{A^k}^*\mu - div_G(\nabla_{A^k}A^k)\mu] \\
&\qquad - div_G a\mu - \nabla_a^*\mu \\
&= \frac{1}{2}[\nabla_{A^k}^*\nabla_{A^k}^*\mu + \nabla_{\nabla_{A^k}A^k}^*\mu] + div_G A^k\nabla_{A^k}^*\mu + \frac{1}{2}[(div_G A^k)(div_G A^k) \\
&\quad + [\nabla_{A^k}div_G A^k - div_G\nabla_{A^k}A^k]]\mu + div_G A^k\nabla_{A^k}^*\mu - div_G a\mu - \nabla_a^*\mu.
\end{aligned} \tag{6.32}$$

The next question we are going to answer is how to formulate the conditions on diffusion coefficients and the initial measure to ensure that the initial measure is invariant with respect to $V^*(t,s)$. For this purpose, we could follow the approach given in the previous section. We start with the case when coefficients

$A(t,x)=A(x)$ and $a(t,x)=a(x)$ do not depend on time variable. Notice that under the above condition, there exists a unique solution to

$$\frac{\partial\mu}{\partial t}=\mathcal{A}^*\mu, \quad \mu_0(dx)=\nu(dx). \tag{6.33}$$

It results that the measure $\mu_t(dy)=V^*(t)\nu(dx)$ is absolutely continuous with respect to ν and hence could be represented in the form $\mu_t(dx)=v(t,x)\nu(dx)$.

Let us derive the equation to govern the function $v(t,x)$. Direct computations show that

$$\frac{\partial[v(t,x)\nu(dx)]}{\partial t}=\frac{\partial v(t,x)}{\partial t}\nu(dx),$$
$$\nabla_z^*[v(t,x)\nu(dx)]=(v'(t,x),z_2)\nu(dx)+v(t,x)\nabla_{z_2}^*\nu(dx),$$

and

$$\begin{aligned}\nabla_{z_1}^*\nabla_{z_2}^*v(t,x)\nu(dx) &= (\nabla_{z_1}v'(t,x),z_2)\nu(dx)\\ &+(v'(t,x),\nabla_{z_1}z_2)\nu(dx)+(v'(t,x),z_2)\nabla_{z_1}^*\nu\\ &+(v'(t,x),z_1)\nabla_{z_2}^*\nu(dx)+v(t,x)\nabla_{z_1}^*\nabla_{z_2}^*\nu(dx).\end{aligned}$$

Finally, recall that $\nabla_z^*\nu=[(\Lambda,z)+div_G z]\nu$, $\nabla_{z_1}^*\nabla_{z_2}^*\nu=[(\nabla_{z_1}\Lambda,z_2)+(\Lambda,\nabla_{z_1}z_2)+\nabla_{z_1}div_G z_2]\nu+[(\Lambda,z_1)+div_G z_1][(\Lambda,z_2)+div_G z_2]\nu$ that yields

$$\begin{aligned}\nabla_{z_1}^*\nabla_{z_2}^*\nu-\nabla_{\nabla_{z_1}z_2}^*\nu &= [(\nabla_{z_1}\Lambda,z_2)+\nabla_{z_1}div_G z_2-div_G\nabla_{z_1}z_2\\ &+ [(\Lambda,z_1)+div_G z_1][(\Lambda,z_2)+div_G z_2]]\nu.\end{aligned}$$

Substituting these relations into (6.33), we show that $v(t,x)$ solves

$$\frac{\partial v}{\partial t}=Fv, \quad v(0,x)=1, \tag{6.34}$$

where

$$\begin{aligned}Fv &= \frac{1}{2}[\nabla_{A^k}\nabla_{A^k}v+\nabla_{\nabla_{A^k}A^k}v+\nabla_{A^k}v(\Lambda,A^k)+div_G A^k\nabla_{A^k}v]-\nabla_a v\\ &+\frac{1}{2}[(\Lambda,\nabla_{A^k}A^k)+\nabla_{A^k}div_G A^k-div_G\nabla_{A^k}A^k+(\Lambda,A^k)(\Lambda,A^k)\\ &+div_G A^k div_G A^k]+\nabla_{A^k}(\Lambda,A^k)+div_G A^k(\Lambda,A^k)-div_G a-(a,\Lambda)]v.\end{aligned} \tag{6.35}$$

Let us construct a probabilistic representation for the solution to (6.34) in terms of a new stochastic process $\gamma(t)$ which solves

$$d\gamma=\exp^M(b(\gamma(t)dt+A(\gamma(t))dw) \tag{6.36}$$

with the same diffusion coefficient $A(x)$ and a new drift coefficient $b(x)$ given by

$$b(x)=[(\Lambda(x),A^k(x))+\nabla_{A^k(x)}+div_G A^k(x)]A^k(x).$$

In terms of this process, the solution to (6.34) has the form

$$v(t,x) = \exp[\int_0^t \alpha(\gamma(\tau))d\tau], \tag{6.37}$$

where

$$\begin{aligned}\alpha(x) &= \frac{1}{2}[(\Lambda(x), \nabla_{A^k(x)}A^k(x)) + \nabla_{A^k(x)}div_G A^k(x) - div_G \nabla_{A^k(x)}A^k(x) \\ &+ (\Lambda(x), A^k(x))(\Lambda(x), A^k(x)) + div_G A^k(x) div_G A^k(x)] \\ &+ \nabla_{A^k(x)}(\Lambda(x), A^k(x))] \\ &+ div_G A^k(x)(\Lambda(x), A^k(x)) - div_G a(x) - (a(x), \Lambda(x))]. \end{aligned} \tag{6.38}$$

Finally, the condition $\alpha(x) = 0$ ensures that $v(t,x) = 1$ satisfies (6.34) and hence measure ν is invariant with respect to the evolution family $V^*(t)$.

In addition, we give another representation for the function $\alpha(x)$ in terms of the initial measure itself rather then its vector logarithmic derivative. In these terms, we derive

$$\begin{aligned}\alpha(x)\nu(dx) &= \frac{1}{2}[\nabla_{A^k}\nabla_{A^k}\nu(dx) + \nabla_{\nabla_{A^k}A^k}\nu(dx)] + \frac{1}{2}[div_G A^k \nabla_{A^k}\nu(dx) \\ &+ \kappa(A^k, A^k)\nu(dx) + (div_G A^k)(div_G A^k)\nu(dx)], \end{aligned} \tag{6.39}$$

where $\kappa(z_1, z_2)$ has the form (6.25).

Let $L_2(M,\nu)$ be a space of square integrable functions on M. Assume that $A(t,x) = A(x)$, $a(t,x) = a(x)$ and the measure ν is invariant with respect to the solution $\xi(t)$ of (6.26) with time homogenous coefficients. Define $V(t)$ and $V^+(t)$ in $L_2(M,\nu)$ by

$$\int_M g(x) \int_M f(y)P(t,x,dy)\nu(dx) = \langle g, V(t)f\rangle = \langle V^+(t)g, f\rangle \tag{6.40}$$

for any measurable bounded functions f and g. A pair of semigroups V_t and V_t^+ is said to be dual with respect to the invariant measure ν if (6.40) holds.

We show that the evolution family $V^+(t)$ coincides with the evolution family $\hat{V}(t)$ generated by the time reversal of $\xi(t)$ and derive the stochastic equation for the time reversal process.

Let $\hat{\xi}(t) = \xi(T-t), \quad t \in [0,T]$. By considerations similar to those used in the last Section, one could check that $\hat{\xi}(t)$ is a diffusion process as well, and $\hat{P}(t,x,dy) = \frac{P(t,y,dx)\nu(dy)}{\nu(dx)}$ is its transition probability. Finally,

$$\begin{aligned}\langle g, V_t f(x)\rangle &= \int_M g(x)[\int_M \frac{P(t,x,dy)}{\nu(dy)}\nu(dy)]\nu(dx) \\ &= \int_M f(y)\int_M g(x)\frac{P(t,x,dy)}{\nu(dy)}\nu(dx) = \int_M \hat{V}_t g(y) f(y)\nu(dy).\end{aligned}$$

Thus, the following assertion is proved.

Theorem 6.2.1 *Let $\nu(dy)$ be an invariant measure of the process $\xi(t) \in M$. Let V_t and $\hat{V}_t$ be evolution families in $L_2(M,\nu)$ generated by the processes $\xi(t)$ and $\hat{\xi}(t)$, respectively. Then $\hat{V}_t = V_t^+$, where V_t^+ is given by (6.40).*

Consider a pair of diffusion processes $\xi(t)$ and $\eta(t)$ in M which solve

$$d\xi = exp^M(a(\xi(t))dt + A(\xi(t))dw), \quad \xi(0) = \xi_0, \tag{6.41}$$

and

$$d\eta = exp^M(\hat{a}(\eta(t))dt + A(\eta(t))dw), \quad \eta(0) = \xi_0, \tag{6.42}$$

respectively, and let the distribution ν of $\xi_0 \in M$ be a smooth measure with vector logarithmic derivative λ. Finally, we state the conditions on drift coefficients a and $\hat{a}$ which ensure that $V_t f(x) = Ef(\xi(t))$ and $\hat{V}_t g(x) = Eg(\eta(t))$ are dual with respect to the invariant measure ν.

Theorem 6.2.2 *Let $\nu \in \mathcal{M}^2(M)$. Then ν is an invariant measure for a pair of diffusion processes $\xi(t)$ and $\eta(t)$ satisfying (6.41) and (6.42) if and only if the drift coefficients $a(x)$ and $\hat{a}(x)$ satisfy*

$$a(x) + \hat{a}(x) = B(x)\Lambda(x) + \nabla_{A^k(x)}A^k(x), \quad B = A^*A, \tag{6.43}$$

and

$$div_G(a(x) - \hat{a}(x)) + (a(x) - \hat{a}(x), \Lambda(x)) = 0. \tag{6.44}$$

PROOF. Recall that generators of the semigroups V_t and $\hat{V}_t$ have the form (6.30). If V_t and $\hat{V}_t$ are dual, then the relation $\langle \hat{\mathcal{A}}g, f\rangle_\mu = \langle g, \mathcal{A}f\rangle_\mu$ should hold. Using integration by part formula, we check that

$$\begin{aligned}
&\int_M [\frac{1}{2}(\nabla_{A^k(x)}\nabla_{A^k(x)} - \nabla_{\nabla_{A^k(x)}A^k(x)})g(x) + \nabla_{\hat{a}(x)}g(x)]f(x)\nu(dx) \\
&= \int_M g[\frac{1}{2}(\nabla_{A^k(x)}\nabla_{A^k(x)} - \nabla_{\nabla_{A^k(x)}A^k(x)})f(x) + div_G(\nabla_{A^k(x)}A^k(x))f(x) \\
&\quad + ((\nabla_{A^k(x)}A^k(x)), \nabla f(x)) - div_G(\hat{a}(x))f(x) - \nabla_{\hat{a}}f(x) \\
&\quad + f(x)[\frac{1}{2}((\nabla A^k(x)\Lambda(x), A^k(x)) + (A^k(x)\Lambda(x), A^k(x)\Lambda(x)) \\
&\quad + div_G(\nabla_{A^k(x)}A^k(x)) + (\nabla_{A^k(x)}A^k(x), \Lambda(x))) - (\hat{a}(x), \Lambda(x))]]\nu(dx) \\
&= \int_M g[\frac{1}{2}(\nabla_{A^k(x)}\nabla_{A^k(x)} - \nabla_{\nabla_{A^k(x)}A^k(x)})f(x) + \nabla_{a(x)}f(x)]\nu(dx).
\end{aligned}$$

As a result,

$$\begin{aligned}
&\frac{1}{2}div\,(\nabla_{A^k(x)}A^k(x)) - div_G(\hat{a}(x)) + \frac{1}{2}((\nabla A^k(x)\Lambda, A^k(x)) \\
&\qquad + (A^k(x)\Lambda, A^k(x)\Lambda) + (\nabla_{A^k(x)}A^k(x), \Lambda)) - (\hat{a}, \Lambda) = 0
\end{aligned}$$

and

$$(\nabla f, B(x)\Lambda + \nabla_{A^k(x)} A^k(x)) = (\hat{a}(x) + a(x), \nabla f).$$

Notice that (6.33) and the invariance of the measure ν with respect to V_t^+ yield (6.43). As far as ν is invariant with respect to V_t as well, we have

$$\begin{aligned} m &= \frac{1}{2}B_{ij}[\nabla_j\Lambda_i + \Lambda_i\Lambda_j] + (\nabla_i B_{ij} - \hat{a}_j)\Lambda_j + \frac{1}{2}\nabla_i\nabla_j B_{ij} - div_G\,\hat{a} \\ &= \frac{1}{2}B_{ij}[\nabla_j\Lambda_i + \Lambda_i\Lambda_j] + (\nabla_i B_{ij} - a_j)\Lambda_j + \frac{1}{2}\nabla_i\nabla_j B_{ij} - div_G\,a = 0 \end{aligned}$$

and hence (6.44) is valid. ■

Acknowledgement. This work was supported by grant RFBR N97-01-01123.

References

1. Belopolskaya, Ya. (1998). Invariant measures of diffusion processes in Hilbert space, *Abstracts of VII Vilnius Conference on Probability Theory and Mathematical Statistics*, 146–147.

2. Belopolskaya, Ya. and Dalecky, Yu. (1990). *Stochastic Equations and Differential Geometry*, Dordrecht: Kluwer Academic Publishers.

3. Dalecky, Yu. Infinite dimensional elliptic operators and parabolic equations associated with them, *Russian Math Surveys*, **22**, 1–53.

4. Dalecky, Yu. and Fomin, S. (1991). *Measures and Differential Equations in Infinite Dimensional Space*, Dordrect: Kluwer Academic Publishers.

5. Dalecky, Yu. and Steblovskaya, V. (1996). On absolutely continuous and invariant evolution of smooth measure in Hilbert space, *C. R. Acad. Sci. Paris* t. 323, Serie 1, 823–827.

6. Follmer, H. and Wakolbinger, A. (1986). Time reversal of infinite dimensional diffusion processes, *Stochastics Processes and their Applications*, **22**, 59–78.

7. Gross, L. Potential theory on Hilbert space, *Journal of Functional Analysis*, **1**, 123–181.

8. Nagasawa, M. (1961). The adjoint process of a diffusion with reflecting barrier, *Kodai Mathematics Seminar, Report 13*, 235–248.

7

Localization of Majorizing Measures

Bettina Bühler, Wenbo V. Li, and Werner Linde

Friedrich Schiller Universität, Jena, Germany
University of Delaware, Newark, Delaware
Friedrich Schiller Universität, Jena, Germany

Abstract: A fundamental result of Fernique and Talagrand characterizes the a.s. boundedness and continuity of Gaussian processes by properties of their covariance. Basic tools are either majorizing measures or quantities defined by partitions and weights. Our objective in this chapter is to investigate generalizations of those expressions. As a consequence, we get a deeper understanding of the structure of precompact metric spaces as well as of Gaussian processes.

Keywords and phrases: Gaussian processes, majorizing measure, metric entropy

7.1 Introduction

Let $Y = (Y_t)_{t\in T}$ be a centered Gaussian process over an arbitrary index set T. Then, Y is (up to equivalence) completely determined by its covariance. It is natural and important to find criteria (only depending on the covariance) which will ensure the existence of a.s. bounded or continuous (if T is metric) versions of Y. By standard methods [Ledoux and Talagrand (1991)], it may be reduced to the special case in the Hilbert space setting. Let H be a separable Hilbert space and let $X = (X_t)_{t\in H}$ be an isonormal Gaussian process defined on H. For example, let X be given by

$$X_t := \sum_{j=1}^{\infty} \xi_j \langle t, f_j \rangle \ , \quad t \in H, \tag{7.1}$$

where $(\xi_j)_{j=1}^{\infty}$ are independent and identically distributed as standard normal and $(f_j)_{j=1}^{\infty}$ is any ONB in H. Then, one needs geometric characterizations of subsets T in H for which $(X_t)_{t\in T}$ has either a bounded or a bounded and uniformly continuous version (in the terminology of Dudley (1967), those sets were

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

called GB– or GC–sets, respectively). Metric entropy conditions are used but it only led to sufficient [cf. Dudley (1967)] or necessary conditions [cf. Sudakov (1969)] due to the fact that it is a very rough tool to describe the detailed structure of a set T precisely. Consequently, finer (and more complicated) quantities were necessary. After basic work of Fernique [cf. Fernique (1975)], the problem was finally solved in Talagrand (1987). For example, the process $X = (X_t)_{t\in T}$ has an a.s. bounded version iff there is a probability measure μ on T (majorizing measure) with

$$\sup_{t\in T}\int_0^\infty \sqrt{\log\frac{1}{\mu(B(t;\varepsilon))}}\,d\varepsilon < \infty \tag{7.2}$$

where $B(t;\varepsilon) := \{s \in T : \|t-s\|_H \le \varepsilon\}$ is the closed ε–ball centered at $t \in T$. Recently, Talagrand [cf. Ledoux (1996) and Talagrand (1996)] presented other conditions which are equivalent to (7.2) and are much easier to handle in concrete situations. For example, a very useful quantity $\Theta(T)$ [see (7.4)] is defined by a countable increasing sequence of partitions of T and corresponding weights. Then the existence of a bounded version of X over T in H is equivalent to $\Theta(T) < \infty$. This is a remarkable result, but it has the disadvantage that it neither tells something about the size of sets T when $\Theta(T) = \infty$, nor does it give any additional information about the structure of T when $\Theta(T) < \infty$. This led us to modify the definition of $\Theta(T)$ in a simple and effective way. Instead of taking partitions and weights starting at a fixed level $N_0(T)$, we investigate partitions and weights ranging from a certain level N to a (possible finite) level M, thus getting quantities $\Theta_N^M(T)$ for any choice of $-\infty < N < M \le \infty$. When $\Theta(T) = \infty$, the behavior of $\Theta_N^M(T) \to \infty$ as $M \to \infty$ leads to detailed information about T. And when $\Theta(T) < \infty$, the speed of convergence of $\Theta_N^\infty(T)$ as $N \to \infty$ provides the desired information. Hence, we finally obtain a tool which describes the finer structure of any precompact $T \subset H$, and this has already turned out to be very useful in the study of the size of the convex hull of a precompact set [cf. Li and Linde (1998)].

The organization of this chapter is as follows. After introducing $\Theta_N^M(T)$ and proving some elementary properties of this quantity, we extend Talagrand's partitioning scheme to our more general setting. As a consequence, we get a majorizing measure characterization for $\Theta_N^M(T)$ similar to (7.2). Although all these quantities originate in the study of Gaussian processes, they also make sense in arbitrary metric spaces (T, d), and we shall investigate these expressions in this more general setting. Only when we treat their relations to Gaussian processes, the metric d is assumed to be generated by a scalar product. In this case, we get a probabilistic characterization of $\Theta_N^M(T)$ completing the classical relation between majorizing measures and Gaussian processes. Finally, some examples show how our results apply in concrete situations.

7.2 Partitions and Weights

Throughout this chapter, (T,d) denotes a metric space with

$$0 < \mathrm{diam}(T) := \sup_{s,t\in T} d(s,t) < \infty .$$

If $B(t;\varepsilon)$ is the closed ε–ball centered in $t \in T$ (w.r.t. the metric d), the covering numbers of a set $A \subseteq T$ are defined by

$$N(A,\varepsilon) := \inf \left\{ n \in \mathbf{N} : \exists\, t_1,\ldots,t_n \in T \ \text{ s.t. } \ A \subseteq \bigcup_{k=1}^{n} B(t_k;\varepsilon) \right\}$$

and its metric entropy is defined by $H(A,\varepsilon) := \log N(A,\varepsilon)$.

Here and in the sequel, q is a sufficiently large fixed number and all constants $c > 0$ (with or without subscript) are assumed to depend on this number q only.

For integers N, M with $-\infty < N < M \le \infty$, let $\mathcal{A} = (\mathcal{A}_{N+1},\ldots,\mathcal{A}_M)$ for $M < \infty$ or $\mathcal{A} = (\mathcal{A}_{N+1},\mathcal{A}_{N+2},\ldots)$ for $M = \infty$ be a sequence of partitions of T possessing the following properties:

(i) Each $\mathcal{A}_j$ is finite and consists of measurable (w.r.t. the Borel–σ–algebra on T) subsets.

(ii) Given $A \in \mathcal{A}_j$, there is a $t \in T$ such that $A \subseteq B(t;q^{-j})$, i.e. we have $N(A,q^{-j}) = 1$ and, hence, $\mathrm{diam}(A) \le 2q^{-j}$.

(iii) Each $\mathcal{A}_{j+1}$ refines $\mathcal{A}_j$.

Let $\mathcal{Z}_N^M(T)$ or $\mathcal{Z}_N^M(T,d)$ be the set of sequences $\mathcal{A}$ which possess properties (i), (ii) and (iii) from above. Given $t \in T$ and $\mathcal{A} \in \mathcal{Z}_N^M(T)$, then for each finite j with $N < j \le M$ we find a unique $A_j(t)$ in $\mathcal{A}_j$ with $t \in A_j(t)$. Note that in view of (iii), necessarily $A_{j+1}(t) \subseteq A_j(t)$.

If $\mathcal{A} \in \mathcal{Z}_N^M(T)$, a sequence $w = (w_{N+1},\ldots,w_M)$ or $w = (w_{N+1},w_{N+2},\ldots)$ is called weight sequence adapted to $\mathcal{A}$ provided

(i) each w_j maps $\mathcal{A}_j$ into $[0,1]$ and

(ii) for all finite j with $N+1 \le j \le M$, we have $\sum_{A\in\mathcal{A}_j} w_j(A) \le 1$.

Denote by $\mathcal{W}(\mathcal{A})$ the set of all those sequences adapted to some $\mathcal{A}$. Basic examples of weight sequences are those defined by probability measures μ on T. Indeed, if $\mathcal{A} \in \mathcal{Z}_N^M(T)$, setting for all j's

$$w_j(A) := \mu(A)\,, \quad A \in \mathcal{A}_j, \tag{7.3}$$

we always obtain a sequence $w \in \mathcal{W}(\mathcal{A})$.

Now we are in position to introduce the main quantity of this chapter: Given $\mathcal{A} \in \mathcal{Z}_N^M(T)$ and $w \in \mathcal{W}(\mathcal{A})$, set

$$\Theta_{\mathcal{A},w}(T) := \sup_{t\in T} \sum_{j=N+1}^{M} q^{-j} \sqrt{\log \frac{1}{w_j(A_j(t))}}$$

with $A_j(t)$ as above. Finally, we define

$$\Theta_N^M(T) := \inf \left\{ \Theta_{\mathcal{A},w}(T) : \mathcal{A} \in \mathcal{Z}_N^M(T) \,,\; w \in \mathcal{W}(\mathcal{A}) \right\}$$

where the numbers N and M indicate that we use partitions from level $N+1$ to level M. Observe (and this is the main difference to former investigations) that the optimal choices for $\mathcal{A}$ and w may heavily depend on N and M. Thus, in general, there is no natural relation between $\mathcal{A}$'s and w's being optimal for different indices.

Two choices of N and M are of special interest:

(1) Let $N_0 = N_0(T)$ be defined by

$$N_0 := \sup \left\{ j \in \mathbf{Z} : N(T, q^{-j}) = 1 \right\} \,.$$

For $N < N_0$ and $N < j \le N_0$, we may choose $\mathcal{A}_j = \{T\}$ and $w_j \equiv 1$, hence $\Theta_N^M(T) = \Theta_{N_0}^M(T)$ in this case. So without loss of any generality, we can always assume $N \ge N_0(T)$. Then for any precompact space T the quantity $\Theta_{N_0}^M(T)$, $M < \infty$, is finite and its behaviour as $M \to \infty$ measures the degree of compactness of T.

(2) Another case of interest is $M = \infty$. Here, one may ask for the behaviour of $\Theta_N^\infty(T)$ as $N \to \infty$.

Combining these two cases, i.e. $N = N_0(T)$ and $M = \infty$, we obtain the classical quantity

$$\Theta(T) := \Theta_{N_0}^\infty(T) \tag{7.4}$$

investigated in Ledoux (1996) and Talagrand (1996).

7.3 Simple Properties of $\Theta_N^M(T)$

As we mentioned above, every probability measure μ on T defines via (7.3) a sequence $w \in \mathcal{W}(\mathcal{A})$. A first result tells us that it suffices to investigate weight sequences generated in this way. Let $\mathcal{P}(T)$ be the set of Borel probability measures on T. Then the following is valid:

Proposition 7.3.1 *We have*

$$\begin{aligned}\Theta_N^M(T) &\le \inf\left\{\sup_{t\in T}\sum_{j=N+1}^{M} q^{-j}\sqrt{\log\frac{1}{\mu(A_j(t))}} : \mathcal{A}\in\mathcal{Z}_N^M(T),\ \mu\in\mathcal{P}(T)\right\}\\ &\le c\cdot\Theta_N^M(T)\,.\end{aligned}$$

PROOF. The left hand inequality follows easily by setting $w_j(A)=\mu(A)$. For the right hand inequality, let $\mathcal{A}\in\mathcal{Z}_N^M(T)$ and $w\in\mathcal{W}(\mathcal{A})$ be arbitrary. We choose some point $t_A\in A$ for any $A\in\mathcal{A}_j$ and set

$$\mu_0=\sum_{j=N+1}^{M}2^{-j+N}\sum_{A\in\mathcal{A}_j}w_j(A)\,\delta_{t_A}.$$

Then we get $\mu_0(T)\le 1$, thus there exists a $\mu\in\mathcal{P}(T)$ with $\mu\ge\mu_0$. Using $N\ge N_0$, we find an $A\in\mathcal{A}_{N+1}$ with $w_{N+1}(A)\le 1/2$, which clearly implies $\Theta_{\mathcal{A},w}(T)\ge q^{-N-1}\sqrt{\ln 2}$. Now using the fact $\mu(A_j(t))\ge 2^{-j+N}w_j(A_j(t))$ and standard methods [cf. Ledoux (1996)], we obtain the desired inequality with $c=q\sum_{j=1}^{\infty}q^{-j}\sqrt{j}+1$. ∎

The next result answers the natural question as to how $\Theta_N^M(T)$ depends on the indices N and M, respectively.

Proposition 7.3.2
(a) *If* $N+1<M\le\infty$, *then* $\Theta_{N+1}^M(T)\le\Theta_N^M(T)\le c\,\Theta_{N+1}^M(T)$. *Thus, the asymptotic behaviour of* $\Theta_N^M(T)$ *as* $M\to\infty$ *does not depend on the special choice of* N.
(b) *For* $N<M<\infty$, *we have*

$$\Theta_N^M(T)\le\Theta_N^{M+1}(T)\le c\,\Theta_N^M(T)+q^{-M-1}\sqrt{H(T,q^{-M-1})}\,.$$

PROOF. Both left hand sides are trivial. For the right hand side of (a), we use Proposition 7.3.1 and $A_{N+2}(t)\subseteq A_{N+1}(t)$, hence $\mu(A_{N+2}(t))\le\mu(A_{N+1}(t))$ for any $\mu\in\mathcal{P}(T)$. To prove the right hand side of (b), let $\mathcal{A}\in\mathcal{Z}_N^M(T)$ and $w\in\mathcal{W}(\mathcal{A})$ be arbitrary. Choose a partition $\mathcal{B}$ of T which is induced by a covering with $N(T,q^{-M-1})$ balls of radius q^{-M-1}. Then we set

$$\mathcal{A}_{M+1}=\{A\cap B\,:\,A\in\mathcal{A}_M,\,B\in\mathcal{B},\,A\cap B\neq\emptyset\}$$

and

$$w_{M+1}(A\cap B)=\frac{w_M(A)}{N(T,q^{-M-1})}\,,$$

and get $\widetilde{\mathcal{A}}=(\mathcal{A}_{N+1},\ldots,\mathcal{A}_{M+1})\in\mathcal{Z}_N^{M+1}(T)$ and $\widetilde{w}=(w_{N+1},\ldots,w_{M+1})$ in $\mathcal{W}(\widetilde{\mathcal{A}})$. Thus we obtain

$$\Theta_{\widetilde{\mathcal{A}},\widetilde{w}}(T)\le(1+1/q)\Theta_{\mathcal{A},w}(T)+q^{-M-1}\sqrt{H(T,q^{-M-1})}$$

as desired. ■

In view of property (ii) in the definition of $\mathcal{A}$'s in $\mathcal{Z}_N^M(T)$, the inclusion $S \subseteq T$ does not necessarily imply $\Theta_N^M(S) \leq \Theta_N^M(T)$. But we have the following result.

Proposition 7.3.3 *For $S \subseteq T$ and $N < M$, we have*

$$\Theta_{N-1}^{M-1}(S) \leq q\ \Theta_N^M(T).$$

PROOF. Let $\mathcal{A} \in \mathcal{Z}_N^M(T)$ and $w \in \mathcal{W}(\mathcal{A})$ be arbitrary. For $N-1 < j \leq M-1$ we define

$$\widetilde{\mathcal{A}}_j = \{A \cap S \,:\, A \in \mathcal{A}_{j+1}\,,\ A \cap S \neq \emptyset\}\,.$$

Then we obtain $A \cap S \subseteq B(s; q^{-j})$ for some $s \in S$ as follows: By definition of $\mathcal{Z}_N^M(T)$, we have $A \subseteq B(t; q^{-j-1})$ for some $t \in T$. Using $A \cap S \neq \emptyset$, we find $s \in S$ with $d(s,t) \leq q^{-j-1}$, so that $A \cap S \subseteq B(s; 2q^{-j-1}) \subseteq B(s; q^{-j})$. Hence, $\widetilde{\mathcal{A}} = (\widetilde{\mathcal{A}}_N, \ldots, \widetilde{\mathcal{A}}_{M-1}) \in \mathcal{Z}_{N-1}^{M-1}(S, d_{|S\times S})$, and by setting $\widetilde{w}_j(A \cap S) = w_{j+1}(A)$, we get $\Theta_{\widetilde{\mathcal{A}},\widetilde{w}}(S) \leq q\ \Theta_{\mathcal{A},w}(T)$ which completes the proof. ■

Let (T,d) be precompact. Then there exists a tight connection between $\Theta_N^M(T)$ and quantities defined by the metric entropy of T. To make this more precise, let $\mathcal{A}_{N+1}$ be generated by an optimal q^{-N-1}–cover, i.e. $\mathcal{A}_{N+1}$ consists of suitable intersections of q^{-N-1}–balls and $\mathrm{card}(\mathcal{A}_{N+1}) = N(T, q^{-N-1})$. If $\mathcal{A}_j$, $j < M$, is already defined, we divide any $A \in \mathcal{A}_j$ in the same way by an optimal q^{-j-1}–cover of A. In this way, we obtain a sequence $\mathcal{A} \in \mathcal{Z}_N^M(T)$ with

$$\mathrm{card}(\mathcal{A}_j) \leq N(T, q^{-N-1}) \cdots N(T, q^{-j})\ .$$

If $j = N+1, \ldots, M$, we define adapted weights by $w_j(A) := (\mathrm{card}(\mathcal{A}_j))^{-1}$ for all $A \in \mathcal{A}_j$, and it is easy to see that

$$\Theta_{\mathcal{A},w}(T) \leq c \int_{q^{-M-1}}^{q^{-N-1}} \sqrt{H(T,\varepsilon)}\, d\varepsilon\ .$$

This is the classical Dudley bound in the case of Gaussian processes when $N = N_0$ and $M = \infty$. Also a lower Sudakov bound is valid in this more general setting. This can be seen using $N(A, q^{-j}) = 1$ for $A \in \mathcal{A}_j$, $\mathcal{A} \in \mathcal{Z}_N^M(T)$, so that we find $t(A) \in T$ with

$$T \subseteq \bigcup_{A \in \mathcal{A}_j} A \subseteq \bigcup_{A \in \mathcal{A}_j} B(t(A); q^{-j}).$$

Thus $N(T, q^{-j}) \leq \mathrm{card}(\mathcal{A}_j)$, and for each j there exists a set $A \in \mathcal{A}_j$ with $w_j(A) \leq (N(T, q^{-j}))^{-1}$. Summing up, the following holds.

Proposition 7.3.4 *We have*

$$\sup_{N+1\le j\le M} q^{-j}\sqrt{H(T,q^{-j})} \le \Theta_N^M(T) \le c \int_{q^{-M-1}}^{q^{-N-1}} \sqrt{H(T,\varepsilon)}\, d\varepsilon\ . \tag{7.5}$$

Let us state a consequence of this proposition, which has been used in Carl, Kyrezi and Pajor (1997) in a weaker form.

Corollary 7.3.1 *Let $\alpha > 0$ and $\beta \in \mathbf{R}$. Then, $H(T,\varepsilon) \le c\,\varepsilon^{-\alpha}(\log(1/\varepsilon))^{\beta}$ iff*

$$\Theta_N^{\infty}(T) \le c\, q^{N(\alpha/2-1)} N^{\beta/2}\ , \quad N \to \infty\,,$$

for $0 < \alpha < 2$, and iff

$$\Theta_N^{M}(T) \le c\, q^{M(\alpha/2-1)} M^{\beta/2}\ , \quad M \to \infty\ ,$$

for $2 < \alpha < \infty$.

Remark. The estimates in (7.5) do not yield similar assertions in the critical case $\alpha = 2$. For special T, this problem has been investigated in Li and Linde (1998).

7.4 Talagrand's Partitioning Scheme

Given a metric space (T,d) and $N < M \le \infty$, it is a highly non–trivial task to construct optimal $\mathcal{A} \in \mathcal{Z}_N^M(T)$ and $w \in \mathcal{W}(\mathcal{A})$. As mentioned before, partitions and weights generated by optimal q^{-j}–covers do not lead to sharp results in general. The deeper reason for this is that those weights do not suffice to describe the finer structure of T. Fortunately, Talagrand found a general scheme for constructing optimal $\mathcal{A}$'s in $\mathcal{Z}_{N_0}^{\infty}(T)$ and $w \in \mathcal{W}(\mathcal{A})$ [cf. Talagrand (1996)]. The same ideas also apply in our more general situation. To be more precise, we need the following.

Given N, M as above, a sequence $\varphi_N, \ldots, \varphi_{M+1}$ of functions from T into $[0,\infty)$ satisfies the Talagrand condition (for some $\kappa > 0$) if they possess the following property:

For $N \le j \le M-1$, all $t \in T$ and any points $t_1, \ldots t_n \in B(t, q^{-j})$ with $d(t_l, t_k) > q^{-j-1}$ for $1 \le k < l \le n$, we have

$$\varphi_j(t) \ge \kappa\, q^{-j}\sqrt{\log n} + \min_{1\le l\le n} \varphi_{j+2}(t_l)\ . \tag{7.6}$$

Theorem 7.4.1 *Let $\varphi = (\varphi_N, \ldots, \varphi_{M+1})$ be a sequence of functions satisfying the Talagrand condition (7.6) with $\kappa > 0$. Let*

$$\|\varphi\| := \sup_{t\in T}\ \sup_{N\le j\le M+1} \varphi_j(t).$$

Then there exist $\mathcal{A} \in \mathcal{Z}_N^M(T)$ *and* $w \in \mathcal{W}(\mathcal{A})$ *such that*

$$\Theta_{\mathcal{A},w}(T) \le c\left(\kappa^{-1}\,\|\varphi\| + q^{-N}\sqrt{H(T,q^{-N})+1}\right) . \tag{7.7}$$

PROOF. The proof follows almost exactly as for $N = N_0$ and $M = \infty$ in Talagrand (1996), but with one important difference. If $N = N_0$, then there exist a natural partition $\mathcal{A}_{N_0} = \{T\}$ and a natural weight $w_{N_0} \equiv 1$ to start with. Then $\mathcal{A}_j$ and w_j, $j > N_0$, are constructed inductively by dividing sets in $\mathcal{A}_{j-1}$ and by splitting w_{j-1}. But, in our situation, $N > N_0$, we have to find a natural partition to start the partitioning procedure. We choose $\mathcal{A}_N$ as partition generated by an optimal q^{-N}–cover and set $w_N = (N(T,q^{-N}))^{-1}$. This leads for $N > N_0(T)$ to the additional term $q^{-N}\sqrt{H(T,q^{-N})}$ on the right hand side of (7.7). Then we may proceed as in the classical case. One should also observe that only $\varphi_N, \ldots, \varphi_{j+1}$ are needed to construct $\mathcal{A}_N, \ldots, \mathcal{A}_j$. ■

7.5 Majorizing Measures

Recall that first characterizations of metric spaces T with $\Theta(T) < \infty$ were via (7.2) by using special probability measures on T (called majorizing measures). Similar results also hold in our more general situation. Given $\mu \in \mathcal{P}(T)$ and $N_0 \le N < M \le \infty$, we define

$$\mathcal{I}_{N,\mu}^M(T) := \sup_{t\in T}\int_{2q^{-M}}^{2q^{-N}} \sqrt{\log\frac{1}{\mu(B(t;\varepsilon))}}\,d\varepsilon$$

and

$$\mathcal{I}_N^M(T) := \inf\left\{\mathcal{I}_{N,\mu}^M(T) : \mu \in \mathcal{P}(T)\right\} . \tag{7.8}$$

Theorem 7.5.1 *There are* $c_1, c_2 > 0$ *such that*

$$c_1\,\mathcal{I}_N^M(T) \le \Theta_N^M(T) \le c_2\,\mathcal{I}_N^{M+1}(T) .$$

PROOF. The first inequality is easy to prove by using $A_j(t) \subseteq B(t;2q^{-j})$ and Proposition 7.3.1. For the second inequality, we set

$$\varphi_j(t) := \sup\left\{\int_{2q^{-M-1}}^{2q^{-j}} \sqrt{\log\frac{1}{\mu(B(s;\varepsilon))}}\,d\varepsilon\,;\; s \in T,\; d(s,t) \le 2q^{-j}\right\}$$

for $N \le j \le M+1$. As in Talagrand (1996), it can be proved that these functions satisfy $\|\varphi\| \le \mathcal{I}_N^{M+1}(T)$ as well as the Talagrand condition with corresponding constant $\kappa = (q-8)/(2q^2)$. Furthermore, it is easy to prove that

$$\mathcal{I}_N^{M+1}(T) \ge c\,q^{-N}\sqrt{H(T,q^{-N})+1},$$

so the result follows by applying Theorem 7.4.1. ■

Remark. Letting

$$\mathcal{I}(T) = \mathcal{I}_{N_0}^{\infty}(T) = \inf_{\mu \in \mathcal{P}(T)} \sup_{t \in T} \int_0^{\infty} \sqrt{\log \frac{1}{\mu(B(t;\varepsilon))}}\, d\varepsilon \,,$$

then the classical majorizing measure theorem reads now

$$c_1\, \mathcal{I}(T) \le \Theta(T) \le c_2\, \mathcal{I}(T). \tag{7.9}$$

7.6 Approximation Properties

Let (T,d) be a metric space with $\Theta(T) < \infty$. Then the behaviour of $\Theta_N^{\infty}(T)$ as $N \to \infty$ should tell us more about the degree of compactness of T. The first aim of this section is to make this more precise.

Lemma 7.6.1 *Suppose that $T = \bigcup_{j=1}^{k} T_j$ for some disjoint T_j's. Then we have*

$$\Theta_N^{\infty}(T) \le c \left(\sup_{1 \le j \le k} \Theta_N^{\infty}(T_j) + q^{-N}\sqrt{\log k} \right) .$$

PROOF. Let $\mu_1, \ldots, \mu_k$ be probability measures on $T_1, \ldots, T_k$. Then we define a probability measure μ on T by $\mu := k^{-1}\sum_{j=1}^{k} \mu_j$. Let $\mathcal{A}^j = \left\{ \mathcal{A}_{N+1}^j, \mathcal{A}_{N+2}^j \ldots \right\}$ be partitions of T_j, $1 \le j \le k$. Setting $\mathcal{A}_i := \bigcup_{j=1}^{k} \mathcal{A}_i^j$, we get a sequence $\mathcal{A} = \{\mathcal{A}_{N+1}, \mathcal{A}_{N+2}, \ldots\}$ in $\mathcal{Z}_N^{\infty}(T)$. If $t \in T_j$, we have

$$A_i(t) = A_i^j(t), \quad i = N+1, N+2, \ldots,$$

hence by using Proposition 7.3.1

$$\begin{aligned} \Theta_N^{\infty}(T) &\le \sup_{1 \le j \le k} \sup_{t \in T_j} \sum_{i=N+1}^{\infty} q^{-i} \sqrt{\log \frac{k}{\mu_j(A_i^j(t))}} \\ &\le \sup_{1 \le j \le k} \sup_{t \in T_j} \sum_{i=N+1}^{\infty} q^{-i} \left(\sqrt{\log \frac{1}{\mu_j(A_i^j(t))}} + \sqrt{\log k} \right) . \end{aligned}$$

Since the partitions and the measures on each T_j were chosen arbitrarily, this completes the proof by taking the infimum over all partitions and measures, and by using Proposition 7.3.1 again. ■

Theorem 7.6.1 *If T is a metric space, then*

$$c_1\, \Theta_N^{\infty}(T) \le \sup_{t \in T} \Theta(B(t;q^{-N})) + q^{-N}\sqrt{H(T,q^{-N})} \le c_2\, \Theta_N^{\infty}(T)\,. \tag{7.10}$$

PROOF. To verify the left hand side, let $T = \bigcup_{j=1}^{k} B(t_j; q^{-N})$ for $k = N(T, q^{-N})$ and $t_1, \ldots, t_k$ in T. Then we obtain disjoint sets $T_j \subseteq B(t_j; q^{-N})$ covering T. Hence, Lemma 7.6.1 implies

$$\begin{aligned} \Theta_N^\infty(T) &\le c \left(\sup_{1 \le j \le k} \Theta(B(t_j; q^{-N})) + q^{-N}\sqrt{H(T, q^{-N})} \right) \\ &\le c \left(\sup_{t \in T} \Theta(B(t; q^{-N})) + q^{-N}\sqrt{H(T, q^{-N})} \right) \end{aligned}$$

as claimed.

Conversely, by Propositions 7.3.3 and 7.3.2 we have

$$\sup_{t \in T} \Theta(B(t; q^{-N})) = \sup_{t \in T} \Theta_N^\infty(B(t; q^{-N})) \le c\, \Theta_N^\infty(T)$$

and Proposition 7.3.4 implies $q^{-N-1}\sqrt{H(T, q^{-N})} \le \Theta_N^\infty(T)$, hence

$$\sup_{t \in T} \Theta(B(t; q^{-N})) + q^{-N}\sqrt{H(T, q^{-N})} \le c\, \Theta_N^\infty(T),$$

completing the proof. ∎

Remark. Note that entropy term in (7.10) is indeed necessary. For example, if all points of T are ε–separated with $\varepsilon > q^{-N}$, then this yields $\Theta(B(t; q^{-N})) = 0$ for $t \in T$, yet $\Theta_N^\infty(T) > 0$ if $\mathrm{card}(T) > 1$.

Corollary 7.6.1 *Let (T, d) be precompact. Then it holds $\lim_{N \to \infty} \Theta_N^\infty(T) = 0$ iff*

$$\lim_{\varepsilon \to 0} \sup_{t \in T} \Theta(B(t; \varepsilon)) = 0 \,.$$

PROOF. We have to show that $\lim_{\varepsilon \to 0} \sup_{t \in T} \Theta(B(t; \varepsilon)) = 0$ implies

$$\lim_{\delta \to 0} \delta\sqrt{H(T, \delta)} = 0.$$

This can be done by using

$$N(T, \delta) \le N(T, \varepsilon) \sup_{t \in T} N(B(t; \varepsilon), \delta)$$

and Proposition 7.3.4 [cf. the proof of Corollary 3.19 in Ledoux and Talagrand (1991)]. ∎

If T is precompact and $\Theta(T) = \infty$, then the behavior of $\Theta_N^M(T)$ as $M \to \infty$ should describe how badly T is behaved, i.e., how far it is away from a set S with $\Theta(S) < \infty$. Before making this more precise, we need the following lemma.

Lemma 7.6.2 *Suppose we have $T = \bigcup_{j=1}^{\infty} T_j$ for some disjoint measurable subsets $T_j \subseteq T$. Let ρ be a metric on $\mathbf{N}$ such that for $B \subseteq \mathbf{N}$ we have*

$$\operatorname{diam}_d(\bigcup_{j\in B} T_j) \le \operatorname{diam}_\rho(B) + 4q^{-M-1} \tag{7.11}$$

for some $M \in \mathbf{Z}$. Then this implies

$$\Theta_N^M(T) \le c\, \Theta_N^{M+1}(\mathbf{N}) \le c\, \Theta_N^\infty(\mathbf{N})\,.$$

PROOF. Let $\mathcal{B} = \{\mathcal{B}_{N+1}, \dots, \mathcal{B}_{M+1}\}$ be a sequence of partitions in $\mathcal{Z}_N^{M+1}(\mathbf{N}, \rho)$ and let $v = (v_{N+1}, \dots, v_{M+1})$ be a sequence of adapted weights. Given $B \in \mathcal{B}_{i+1}$ for some $i \in \{N, \dots, M\}$, we define a subset $A_B := \bigcup_{j\in B} T_j \subseteq T$. Then by (7.11), $\operatorname{diam}_d(A_B) \le 2q^{-i-1} + 4q^{-M-1} \le q^{-i}$ provided $q \ge 6$. Defining now

$$\mathcal{A}_i := \{A_B : B \in \mathcal{B}_{i+1}\} \quad \text{for} \quad i = N+1, \dots, M,$$

we get a sequence $\mathcal{A} \in \mathcal{Z}_N^M(T,d)$. Next let $w_i(A_B) := v_{i+1}(B)$ for $B \in \mathcal{B}_{i+1}$. Given $t \in T$, there is a unique $j \in \mathbf{N}$ such that $t \in T_j$. Hence, if $B_{i+1}(j)$ is the unique set in $\mathcal{B}_{i+1}$ such that $j \in B_{i+1}(j)$, we have $A_i(t) = A_{B_{i+1}(j)}$ by the construction. Consequently, for each fixed $t \in T$

$$\sum_{i=N+1}^{M} q^{-i}\sqrt{\log\frac{1}{w_i(A_i(t))}} \;\le\; q \sup_{j\in\mathbf{N}} \sum_{i=N+1}^{M+1} q^{-i}\sqrt{\log\frac{1}{v_i(B_i(j))}}\,,$$

so after taking the supremum over all $t \in T$ on the left-hand side, we get the desired estimate by taking the infimum over all partitions and weights on the right-hand side. ∎

Theorem 7.6.2 *Let T be a subset of a metric space (E,d) and suppose that there is a countable set $S \subset E$ such that*

$$T \subseteq \bigcup_{s\in S} B(s; 2q^{-M-1})$$

for some $M \in \mathbf{Z}$. Then this implies

$$\Theta_N^M(T) \le c\, \Theta_N^{M+1}(S) \le c\, \Theta_N^\infty(S)\,.$$

PROOF. Of course, we may assume $B(s; 2q^{-M-1}) \cap T \ne \emptyset$ for any $s \in S$. Writing $S = \{s_1, s_2, \dots\}$, there exist disjoint subsets $T_j \subseteq B(s_j; 2q^{-M-1})$ such that $T = \bigcup_{j\in\mathbf{N}} T_j$. Define a distance ρ on $\mathbf{N}$ by

$$\rho(i,j) := d(s_i, s_j) \quad \text{for } i,j \in \mathbf{N}\,.$$

Then estimate (7.11) holds by the construction and we obtain

$$\Theta_N^M(T) \le c\, \Theta_N^{M+1}(\mathbf{N}) = c \cdot \Theta_N^{M+1}(S)$$

completing the proof. ∎

Next, we view our results in a uniform way.

Proposition 7.6.1 *We have*

$$\begin{aligned} c_1\ \Theta_N^M(T) &\le \inf\left\{\Theta_N^\infty(S) : T \subseteq \bigcup_{s\in S} B(s;2q^{-M-1})\ ,\ S \textit{ countable}\right\} \\ &\le \inf\left\{\Theta_N^\infty(S) : T \subseteq \bigcup_{s\in S} B(s;2q^{-M-1})\ ,\ S\subseteq T\ ,\ S \textit{ countable}\right\} \\ &\le \inf\left\{\Theta_N^\infty(S) : T \subseteq \bigcup_{s\in S} B(s;2q^{-M-1})\ ,\ S\subseteq T\ ,\ S \textit{ finite}\right\} \\ &\le c_2\ \Theta_N^{M+1}(T)\ . \end{aligned}$$

PROOF. It only remains to prove the last inequality. For that purpose, let $\mathcal{A} = (\mathcal{A}_{N+1}, \dots, \mathcal{A}_{M+1}) \in \mathcal{Z}_N^{M+1}(T)$ and $w = (w_{N+1}, \dots, w_{M+1}) \in \mathcal{W}(\mathcal{A})$ be arbitrary. For any $A \in \mathcal{A}_{M+1}$, we choose an element $t_A \in A$ and denote the set $\{t_A : A \in \mathcal{A}_{M+1}\}$ by S. Obviously $S \subseteq T$, S finite and $\operatorname{diam}(A) \le 2q^{-M-1}$ so that $T \subseteq \bigcup_{s\in S} B(s;2q^{-M-1})$. Thus, it suffices to prove $\Theta_{\widetilde{\mathcal{A}},\widetilde{w}}(S) \le c\cdot\Theta_{\mathcal{A},w}(T)$ for suitable $\widetilde{\mathcal{A}} \in \mathcal{Z}_N^\infty(S, d_{|S\times S})$ and $\widetilde{w} \in \mathcal{W}(\widetilde{\mathcal{A}})$. We define

$$\widetilde{\mathcal{A}}_j = \{A\cap S : A \in \mathcal{A}_{j+1}\} \quad \text{for } N < j \le M$$

and

$$\widetilde{\mathcal{A}}_j = \{\{s\} : s \in S\} \quad \text{for } M < j < \infty$$

and claim that the $\widetilde{\mathcal{A}}_j$'s generate a sequence of partitions in $\mathcal{Z}_N^\infty(S)$. Indeed, if $j \le M$ and $A \in \mathcal{A}_{j+1}$, then $A\cap S \subseteq B(s;q^{-j})$ for a suitable $s \in S$. This easily follows from $A \cap S \subseteq B(t;q^{-j-1})$ for some $t \in T$. Set $\widetilde{w}_j(A\cap S) = w_{j+1}(A)$ for $N < j \le M$, $A \in \mathcal{A}_{j+1}$ and $\widetilde{w}_j(\{s\}) = w_{M+1}(A)$ for $j > M$ and $s = t_A \in S$. Thus,

$$\begin{aligned} &\Theta_{\widetilde{\mathcal{A}},\widetilde{w}}(S) \\ &\le \sup_{s\in S}\left(q\sum_{j=N+2}^{M+1} q^{-j}\sqrt{\log\frac{1}{w_j(A_j(s))}} + \frac{q^{-M-1}}{1-1/q}\sqrt{\log\frac{1}{w_{M+1}(A_{M+1}(s))}}\right) \\ &\le c\ \Theta_{\mathcal{A},w}(T) \end{aligned}$$

completing the proof. ■

Corollary 7.6.2 *For every finite S there exists some $M < \infty$ (depending on S) with*

$$\Theta_N^\infty(S) \le c\cdot\Theta_N^M(S)\ .$$

7.7 Gaussian Processes

As already mentioned, the quantities $\Theta(T)$ and $\mathcal{I}(T)$ originate in the theory of Gaussian processes. More precisely, if $T \subseteq H$, H separable Hilbert space, and

$$d(s,t) := \|t-s\|_H \ , \quad s,t \in T \ ,$$

then the famous theorem of Fernique and Talagrand [cf. Ledoux (1996) or Talagrand (1996)] asserts the following.

Theorem 7.7.1 *Let $(X_t)_{t\in H}$ be the isonormal process on H defined in* (7.1). *Then, we have*

$$c_1\ \Theta(T) \le \mathbf{E} \sup_{t\in T} X_t \le c_2\ \Theta(T) \tag{7.12}$$

with some universal $c_1, c_2 > 0$.

The expression $\mathbf{E}\ \sup_{t\in T} X_t$ in (7.12) should be understood throughout as

$$\sup \left\{ \mathbf{E} \sup_{s\in S} X_s : S \subseteq T \ , \ S \text{ finite} \right\} \ .$$

In view of (7.9), the estimate (7.12) remains true with $\mathcal{I}(T)$ instead of $\Theta(T)$, which was the original form of Theorem 7.7.1 in the language of majorizing measures [cf. Fernique (1975) and Talagrand (1987)]. Let us recall the important and well–known fact that Theorem 7.7.1 applies by standard arguments [cf. Ledoux and Talagrand (1991)] to any centered Gaussian process $Y = (Y_t)_{t\in T}$ provided the (pseudo)–metric d on T is generated by Y via $d(s,t) = (\mathbf{E}\ |Y_t - Y_s|^2)^{1/2}$.

The aim of this section is to compare $\Theta_N^M(T)$ with quantities generated by $(X_t)_{t\in T}$, similarly as in (7.12). Let us start with the case $N \ge N_0(T)$ and $M = \infty$. If $t \in T$, the modulus of continuity (w.r.t. $(X_t)_{t\in T}$) at t is defined as function of $\varepsilon > 0$ below:

$$\omega_T(t;\varepsilon) := \mathbf{E} \sup_{\substack{s\in T \\ d(s,t)\le\varepsilon}} X_s = \mathbf{E} \sup_{\substack{s\in T \\ d(s,t)\le\varepsilon}} (X_s - X_t) \ .$$

Similarly, the modulus of uniform continuity may be defined by

$$u_T(\varepsilon) := \mathbf{E} \sup_{\substack{s,t\in T \\ d(s,t)\le\varepsilon}} |X_t - X_s| = \mathbf{E} \sup_{\substack{s,t\in T \\ d(s,t)\le\varepsilon}} (X_t - X_s) \ . \tag{7.13}$$

Theorem 7.7.2 *For any $N \ge N_0(T)$,*

$$c_1\ \Theta_N^\infty(T) \ \le \ \sup_{t\in T} \omega_T(t;q^{-N}) + q^{-N}\sqrt{H(T,q^{-N})} \tag{7.14}$$

$$\le \ u_T(2q^{-N}) + q^{-N}\sqrt{H(T,q^{-N})} \le c_2\ \Theta_N^\infty(T) \ . \tag{7.15}$$

PROOF. The first inequality is a direct consequence of Theorem 7.6.1 and

$$\Theta(B(t;q^{-N})) \leq c\, \omega_T(t;q^{-N}),$$

which follows from Theorem 7.7.1. Of course, $\sup_{t\in T} \omega_T(t;\varepsilon) \leq u_T(\varepsilon) \leq u_T(2\varepsilon)$, hence it remains to prove the last estimate. To do so, we use the following result which is a direct consequence of (5.2.6) in Fernique (1997): If $\mu \in \mathcal{P}(T)$ and $\varepsilon > 0$, then

$$u_T(\varepsilon) = \mathbf{E} \sup_{\substack{s,t\in T\\ d(s,t)\leq\varepsilon}} |X_t - X_s| \leq c \sup_{t\in T} \int_0^\varepsilon \sqrt{\log \frac{1}{\mu(B(t;\delta))}} d\delta\ ,$$

i.e., we have $u_T(2q^{-N}) \leq c\ \mathcal{I}_N^\infty(T)$ where $\mathcal{I}_N^\infty(T)$ was defined in (7.8). Using Proposition 7.3.4 and Theorem 7.5.1, we have the proof. ■

Remark. Since the term $q^{-N}\sqrt{H(T,q^{-N})}$ in (7.14) and (7.15) is indeed necessary, $\Theta_N^\infty(T)$ is a combination of the local quantities $\sup_{t\in T} \omega_T(t;q^{-N})$ or $u_T(2q^{-N})$ and of $q^{-N}\sqrt{H(T,q^{-N})}$ measuring the global size of T.

Combining Theorem 7.7.2 with Corollary 7.6.1, the following holds.

Corollary 7.7.1 *Let $T \subset H$ be precompact. Then the following are equivalent:*

(i) $\lim_{N\to\infty} \Theta_N^\infty(T) = 0$,

(ii) $\lim_{\varepsilon\to 0} \sup_{t\in T} \omega_T(t;\varepsilon) = 0$,

(iii) $\lim_{\varepsilon\to 0} u_T(\varepsilon) = 0$ *and*

(iv) $(X_t)_{t\in T}$ *has an almost surely uniformly continuous version.*

Our next objective is to investigate the case $M < \infty$. A first natural question is about the relation between $\Theta_N^\infty(T)$ and $\sup_{M<\infty} \Theta_N^M(T)$. Of course, restricting any $\mathcal{A} \in \mathcal{Z}_N^\infty(T)$ and $w \in \mathcal{W}(\mathcal{A})$, we obtain partitions in $\mathcal{Z}_N^M(T)$ for any $M < \infty$ (and adapted weights), i.e. it always holds

$$\sup_{M<\infty} \Theta_N^M(T) \leq \Theta_N^\infty(T)\ . \tag{7.16}$$

But it is far from clear whether or not the converse of (7.16) is true as well. To verify this, one has to construct optimal $\mathcal{A}$'s in $\mathcal{Z}_N^\infty(T)$ out of optimal partitions in $\mathcal{Z}_N^M(T)$ for every $M < \infty$. This seems to be complicated, so we use a different approach.

Proposition 7.7.1 *There is a constant $c > 0$ such that*

$$\Theta_N^\infty(T) \le c \sup_{M<\infty} \Theta_N^M(T) .$$

Especially, we have $\Theta(T) < \infty$ iff $\sup_{M<\infty} \Theta_N^M(T) < \infty$ for one (each) N.

PROOF. Of course,

$$u_T(\varepsilon) = \sup \{u_S(\varepsilon) : S \subseteq T\,,\ \ S \text{ finite}\}\,;$$

hence by Theorem 7.7.2 it follows

$$\Theta_N^\infty(T) \le c \left(\sup\left\{u_S(2q^{-N}) : S \subseteq T\,,\ \ S \text{ finite}\right\} + q^{-N}\sqrt{H(T,q^{-N})}\right) .$$

Proposition 7.3.4 implies $\Theta_N^M(T) \ge q^{-N-1}\sqrt{H(T,q^{-N})}$ for any $M > N$. Thus, it remains to show that

$$\sup\left\{u_S(2q^{-N}) \ :\ S \subseteq T, S \text{ finite}\right\} \le c \sup_{M<\infty} \Theta_N^M(T) .$$

Applying Theorem 7.7.2 for S, we obtain $u_S(2q^{-N}) \le c\ \Theta_N^\infty(S)$. Since S is finite, by Corollary 7.6.2 we find an $M < \infty$ such that $\Theta_N^\infty(S) \le c\ \Theta_N^M(S)$. Now, Proposition 7.3.3 completes the proof. ■

Remark. The last proof depends heavily on the special choice of T as subset of a Hilbert space, i.e. on the fact that the metric d on T is generated by a scalar product. Using refined methods [cf. Bühler (1998)], one can show that Proposition 7.7.1 remains valid for general metric spaces (T, d). Note also that one big advantage of Proposition 7.7.1 is that the weights can depend on M and sometimes this make the construction much easier. An explicit example is given in the paper of Li and Linde (1998).

Let us formulate, for simplicity, the next result for $N = N_0(T)$ only. Here, $B(\varepsilon) = B(0;\varepsilon)$ denotes the closed ε–ball in H centered at zero.

Theorem 7.7.3 *If $M < \infty$, then*

$$c_1 \inf\left\{\mathbf{E}\sup_{s\in S} X_s : T \subseteq S + B(2q^{-M})\right\}$$
$$\le\ \Theta_{N_0}^M(T) \le c_2 \inf\left\{\mathbf{E}\sup_{r\in R} X_r : T \subseteq R + B(2q^{-M-1})\right\} ,$$

where the sets S and R may be chosen as finite or countable subsets of H or T, respectively.

PROOF. Applying Proposition 7.6.1, we obtain

$$\Theta_{N_0}^M(T) \le c \ \inf\left\{\Theta_{N_0}^\infty(R) \ : T \subseteq R + B(2q^{-M-1}),\ R \text{ countable}\right\}$$

and

$$\Theta_{N_0}^M(T) \ge c \ \inf\left\{\Theta_{N_0}^\infty(S) \ : T \subseteq S + B(2q^{-M}),\ S \subseteq T,\ S \text{ finite}\right\}.$$

By using Theorem 7.7.2 for R and $N_1 = N_0(R)$, we get

$$c_1 \ \Theta_{N_1}^\infty(R) \le \mathbf{E} \sup_{r\in R} X_r \le \mathbf{E} \sup_{r,t\in R} |X_r - X_t| \le c_2 \ \Theta_{N_1}^\infty(R),$$

analogous for S and $N_2 = N_0(S)$. Thus, it remains to show that

$$\Theta_{N_0}^\infty(R) \le c \ \Theta_{N_1}^\infty(R) \ \text{ and } \ \Theta_{N_0}^\infty(S) \ge c \ \Theta_{N_2}^\infty(S).$$

By definition of N_0, we find $t_1, t_2 \in T$ satisfying

$$T \subseteq B(t_1; q^{-N_0}) \ \text{ and } \ d(t_1, t_2) > q^{-N_0-1}.$$

We choose $r_1, r_2 \in R$ with $d(r_i, t_i) \le 2q^{-M-1}$ and obtain $d(r_1, r_2) > 2q^{-N_0-2}$, thus $N_1 < N_0 + 2$. By using Proposition 7.3.2, this yields $\Theta_{N_0}^\infty(R) \le c \ \Theta_{N_1}^\infty(R)$. Furthermore, we get $S \subseteq B(s, q^{-N_0} + 2q^{-M}) \subseteq B(s, q^{-N_0+1})$ for some $s \in S$ satisfying $d(s, t_1) \le 2q^{-M}$. Hence, $N_2 \ge N_0 - 1$ and by using Proposition 7.3.2 again this completes the proof. ■

7.8 Examples

We first treat the case

$$T = \{\alpha_j e_j : j = 1, 2, \ldots\} \cup \{0\} \subset H,$$

where $\alpha_1 > \alpha_2 > \cdots > 0$ tends to zero and $(e_j)_{j=1}^\infty$ is an ONB in H. Note that this set corresponds to the stochastic process $Y = (Y_n)_{n\ge1}$ with $Y_n = \alpha_n \xi_n$, where $\xi_1, \xi_2, \ldots$ are i.i.d. $\mathcal{N}(0,1)$. Instead of T, we may use $\mathbf{N} \cup \{\infty\}$ with

$$d(i,j) = (\alpha_i^2 + \alpha_j^2)^{1/2} \quad \text{for } i \ne j \quad \text{and } d(i,\infty) = \alpha_i. \tag{7.17}$$

Proposition 7.8.1 *If $T = \mathbf{N} \cup \{\infty\}$ is endowed with metric d defined by* (7.17), *then for $N < M \le \infty$ we have*

$$\Theta_N^M(T) \ \le \ c \left(\sup\left\{\alpha_j \sqrt{\log j + 1} : q^{-M-1} < \alpha_j \le q^{-N}\right\} + q^{-N} \sqrt{\log \operatorname{card}(\{j : \alpha_j > q^{-N}\}) + 1} \right).$$

PROOF. Let

$$\sigma(K) := \max\left\{j : \alpha_j > q^{-K}\right\}$$

and define $S \subseteq T$ by $S = \{1, \ldots, \sigma(M+1), \infty\}$. Then we have the inclusion $T \subseteq S + B(2q^{-M-1})$, hence Proposition 7.6.1 implies $\Theta_N^M(T) \leq c\, \Theta_N^\infty(S)$. Now using Theorem 7.7.2 for the metric space $(S, d_{|S\times S})$, we get

$$\Theta_N^\infty(S) \leq c \left(\sup_{j\in S} \omega_S(j; q^{-N}) + q^{-N}\sqrt{H(S, d_{|S\times S}, q^{-N})}\right) .$$

For $j \in S$ with $\alpha_j > q^{-N}$ (that means for $j = 1, \ldots, \sigma(N)$), the q^{-N}–ball around j consists of j only, thus $\omega_S(j; q^{-N}) = 0$ in this case. Furthermore, it holds $B(\infty, q^{-N}) \cap S = \{\sigma(N)+1, \ldots, \sigma(M+1), \infty\}$. Hence,

$$H(S, d_{|S\times S}, q^{-N}) = \log(\sigma(N) + 1)$$

and

$$\Theta_N^\infty(S) \leq c \left(\mathbf{E} \sup_{\sigma(N)+1\leq j\leq \sigma(M+1)} \alpha_j\, |\xi_j| + q^{-N}\sqrt{\log(\sigma(N)+1)}\right) . \tag{7.18}$$

By Theorem 9 in Linde and Pietsch (1974) and by the closed graph theorem, we have

$$\mathbf{E} \sup_{1\leq k\leq n} \beta_k\, |\xi_k| \leq c \sup_{1\leq k\leq n} \sqrt{\log(k+1)}\, \beta_k$$

for any $\beta_1 \geq \cdots \geq \beta_n \geq 0$. Applying this to (7.18), the proof is complete. ■

Remark. Observe that $H(T, q^{-k}) = \log(\sigma(k) + 1)$, hence Proposition 7.8.1 yields

$$\begin{aligned}\Theta_N^M(T) &\leq c \left(\sup\left\{\alpha_j\sqrt{\log(j+1)}\ : q^{-l-1} < \alpha_j \leq q^{-l},\ N \leq l \leq M\right\}\right. \\ &\qquad \left. + q^{-N}\sqrt{H(T, q^{-N})}\right) \\ &\leq c \sup_{N\leq l\leq M+1} q^{-l}\sqrt{H(T, q^{-l})}.\end{aligned}$$

In view of Proposition 7.3.4, this tells us that the estimate in Proposition 7.8.1 is nearly optimal.

Our next example treats the set

$$T = \{(\varepsilon_1\alpha_1, \varepsilon_2\alpha_2, \ldots) : \varepsilon_j = \pm 1\} \tag{7.19}$$

for a non–increasing square summable sequence $(\alpha_j)_{j=1}^\infty$ of positive real numbers. In different words, T consists of the corners of an infinite dimensional block in l_2. A corresponding Gaussian process is $Y = (Y_e)_{e\in E}$ with $E = \{-1,1\}^\infty$ and

$$Y_e = \sum_{j=1}^{\infty} \alpha_j \varepsilon_j \xi_j \quad \text{for } e = (\varepsilon_j)_{j=1}^{\infty} \in E\ .$$

Lemma 7.8.1 *Let T be defined by* (7.19) *with α_j's satisfying the regularity condition*

$$\alpha_j \le \gamma\ \alpha_{2j} \tag{7.20}$$

for some $\gamma \ge 1$. Then we have

$$\int_0^\varepsilon \sqrt{H(T,\delta)}\, d\delta \le c \sum_{k=\sigma(\varepsilon)}^{\infty} \alpha_k,$$

where

$$\sigma(\varepsilon) = \text{card}\left\{ n : \sum_{k=n}^{\infty} \alpha_k^2 > \varepsilon^2 \right\}\ . \tag{7.21}$$

Proof. If we define the strictly decreasing sequence $(\beta_n)_{n=1}^{\infty}$ by

$$\beta_n := \sum_{k=n}^{\infty} \alpha_k^2\ ,$$

it follows $\beta_{\sigma(\varepsilon)+1} \le \varepsilon^2$ as well as $N(T,2\varepsilon) \le 2^{\sigma(\varepsilon)}$, hence $H(T,2\varepsilon) \le c\ \sigma(\varepsilon)$. This implies

$$\begin{aligned}
\int_0^\varepsilon \sqrt{H(T,\delta)}\, d\delta &\le 2 \sum_{k=\sigma(\varepsilon)}^{\infty} \int_{\sqrt{\beta_{k+1}}}^{\sqrt{\beta_k}} \sqrt{H(T,2\delta)}\, d\delta \\
&\le 2 \sum_{k=\sigma(\varepsilon)}^{\infty} \left(\sqrt{\beta_k} - \sqrt{\beta_{k+1}}\right) \sqrt{H(T,2\sqrt{\beta_{k+1}})} \\
&\le c \sum_{k=\sigma(\varepsilon)}^{\infty} \sqrt{k}\left(\sqrt{\beta_k} - \sqrt{\beta_{k+1}}\right) \\
&\le c \sum_{k=\sigma(\varepsilon)}^{\infty} \frac{\alpha_k^2}{\left(\frac{1}{k}\sum_{j=k+1}^{\infty} \alpha_j^2\right)^{1/2}} \le c \sum_{k=\sigma(\varepsilon)}^{\infty} \alpha_k
\end{aligned}$$

in view of (7.20) and completes the proof. ∎

Lemma 7.8.2 *Let $T \subset l_2$ be defined by* (7.19) *with α_j's as before (we do not suppose* (7.20) *here). Then for this set T, its modulus of uniform continuity u_T (cf.* (7.13) *for the definition) satisfies*

$$u_T(2\varepsilon) \ge c \sum_{k=\sigma(\varepsilon)+1}^{\infty} \alpha_k$$

with $\sigma(\varepsilon)$ given by (7.21).

PROOF. By the definition

$$u_T(2\varepsilon) = \mathbf{E} \sup_{\substack{t,s\in T \\ d(t,s)\le 2\varepsilon}} |X_t - X_s| = \mathbf{E} \sup_{v\in A_\varepsilon} |X_v| ,$$

where

$$A_\varepsilon := \left\{ (\delta_1\alpha_1, \delta_2\alpha_2, \ldots) : \delta_j \in \{-1,0,1\} \, , \sum_{\delta_j\neq 0} \alpha_j^2 \le \varepsilon^2 \right\} .$$

If we set

$$B_\varepsilon := \left\{ (\delta_1\alpha_1, \delta_2\alpha_2, \ldots) : \delta_1 = \cdots = \delta_{\sigma(\varepsilon)} = 0 \, , \; \delta_j = \pm 1 \, , \, j > \sigma(\varepsilon) \right\} ,$$

then the choice of $\sigma(\varepsilon)$ implies $B_\varepsilon \subseteq A_\varepsilon$. Hence

$$u_T(2\varepsilon) \ge \mathbf{E} \sup_{v\in B_\varepsilon} |X_v| = \sqrt{\frac{2}{\pi}} \sum_{k=\sigma(\varepsilon)+1}^{\infty} \alpha_k$$

which completes the proof. ■

Proposition 7.8.2 *Let T be as in* (7.19) *with positive non–increasing α_j's satisfying* (7.20). *If*

$$\sigma_N := \sigma(q^{-N}) = \operatorname{card}\left\{ n : \sum_{k=n}^{\infty} \alpha_k^2 > q^{-2N} \right\} ,$$

then

$$c_1 \sum_{k=\sigma_N+1}^{\infty} \alpha_k \le \Theta_N^\infty(T) \le c_2 \sum_{k=\sigma_N+1}^{\infty} \alpha_k .$$

PROOF. This follows directly by combining Lemma 7.8.1 and 7.8.2 with Theorem 7.7.2 and Proposition 7.3.4. ■

Acknowledgement. This research was supported in part by the NSF.

References

1. Bühler, B. (1998). Finite majorizing measures, *Ph.D. Thesis*, University of Jena.

2. Carl, B., Kyrezi, I. and Pajor, A. (1997). Metric entropy of convex hulls in Banach spaces, Preprint.

3. Dudley, R. M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes, *Journal of Functional Analysis*, **1**, 290–330.

4. Fernique, X. (1975). Régularité des trajectoires des fonctions aléatoires gaussiennes, École Été de Probabilités de St.–Flour 1974, *Lecture Notes in Mathematics*, **480**, pp. 1–96, New York: Springer-Verlag.

5. Fernique, X. (1997). *Fonctions aléatoires gaussiennes vecteurs aléatoires gaussiens*, Montreal: Les Publications CRM.

6. Ledoux, M. (1996). Isoperimetry and Gaussian Analysis. Lectures on Probability Theory and Statistics, *Lecture Notes in Mathematics*, **1648**, pp. 165–294, New York: Springer-Verlag.

7. Ledoux, M. and Talagrand, M. (1991). *Probability in a Banach Space*, New York: Springer-Verlag.

8. Li, W. V. and Linde, W. (1998). Metric entropy of convex hulls in Hilbert spaces, Preprint.

9. Linde, W. and Pietsch, A. (1974). Mappings of Gaussian measures in Banach spaces, *Theory of Probability and Its Applications*, **19**, 445–460.

10. Sudakov, V. N. (1969). Gaussian measures, Cauchy measures and ε–entropy, *Soviet Math. Dokl.*, **10**, 310–313.

11. Talagrand, M. (1987). Regularity of Gaussian processes, *Acta Mathematica Sinica*, **159**, 99–149.

12. Talagrand, M. (1992). A simple proof of the majorizing measure theorem, *Geometric and Functional Analysis*, **2**, 119–125.

13. Talagrand, M. (1996). Majorizing measures: The generic chaining, *Annals of Probability*, **24**, 1049–1103.

8

Multidimensional Hungarian Construction for Vectors With Almost Gaussian Smooth Distributions

F. Götze and A. Yu. Zaitsev

University of Bielefeld, Bielefeld, Germany
Steklov Mathematical Institute, St. Petersburg, Russia

Abstract: A multidimensional version of the results of Komlós, Major, and Tusnády for sums of independent random vectors with finite exponential moments is obtained in the particular case when the summands have smooth distributions which are close to Gaussian ones. The bounds obtained reflect this closeness. Furthermore, the results provide sufficient conditions for the existence of i.i.d. vectors $X_1, X_2, \ldots$ with given distributions and corresponding i.i.d. Gaussian vectors $Y_1, Y_2, \ldots$ such that, for given small ε,

$$\mathbf{P}\Big\{\limsup_{n\to\infty} \frac{1}{\log n}\Big|\sum_{j=1}^{n} X_j - \sum_{j=1}^{n} Y_j\Big| \le \varepsilon\Big\} = 1.$$

Keywords and phrases: Multidimensional invariance principle, strong approximation, sums of independent random vectors, Central Limit Theorem

8.1 Introduction

This chapter is devoted to an improvement of a multidimensional version of strong approximation results of Komlós, Major and Tusnády (KMT) for sums of independent random vectors with finite exponential moments and with smooth distributions which are close to Gaussian ones.

Let $\mathcal{F}_d$ be the set of all d-dimensional probability distributions defined on the σ-algebra $\mathcal{B}_d$ of Borel subsets of $\mathbf{R}^d$. By $\widehat{F}(t)$, $t \in \mathbf{R}^d$, we denote the characteristic function of a distribution $F \in \mathcal{F}_d$. The product of measures is understood as their convolution, i.e., $FG = F * G$. The distribution and the

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

corresponding covariance operator of a random vector ξ will be denoted by $\mathcal{L}(\xi)$ and $\operatorname{cov}\xi$ (or $\operatorname{cov} F$, if $F = \mathcal{L}(\xi)$). The symbol $\mathbf{I}_d$ will be used for the identity operator in $\mathbf{R}^d$. For $b > 0$, we denote $\log^* b = \max\{1, \log b\}$. Writing $z \in \mathbf{R}^d$ (resp. $\mathbf{C}^d$), we shall use the representation $z = (z_1, \dots, z_d) = z_1 e_1 + \dots + z_d e_d$, where $z_j \in \mathbf{R}^1$ (resp. $\mathbf{C}^1$) and the e_j, are the standard orthonormal vectors. The scalar product is denoted by $\langle x, y\rangle = x_1\overline{y}_1 + \dots + x_d\overline{y}_d$. We shall use the Euclidean norm $\|z\| = \langle z, z\rangle^{1/2}$ and the maximum norm $|z| = \max\limits_{1\le j\le k} |z_j|$. The symbols $c, c_1, c_2, \dots$ will be used for absolute positive constants. The letter c may denote different constants when we do not need to fix their numerical values.

Let us consider the definition and some useful properties of classes of distributions $\mathcal{A}_d(\tau) \subset \mathcal{F}_d$, $\tau \ge 0$, introduced in Zaitsev (1986); also see Zaitsev (1995, 1996, 1998a). The class $\mathcal{A}_d(\tau)$ (with a fixed $\tau \ge 0$) consists of distributions $F \in \mathcal{F}_d$ for which the function

$$\varphi(z) = \varphi(F, z) = \log \int_{\mathbf{R}^d} e^{\langle z, x\rangle} F\{dx\} \qquad (\varphi(0) = 0)$$

is defined and analytic for $\|z\|\,\tau < 1$, $z \in \mathbf{C}^d$, and

$$\left|d_u d_v^2\, \varphi(z)\right| \le \|u\|\,\tau\,\langle \mathbf{D}\, v, v\rangle \qquad \text{for all } u, v \in \mathbf{R}^d \text{ and } \|z\|\,\tau < 1,$$

where $\mathbf{D} = \operatorname{cov} F$, and the derivative $d_u\varphi$ is given by

$$d_u\varphi(z) = \lim_{\beta\to 0} \frac{\varphi(z + \beta u) - \varphi(z)}{\beta}\,.$$

It is easy to see that $\tau_1 < \tau_2$ implies $\mathcal{A}_d(\tau_1) \subset \mathcal{A}_d(\tau_2)$. Moreover, the class $\mathcal{A}_d(\tau)$ is closed with respect to convolution: if $F_1, F_2 \in \mathcal{A}_d(\tau)$, then $F_1 F_2 \in \mathcal{A}_d(\tau)$. The class $\mathcal{A}_d(0)$ coincides with the class of all Gaussian distributions in $\mathbf{R}^d$. The following inequality can be considered as an estimate of the stability of this characterization: if $F \in \mathcal{A}_d(\tau)$, $\tau > 0$, then

$$\pi\big(F, \Phi(F)\big) \le c d^2 \tau \,\log^*(\tau^{-1}), \tag{8.1}$$

where $\pi(\cdot\,,\cdot)$ is the Prokhorov distance and $\Phi(F)$ denotes the Gaussian distribution whose mean and covariance operator are the same as those of F. Moreover, for all $X \in \mathcal{B}_d$ and all $\lambda > 0$, we have

$$F\{X\} \;\le\; \Phi(F)\{X^\lambda\} + c d^2 \exp\Big(-\frac{\lambda}{c d^2 \tau}\Big), \tag{8.2}$$

$$\Phi(F)\{X\} \;\le\; F\{X^\lambda\} + c d^2 \exp\Big(-\frac{\lambda}{c d^2 \tau}\Big), \tag{8.3}$$

where $X^\lambda = \big\{y \in \mathbf{R}^d : \inf\limits_{x\in X} \|x - y\| < \lambda\big\}$ is the λ-neighborhood of the set X; see Zaitsev (1986).

The classes $\mathcal{A}_d(\tau)$ are closely connected with other natural classes of multidimensional distributions. In particular, by the definition of $\mathcal{A}_d(\tau)$, any distribution $\mathcal{L}(\xi)$ from $\mathcal{A}_d(\tau)$ has finite exponential moments $\mathbf{E}\, e^{\langle h,\xi\rangle}$, for $\|h\|\,\tau < 1$. This leads to exponential estimates for the tails of distributions; see, for example, Lemma 8.3.3 below. On the other hand, if $\mathbf{E}\, e^{\langle h,\xi\rangle} < \infty$, for $h \in A \subset \mathbf{R}^d$, where A is a neighborhood of zero, then $F = \mathcal{L}(\xi) \in \mathcal{A}_d(\tau(F))$ with some $\tau(F)$ depending on F only.

Throughout we assume that $\tau \geq 0$ and $\xi_1, \xi_2, \ldots$ are random vectors with given distributions $\mathcal{L}(\xi_k) \in \mathcal{A}_d(\tau)$ such that $\mathbf{E}\,\xi_k = 0$, $\operatorname{cov}\xi_k = \mathbf{I}_d$, $k = 1, 2, \ldots$. The problem is to construct, for a given n, $1 \leq n \leq \infty$, on a probability space a sequence of independent random vectors $X_1, \ldots, X_n$ and a sequence of i.i.d. Gaussian random vectors $Y_1, \ldots, Y_n$ with $\mathcal{L}(X_k) = \mathcal{L}(\xi_k)$, $\mathbf{E}\, Y_k = 0$, $\operatorname{cov} Y_k = \mathbf{I}_d$, $k = 1, \ldots, n$, such that, with large probability,

$$\Delta(n) = \max_{1\leq r\leq n} \Big| \sum_{k=1}^{r} X_k - \sum_{k=1}^{r} Y_k \Big|$$

is as small as possible.

The aim of this Chapter is to provide sufficient conditions for the following Assertion A.

Assertion A *There exist absolute positive constants c_1, c_2 and c_3 such that, for $\tau\, d^{3/2} \leq c_1$, there exists a construction with*

$$\mathbf{E}\, \exp\Big(\frac{c_2\, \Delta(n)}{d^{3/2}\,\tau} \Big) \leq \exp\big(c_3 \log^* d\, \log^* n\big). \tag{8.4}$$

Using the exponential Chebyshev inequality, we see that (8.4) implies

$$\mathbf{P}\left\{ c_2\, \Delta(n) \geq \tau\, d^{3/2} (c_3 \log^* d\, \log^* n + x) \right\} \leq e^{-x}, \qquad x \geq 0. \tag{8.5}$$

Therefore, Assertion A can be considered as a generalization of the classical result of Komlós, Major and Tusnády (1975, 1976). Assertion A provides a supplement to an improvement of a multidimensional KMT-type result of Einmahl (1989) presented by Zaitsev (1995, 1998a) which differs from Assertion A by the restriction $\tau \geq 1$ and by another explicit power-type dependence of the constants on the dimension d. In a particular case, when $d = 1$ and all summands have a common variance, the result of Zaitsev is equivalent to the main result of Sakhanenko (1984), who extended the KMT construction to the case of non-identically distributed summands and stated the dependence of constants on the distributions of the summands belonging to a subclass of $\mathcal{A}_1(\tau)$. The main difference between Assertion A and the aforementioned results consists in the fact that in Assertion A we consider "small" τ, $0 \leq \tau \leq c_1\, d^{-3/2}$. In previous results, the constants are separated from zero by quantities which are larger than some absolute constants. In Komlós, Major and Tusnády (1975, 1976), the dependence of the constants on the distributions is not specified. From the conditions (1) and (4) in Sakhanenko (1984, Section 1), it follows that $\operatorname{Var}\xi_k \le \lambda^{-2}$ (λ^{-1} plays in Sakhanenko's paper the role of τ) and, if $\operatorname{Var}\xi_k = 1$, then $\lambda^{-1} \ge 1$. This corresponds to the restrictions $\alpha^{-1} \ge 2$ in Einmahl [1989, conditions (3.6) and (4.3)] and $\tau \ge 1$ in Zaitsev (1995, 1998a, Theorem 1).

Note that in Assertion A we do not require that the distributions $\mathcal{L}(\xi_k)$ are identical, but we assume that they have the same covariance operators; see Einmahl (1989) and Zaitsev (1995, 1998a). A generalization of the results of Zaitsev (1995, 1998a) and of this chapter to the case of non-identical covariance operators appeared recently in the preprint Zaitsev (1998b).

According to (8.1)–(8.3), the condition $\mathcal{L}(\xi_k) \in \mathcal{A}_d(\tau)$ with small τ means that $\mathcal{L}(\xi_k)$ are close to the corresponding Gaussian laws. It is easy to see that Assertion A becomes stronger for small τ (see as well Theorem 8.1.4 below). Passing to the limit as $\tau \to 0$, we obtain a spectrum of statements with the trivial limiting case: if $\tau = 0$ (and, hence, $\mathcal{L}(\xi_k)$ are Gaussian), we can take $X_k = Y_k$ and $\Delta(n) = 0$.

We show that *Assertion* A *is valid under some additional smoothness-type restrictions on* $\mathcal{L}(\xi_k)$. The question about the necessity of these conditions remains open. The case $\tau \ge 1$ considered by Zaitsev (1995, 1998a, Theorem 1) does not need conditions of such kind. The formulation of our main result—Theorem 8.2.1—includes some additional notation. In order to show that the conditions of Theorem 8.2.1 can be verified in some concrete simple situations, we consider at first three particular applications—Theorems 8.1.1, 8.1.2 and 8.1.3.

Theorem 8.1.1 *Assume that the distributions* $\mathcal{L}(\xi_k) \in \mathcal{A}_d(\tau)$ *can be represented in the form*

$$\mathcal{L}(\xi_k) = H_k G, \qquad k = 1, \ldots, n,$$

where G *is a Gaussian distribution with covariance operator* $\operatorname{cov} G = b^2\,\mathbf{I}_d$ *and* $b^2 \ge 2^{10}\,\tau^2 d^3 \log^* \frac{1}{\tau}$. *Then, Assertion* A *is valid.*

The following example deals with a non-convolution family of distributions approximating a Gaussian distribution for small τ.

Theorem 8.1.2 *Let* η *be a random vector with an absolutely continuous distribution and density*

$$p_\tau(x) = \frac{(4 + \tau^2\,\|x\|^2\,)\,\exp\,(\,-\,\|x\|^2/2)}{(2\pi)^{d/2}\,(4 + \tau^2 d)}, \qquad x \in \mathbf{R}^d. \tag{8.6}$$

Assume that $\mathcal{L}(\xi_k) = \mathcal{L}(\eta/\gamma)$, $k = 1, \ldots, n$, *where*

$$\gamma^2 = \frac{(4 + \tau^2\,(d+2))}{(4 + \tau^2 d)}, \qquad \gamma > 0. \tag{8.7}$$

Then, Assertion A *is valid.*

The proof of Theorem 8.1.2 can be apparently extended to the distributions with some more general densities of type $P(\tau^2 \|x\|^2)\ \exp(-c\ \|x\|^2)$, where $P(\cdot)$ is a suitable polynomial.

Theorem 8.1.3 *Assume that a random vector ζ satisfies the relations*

$$\mathbf{E}\,\zeta = 0, \qquad \mathbf{P}\{\ \|\zeta\| \le b_1\} = 1, \qquad H := \mathcal{L}(\zeta) \in \mathcal{A}_d(b_2) \tag{8.8}$$

and admits a differentiable density $p(\cdot)$ such that

$$\sup_{x \in \mathbf{R}^d} |d_u\, p(x)| \le b_3\ \|u\|, \qquad \text{for all} \quad u \in \mathbf{R}^d, \tag{8.9}$$

with some positive b_1, b_2 and b_3. Let $\zeta_1, \zeta_2, \ldots$ be independent copies of ζ. Write

$$\tau = b_2 m^{-1/2}, \tag{8.10}$$

where m is a positive integer. Assume that the distributions $\mathcal{L}(\xi_k)$ can be represented in the form

$$\mathcal{L}(\xi_k) = L^{(k)} P, \qquad k = 1, \ldots, n, \tag{8.11}$$

where

$$L^{(k)} \in \mathcal{A}_d(\tau) \quad \text{and} \quad P = \mathcal{L}((\zeta_1 + \ldots + \zeta_m)/\sqrt{m}). \tag{8.12}$$

Then there exist a positive b_4 depending on H only and such that $m \ge b_4$ implies Assertion A.

Remark 8.1.1 If all the distributions $L^{(k)}$ are concentrated at zero, then the statement of Theorem 8.1.3 (for $\tau = bm^{-1/2}$ with some $b = b(H)$) can be derived from the main results of Komlós, Major and Tusnády (1975, 1976) (for $d = 1$) and of Zaitsev (1995, 1998a) (for $d \ge 1$).

A consequence of Assertion A is given in Theorem 8.1.4 below.

Theorem 8.1.4 *Assume that ξ, $\xi_1, \xi_2, \ldots$ are i.i.d. with a common distribution $\mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$. Let Assertion* A *be satisfied for $\xi_1, \ldots, \xi_n$ for all n with some c_1, c_2 and c_3 independent of n. Suppose that $\tau\, d^{3/2} \le c_1$. Then, there exists a construction such that*

$$\mathbf{P}\Big\{ \limsup_{n\to\infty} \frac{1}{\log n} \Big| \sum_{j=1}^n X_j - \sum_{j=1}^n Y_j \Big| \le c_4\, \tau\, d^{3/2} \log^* d \Big\} = 1 \tag{8.13}$$

with some constant $c_4 = c_4(c_2, c_3)$.

From a result of Bártfai (1966), it follows that the rate $O(\log n)$ in (8.13) is the best possible if $\mathcal{L}(\xi)$ is non-Gaussian. In the case of distributions with finite exponential moments, this rate was established by Zaitsev (1995, 1998a, Corollary 1). Theorems 8.1.1–8.1.3 and 8.2.1 provide examples of smooth distributions which are close to Gaussian ones and for which the constants corresponding to this rate are arbitrarily small. The existence of such examples has been already mentioned in the one-dimensional case; for example, see Major (1978, p. 498).

This Chapter is organized as follows. In Section 8.2 we formulate Theorem 8.2.1. To this end, we define at first a class of distributions $\overline{\mathcal{A}}_d(\tau,\rho)$ used in Theorem 8.2.1. The definition of this class is given in terms of smoothness conditions on the so-called conjugate distributions. Then we describe a multidimensional version of the KMT dyadic scheme, cf. Einmahl (1989). We prove Theorem 8.2.1 in Section 8.3. Section 8.4 is devoted to the proofs of Theorems 8.1.1–8.1.4.

A preliminary version of this work appeared as a preprint of Götze and Zaitsev (1997).

8.2 The Main Result

Let $F = \mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$, $\|h\|\,\tau < 1$, $h \in \mathbf{R}^d$. The conjugate distribution $\overline{F} = \overline{F}(h)$ is defined by

$$\overline{F}\{dx\} = (\mathbf{E}\, e^{\langle h,\xi\rangle})^{-1} e^{\langle h,x\rangle} F\{dx\}. \tag{8.14}$$

Sometimes, we shall write $F_h = \overline{F}(h)$. It is clear that $\overline{F}(0) = F$. Denote by $\overline{\xi}(h)$ a random vector with $\mathcal{L}(\overline{\xi}(h)) = \overline{F}(h)$. From (8.14), it follows that

$$\mathbf{E}\, f(\overline{\xi}(h)) = (\mathbf{E}\, e^{\langle h,\xi\rangle})^{-1}\, \mathbf{E}\, f(\xi)\, e^{\langle h,\xi\rangle} \tag{8.15}$$

provided that $\mathbf{E}\,|\,f(\xi)\, e^{\langle h,\xi\rangle}\,| < \infty$. It is easy to see that

$$\text{if} \quad U_1, U_2 \in \mathcal{A}_d(\tau), \quad U = U_1 U_2, \quad \text{then} \quad \overline{U}(h) = \overline{U}_1(h)\, \overline{U}_2(h). \tag{8.16}$$

Below we shall also use the following subclasses of $\mathcal{A}_d(\tau)$ containing distributions satisfying some special smoothness-type restrictions. Let $\tau \geq 0$, $\delta > 0$, $\rho > 0$, $h \in \mathbf{R}^d$. Consider the conditions

$$\int\limits_{\rho\|t\|\tau d \geq 1} |\widehat{F}_h(t)|\, dt \leq \frac{(2\pi)^{d/2}\, \tau\, d^{3/2}}{\sigma\, (\det \mathbf{D})^{1/2}}, \tag{8.17}$$

$$\int\limits_{\rho\|t\|\tau d \geq 1} |\widehat{F}_h(t)|\, dt \leq \frac{(2\pi)^{d/2}\, \tau^2\, d^{2}}{\sigma^2\, (\det \mathbf{D})^{1/2}}, \tag{8.18}$$

$$\int\limits_{\rho\|t\|\tau d\geq 1} |\langle t, v\rangle \widehat{F}_h(t)|\,dt \leq \frac{(2\pi)^{d/2}\langle \mathbf{D}^{-1}v, v\rangle^{1/2}}{\delta\,(\det \mathbf{D})^{1/2}}, \qquad \text{for all} \quad v \in \mathbf{R}^d, \tag{8.19}$$

where $F_h = \overline{F}(h)$ and $\sigma^2 = \sigma^2(F) > 0$ is the minimal eigenvalue of $\mathbf{D} = \operatorname{cov} F$. Denote by $\overline{\mathcal{A}}_d(\tau, \rho)$ (resp. $\mathcal{A}_d^*(\tau, \delta, \rho)$) the class of distributions $F \in \mathcal{A}_d(\tau)$ such that the condition (8.17) [resp. (8.18) and (8.19)] is satisfied for $h \in \mathbf{R}^d$, $\|h\|\,\tau < 1$. It is easy to see that

$$\mathcal{A}_d^*(\tau, \delta, \rho) \subset \overline{\mathcal{A}}_d(\tau, \rho) \qquad \text{if} \quad \frac{\tau\, d^{1/2}}{\sigma} \leq 1. \tag{8.20}$$

In the present work, the class $\overline{\mathcal{A}}_d(\tau, \rho)$ plays the role of the class $\mathcal{A}_d^*(\tau, \delta, \rho)$ which was used by Zaitsev (1995, 1998a); see also Sakhanenko [1984, inequality (49), p. 9] or Einmahl [1989, inequality (1.5)]. Note that (8.15) implies

$$\widehat{F}_h(t) = \mathbf{E}\, e^{\langle it, \overline{\xi}(h)\rangle} = \left(\mathbf{E}\, e^{\langle h, \xi\rangle}\right)^{-1} \mathbf{E}\, e^{\langle h+it, \xi\rangle}. \tag{8.21}$$

The dyadic scheme. Let N be a positive integer and $\{\xi_1, \dots, \xi_{2^N}\}$ be a collection of d-dimensional independent random vectors. Denote

$$\widetilde{S}_0 = 0; \qquad \widetilde{S}_k = \sum_{l=1}^{k} \xi_l, \quad 1 \leq k \leq 2^N; \tag{8.22}$$

$$U_{m,k}^* = \widetilde{S}_{(k+1)\cdot 2^m} - \widetilde{S}_{k\cdot 2^m}, \qquad 0 \leq k < 2^{N-m}, \quad 0 \leq m \leq N. \tag{8.23}$$

In particular, $U_{0,k}^* = \xi_{k+1}$, $U_{N,0}^* = \widetilde{S}_{2^N} = \xi_1 + \ldots + \xi_{2^N}$. In the sequel, we shall use the term *block of summands* for a collection of summands with indices of the form $k\cdot 2^m + 1, \dots, (k+1)\cdot 2^m$, where $0 \leq k < 2^{N-m}$, $0 \leq m \leq N$. Thus, $U_{m,k}^*$ is the sum over a block containing 2^m summands. Put

$$\widetilde{U}_{n,k}^* = U_{n-1,2k}^* - U_{n-1,2k+1}^*, \qquad 0 \leq k < 2^{N-n}, \quad 1 \leq n \leq N. \tag{8.24}$$

Note that

$$U_{n-1,2k}^* + U_{n-1,2k+1}^* = U_{n,k}^*, \qquad 0 \leq k < 2^{N-n}, \quad 1 \leq n \leq N. \tag{8.25}$$

Introduce the vectors

$$\widetilde{\mathbf{U}}_{n,k}^* = \left(U_{n-1,2k}^*, U_{n-1,2k+1}^*\right) \in \mathbf{R}^{2d}, \qquad 0 \leq k < 2^{N-n}, \quad 1 \leq n \leq N, \tag{8.26}$$

with the first d coordinates coinciding with those of the vectors $U_{n-1,2k}^*$ and with the last d coordinates coinciding with those of the vectors $U_{n-1,2k+1}^*$. Similarly, denote

$$\mathbf{U}_{n,k}^* = \left(U_{n,k}^*, \widetilde{U}_{n,k}^*\right) \in \mathbf{R}^{2d}, \qquad 0 \leq k < 2^{N-n}, \quad 1 \leq n \leq N. \tag{8.27}$$

Introduce now the projectors $\mathbf{P}_i : \mathbf{R}^s \to \mathbf{R}^1$ and $\overline{\mathbf{P}}_j : \mathbf{R}^s \to \mathbf{R}^j$, for $i, j = 1, \ldots, s$, by the relations $\mathbf{P}_i x = x_i$, $\overline{\mathbf{P}}_j x = (x_1, \ldots, x_j)$, where $x = (x_1, \ldots, x_s) \in \mathbf{R}^s$ (we shall use this notation for $s = d$ or $s = 2d$).

It is easy to see that, according to (8.24)–(8.27),

$$\mathbf{U}^*_{n,k} = \mathbf{A}\,\widetilde{\mathbf{U}}^*_{n,k} \in \mathbf{R}^{2d}, \qquad 0 \le k < 2^{N-n}, \quad 1 \le n \le N, \tag{8.28}$$

where $\mathbf{A} : \mathbf{R}^{2d} \to \mathbf{R}^{2d}$ is a linear operator defined, for $x = (x_1, \ldots, x_{2d}) \in \mathbf{R}^{2d}$, as follows:

$$\begin{aligned} \mathbf{P}_j\,\mathbf{A}\,x &= x_j + x_{d+j}, && j = 1, \ldots, d, \\ \mathbf{P}_j\,\mathbf{A}\,x &= x_j - x_{d+j}, && j = d+1, \ldots, 2d. \end{aligned} \tag{8.29}$$

Denote

$$\left.\begin{aligned} \mathbf{U}^{*(j)}_{n,k} &= \mathbf{P}_j\,\mathbf{U}^*_{n,k}, \\ \mathbf{U}^{*j}_{n,k} &= \big(\mathbf{U}^{*(1)}_{n,k}, \ldots, \mathbf{U}^{*(j)}_{n,k}\big) = \overline{\mathbf{P}}_j\,\mathbf{U}^*_{n,k} \in \mathbf{R}^j, \end{aligned}\right. \qquad j = 1, \ldots, 2d. \tag{8.30}$$

Now we can formulate the main result of the Chapter.

Theorem 8.2.1 *Let the conditions described in (8.22)–(8.30) be satisfied,* $\tau \ge 0$ *and* $\mathbf{E}\,\xi_k = 0$, $\operatorname{cov}\xi_k = \mathbf{I}_d$, $k = 1, \ldots, 2^N$. *Assume that*

$$\mathcal{L}(\mathbf{U}^{*j}_{n,k}) \in \overline{\mathcal{A}}_j(\tau, 4) \quad \text{for} \quad 0 \le k < 2^{N-n}, \quad 1 \le n \le N, \ d \le j \le 2d, \tag{8.31}$$

and

$$\mathcal{L}(\mathbf{U}^{*j}_{N,0}) \in \overline{\mathcal{A}}_j(\tau, 4) \qquad \text{for} \quad 1 \le j \le 2d. \tag{8.32}$$

Then there exist absolute positive constants c_5, c_6 *and* c_7 *such that, for* $\tau\, d^{3/2} \le c_5$, *one can construct on a probability space sequences of independent random vectors* $X_1, \ldots, X_{2^N}$ *and i.i.d. Gaussian random vectors* $Y_1, \ldots, Y_{2^N}$ *so that*

$$\mathcal{L}(X_k) = \mathcal{L}(\xi_k), \quad \mathbf{E}\,Y_k = 0, \quad \operatorname{cov} Y_k = \mathbf{I}_d, \qquad k = 1, \ldots, 2^N, \tag{8.33}$$

and

$$\mathbf{E}\,\exp\Big(\frac{c_6\,\Delta(2^N)}{d^{3/2}\,\tau}\Big) \le \exp\big(c_7 N\,\log^* d\big), \tag{8.34}$$

where $\Delta(2^N) = \max\limits_{1 \le r \le 2^N} \Big|\sum\limits_{k=1}^{r} X_k - \sum\limits_{k=1}^{r} Y_k\Big|$.

Theorem 8.2.1 says that the conditions (8.31) and (8.32) suffice for Assertion A. However, these conditions require that the number of summands is 2^N.

For an arbitrary number of summands, one should consider additional (for simplicity, Gaussian) summands in order to apply Theorem 8.2.1.

Below, we shall prove Theorem 8.2.1. Suppose that its conditions are satisfied.

At first, we describe a procedure of constructing the random vectors $\{U_{n,k}\}$ with $\mathcal{L}(\{U_{n,k}\}) = \mathcal{L}(\{U^*_{n,k}\})$, provided that the vectors $Y_1, \ldots, Y_{2^N}$ are already constructed (then we shall define $X_k = U_{0,k-1}$, $k = 1, \ldots, 2^N$). This procedure is an extension of the Komlós, Major and Tusnády (1975, 1976) dyadic scheme to the multivariate case due to Einmahl (1989). For this purpose, we shall use the so-called Rosenblatt quantile transformation [see Rosenblatt (1952) and Einmahl (1989)].

Denote by $F^{(1)}_{N,0}(x_1) = \mathbf{P}\left\{\mathbf{P}_1 U^*_{N,0} < x_1\right\}$, $x_1 \in \mathbf{R}^1$, the distribution function of the first coordinate of the vector $U^*_{N,0}$. Introduce the conditional distributions, denoting by $F^{(j)}_{N,0}(\,\cdot\,|x_1, \ldots, x_{j-1})$, $2 \le j \le d$, the regular conditional distribution function (r.c.d.f.) of $\mathbf{P}_j U^*_{N,0}$, given $\overline{\mathbf{P}}_{j-1} U^*_{N,0} = (x_1, \ldots, x_{j-1})$. Denote by $\widetilde{F}^{(j)}_{n,k}(\,\cdot\,|x_1, \ldots, x_{j-1})$ the r.c.d.f. of $\mathbf{P}_j \mathbf{U}^*_{n,k}$, given $\overline{\mathbf{P}}_{j-1} \mathbf{U}^*_{n,k} = (x_1, \ldots, x_{j-1})$, for $0 \le k < 2^{N-n}$, $1 \le n \le N$, $d+1 \le j \le 2d$. Put

$$T_k = \sum_{l=1}^{k} Y_l, \qquad 1 \le k \le 2^N; \tag{8.35}$$

$$V_{m,k} = \left(V^{(1)}_{m,k}, \ldots, V^{(d)}_{m,k}\right) = T_{(k+1)\cdot 2^m} - T_{k\cdot 2^m}, \qquad 0 \le k < 2^{N-m},\ 0 \le m \le N; \tag{8.36}$$

$$\widetilde{\mathbf{V}}_{n,k} = \left(V_{n-1,2k},\, V_{n-1,2k+1}\right) = \left(\widetilde{\mathbf{V}}^{(1)}_{n,k}, \ldots, \widetilde{\mathbf{V}}^{(2d)}_{n,k}\right) \in \mathbf{R}^{2d}, \qquad 0 \le k < 2^{N-n},\ 1 \le n \le N; \tag{8.37}$$

and

$$\mathbf{V}_{n,k} = \left(\mathbf{V}^{(1)}_{n,k}, \ldots, \mathbf{V}^{(2d)}_{n,k}\right) = \mathbf{A}\,\widetilde{\mathbf{V}}_{n,k} \in \mathbf{R}^{2d}, \qquad 0 \le k < 2^{N-n},\ 1 \le n \le N. \tag{8.38}$$

Note that, according to the definition of the operator $\mathbf{A}$, we have [see (8.24)–(8.29) and (8.35)–(8.38)]

$$\mathbf{V}_{n,k} = \left(V_{n,k},\, \widetilde{V}_{n,k}\right) \in \mathbf{R}^{2d}, \qquad 0 \le k < 2^{N-n},\ 1 \le n \le N, \tag{8.39}$$

where

$$\begin{aligned} V_{n,k} &= V_{n-1,2k} + V_{n-1,2k+1}, \\ \widetilde{V}_{n,k} &= V_{n-1,2k} - V_{n-1,2k+1}, \end{aligned} \qquad 0 \le k < 2^{N-n},\ 1 \le n \le N, \tag{8.40}$$

and

$$V_{N,0} = Y_1 + \ldots + Y_{2^N}. \tag{8.41}$$

Thus, the vectors $V_{m,k}$, $\widetilde{\mathbf{V}}_{n,k}$ and $\mathbf{V}_{n,k}$ can be constructed from the vectors Y_1, $\ldots$, Y_{2^N} by the same linear procedure which was used for constructing the vectors $U^*_{m,k}$, $\widetilde{\mathbf{U}}^*_{n,k}$ and $\mathbf{U}^*_{n,k}$ from the vectors $\xi_1, \ldots, \xi_{2^N}$.

It is obvious that, for fixed n and k,

$$\operatorname{cov} \mathbf{U}^*_{n,k} = \operatorname{cov} \mathbf{V}_{n,k} = 2^n \, \mathbf{I}_{2d} \tag{8.42}$$

and, hence, the coordinates of the Gaussian vector $\mathbf{V}_{n,k}$ are independent with the same distribution function $\Phi_{2^{n/2}}(\cdot)$; here and below,

$$\Phi_\sigma(x) = \int\limits_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\Big(-\frac{y^2}{2\sigma^2}\Big)\, dy, \qquad x \in \mathbf{R}^1, \quad \sigma > 0,$$

is the distribution function of the normal law with mean zero and variance σ^2.

Denote now the new collection of random vectors X_k as follows. At first, we define

$$\begin{aligned} U^{(1)}_{N,0} &= (F^{(1)}_{N,0})^{-1}(\Phi_{2^{N/2}}(V^{(1)}_{N,0})) \qquad \text{and, for } 2 \le j \le d, \\ U^{(j)}_{N,0} &= (F^{(j)}_{N,0})^{-1}(\Phi_{2^{N/2}}(V^{(j)}_{N,0}) \,|\, U^{(1)}_{N,0}, \ldots, U^{(j-1)}_{N,0}) \end{aligned} \tag{8.43}$$

(here $(F^{(1)}_{N,0})^{-1}(t) = \sup\,\{x : F^{(1)}_{N,0}(x) \le t\}$, $0 < t < 1$, and so on). Taking into account that the distributions of the random vectors $\xi_1, \ldots, \xi_{2^N}$ are absolutely continuous, we see that this formula can be rewritten in a more natural form [see Sakhanenko (1984, p. 30–31)] as

$$\begin{aligned} F^{(1)}_{N,0}(U^{(1)}_{N,0}) &= \Phi_{2^{N/2}}(V^{(1)}_{N,0}), \\ F^{(j)}_{N,0}(U^{(j)}_{N,0} \,|\, U^{(1)}_{N,0}, \ldots, U^{(j-1)}_{N,0}) &= \Phi_{2^{N/2}}(V^{(j)}_{N,0}), \qquad \text{for} \quad 2 \le j \le d. \end{aligned} \tag{8.44}$$

Suppose that the random vectors

$$U_{n,k} = (U^{(1)}_{n,k}, \ldots, U^{(d)}_{n,k}), \qquad 0 \le k < 2^{N-n}, \tag{8.45}$$

corresponding to blocks containing each 2^n summands with fixed n, $1 \le n \le N$, are already constructed. Now our aim is to construct the blocks containing each 2^{n-1} summands. To this end, we define

$$\mathbf{U}^{(j)}_{n,k} = \mathbf{P}_j \, U_{n,k} = U^{(j)}_{n,k}, \quad 1 \le j \le d, \tag{8.46}$$

and, for $d+1 \le j \le 2d$,

$$\mathbf{U}_{n,k}^{(j)} = (\widetilde{F}_{n,k}^{(j)})^{-1}(\Phi_{2^{n/2}}(\mathbf{V}_{n,k}^{(j)})\,|\,\mathbf{U}_{n,k}^{(1)},\dots,\mathbf{U}_{n,k}^{(j-1)}). \tag{8.47}$$

It is clear that (8.47) can be rewritten in a form similar to (8.44). Then, we put

$$\begin{aligned} \mathbf{U}_{n,k} &= (\mathbf{U}_{n,k}^{(1)},\dots,\mathbf{U}_{n,k}^{(2d)}) \in \mathbf{R}^{2d}, \\ \mathbf{U}_{n,k}^{j} &= (\mathbf{U}_{n,k}^{(1)},\dots,\mathbf{U}_{n,k}^{(j)}) = \overline{\mathbf{P}}_j\,\mathbf{U}_{n,k} \in \mathbf{R}^{j}, \qquad j=1,\dots,2d, \\ \widetilde{U}_{n,k}^{(j)} &= \mathbf{U}_{n,k}^{(j+d)}, \qquad j=1,\dots,d, \\ \widetilde{U}_{n,k} &= (\widetilde{U}_{n,k}^{(1)},\dots,\widetilde{U}_{n,k}^{(d)}) \in \mathbf{R}^{d} \end{aligned} \tag{8.48}$$

and

$$\begin{aligned} U_{n-1,2k} &= (U_{n,k}+\widetilde{U}_{n,k})/2, \\ U_{n-1,2k+1} &= (U_{n,k}-\widetilde{U}_{n,k})/2. \end{aligned} \tag{8.49}$$

Thus, we have constructed the random vectors $U_{n-1,k}$, $0 \le k < 2^{N-n+1}$. After N steps, we obtain the random vectors $U_{0,k}$, $0 \le k < 2^N$. Now we set

$$X_k = U_{0,k-1}, \qquad S_0 = 0, \quad S_k = \sum_{l=1}^{k} X_l, \qquad 1 \le k \le 2^N. \tag{8.50}$$

Lemma 8.2.1 [Einmahl (1989)] *The joint distribution of the constructed vectors $U_{n,k}$ and $\mathbf{U}_{n,k}$ coincides with that of the vectors $U_{n,k}^*$ and $\mathbf{U}_{n,k}^*$. In particular, X_k, $k=1,\dots,2^N$, are independent and $\mathcal{L}(X_k)=\mathcal{L}(\xi_k)$.*

Moreover, according to (8.24) and (8.25), we have

$$\begin{aligned} \widetilde{U}_{n,k} &= U_{n-1,2k} - U_{n-1,2k+1}, \\ U_{n,k} &= U_{n-1,2k} + U_{n-1,2k+1} = S_{(k+1)\cdot 2^n} - S_{k\cdot 2^n}, \end{aligned} \tag{8.51}$$

for $0 \le k < 2^{N-n}$, $1 \le n \le N$ [it is clear that (8.51) follows from (8.49)]. Furthermore, putting

$$\widetilde{\mathbf{U}}_{n,k} = (U_{n-1,2k},\, U_{n-1,2k+1}) \in \mathbf{R}^{2d}, \qquad 0 \le k < 2^{N-n},\ 1 \le n \le N, \tag{8.52}$$

we have [see (8.26) and (8.28)]

$$\mathbf{U}_{n,k} = \mathbf{A}\,\widetilde{\mathbf{U}}_{n,k} \in \mathbf{R}^{2d}, \qquad 0 \le k < 2^{N-n},\ 1 \le n \le N. \tag{8.53}$$

Note that it is not difficult to verify that, according to (8.29),

$$\|\mathbf{A}\| = \frac{1}{\|\mathbf{A}^{-1}\|} = \|\mathbf{A}^*\| = \frac{1}{\|(\mathbf{A}^*)^{-1}\|} = \sqrt{2}, \tag{8.54}$$

where the asterisk is used to denote the adjoint operator $\mathbf{A}^*$ for the operator $\mathbf{A}$.

Remark 8.2.1 The conditions of Theorem 8.2.1 imply the coincidence of the corresponding first and second moments of the vectors $\mathbf{U} = \{U_{n,k}, \widetilde{\mathbf{U}}_{n,k}, \mathbf{U}_{n,k}\}$ and $\mathbf{V} = \{V_{n,k}, \widetilde{\mathbf{V}}_{n,k}, \mathbf{V}_{n,k}\}$ since the vectors $\mathbf{U}$ can be restored from vectors $X_1, \ldots, X_{2^N}$ by the same linear procedure which is used for reconstruction of the vectors $\mathbf{V}$ from $Y_1, \ldots, Y_{2^N}$. In particular, $\mathbf{E}\,\mathbf{U} = \mathbf{E}\,\mathbf{V} = 0$.

Lemma 8.2.2 [Einmahl (1989) Lemma 5, p. 55] *Let* $1 \le m = (2s+1)\cdot 2^r \le 2^N$, *where* s, r *are non-negative integers. Then,*

$$S_m = \frac{m}{2^N} S_{2^N} + \sum_{n=r+1}^{N} \gamma_n \, \widetilde{U}_{n,l_{n,m}}, \tag{8.55}$$

where $\gamma_n = \gamma_n(m) \in [0, 1/2]$ *and the integers* $l_{n,m}$ *are defined by*

$$l_{n,m} \cdot 2^n < m \le (l_{n,m} + 1) \cdot 2^n. \tag{8.56}$$

The shortest proof of Lemma 8.2.2 can be obtained with the help of a geometrical approach due to Massart (1989, p. 275).

Remark 8.2.2 The inequalities (8.56) give a formal definition of $l_{n,m}$. To understand better the mechanism of the dyadic scheme, one should remember another characterization of these numbers: $U_{n,l_{n,m}}$ is the sum over the block of 2^n summands which contains X_m, the last summand in the sum S_m.

Corollary 8.2.1 *Under the conditions of Lemma* 8.2.2,

$$|\,S_m - T_m\,| \le |\,U_{N,0} - V_{N,0}\,| + \frac{1}{2} \sum_{n=r+1}^{N} |\,\widetilde{U}_{n,l_{n,m}} - \widetilde{V}_{n,l_{n,m}}\,|, \;\; m = 1, \ldots, 2^N.$$

This statement evidently follows from Lemmas 8.2.1 and 8.2.2 and from the relations (8.22)–(8.25), (8.35) and (8.36).

8.3 Proof of Theorem 8.2.1

In the proof of Theorem 8.2.1, we shall use the following auxiliary Lemmas 8.3.1–8.3.4 [Zaitsev (1995, 1996, 1998a)].

Lemma 8.3.1 *Suppose that* $\mathcal{L}(\xi) \in \mathcal{A}_d(\tau)$, $y \in \mathbf{R}^m$, $\alpha \in \mathbf{R}^1$. *Let* $\mathbf{M} : \mathbf{R}^d \to \mathbf{R}^m$ *be a linear operator and* $\widetilde{\xi} \in \mathbf{R}^k$ *be the vector consisting of a subset of coordinates of the vector* ξ. *Then,*

$$\mathcal{L}(\mathbf{M}\,\xi + y) \in \mathcal{A}_m(\,\|\mathbf{M}\|\,\tau), \qquad \textit{where} \quad \|\mathbf{M}\| = \sup_{\|x\| \le 1} \|\mathbf{M}x\|,$$

$$\mathcal{L}(\alpha\xi) \in \mathcal{A}_d(|\alpha|\,\tau), \qquad \mathcal{L}(\widetilde{\xi}) \in \mathcal{A}_k(\tau).$$

Lemma 8.3.2 *Suppose that independent random vectors $\xi^{(k)}$, $k = 1, 2$, satisfy the condition $\mathcal{L}(\xi^{(k)}) \in \mathcal{A}_{d_k}(\tau)$. Let $\xi = (\xi^{(1)}, \xi^{(2)}) \in \mathbf{R}^{d_1+d_2}$ be the vector with the first d_1 coordinates coinciding with those of $\xi^{(1)}$ and with the last d_2 coordinates coinciding with those of $\xi^{(2)}$. Then, $\mathcal{L}(\xi) \in \mathcal{A}_{d_1+d_2}(\tau)$.*

Lemma 8.3.3 (Bernstein-type inequality) *Suppose that $\mathcal{L}(\xi) \in \mathcal{A}_1(\tau)$, $\mathbf{E}\,\xi = 0$ and $\mathbf{E}\,\xi^2 = \sigma^2$. Then,*

$$\mathbf{P}\left\{|\xi| \geq x\right\} \leq 2\,\max\left\{\exp\left(-x^2/4\sigma^2\right), \exp\left(-x/4\tau\right)\right\}, \qquad x \geq 0.$$

Lemma 8.3.4 *Let the distribution of a random vector $\xi \in \mathbf{R}^d$ with $\mathbf{E}\,\xi = 0$ satisfy the condition $\mathcal{L}(\xi) \in \overline{\mathcal{A}}_d(\tau, 4)$, $\tau \geq 0$. Assume that the variance $\sigma^2 = \mathbf{E}\,\xi_d^2 > 0$ of the last coordinate ξ_d of the vector ξ is the minimal eigenvalue of $\operatorname{cov}\xi$. Then, there exist absolute positive constants $c_8, \ldots, c_{12}$ such that the following assertions hold:*

a) *Let $d \geq 2$. Assume that ξ_d is not correlated with previous coordinates $\xi_1, \ldots, \xi_{d-1}$ of the vector ξ. Define $\mathbf{B} = \operatorname{cov}\overline{\mathbf{P}}_{d-1}\xi$ and denote by $F(z\,|\,x)$, $z \in \mathbf{R}^1$, the r.c.d.f. of ξ_d for a given value of $\overline{\mathbf{P}}_{d-1}\xi = x \in \mathbf{R}^{d-1}$. Let $\mathcal{L}(\overline{\mathbf{P}}_{d-1}\xi) \in \overline{\mathcal{A}}_{d-1}(\tau, 4)$. Then there exists $y \in \mathbf{R}^1$ such that*

$$|y| \leq c_8\,\tau\,\|\mathbf{B}^{-1/2}x\,\|^2 \leq c_8\,\tau\,\frac{\|x\|^2}{\sigma^2} \tag{8.57}$$

and

$$\Phi_\sigma(z - \gamma(z)) < F(z + y\,|\,x) < \Phi_\sigma(z + \gamma(z)) \tag{8.58}$$

for $\frac{\tau\,d^{3/2}}{\sigma} \leq c_9$, $|\mathbf{B}^{-1/2}x| \leq \frac{c_{10}\,\sigma}{d^{3/2}\tau}$, $|z| \leq \frac{c_{11}\,\sigma^2}{d\tau}$, *where*

$$\gamma(z) = c_{12}\,\tau\left(d^{3/2} + d\,\delta\left(1 + \frac{|z|}{\sigma}\right) + \frac{z^2}{\sigma^2}\right), \qquad \delta = \|\mathbf{B}^{-1/2}x\,\|. \tag{8.59}$$

b) *The assertion* a) *remains valid for $d = 1$ with $F(z\,|\,x) = \mathbf{P}\left\{\xi_1 < z\right\}$ and $y = \delta = 0$ without any restrictions on $\mathbf{B}$, $\overline{\mathbf{P}}_{d-1}\xi$ and x.*

Remark 8.3.1 In Zaitsev (1995, 1996), the formulation of Lemma 8.3.4 is in some sense weaker; see Zaitsev (1995, 1996, Lemmas 6.1 and 8.1). In particular, instead of the conditions

$$\mathcal{L}(\xi) \in \overline{\mathcal{A}}_d(\tau, 4) \qquad \text{and} \qquad \mathcal{L}(\overline{\mathbf{P}}_{d-1}\xi) \in \overline{\mathcal{A}}_{d-1}(\tau, 4), \tag{8.60}$$

the stronger conditions

$$\mathcal{L}(\xi) \in \mathcal{A}_d^*(\tau, 4, 4) \qquad \text{and} \qquad \mathcal{L}(\overline{\mathbf{P}}_{d-1}\xi) \in \mathcal{A}_{d-1}^*(\tau, 4, 4) \tag{8.61}$$

are used. However, in the proof of (8.57) and (8.58) only the conditions (8.60) are applied. The conditions (8.61) are necessary for the investigation of quantiles of conditional distributions corresponding to random vectors having coinciding moments up to third order which has been done in Zaitsev (1995, 1996) simultaneously with the proof of (8.57) and (8.58).

Lemma 8.3.5 *Let $S_k = X_1 + \ldots + X_k$, $k = 1, \ldots, n$, be sums of independent random vectors $X_j \in \mathbf{R}^d$ and let $q(\cdot)$ be a semi-norm in $\mathbf{R}^d$. Then,*

$$\mathbf{P}\{\max_{1\le k\le n} q(S_k) > 3t\} \le 3 \max_{1\le k\le n} \mathbf{P}\{q(S_k) > t\}, \qquad t \ge 0. \tag{8.62}$$

Lemma 8.3.5 is a version of the Ottaviani inequality; see Dudley (1989, p. 251) or Hoffmann-Jørgensen (1994, p. 472). In the form (8.62), this inequality can be found in Etemadi (1985) with 4 instead of 3 (twice). The proof of Lemma 8.3.5 repeats those from the references above and is therefore omitted.

Lemma 8.3.6 *Let the conditions of Theorem* 8.2.1 *be satisfied and assume that the vectors X_k, $k = 1, \ldots, 2^N$, are constructed by the dyadic procedure described in* (8.35)–(8.50). *Then there exist absolute positive constants $c_{13}, \ldots, c_{17}$ such that:*

a) *If $\tau d^{3/2}/2^{N/2} \le c_9$, then*

$$|U_{N,0} - V_{N,0}| \le c_{13} d^{3/2} \tau \left(1 + 2^{-N} |U_{N,0}|^2\right) \tag{8.63}$$

provided that $|U_{N,0}| \le \frac{c_{14}\cdot 2^N}{d^{3/2}\tau}$;

b) *If $1 \le n \le N$, $0 \le k < 2^{N-n}$, $\tau d^{3/2}/2^{n/2} \le c_{15}$, then*

$$|\widetilde{U}_{n,k} - \widetilde{V}_{n,k}| \le c_{16} d^{3/2} \tau \left(1 + 2^{-n} |\mathbf{U}_{n,k}|^2\right) \tag{8.64}$$

provided that $|\mathbf{U}_{n,k}| \le \frac{c_{17}\cdot 2^n}{d^{3/2}\tau}$.

In the proof of Lemma 8.3.6, we need the following auxiliary Lemma 8.3.7 which is useful for the application of Lemma 8.3.4 to the conditional distributions involved in the dyadic scheme.

Lemma 8.3.7 *Let $F(\cdot)$ denote a continuous distribution function and $G(\cdot)$ an arbitrary distribution function satisfying for $z \in B \in \mathcal{B}_1$ the inequality*

$$G(z - f(z)) < F(z + w) < G(z + f(z))$$

with some $f : B \to \mathbf{R}^1$ and $w \in \mathbf{R}^1$. Let $\eta \in \mathbf{R}^1$, $0 < G(\eta) < 1$ and $\xi = F^{-1}(G(\eta))$, where $F^{-1}(x) = \sup\{u : F(u) \le x\}$, $0 < x < 1$. Then,

$$|\xi - \eta| < f(\xi - w) + |w| \qquad \text{if} \quad \xi - w \in B.$$

Proof. Put $\zeta = \xi - w$. The continuity of F implies that $F(F^{-1}(x)) \equiv x$ for $0 < x < 1$. Therefore,

$$\zeta \in B \Rightarrow G(\zeta - f(\zeta)) < F(\xi) = G(\eta) \Rightarrow \zeta - f(\zeta) < \eta \Rightarrow \xi - \eta < f(\zeta) + w$$

and

$$\zeta \in B \Rightarrow G(\eta) = F(\xi) < G(\zeta + f(\zeta)) \Rightarrow \eta < \zeta + f(\zeta) \Rightarrow \eta - \xi < f(\zeta) - w.$$

This completes the proof of the lemma. ■

PROOF OF LEMMA 8.3.6. At first, we note that the conditions of Theorem 8.2.1 imply that

$$\operatorname{cov} \mathbf{U}_{n,k} = 2^n \mathbf{I}_{2d}, \qquad \text{for} \quad 1 \le n \le N, \quad 0 \le k < 2^{N-n},$$

and, hence [see (8.42)],

$$\operatorname{cov} \mathbf{U}_{n,k}^j = 2^n \mathbf{I}_j, \qquad \text{for} \quad 1 \le j \le 2d. \tag{8.65}$$

Let us prove the assertion a). Introduce the vectors

$$U_{N,0}^j = (U_{N,0}^{(1)}, \ldots, U_{N,0}^{(j)}), \qquad V_{N,0}^j = (V_{N,0}^{(1)}, \ldots, V_{N,0}^{(j)}) \tag{8.66}$$

consisting of the first j coordinates of the vectors $U_{N,0}$, $V_{N,0}$, respectively. By (8.65), (8.46) and (8.48),

$$U_{N,0} = \overline{\mathbf{P}}_d \, \mathbf{U}_{N,0} \tag{8.67}$$

and

$$U_{N,0}^j = \mathbf{U}_{N,0}^j, \qquad \operatorname{cov} U_{N,0}^j = 2^n \mathbf{I}_j, \qquad \text{for} \quad 1 \le j \le d. \tag{8.68}$$

Moreover, according to Lemma 8.2.1, Remark 8.2.1, (8.68) and (8.32), the distributions $\mathcal{L}(U_{N,0}^j)$, $j = 1, \ldots, d$, satisfy in the j-dimensional case the conditions of Lemma 8.3.4 with $\sigma^2 = 2^N$ and $\mathbf{B} = \operatorname{cov} U_{N,0}^{j-1} = 2^N \mathbf{I}_{j-1}$ (the last equality for $j \ge 2$).

Taking into account (8.43) and applying Lemmas 8.3.4 and 8.3.7, we obtain that

$$|U_{N,0}^{(1)} - V_{N,0}^{(1)}| \le c_{12} \tau \left(1 + \frac{|U_{N,0}^{(1)}|^2}{2^N}\right) \tag{8.69}$$

if $\frac{\tau}{2^{N/2}} \le c_9$, $|U_{N,0}^{(1)}| \le \frac{c_{11} \cdot 2^N}{\tau}$. Furthermore,

$$\begin{aligned} |U_{N,0}^{(j)} - V_{N,0}^{(j)}| \;\; &\le \;\; c_{12} \tau \left(j^{3/2} + j^{3/2} \frac{|U_{N,0}^{j-1}|}{2^{N/2}} \left(1 + \frac{|U_{N,0}^{(j)} - y_j|}{2^{N/2}}\right) \right. \\ &\qquad \left. + \frac{|U_{N,0}^{(j)} - y_j|^2}{2^N} \right) + |y_j| \end{aligned} \tag{8.70}$$

if

$$\frac{\tau j^{3/2}}{2^{N/2}} \le c_9, \qquad \frac{|U_{N,0}^{j-1}|}{2^{N/2}} \le \frac{c_{10}\cdot 2^{N/2}}{j^{3/2}\tau},$$

$$|U_{N,0}^{(j)} - y_j| \le \frac{c_{11}\cdot 2^N}{j\tau}, \qquad 2 \le j \le d, \tag{8.71}$$

where

$$|y_j| \le c_8 \tau j \,\frac{|U_{N,0}^{j-1}|^2}{2^N}, \qquad 2 \le j \le d. \tag{8.72}$$

Obviously,

$$|U_{N,0}^{(1)}| \le \max\{|U_{N,0}^{j-1}|, |U_{N,0}^{(j)}|\} = |U_{N,0}^{j}| \le |U_{N,0}|, \qquad 2 \le j \le d \tag{8.73}$$

see (8.45) and (8.66). Using (8.69), (8.70), (8.72) and (8.73), we see that one can choose c_{13} to be so large and c_{14} to be so small that

$$|U_{N,0}^{(j)} - V_{N,0}^{(j)}| \le c_{13} d^{3/2}\tau\,(1 + 2^{-N}|U_{N,0}|^2) \tag{8.74}$$

if $\frac{\tau d^{3/2}}{2^{N/2}} \le c_9$, $|U_{N,0}| \le \frac{c_{14}\cdot 2^N}{d^{3/2}\tau}$, $1 \le j \le d$. The inequality (8.63) immediately follows from (8.74), (8.36) and (8.45).

Now we shall prove item b). According to Lemma 8.2.1, Remark 8.2.1, (8.31), (8.45) and (8.65), the distributions $\mathcal{L}(\mathbf{U}_{n,k}^{j})$, $j = d+1, \ldots, 2d$, satisfy in the j-dimensional case the conditions of Lemma 8.3.4 with $\sigma^2 = 2^n$, $\mathbf{B} = \operatorname{cov}\mathbf{U}_{n,k}^{j-1} = 2^n\mathbf{I}_{j-1}$.

Using (8.47) and applying Lemmas 8.3.4 and 8.3.7, we obtain that

$$\begin{aligned}|\mathbf{U}_{n,k}^{(j)} - \mathbf{V}_{n,k}^{(j)}| \;\le\; & c_{12}\tau\Big(j^{3/2} + j^{3/2}\,\frac{|\mathbf{U}_{n,k}^{j-1}|}{2^{n/2}}\Big(1 + \frac{|\mathbf{U}_{n,k}^{(j)} - y_j|}{2^{n/2}}\Big) \\ & + \frac{|\mathbf{U}_{n,k}^{(j)} - y_j|^2}{2^n}\Big) + |y_j| \end{aligned} \tag{8.75}$$

if

$$\frac{\tau j^{3/2}}{2^{n/2}} \le c_9, \qquad \frac{|\mathbf{U}_{n,k}^{j-1}|}{2^{n/2}} \le \frac{c_{10}\cdot 2^{n/2}}{j^{3/2}\tau}, \qquad |\mathbf{U}_{n,k}^{(j)} - y_j| \le \frac{c_{11}\cdot 2^n}{j\tau}, \tag{8.76}$$

where

$$|y_j| \le c_8\tau j\,\frac{|\mathbf{U}_{n,k}^{j-1}|^2}{2^n}, \qquad d+1 \le j \le 2d. \tag{8.77}$$

Obviously,

$$\max\{|\mathbf{U}_{n,k}^{j-1}|, |\mathbf{U}_{n,k}^{(j)}|\} = |\mathbf{U}_{n,k}^{j}| \le |\mathbf{U}_{n,k}| \tag{8.78}$$

see (8.48). Using (8.75), (8.77) and (8.78), we see that one can choose c_{15} and c_{17} to be so small and c_{16} to be so large that

$$|\mathbf{U}_{n,k}^{(j)} - \mathbf{V}_{n,k}^{(j)}| \le c_{16}\, d^{3/2}\tau\,(1 + 2^{-n}|\mathbf{U}_{n,k}|^2) \tag{8.79}$$

if $\frac{\tau d^{3/2}}{2^{n/2}} \le c_{15}$, $|\mathbf{U}_{n,k}| \le \frac{c_{17}\cdot 2^n}{d^{3/2}\tau}$, $d+1 \le j \le 2d$. The inequality (8.64) immediately follows from (8.79), (8.38), (8.39) and (8.48). ■

PROOF OF THEOREM 8.2.1. Let X_k, $k = 1,\dots,2^N$, denote the vectors constructed by the dyadic procedure described in (8.35)–(8.50). Denote

$$\Delta = \Delta(2^N) = \max_{1\le k\le 2^N} |S_k - T_k|, \tag{8.80}$$

$$c_5 = \min\{c_9,\, c_{15}\}, \quad c_{18} = \min\{c_{14},\, c_{17},\, 1\}, \quad y := \frac{c_{18}}{d^{3/2}\tau} \le \frac{1}{\tau}, \tag{8.81}$$

fix some $x > 0$ and choose the integer M such that

$$x < 4y\cdot 2^M \le 2x. \tag{8.82}$$

We shall estimate $\mathbf{P}\{\Delta \ge x\}$. Consider separately two possible cases: $M \ge N$ and $M < N$. Let, at first, $M \ge N$. Denote

$$\Delta_1 = \max_{1\le k\le 2^N} |S_k|, \qquad \Delta_2 = \max_{1\le k\le 2^N} |T_k|. \tag{8.83}$$

It is easy to see that $\Delta \le \Delta_1 + \Delta_2$ and, hence,

$$\mathbf{P}\{\Delta \ge x\} \le \mathbf{P}\{\Delta_1 \ge x/2\} + \mathbf{P}\{\Delta_2 \ge x/2\}. \tag{8.84}$$

Taking into account the completeness of classes $\mathcal{A}_d(\tau)$ with respect to convolution, applying Lemmas 8.3.5, 8.3.1 and 8.3.3 and using (8.81) and (8.82), we obtain that $2^N \le 2^M \le x/2y$ and

$$\begin{aligned}\mathbf{P}\{\Delta_1 \ge x/2\} &\le 3 \max_{1\le k\le 2^N} \mathbf{P}\{|S_k| \ge x/6\}\\ &\le 6d\,\exp\Big(-\min\Big\{\frac{x^2}{144\cdot 2^N},\, \frac{x}{24\tau}\Big\}\Big)\\ &\le 6d\,\exp\Big(-\frac{c_{19}\,x}{d^{3/2}\tau}\Big).\end{aligned} \tag{8.85}$$

Since all d-dimensional Gaussian distributions belong to all classes $\mathcal{A}_d(\tau)$, $\tau \ge 0$, we automatically obtain that

$$\mathbf{P}\{\Delta_2 \ge x/2\} \le 6d\,\exp\Big(-\frac{c_{19}\,x}{d^{3/2}\tau}\Big). \tag{8.86}$$

From (8.84)–(8.86), it follows in the case $M \geq N$ that

$$\mathbf{P}\{\Delta \geq x\} \leq 12d\exp\Big(-\frac{c_{19}\,x}{d^{3/2}\tau}\Big). \tag{8.87}$$

Let now $M < N$. Denote

$$L = \max\{0,\, M\} \tag{8.88}$$

and

$$\Delta_3 = \max_{0\leq k<2^{N-L}}\ \max_{1\leq l\leq 2^L} |S_{k\cdot 2^L+l} - S_{k\cdot 2^L}|, \tag{8.89}$$

$$\Delta_4 = \max_{0\leq k<2^{N-L}}\ \max_{1\leq l\leq 2^L} |T_{k\cdot 2^L+l} - T_{k\cdot 2^L}|, \tag{8.90}$$

$$\Delta_5 = \max_{1\leq k\leq 2^{N-L}} |S_{k\cdot 2^L} - T_{k\cdot 2^L}|. \tag{8.91}$$

Introduce the event

$$A = \{\omega : |U_{L,k}| < y\cdot 2^L,\ 0\leq k < 2^{N-L}\} \tag{8.92}$$

(we assume that all considered random vectors are measurable mappings of $\omega \in \Omega$). For the complementary event we use the notation $\overline{A} = \Omega \setminus A$.

We consider separately two possible cases: $L = M$ and $L = 0$. Let $L = M$. It is evident that in this case

$$\Delta \leq \Delta_3 + \Delta_4 + \Delta_5. \tag{8.93}$$

Moreover, by virtue of (8.93), (8.82), (8.89) and (8.92), we have

$$\overline{A} \subset \{\omega : \Delta_3 \geq x/4\}. \tag{8.94}$$

From (8.93) and (8.94), it follows that

$$\mathbf{P}\{\Delta \geq x\} \leq \mathbf{P}\{\Delta_3 \geq x/4\} + \mathbf{P}\{\Delta_4 \geq x/4\} + \mathbf{P}\{\Delta_5 \geq x/2,\, A\}. \tag{8.95}$$

Using Lemmas 8.3.5, 8.3.1 and 8.3.3, the completeness of classes $\mathcal{A}_d(\tau)$ with respect to convolution and the relations (8.81) and (8.82), we obtain, for $0 \leq k < 2^{N-L}$, that $2^L = 2^M \leq x/2y$ and

$$\begin{aligned}
&\mathbf{P}\{\max_{1\leq l\leq 2^L} |S_{k\cdot 2^L+l} - S_{k\cdot 2^L}| \geq x/4\}\\
&\leq 3\max_{1\leq l\leq 2^L} \mathbf{P}\{|S_{k\cdot 2^L+l} - S_{k\cdot 2^L}| \geq x/12\}\\
&\leq 6d\exp\Big(-\min\Big\{\frac{x^2}{576\cdot 2^L},\ \frac{x}{48\tau}\Big\}\Big)\\
&\leq 6d\exp\Big(-\frac{c_{20}\,x}{d^{3/2}\tau}\Big).
\end{aligned} \tag{8.96}$$

Since all d-dimensional Gaussian distributions belong to classes $\mathcal{A}_d(\tau)$ for all $\tau \geq 0$, we immediately obtain that

$$\mathbf{P}\Big\{ \max_{1\leq l\leq 2^L} |\, T_{k\cdot 2^L+l} - T_{k\cdot 2^L}\,| \geq x/4\Big\} \leq 6d \exp\Big(- \frac{c_{20}\, x}{d^{3/2}\tau}\Big). \tag{8.97}$$

From (8.89), (8.90), (8.96) and (8.97), it follows that

$$\mathbf{P}\big\{\Delta_3 \geq x/4\big\} + \mathbf{P}\big\{\Delta_4 \geq x/4\big\} \leq 2^N \cdot 12d \exp\Big(- \frac{c_{20}\, x}{d^{3/2}\tau}\Big). \tag{8.98}$$

Assume that $L = 0$. Then, according to (8.80) and (8.91), $\Delta = \Delta_5$ and, hence, we have the rough bound

$$\mathbf{P}\big\{\Delta \geq x\big\} \leq \mathbf{P}\big\{\overline{A}\big\} + \mathbf{P}\big\{\Delta_5 \geq x/2,\, A\big\}. \tag{8.99}$$

In this case, $U_{L,k} = X_{k+1}$, $2^L = 1 \geq 2^M$, $y > x/4$ [see (8.81), (8.82) and (8.88)]. Therefore, by (8.92) and by Lemmas 8.3.1 and 8.3.3,

$$\begin{aligned}
\mathbf{P}\big\{\overline{A}\big\} &\leq \sum_{k=0}^{2^N-1} \mathbf{P}\big\{|U_{L,k}| \geq y\cdot 2^L\big\} = \sum_{k=1}^{2^N} \mathbf{P}\big\{|X_k| \geq y\big\} \\
&\leq 2^{N+1} d \exp\Big(- \min\Big\{\frac{y^2}{4},\, \frac{y}{4\tau}\Big\}\Big) \\
&\leq 2^{N+1} d \exp\Big(- \min\Big\{\frac{xy}{16},\, \frac{x}{16\tau}\Big\}\Big) \\
&\leq 2^{N+1} d \exp\Big(- \frac{c_{21}\, x}{d^{3/2}\tau}\Big).
\end{aligned} \tag{8.100}$$

It remains to estimate $\mathbf{P}\big\{\Delta_5 \geq x/2,\, A\big\}$ in both cases: $L = M$ and $L = 0$ [see (8.95) and (8.98)–(8.100)]. Let L defined by (8.88) be arbitrary. Fix an integer k satisfying $1 \leq k \leq 2^{N-L}$ and denote for simplicity

$$j = j(k) := k \cdot 2^L. \tag{8.101}$$

By Corollary 8.2.1, we have

$$|\, S_{k\cdot 2^L} - T_{k\cdot 2^L}\,| = |\, S_j - T_j\,| \leq |\, U_{N,0} - V_{N,0}\,| + \frac{1}{2}\sum_{n=L+1}^{N} |\,\widetilde{U}_{n,l_{n,j}} - \widetilde{V}_{n,l_{n,j}}\,|, \tag{8.102}$$

where $l_{n,j}$ are integers, defined by $l_{n,j}\cdot 2^n < j \leq (l_{n,j}+1)\cdot 2^n$ [see (8.56)].

By virtue of (8.81) and (8.92), for $\omega \in A$ we have

$$|U_{L,l}| < y \cdot 2^L = \frac{c_{18}\cdot 2^L}{d^{3/2}\tau} \leq \frac{\min\{c_{14},\, c_{17}\}\cdot 2^L}{d^{3/2}\tau}, \qquad 0 \leq l < 2^{N-L}, \tag{8.103}$$

and, by (8.49)–(8.93), $U_{L,l}$ are sums over blocks consisting of 2^L summands. Moreover, $U_{n,l}$ (resp. $\widetilde{U}_{n,l}$), $L+1 \le n \le N$, $0 \le l < 2^{N-n}$, are sums (resp. differences) of two sums over blocks containing each 2^{n-1} summands. These sums and differences can be represented as linear combinations (with coefficients ± 1) of 2^{n-L} sums over blocks containing each 2^L summands and satisfying (8.103). Therefore, for $\omega \in A$, $L+1 \le n \le N$, $0 \le l < 2^{N-n}$, we have [see (8.46) and (8.48)]

$$|\mathbf{U}_{n,l}| = \max\{|U_{n,l}|, |\widetilde{U}_{n,l}|\} \le 2^{n-L} y \cdot 2^L = y \cdot 2^n \le \frac{\min\{c_{14}, c_{17}\} \cdot 2^n}{d^{3/2}\tau}. \tag{8.104}$$

Using (8.104), we see that if $\omega \in A$, the conditions of Lemma 8.3.6 are satisfied for τ, $U_{N,0}$ and $\mathbf{U}_{n,l}$, if $L+1 \le n \le N$, $0 \le l < 2^{N-n}$. By (8.102), (8.104) and by Lemma 8.3.6, for $\omega \in A$ we have

$$\begin{aligned} |S_j - T_j| &\le c_{13} d^{3/2} \tau \left(1 + 2^{-N} |U_{N,0}|^2\right) \\ &\quad + \sum_{n=L+1}^{N} c_{16} d^{3/2} \tau \left(1 + 2^{-n} \max\{|U_{n,l_{n,j}}|^2, |\widetilde{U}_{n,l_{n,j}}|^2\}\right) \\ &\le c d^{3/2} \tau \left(N + 1 + 2^{-N} |U_{N,0}|^2 + \sum_{n=L}^{N-1} 2^{-n} \left(|U^{(n)}|^2 + |U_{(n)}|^2\right)\right), \end{aligned} \tag{8.105}$$

where

$$U^{(n)} = U_{n,l_{n,j}}, \qquad U_{(n)} = U_{n,\widetilde{l}_{n,j}}, \tag{8.106}$$

and

$$\widetilde{l}_{n-1,j} = \begin{cases} 2l_{n,j}, & \text{if } l_{n-1,j} = 2l_{n,j} + 1, \\ 2l_{n,j} + 1, & \text{if } l_{n-1,j} = 2l_{n,j}, \end{cases} \qquad L < n \le N \tag{8.107}$$

(it is easy to see that $l_{n-1,j}$ can be equal either to $2l_{n,j}$ or to $2l_{n,j}+1$, for given $l_{n,j}$). In other words, $U^{(n)}$, $L \le n \le N$, is the sum over the block of 2^n summands which contains X_j. The sum $U_{(n)}$ does not contain X_j and

$$U^{(n+1)} = U^{(n)} + U_{(n)}, \qquad L \le n < N \tag{8.108}$$

[see (8.93)]. The equality (8.108) implies

$$U^{(n)} = U^{(L)} + \sum_{s=0}^{n-L-1} U_{(L+s)}, \qquad L \le n \le N. \tag{8.109}$$

It is important that all summands in the right-hand side of (8.109) are the sums of disjoint blocks of independent summands. Therefore, they are independent.

Put $\beta = 1/\sqrt{2}$. Then, using (8.109) and the Hölder inequality, one can easily derive that, for $L \le n \le N$,

$$|U^{(n)}|^2 \le c_{22} \left(\beta^{-(n-L)} |U^{(L)}|^2 + \sum_{s=0}^{n-L-1} \beta^{-(n-L-1)+s} |U_{(L+s)}|^2 \right) \quad (8.110)$$

with $c_{22} = \sum_{j=0}^{\infty} \beta^j = \frac{\sqrt{2}}{\sqrt{2}-1}$. It is easy to see that

$$\sum_{n=L}^{N} 2^{-n} \beta^{-(n-L)} |U^{(L)}|^2 \le c_{22} \cdot 2^{-L} |U^{(L)}|^2. \quad (8.111)$$

Moreover,

$$\begin{aligned}
&\sum_{n=L+1}^{N} \sum_{s=0}^{n-L-1} 2^{-n} \beta^{-(n-L-1)+s} |U_{(L+s)}|^2 \\
&\quad = \sum_{s=0}^{N-L-1} \sum_{n=L+1+s}^{N} 2^{-n} \beta^{-(n-L-1)+s} |U_{(L+s)}|^2 \\
&\quad \le c_{22} \sum_{s=0}^{N-L-1} 2^{-(L+1+s)} |U_{(L+s)}|^2. \quad (8.112)
\end{aligned}$$

It is clear that the inequalities (8.110)–(8.112) imply

$$\begin{aligned}
&2^{-N} |U_{N,0}|^2 + \sum_{n=L}^{N-1} 2^{-n} (|U^{(n)}|^2 + |U_{(n)}|^2) \\
&\le c_{22} \left(\frac{|U^{(L)}|^2}{2^L} + \sum_{s=0}^{N-L-1} \frac{|U_{(L+s)}|^2}{2^{L+1+s}} \right) + \sum_{n=L}^{N-1} \frac{|U_{(n)}|^2}{2^n} \\
&\le c \left(\frac{|U^{(L)}|^2}{2^L} + \sum_{n=L}^{N-1} \frac{|U_{(n)}|^2}{2^n} \right). \quad (8.113)
\end{aligned}$$

From (8.105) and (8.113), it follows that for $\omega \in A$ we have

$$|S_j - T_j| \le c_{23} d^{3/2} \tau \left(N + 1 + \frac{|U^{(L)}|^2}{2^L} + \sum_{n=L}^{N-1} \frac{|U_{(n)}|^2}{2^n} \right). \quad (8.114)$$

Denote (for $0 \le n \le N,\ 0 \le l < 2^{N-n}$)

$$W_{n,l} = \begin{cases} 2^{-n} |U_{n,l}|^2, & \text{if } |U_{n,l}| \le y \cdot 2^n, \\ 0, & \text{otherwise.} \end{cases} \quad (8.115)$$

Let us show that

$$\mathbf{E} \exp(t W_{n,l}) \le 2d + 1 \qquad \text{for} \quad 0 \le t \le \frac{1}{8}. \quad (8.116)$$

Indeed, integrating by parts, we obtain

$$\begin{aligned}\mathbf{E}\,\exp(tW_{n,l}) &= 1+\int_0^{y^2\cdot 2^n} t\exp(tu)\,\mathbf{P}\{W_{n,l}\ge u\}\,du\\ &\le 1+\frac{1}{8}\int_0^{y^2\cdot 2^n}\exp(u/8)\,\mathbf{P}\{|U_{n,l}|\ge 2^{n/2}\sqrt{u}\}\,du.\end{aligned}\tag{8.117}$$

Taking into account (8.93), (8.81) and using Lemmas 8.3.1 and 8.3.3, we obtain that

$$\begin{aligned}\mathbf{P}\{|U_{n,l}| &\ge 2^{n/2}\sqrt{u}\}\le 2d\,\exp\Big(-\min\Big\{\frac{2^n u}{4\cdot 2^n},\,\frac{2^{n/2}\sqrt{u}}{4\tau}\Big\}\Big)\\ &\le 2d\,\exp\Big(-\min\Big\{\frac{u}{4},\,\frac{u}{4y\tau}\Big\}\Big)\\ &= 2d\,\exp\Big(-\frac{u}{4}\Big)\end{aligned}\tag{8.118}$$

if $0\le u\le y^2\cdot 2^n$. The relation (8.116) immediately follows from (8.117) and (8.118).

The relations (8.103), (8.104) and (8.115) imply that, for $L\le n\le N$, $0\le l<2^{N-n}$, $\omega\in A$,

$$2^{-n}|U_{n,l}|^2=W_{n,l}.\tag{8.119}$$

Thus, according to (8.106), we can rewrite (8.114) in the form

$$|S_j-T_j|\le c_{23}d^{3/2}\tau\Big(N+1+W^{(L)}+\sum_{n=L}^{N-1}W_{(n)}\Big),\qquad \omega\in A,\tag{8.120}$$

where

$$W^{(L)}=W_{L,l_{L,j}},\qquad W_{(n)}=W_{n,\widetilde{l}_{n,j}}.\tag{8.121}$$

Putting now $t^*=(8\,c_{23}d^{3/2}\tau)^{-1}$ and $t=t^*\cdot c_{23}d^{3/2}\tau=1/8$, taking into account that the random variables $W^{(L)}$, $W_{(L)}$, ..., $W_{(N-1)}$ are independent and applying (8.116), (8.120) and (8.121), we obtain

$$\begin{aligned}&\mathbf{P}\Big\{\{\omega:|S_j-T_j|\ge x/2\}\cap A\Big\}\\ &\le \mathbf{P}\Big\{c_{23}d^{3/2}\tau\Big(N+1+W^{(L)}+\sum_{n=L}^{N-1}W_{(n)}\Big)\ge x/2\Big\}\\ &\le \mathbf{P}\Big\{t\Big(W^{(L)}+\sum_{n=L}^{N-1}W_{(n)}\Big)\ge t^*x/2-t(N+1)\Big\}\end{aligned}$$

$$\begin{aligned}
&\leq \mathbf{E}\exp\Big(t\,\Big(W^{(L)}+\sum_{n=L}^{N-1}W_{(n)}\Big)\Big)\Big/\exp\left(t^{*}x/2-t\,(N+1)\right)\\
&= \mathbf{E}\exp\left(t\,W^{(L)}\right)\prod_{n=L}^{N-1}\mathbf{E}\exp\left(t\,W_{(n)}\right)\Big/\exp\left(t^{*}x/2-t\,(N+1)\right)\\
&\leq (3d)^{N+1}\exp\Big(\frac{N+1}{8}-\frac{x}{16\,c_{23}\,d^{3/2}\,\tau}\Big). \qquad (8.122)
\end{aligned}$$

From (8.91), (8.101), and (8.122) it follows that

$$\mathbf{P}\left\{\Delta_5\geq x/2,\,A\right\}\leq 2^N\,(3d)^{N+1}\exp\Big(\frac{N+1}{8}-\frac{x}{16\,c_{23}\,d^{3/2}\,\tau}\Big). \qquad (8.123)$$

Using (8.87), (8.95), (8.98)–(8.100) and (8.123), we obtain that

$$\mathbf{P}\left\{\Delta\geq x\right\}\leq (19d)^{N+1}\exp\Big(-\frac{x}{c_{24}\,d^{3/2}\,\tau}\Big),\qquad x\geq 0, \qquad (8.124)$$

where we can take $c_{24}=\max\left\{16c_{23},\,c_{19}^{-1},\,c_{20}^{-1},\,c_{21}^{-1},\,2\right\}$. Let the quantities $\varepsilon,\,x_0>0$ be defined by the relations

$$\varepsilon=\frac{1}{2\,c_{24}\,d^{3/2}\,\tau}\leq\frac{1}{4\tau},\qquad e^{\varepsilon x_0}=(19d)^{N+1}. \qquad (8.125)$$

Integrating by parts and using (8.124) and (8.125), we obtain

$$\mathbf{E}\,e^{\varepsilon\Delta}=\int_0^\infty\varepsilon\,e^{\varepsilon x}\,\mathbf{P}\left\{\Delta\geq x\right\}dx+1,$$

$$\int_0^{x_0}\varepsilon\,e^{\varepsilon x}\,\mathbf{P}\left\{\Delta\geq x\right\}dx\leq\int_0^{x_0}\varepsilon\,e^{\varepsilon x}\,dx=e^{\varepsilon x_0}-1=(19d)^{N+1}-1,$$

$$\int_{x_0}^\infty\varepsilon\,e^{\varepsilon x}\,\mathbf{P}\left\{\Delta\geq x\right\}dx\leq\int_{x_0}^\infty\varepsilon\,e^{-\varepsilon(x-x_0)}\,dx=1,$$

and, hence,

$$\mathbf{E}\,e^{\varepsilon\Delta}\leq(19d)^{N+1}+1\leq(20d)^{N+1}.$$

Together with (8.80) and (8.125), this completes the proof of Theorem 8.2.1. ■

8.4 Proofs of Theorems 8.1.1–8.1.4

We start the proofs of Theorems 8.1.1–8.1.3 with the following common part.

BEGINNING OF THE PROOFS OF THEOREMS 8.1.1, 8.1.2 AND 8.1.3. At first, we shall verify that under the conditions of Theorems 8.1.2 or 8.1.3 we have

$\mathcal{L}(\xi_k) \in \mathcal{A}_d(\tau)$. For Theorem 8.1.3, this relation is an immediate consequence of Lemma 8.3.1, of the completeness of classes $\mathcal{A}_d(\tau)$ with respect to convolution, and of the conditions (8.8) and (8.10)–(8.12). In the case of Theorem 8.1.2, we denote $K = \mathcal{L}(\eta)$. One can easily verify that $\mathbf{B} = \operatorname{cov} K = \gamma^2 \mathbf{I}_d$, where γ^2 is defined by (8.7) and, hence,

$$1 \le \gamma^2 \le 3. \tag{8.126}$$

Moreover,

$$\varphi(K,z) = \log \mathbf{E}\, e^{\langle z,\eta\rangle} = \log \frac{(4+\tau^2(d+\langle z,\overline{z}\rangle))\exp(\langle z,\overline{z}\rangle/2)}{(4+\tau^2 d)}, \quad z \in \mathbf{C}^d. \tag{8.127}$$

Using (8.126) and (8.127), we obtain

$$\begin{aligned} |d_u d_v^2 \varphi(K,z)| &= |d_u d_v^2 \log(4+\tau^2(d+\langle z,\overline{z}\rangle))| \\ &\le c\tau^3 \|u\|\,\|v\|^2 \le \|u\|\,\tau\langle \mathbf{B}\,v,v\rangle \end{aligned} \tag{8.128}$$

for $\|z\|\,\tau \le 1$, if c_1 involved in Assertion A is sufficiently small. This means that $K = \mathcal{L}(\eta) \in \mathcal{A}_d(\tau)$. The relation $\mathcal{L}(\xi_k) = \mathcal{L}(\eta/\gamma) \in \mathcal{A}_d(\tau)$, $k = 1,\dots,n$, follows from (8.126) and from Lemma 8.3.1.

The text below is related to Theorems 8.1.1, 8.1.2 and 8.1.3 simultaneously. Without loss of generality we assume that the amount of summands is equal to 2^N with some positive integer N. It suffices to show that the dyadic scheme related to the vectors $\xi_1,\dots,\xi_{2^N}$ satisfies the conditions of Theorem 8.2.1 with $\tau^* = \sqrt{2}\,\tau$ instead of τ. According to Lemma 8.2.1, we can verify the conditions (8.31) and (8.32) for the vectors $\mathbf{U}^j_{n,k}$ and $\mathbf{U}^j_{N,0}$ instead of $\mathbf{U}^{*j}_{n,k}$ and $\mathbf{U}^{*j}_{N,0}$. To this end, we shall show that

$$\mathcal{L}(\mathbf{U}^j_{n,k}) \in \overline{\mathcal{A}}_j(\sqrt{2}\,\tau,4) \qquad \text{for} \quad 0 \le k < 2^{N-n}, \quad 1 \le n \le N, \quad 1 \le j \le 2d. \tag{8.129}$$

Recall that $\mathbf{U}_{n,k} = \mathbf{A}\,\widetilde{\mathbf{U}}_{n,k}$, where $\mathbf{A}$ is the linear operator defined by (8.29) and satisfying (8.54). Furthermore, $\widetilde{\mathbf{U}}_{n,k} = (U_{n-1,2k}, U_{n-1,2k+1}) \in \mathbf{R}^{2d}$, where the d-dimensional vectors $U_{n-1,2k}$ and $U_{n-1,2k+1}$ are independent. The relation $\mathcal{L}(\mathbf{U}_{n,k}) \in \mathcal{A}_{2d}(\sqrt{2}\,\tau)$ can be therefore easily derived from the conditions of Theorems 8.1.1, 8.1.2 and 8.1.3 with the help of Lemmas 8.2.1, 8.3.1 and 8.3.2 [see (8.54)] if we take into account the completeness of classes $\mathcal{A}_d(\tau)$ with respect to convolution and their monotonicity with respect to τ. It is easy to see that $\mathbf{U}^j_{n,k} = \overline{\mathbf{P}}_j\,\mathbf{U}_{n,k}$, where the projector $\overline{\mathbf{P}}_j : \mathbf{R}^{2d} \to \mathbf{R}^j$ can be considered as a linear operator with $\|\overline{\mathbf{P}}_j\| = 1$ [see (8.48)]. Applying Lemma 8.3.1 again, we obtain the relations $\mathcal{L}(\mathbf{U}^j_{n,k}) \in \mathcal{A}_j(\sqrt{2}\,\tau)$, $1 \le j \le 2d$.

It remains to verify that, for $h \in \mathbf{R}^j$, $\|h\|\,\sqrt{2}\,\tau < 1$, the following inequality holds

$$\int\limits_T |\widehat{F}_h(t)|\,dt \le \frac{(2\pi)^{j/2}\,\sqrt{2}\,\tau\,j^{3/2}}{\sigma\,(\det \mathbf{D})^{1/2}}, \tag{8.130}$$

$$T = \{t \in \mathbf{R}^j : 4\,\|t\|\,\sqrt{2}\,\tau j \ge 1\}, \tag{8.131}$$

where $F = \mathcal{L}(\mathbf{U}^j_{n,k})$, and σ^2 is the minimal eigenvalue of $\mathbf{D} = \operatorname{cov}\mathbf{U}^j_{n,k}$. Note that, according to (8.65), we have

$$\mathbf{D} = 2^n\,\mathbf{I}_j, \quad \sigma^2 = 2^n, \quad \det\mathbf{D} = 2^{nj}. \tag{8.132}$$

Introduce 2^{n-1} random vectors

$$\mathbf{X}_r = (X_r,\, X_{2^{n-1}+r}) \in \mathbf{R}^{2d}, \qquad r = 2^{n-1}\cdot 2k+1,\ldots,2^{n-1}(2k+1). \tag{8.133}$$

Obviously, these vectors are independent. According to (8.50), (8.162) and (8.133),

$$\widetilde{\mathbf{U}}_{n,k} = (U_{n-1,2k},\, U_{n-1,2k+1}) = \sum_{r=2^{n-1}\cdot 2k+1}^{2^{n-1}(2k+1)} \mathbf{X}_r. \tag{8.134}$$

Denote $R_h^{(s)} = \overline{\mathcal{L}(X_s)}(h)$, for $s = 1,\ldots,2^N$, $h \in \mathbf{R}^d$, and $M_h^{(r)} := \overline{\mathcal{L}(\mathbf{X}_r)}(h)$, $Q_h^{(r)} := \overline{\mathcal{L}(\mathbf{A}\,\mathbf{X}_r)}(h)$, for $r = 2^{n-1}\cdot 2k+1,\ldots,2^{n-1}(2k+1)$, $h \in \mathbf{R}^{2d}$. As usually, we consider only such h for which these distributions exist. Using (8.21), we see that, for all $t \in \mathbf{R}^{2d}$,

$$\begin{aligned}\widehat{Q}_h^{(r)}(t) = \frac{\mathbf{E}\,\exp\left(\langle h+it,\mathbf{A}\mathbf{X}_r\rangle\right)}{\mathbf{E}\,\exp\left(\langle h,\mathbf{A}\mathbf{X}_r\rangle\right)} &= \frac{\mathbf{E}\,\exp\left(\langle \mathbf{A}^*h+i\mathbf{A}^*t,\mathbf{X}_r\rangle\right)}{\mathbf{E}\,\exp\left(\langle \mathbf{A}^*h,\mathbf{X}_r\rangle\right)} \\ &= \widehat{M}^{(r)}_{\mathbf{A}^*h}(\mathbf{A}^*t).\end{aligned} \tag{8.135}$$

By (8.16) and (8.134), we have (for $j = 2d$)

$$|\widehat{F}_h(t)| = \prod_{r=2^{n-1}\cdot 2k+1}^{2^{n-1}(2k+1)} |\widehat{Q}_h^{(r)}(t)|. \tag{8.136}$$

Split $t = (t_1,\ldots,t_{2d}) \in \mathbf{R}^{2d}$ as $t = (t^{(1)},\, t^{(2)})$, where $t^{(1)} = (t_1,\ldots,t_d)$ and $t^{(2)} = (t_{d+1},\ldots,t_{2d}) \in \mathbf{R}^d$. Using (8.133), (8.21) and introducing a similar notation for $h \in \mathbf{R}^{2d}$, it is easy to check that

$$\widehat{M}_h^{(r)}(t) = \widehat{R}^{(r)}_{h^{(1)}}(t^{(1)})\,\widehat{R}^{(2^{n-1}+r)}_{h^{(2)}}(t^{(2)}). \tag{8.137}$$

Note that

$$\|t\|^2 = \|t^{(1)}\|^2 + \|t^{(2)}\|^2 . \tag{8.138}$$

END OF THE PROOF OF THEOREM 8.1.1. Let now the distributions $\mathcal{L}(\xi_s)$ satisfy the conditions of Theorem 8.1.1. In this case, according to (8.16), we have $R_h^{(s)} = \overline{H}_s(h)\,\overline{G}(h)$. It is well-known that the conjugate distributions $\overline{G}(h)$ of the Gaussian distribution G are also Gaussian with covariance operator $\operatorname{cov}\overline{G}(h) = \operatorname{cov} G = b^2\,\mathbf{I}_d$. Therefore,

$$|\widehat{R}_h^{(s)}(t)| \leq \exp(-b^2\,\|t\|^2/2), \qquad t, h \in \mathbf{R}^d, \ \ \|h\|\,\tau < 1. \tag{8.139}$$

Using (8.137)–(8.139), we get, for $t, h \in \mathbf{R}^{2d}$, $\|h\|\,\tau < 1$,

$$|\widehat{M}_h^{(s)}(t)| \leq \prod_{\mu=1}^{2} \exp(-b^2\,\|t^{(\mu)}\|^2/2) = \exp(-b^2\,\|t\|^2/2). \tag{8.140}$$

Applying (8.54), (8.135) and (8.140) with $t = \mathbf{A}^* u$ and $h = \mathbf{A}^*\gamma$, we see that

$$|\widehat{Q}_\gamma^{(s)}(u)| \leq \exp(-b^2\,\|\mathbf{A}^* u\|^2/2) \leq \exp(-b^2\,\|u\|^2) \tag{8.141}$$

for $u, \gamma \in \mathbf{R}^{2d}$, $\|\gamma\|\,\sqrt{2}\,\tau < 1$. The relations (8.136) and (8.141) imply that

$$|\widehat{F}_h(t)| \leq \exp(-b^2\,\|t\|^2 \cdot 2^{n-2}), \qquad t, h \in \mathbf{R}^j, \ \ \|h\|\,\sqrt{2}\,\tau < 1. \tag{8.142}$$

It is clear that it suffices to verify (8.142) for $j = 2d$ [for $1 \leq j < 2d$, one should apply (8.142) for $j = 2d$ and for $t, h \in \mathbf{R}^{2d}$, with $h_m = t_m = 0$, $m = j+1, \ldots, 2d$].

Using (8.131), (8.132) and (8.142), we see that

$$\begin{aligned}
\int_T |\widehat{F}_h(t)|\,dt \ &\leq \ \exp\Big(-\frac{b^2 \cdot 2^{n-3}}{32\,\tau^2 j^2}\Big) \int_{\mathbf{R}^j} \exp(-b^2\,\|t\|^2 \cdot 2^{n-3})\,dt \\
&= \ \frac{(2\,\pi)^{j/2}}{(b^2 \cdot 2^{n-2})^{j/2}}\, \exp\Big(-\frac{b^2 \cdot 2^n}{2^8\,\tau^2 j^2}\Big) \\
&\leq \ \frac{(2\,\pi)^{j/2}\,\tau^{4j\cdot 2^n}}{(\det \mathbf{D})^{1/2}\,\tau^{2j}} \leq \frac{(2\,\pi)^{j/2}\,\tau}{2^{n/2}\,(\det \mathbf{D})^{1/2}}
\end{aligned} \tag{8.143}$$

if c_1 is small enough. The relations (8.132) and (8.143) imply (8.130). It remains to apply Theorem 8.2.1 to complete the proof of Theorem 8.1.1. ■

END OF THE PROOF OF THEOREM 8.1.2. Let now the distributions $\mathcal{L}(\xi_s)$ satisfy the conditions of Theorem 8.1.2. In this case, according to (8.21) and (8.127), we have

$$\begin{aligned} |\widehat{R}_h^{(s)}(t)| &= \left| \frac{\left(4+\tau^2 (d+\|h\|^2+2i\langle h,t\rangle-\|t\|^2)\right) \exp\left((\|h\|^2+2i\langle h,t\rangle-\|t\|^2)/2\right)}{\left(4+\tau^2 (d+\|h\|^2)\right) \exp\left(\|h\|^2/2\right)} \right| \\ &\le (2+\|t\|^2)\exp(-\|t\|^2/2) \\ &\le c_{25}\exp(-\|t\|^2/4), \qquad \|h\|\,\tau<1. \end{aligned} \tag{8.144}$$

The rest of the proof is omitted. It is similar to that of Theorem 8.1.1 with $b^2 = 1/2$. The presence of c_{25} in the right-hand side of (8.144) can be easily compensated by choosing c_1 to be sufficiently small. ■

End of the proof of Theorem 8.1.3. Consider the dyadic scheme with

$$\mathcal{L}(\xi_s) = \mathcal{L}(X_s) = L^{(s)}P, \qquad s = 1,\ldots,2^N. \tag{8.145}$$

Putting $H := \mathcal{L}(\zeta)$, $\psi_h(x) = e^{\langle h,x\rangle}\,p(x)$, $h,x \in \mathbf{R}^d$, and integrating by parts, we see that (for $t \in \mathbf{R}^d$, $t \neq 0$)

$$\begin{aligned} \widehat{H}_h(t) &= (\mathbf{E}\,e^{\langle h,\zeta\rangle})^{-1} \int\limits_{\|x\|\le b_1} e^{i\langle t,x\rangle}\,\psi_h(x)\,dx \\ &= -(\mathbf{E}\,e^{\langle h,\zeta\rangle})^{-1} \int\limits_{\|x\|\le b_1} \frac{e^{i\langle t,x\rangle}}{i\,\|t\|^2}\,d_t\psi_h(x)\,dx, \end{aligned} \tag{8.146}$$

where $H_h = \overline{H}(h)$. Besides, using (8.9), we see that

$$\sup_{\|x\|\le b_1}\ \sup_{\|h\|\,b_2\le 1} |d_t\psi_h(x)| \le b_5\,\|t\|. \tag{8.147}$$

As in the formulation of Theorem 8.1.3, we denote by b_m different positive quantities depending on H. Note that the quantities depending on the dimension d can be considered as depending on H only as well. From (8.146) and (8.147), it follows that

$$\sup_{\|h\|\,b_2\le 1} |\widehat{H}_h(t)| \le b_6\,\|t\|^{-1} \tag{8.148}$$

(note that, by the Jensen inequality, $\mathbf{E}\,e^{\langle h,\zeta\rangle} \ge e^{\mathbf{E}\langle h,\zeta\rangle} = 1$). The inequality (8.148) implies that

$$\sup_{\|h\|\,b_2\le 1} |\widehat{H}_h(t)| \le \left(1+\frac{\|t\|}{b_7}\right)^{-1} \qquad \text{for} \quad \|t\| \ge b_7 = 2b_6 \tag{8.149}$$

and

$$\sup_{\|h\|\,b_2\le 1}\ \sup_{\|t\|\ge b_7} |\widehat{H}_h(t)| \le 1/2. \tag{8.150}$$

Since the distributions H_h are absolutely continuous, the relation $|\widehat{H}_h(t)| = 1$ can be valid for $t = 0$ only. Furthermore, the function $|\widehat{H}_h(t)|$ considered as a function of two variables h and t is continuous for all $h, t \in \mathbf{R}^d$. Therefore,

$$\sup_{\|h\|\,b_2\leq 1}\ \sup_{b_8\leq\|t\|\leq b_7} |\widehat{H}_h(t)| \leq b_9 < 1, \tag{8.151}$$

where

$$b_8 = (4\sqrt{2}\,b_2\,d)^{-1} \quad \text{and} \quad b_9 \geq 1/2. \tag{8.152}$$

The inequalities (8.150) and (8.151) imply that

$$\sup_{\|h\|\,b_2\leq 1}\ \sup_{\|t\|\geq b_8} |\widehat{H}_h(t)| \leq b_9 := e^{-b_{10}} < 1. \tag{8.153}$$

Denoting $L_h^{(s)} = \overline{L}^{(s)}(h)$, $h \in \mathbf{R}^d$, $s = 1, \ldots, 2^N$, and using (8.11), (8.12), (8.16) and (8.21), it is easy to see that

$$\widehat{R}_h^{(s)}(t) = \big(\widehat{H}_{h/\sqrt{m}}(t/\sqrt{m})\big)^m\, \widehat{L}_h^{(s)}(t). \tag{8.154}$$

The relations (8.10), (8.149), (8.153) and (8.154) imply that

$$\sup_{\|h\|\,\mathcal{T}\leq 1} |\widehat{R}_h^{(s)}(t)| \leq \Big(1 + \frac{\|t\|}{b_7\sqrt{m}}\Big)^{-m} \qquad \text{for} \quad \|t\| \geq b_7\sqrt{m} \tag{8.155}$$

and

$$\sup_{\|h\|\,\mathcal{T}\leq 1}\ \sup_{\|t\|\geq b_8\sqrt{m}} |\widehat{R}_h^{(s)}(t)| \leq e^{-m b_{10}}. \tag{8.156}$$

Using (8.137), (8.138), (8.145) and (8.155), we get, for $r = 2^{n-1}\cdot 2k + 1, \ldots, 2^{n-1}\,(2k+1)$, $\|t\| \geq b_7\sqrt{2m}$, $t \in \mathbf{R}^{2d}$,

$$\sup_{\|h\|\mathcal{T}\leq 1} |\widehat{M}_h^{(r)}(t)| \leq \min_{\mu=1,2}\Big(1 + \frac{\|t^{(\mu)}\|}{b_7\sqrt{m}}\Big)^{-m} \leq \Big(1 + \frac{\|t\|}{b_7\sqrt{2m}}\Big)^{-m}. \tag{8.157}$$

Moreover,

$$\sup_{\|h\|\mathcal{T}\leq 1}\ \sup_{\|t\|\geq b_8\sqrt{2m}} |\widehat{M}_h^{(r)}(t)| \leq e^{-m b_{10}}. \tag{8.158}$$

Using (8.54), (8.135), (8.157) and (8.158), we see that, for the same r and for $t \in \mathbf{R}^{2d}$, $\|t\| \geq b_7\sqrt{m}$,

$$\sup_{\|h\|\mathcal{T}\sqrt{2}\leq 1} |\widehat{Q}_h^{(r)}(t)| \leq \Big(1 + \frac{\|t\|}{b_7\sqrt{m}}\Big)^{-m} \tag{8.159}$$

and

$$\sup_{\|h\|\tau\sqrt{2}\le 1}\ \sup_{\|t\|\ge b_8\sqrt{m}} |\widehat{Q}_h^{(r)}(t)| \le e^{-mb_{10}}. \tag{8.160}$$

It is easy to see that the relations (8.136), (8.159) and (8.160) imply that, for $h \in \mathbf{R}^j$, $\|h\|\sqrt{2}\,\tau < 1$, and for $t \in \mathbf{R}^j$, $\|t\| \ge b_7\sqrt{m}$,

$$|\widehat{F}_h(t)| \le \Big(1 + \frac{\|t\|}{b_7\sqrt{m}}\Big)^{-m\cdot 2^{n-1}} \tag{8.161}$$

and

$$\sup_{\|t\|\ge b_8\sqrt{m}} |\widehat{F}_h(t)| \le e^{-mb_{10}\cdot 2^{n-1}}. \tag{8.162}$$

It suffices to prove (8.161) and (8.162) for $j = 2d$ [for $1 \le j < 2d$, one should apply (8.161) and (8.162) for $j = 2d$ and for $h \in \mathbf{R}^{2d}$, $\|h\|\sqrt{2}\,\tau < 1$, $t \in \mathbf{R}^{2d}$ with $h_m = t_m = 0$, $m = j+1, \dots, 2d$].

Note now that the set T defined in (8.131) satisfies the relation

$$T \subset \{t \in \mathbf{R}^j : \|t\| \ge b_8\sqrt{m}\} \tag{8.163}$$

[see (8.10) and (8.152)]. Below [in the proof of (8.130)] we assume that $\|h\|\sqrt{2}\,\tau < 1$. According to (8.162) and (8.163), for $t \in T$ we have

$$|\widehat{F}_h(t)|^{1/2} \le e^{-mb_{10}\cdot 2^{n-2}}. \tag{8.164}$$

Taking into account that $|\widehat{F}_h(t)| \le 1$, and $m \ge b_4$, choosing b_4 to be sufficiently large and using (8.10), (8.132), (8.161) and (8.164), we obtain

$$\begin{aligned}
&\int_T |\widehat{F}_h(t)|\,dt \\
&\le \exp(-mb_{10}\cdot 2^{n-2})\Big(\int_{\mathbf{R}^j}\Big(1 + \frac{\|t\|}{b_7\sqrt{m}}\Big)^{-m\cdot 2^{n-2}} dt + b_{11}m^{d/2}\Big) \\
&\le b_{12}\, m^{d/2}\exp(-mb_{10}\cdot 2^{n-2}) \\
&\le \frac{(2\pi)^{j/2}\sqrt{2}\, b_2 j^{3/2}}{m^{1/2}\cdot 2^{n/2}\cdot 2^{nj/2}} = \frac{(2\pi)^{j/2}\sqrt{2}\,\tau j^{3/2}}{\sigma\,(\det\mathbf{D})^{1/2}}.
\end{aligned} \tag{8.165}$$

The inequality (8.130) follows from (8.165) immediately. It remains to apply Theorem 8.2.1. ■

PROOF OF THEOREM 8.1.4. Define $m_0, m_1, m_2, \dots$ and $n_1, n_2, \dots$ by

$$m_0 = 0, \quad m_s = 2^{2^s}, \qquad n_s = m_s - m_{s-1}, \qquad s = 1, 2, \dots. \tag{8.166}$$

It is easy to see that

$$\log n_s \le \log m_s = 2^s \log 2, \qquad s = 1, 2, \dots . \tag{8.167}$$

By Assertion A [see (8.5)], for any $s = 1, 2, \dots$ one can construct on a probability space a sequence of i.i.d. $X_1^{(s)}, \dots, X_{n_s}^{(s)}$ and a sequence of i.i.d. Gaussian $Y_1^{(s)}, \dots, Y_{n_s}^{(s)}$ so that $\mathcal{L}(X_k^{(s)}) = \mathcal{L}(\xi)$, $\mathbf{E}\, Y_k^{(s)} = 0$, $\operatorname{cov} Y_k^{(s)} = \mathbf{I}_d$, and

$$\mathbf{P}\,\{\, c_2\, \Delta_s \ge \tau\, d^{3/2} (c_3 \log^* d \, \log n_s + x)\,\} \le e^{-x}, \qquad x \ge 0, \tag{8.168}$$

where

$$\Delta_s = \max_{1 \le r \le n_s} \Big|\sum_{k=1}^{r} X_k^{(s)} - \sum_{k=1}^{r} Y_k^{(s)}\Big|. \tag{8.169}$$

It is clear that we can define all the vectors mentioned above on the same probability space so that the collections $\Xi_s = \{X_1^{(s)}, \dots, X_{n_s}^{(s)}; Y_1^{(s)}, \dots, Y_{n_s}^{(s)}\}$, $s = 1, 2, \dots$ are jointly independent. Then, we define $X_1, X_2, \dots$ and $Y_1, Y_2, \dots$ by

$$\begin{aligned} X_{m_{s-1}+k} &= X_k^{(s)}, \\ Y_{m_{s-1}+k} &= Y_k^{(s)}, \end{aligned} \qquad k = 1, \dots, n_s, \quad s = 1, 2, \dots . \tag{8.170}$$

In order to show that these sequences satisfy the assertion of Theorem 8.1.4, it remains to verify the equality (8.13).

Put

$$c_{25} = \frac{(c_3 \log 2 + 1)}{c_2}, \qquad c_{26} = c_{25} \sum_{l=0}^{\infty} 2^{-l/2} = \frac{c_{25}\sqrt{2}}{\sqrt{2} - 1}, \tag{8.171}$$

and introduce the events

$$A_l = \{\, \omega : \Delta^{(l)} \ge 2^l \, c_{26}\, \tau\, d^{3/2} \log^* d \,\}, \qquad l = 1, 2, \dots, \tag{8.172}$$

where

$$\Delta^{(l)} = \max_{1 \le r \le m_l} \Big|\sum_{j=1}^{r} X_j - \sum_{j=1}^{r} Y_j\Big|. \tag{8.173}$$

According to (8.169), (8.170) and (8.173), we have

$$\Delta^{(l)} \le \Delta_1 + \dots + \Delta_l. \tag{8.174}$$

Taking into account the relations (8.167), (8.171), (8.172), (8.174) and applying the inequality (8.168) with $x = 2^{(s+l)/2}$, we get

$$\begin{aligned} \mathbf{P}\,\{A_l\} &\le \sum_{s=1}^{l} \mathbf{P}\,\{\, \Delta_s \ge 2^{(s+l)/2} \, c_{25}\, \tau\, d^{3/2} \log^* d \,\} \\ &\le \sum_{s=1}^{l} \exp\big(-2^{(s+l)/2} \big) \le c \exp\big(-2^{l/2} \big). \end{aligned} \tag{8.175}$$

The inequality (8.175) implies that $\sum_{l=1}^{\infty} \mathbf{P}\{A_l\} < \infty$, Hence, by the Borel–Cantelli lemma, with probability one a finite number of the events A_l occurs only. This implies the equality (8.13) with $c_4 = 2c_{26}/\log 2$ [see (8.166), (8.172) and (8.173)]. ■

Acknowledgements. The authors would like to thank V. Bentkus for very useful discussions.

The first author acknowledges research supported by the SFB 343 in Bielefeld by grant INTAS-RFBR 95-0099 and by grant RFBR-DFG 96-01-00096-ge. The second author acknowledges research supported by the SFB 343 in Bielefeld, by the Russian Foundation of Basic Research, grant 96-01-00672, by grant INTAS-RFBR 95-0099 and by grant RFBR-DFG 96-01-00096-ge.

References

1. Bártfai, P. (1966). Die Bestimmung der zu einem wiederkehrenden Prozess gehörenden Verteilungfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation, *Studia Sci. Math. Hungar.*, **1**, 161–168.

2. Dudley, R. M. (1989). *Real Analysis and Probability*, Pacific Grove, California: Wadsworth & Brooks/Cole.

3. Einmahl, U. (1989). Extensions of results of Komlós, Major and Tusnády to the multivariate case, *Journal of Multivariate Analysis*, **28**, 20–68.

4. Etemadi, N. (1985). On some classical results in probability theory, *Sankhyā, Series A*, **47**, 215–221.

5. Hoffmann-Jørgensen, J. (1994). *Probability with a View Toward Statistics*, I, New York: Chapman and Hall.

6. Götze, F. and Zaitsev A. Yu. (1997). Multidimensional Hungarian construction for vectors with almost Gaussian smooth distributions, *Preprint 97-071 SFB 343*, Universität Bielefeld.

7. Komlós, J., Major, P., Tusnády, G. (1975-76). An approximation of partial sums of independent RV'-s and the sample DF. I; II, *Z. Wahrscheinlichkeitstheor. verw. Geb.*, **32**, 111–131; **34**, 34–58.

8. Major, P. (1978). On the invariance principle for sums of independent identically distributed random variables, *Journal of Multivariate Analysis*, **8**, 487–517.

9. Massart, P. (1989). Strong approximation for multivariate empirical and related processes, via KMT construction, *Annals of Probability*, **17**, 266–291.

10. Rosenblatt, M. (1952). Remarks on a multivariate transformation, *Annals of Statistics*, **23**, 470–472.

11. Sakhanenko, A. I. (1984). Rate of convergence in the invariance principles for variables with exponential moments that are not identically distributed, In *Trudy Inst. Mat. SO AN SSSR*, **3**, pp. 4–49, Novosibirsk: Nauka (in Russian).

12. Zaitsev, A. Yu. (1986). Estimates of the Lévy–Prokhorov distance in the multivariate central limit theorem for random variables with finite exponential moments, *Theory of Probability and its Applications*, **31**, 203–220.

13. Zaitsev, A. Yu. (1995). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments, *Preprint 95-055 SFB 343*, Universität Bielefeld.

14. Zaitsev, A. Yu. (1996). Estimates for quantiles of smooth conditional distributions and multidimensional invariance principle, *Siberian Math. J.*, **37**, 807–831 (in Russian).

15. Zaitsev, A. Yu. (1998a). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments, *ESAIM : Probability and Statistics*, **2**, 41–108.

16. Zaitsev, A. Yu. (1998b). Multidimensional version of the results of Sakhanenko in the invariance principle for vectors with finite exponential moments, *Preprint 98-045 SFB 343*, Universität Bielefeld.

9

On the Existence of Weak Solutions for Stochastic Differential Equations With Driving L^2-Valued Measures

V. A. Lebedev

M. V. Lomonosov Moscow State University, Moscow, Russia

Abstract: For Eq. (9.4) with a σ-finite L^2-valued random measure θ in the sense of Bichteler and Jacod (1983), a theorem on the existence of its weak solution in terms of the decomposition of θ according to Theorem 1 of Lebedev (1995) is proved.

Keywords and phrases: σ-finite L^p-valued random measure, stochastic differential equation, weak solution, extension of a stochastic basis

9.1 Basic Properties of σ-Finite L^p-Valued Random Measures

Let $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ be a stochastic basis consisting of a probability space $(\Omega, \mathcal{F}, \mathrm{P})$ and of a right-continuous filtration $\mathbf{F} = (\mathcal{F}_t)_{t \in \mathsf{R}_+}$, and let $\mathcal{O}$ and $\mathcal{P}$ be the $\mathbf{F}$-optional and the $\mathbf{F}$-predictable σ-algebra on $\Omega \times \mathsf{R}_+$, respectively.

Let $(E, \mathcal{E})$ be a measurable space and θ be a *σ-finite L^p-valued random measure on* $(\Omega \times \mathsf{R}_+ \times E, \mathcal{P} \otimes \mathcal{E})$ in the sense of Bichteler and Jacod (1983) for some $p \geq 0$, i.e., a family $\theta = (\theta_t)_{t \in \mathsf{R}_+}$ satisfying the following conditions:

(a) for every $t \in \mathsf{R}_+$ θ_t is a σ-finite measure on $(\Omega \times \mathsf{R}_+ \times E, \mathcal{P} \otimes \mathcal{E})$ with values in $L^p(\Omega, \mathcal{F}_t, \mathrm{P})$, i.e., there is a strictly positive $\mathcal{P} \otimes \mathcal{E}$-measurable function V on $\Omega \times \mathsf{R}_+ \times E$ such that if $(\mathcal{P} \otimes \mathcal{E})_V$=$\{\varphi$: $\mathcal{P} \otimes \mathcal{E}$-measurable, φ/V is bounded$\}$ then we have:

[(a-1)] θ_t is a linear mapping from $(\mathcal{P} \otimes \mathcal{E})_V$ into $L^p(\Omega, \mathcal{F}_t, \mathrm{P})$,

[(a-2)] if (φ_n) is a sequence in $(\mathcal{P} \otimes \mathcal{E})_V$ with $|\varphi_n| \leq V$ converging pointwise to 0 then $\theta_t(\varphi_n) \to 0$ in $L^p(\Omega, \mathcal{F}_t, \mathrm{P})$;

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

(b) $\theta_s(\varphi) = \theta_t(\varphi 1_{[0,s]})$ for all $\varphi \in (\mathcal{P} \otimes \mathcal{E})_V$ and $s \le t$;

(c) $\theta_t(\varphi 1_{A\times I\times E}) = 1_A \theta_t(\varphi 1_{\Omega\times I\times E})$ for all $\varphi \in (\mathcal{P}\otimes\mathcal{E})_V$ and $t \in \mathsf{R}_+$ if $A \in \mathcal{F}_0$ and $I = \mathsf{R}_+$ or if $A \in \mathcal{F}_s$ and $I =]s, s']$ with $s < s'$.

In particular when $V \equiv 1$ θ is called a *finite L^p-valued random measure.* Let us denote by $\tilde{\mathcal{S}}^p_\sigma$ (respectively by $\tilde{\mathcal{S}}^p$) the space of all σ-finite (finite) L^p-valued random measures on $(\Omega \times \mathsf{R}_+ \times E, \mathcal{P} \otimes \mathcal{E})$.

We put for $\theta \in \tilde{\mathcal{S}}^p_\sigma$

$$\|\varphi\|_{L^{1,p}(\theta)} = \sum_{N=1}^{\infty} (1 \wedge \sup_{\psi\in(\mathcal{P}\otimes\mathcal{E})_V,\ |\psi|\le|\varphi|} \|\theta_N(\psi)\|_p)$$

and denote by $L^{1,p}(\theta)$ the set of $\mathcal{P}\otimes\mathcal{E}$-measurable functions φ for each of which there is a sequence $(\varphi_n) \subset (\mathcal{P} \otimes \mathcal{E})_V$ with $\|\varphi_n - \varphi\|_{L^{1,p}(\theta)} \to 0$. Then for each $t \in \mathsf{R}_+$, $\theta_t(\varphi_n)$ tends to a limit in $L^p(\Omega, \mathcal{F}_t, \mathrm{P})$ which does not depend on the choice of (φ_n) for the given φ and is denoted by $\theta_t(\varphi)$. In addition, if $p \le q$ and $\theta \in \tilde{\mathcal{S}}^q_\sigma$, then $\theta \in \tilde{\mathcal{S}}^p_\sigma$ and $L^{1,q}(\theta) \subset L^{1,p}(\theta)$.

Let us introduce an example which is important for the further development of the theory. Let $E = \{1\}$, i.e., let it consist of one point, so that we consider random measures on $(\Omega \times \mathsf{R}_+, \mathcal{P})$. Then, by the Dellacherie–Mokobodzki–Bichteler theorem [for example Bichteler (1981, Theorem 7.6)], there is a bijective correspondence between the sets of finite L^0-valued random measures on $(\Omega \times \mathsf{R}_+, \mathcal{P})$ and of defined up to indistinguishability semimartingales by the formula

$$\theta_t(H) = H \cdot X_t \tag{9.1}$$

for any bounded predictable H at every $t \in \mathsf{R}_+$. As far as σ-finite L^0-valued random measures are concerned, they are called usually formal semimartingales.

Let $\theta \in \tilde{\mathcal{S}}^p_\sigma$ and $\varphi \in L^{1,p}(\theta)$. Then the equality

$$(\varphi * \theta)_t(H) = \theta_t(H\varphi) \tag{9.2}$$

for $H \in \mathcal{P}_1$ defines the family $\varphi * \theta$ as a finite L^p-valued measure on $(\Omega \times \mathsf{R}_+, \mathcal{P})$ for which by the preceding example there is a semimartingale denoted also by $\varphi * \theta$ and called *the stochastic integral process for φ with respect to θ,* and by (9.1)

$$\varphi * \theta_t = (\varphi * \theta)_t(1) = \theta_t(\varphi).$$

We can define integrals with respect to $\theta \in \tilde{\mathcal{S}}^p_\sigma$ for a wider class of $\mathcal{P} \otimes \mathcal{E}$-measurable functions than $L^{1,p}(\theta)$. Let us introduce the set $L^p_\sigma(\theta)=\{\varphi\colon \mathcal{P} \otimes \mathcal{E}$-measurable, and there is a strictly positive predictable process K such that $K\varphi \in L^{1,p}(\theta)\}$. Then the equality (9.2) for predictable H with bounded H/K defines $\varphi * \theta$ as a σ-finite L^p-valued random measure on $(\Omega \times \mathsf{R}_+, \mathcal{P})$, i.e., as a formal semimartingale. Now we select the set $\hat{L}^p(\theta) \subset L^p_\sigma(\theta)$ of such φ for which

$\varphi * \theta$ is a finite L^p-valued random measure on $(\Omega \times \mathsf{R}_+, \mathcal{P})$, i.e. a semimartingale. Obviously, $L^{1,p}(\theta) \subset \hat{L}^p(\theta)$ and $\hat{L}^q(\theta) \subset \hat{L}^p(\theta)$ for $p \leq q$.

So let $\theta \in \tilde{\mathcal{S}}^2_\sigma$ be given on $(\Omega \times \mathsf{R}_+ \times E, \mathcal{P} \otimes \mathcal{E})$ in the case when $(E, \mathcal{E})$ is a Lusin space with its Borel σ-algebra. Then according to Theorem 1 of Lebedev (1995), there exist a predictable increasing process X^1 and starting at 0 square-integrable martingales X^i for $i \geq 2$ with $\langle X^i, X^j \rangle \equiv 0$ for $i \neq j$ and regular (signed) transition measures ρ^i from $(\Omega \times \mathsf{R}_+, \mathcal{P})$ to $(E, \mathcal{E})$ such that for every $\varphi \in \hat{L}^2(\theta)$ its sections are ρ^i-integrable almost everywhere in the measures $\mathrm{P} \times dX^1$ and $\mathrm{P} \times d\langle X^i, X^i \rangle$ for $i \geq 2$ respectively, and

$$\varphi * \theta = \sum_{i=1}^{\infty} \int_E \varphi \, \rho^i_s(\omega, du) \cdot X^i, \tag{9.3}$$

the series converging unconditionally in $\mathcal{S}^2$, i.e. in $\mathcal{H}^2$ from Emery (1979) on each finite interval. Besides, we choose the martingales X^i for $i = 3k - 1$ to be continuous, for $i = 3k$ to be purely discontinuous and quasi-left-continuous, and for $i = 3k + 1$ to be purely discontinuous and accessible. Let B be some predictable increasing process with respect to which X^1 and $\langle X^i, X^i \rangle$ are absolutely continuous for all $i \geq 2$. The main result of this Chapter will be formulated just in these terms.

9.2 Formulation and Proof of the Main Result

Now let $\check{\Omega}$ be the set of all R^d-valued functions on R_+, $\check{\mathcal{F}}$ be its Borel σ-algebra for the Tikhonov topology, $\check{\mathbf{F}}$ be the filtration of σ-algebras $\check{\mathcal{F}}_t$ for $t \in \mathsf{R}_+$ each of which is the intersection for $s > t$ of sub-σ-algebras of $\check{\mathcal{F}}$ generated by restrictions to $[0, s]$ of functions from $\check{\Omega}$. Then, let X be the canonical process on $\check{\Omega}$, i.e. $X_t(\check{\omega}) = \check{\omega}_t$. Let also $\bar{\Omega} = \Omega \times \check{\Omega}$, $\bar{\mathcal{F}} = \mathcal{F} \otimes \check{\mathcal{F}}$, $\bar{\mathcal{F}}_t = \bigcap_{s>t}(\mathcal{F}_s \otimes \check{\mathcal{F}}_s)$, and $\bar{\mathbf{F}} = (\bar{\mathcal{F}}_t)_{t \in \mathsf{R}_+}$.

Now, let us consider a $\mathcal{P}(\bar{\mathbf{F}}) \otimes \mathcal{E}$-measurable d-vector function h on $\bar{\Omega} \times E$ and the equation

$$X = N + h * \theta \tag{9.4}$$

for $\theta \in \tilde{\mathcal{S}}^0_\sigma$, where N is a given R^d-valued $\mathbf{F}$-progressively measurable process playing the role of an initial condition. Then analogous to Lebedev (1983), Lebedev (1996) or Jacod and Mémin (1981), we can define a weak solution of (9.4). Namely, a *solution-measure* (or a *weak solution*) of (9.4) is a probability measure $\bar{\mathrm{P}}$ on $(\bar{\Omega}, \bar{\mathcal{F}})$ such that its Ω-marginal $\bar{\mathrm{P}}|_\Omega$ is equal to P, Eq. (9.4) keeps its sense on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{F}}, \bar{\mathrm{P}})$, and the canonical process X, being substituted into h instead of $\check{\omega} \in \check{\Omega}$, is a solution-process of (9.4). Keeping the sense by Eq. (9.4) means holding the following two conditions:

(a) θ admits an extension to $(\bar{\Omega} \times \mathsf{R}_+ \times E, \mathcal{P}(\bar{\mathbf{F}}) \otimes \mathcal{E})$ as a σ-finite $L^0(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathrm{P}})$-valued random measure, i.e., there exists a strictly positive $\mathcal{P} \otimes \mathcal{E}$-measurable function V on $\Omega \times \mathsf{R}_+ \times E$ belonging to $L^{1,0}(\theta, \bar{\mathbf{F}})$ and such that for all $\mathcal{P} \otimes \mathcal{E}$-measurable φ on $\Omega \times \mathsf{R}_+ \times E$ with $|\varphi| \leq V$ the stochastic integral processes $\varphi * \theta$ on $\mathbf{F}$ and $\bar{\mathbf{F}}$ coincide;

(b) if a $\mathcal{P} \otimes \mathcal{E}$-measurable function φ on $\Omega \times \mathsf{R}_+ \times E$ belongs to $\hat{L}^0(\theta, \bar{\mathbf{F}})$, then it belongs also to $\hat{L}^0(\theta, \mathbf{F})$ and besides the stochastic integral processes $\varphi * \theta$ on $\mathbf{F}$ and $\bar{\mathbf{F}}$ also coincide.

According to Theorem 3 of Lebedev (1995), Condition (b) holds in particular when $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{F}}, \bar{\mathrm{P}})$ is a very good extension of the stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$, i.e., when any $\mathbf{F}$-martingale is a $\bar{\mathbf{F}}$-martingale. In this case the solution-measure itself is also said to be *very good.*

Let us formulate the main result of this Chapter.

Theorem 9.2.1 *Let for $\theta \in \tilde{S}^2_\sigma$ and for Eq.* (9.4) *the following assumptions hold:*

(1)

$$\left| \int_E h\,\rho^1(du) \right| \frac{dX^1}{dB} + \sum_{i=2}^{\infty} \left| \int_E h\,\rho^i(du) \right|^2 \frac{d\langle X^i, X^i \rangle}{dB} \leq c$$

$\mathrm{P} \times dB$-a.e. on $\Omega \times \mathsf{R}_+$ at all $\check{\omega} \in \check{\Omega}$ for a $\mathcal{P}$-measurable process c with a P-a.s. finite for any $t \in \mathsf{R}_+$ integral $c \cdot B_t$;

(2) *the functions $\int_E h\,\rho^i(du)$ are continuous in $\check{\omega} \in \check{\Omega}$ for the U-topology almost everywhere in the measures $\mathrm{P} \times dX^1$ and $\mathrm{P} \times d\langle X^i, X^i \rangle$ for $i \geq 2$ on $\Omega \times \mathsf{R}_+$ respectively, and $\mathrm{P} \times dB$-a.e. on $\Omega \times \mathsf{R}_+$ the series*

$$\sum_{k=1}^{\infty} \left| \int_E h\,\rho^{3k-1}(du) \right|^2 \frac{d\langle X^{3k-1}, X^{3k-1} \rangle}{dB}$$

and

$$\sum_{k=1}^{\infty} \left| \int_E h\,\rho^{3k}(du) \right|^2 \frac{d\langle X^{3k}, X^{3k} \rangle}{dB}$$

converge uniformly in $\check{\omega}$ on each subset of $\check{\Omega}$ of the form $N(\omega) + K$, where K is compact respectively for the U- and J_1-topology respectively.

Then for Eq. (9.4)*, there exists a very good solution-measure.*

Let us note that for this solution-measure $\bar{\mathrm{P}}$ the extension of the measure θ to $(\bar{\Omega} \times \mathsf{R}_+ \times E, \mathcal{P}(\bar{\mathbf{F}}) \otimes \mathcal{E})$ by Theorem 3 of Lebedev (1995) is also an L^2-valued random measure.

The proof of the theorem is carried out quite analogously to Theorem 9.20 of Lebedev (1996) or Theorem 1 of Lebedev (1983). First of all, analogous to Lemma 9.22 of Lebedev (1996) or Lemma 1 of Lebedev (1983) the process N can be eliminated from Eq. (9.4). The following generalization of Lemma 9.24 of Lebedev (1996) or Lemma 3 of Lebedev (1983) is less trivial.

Lemma 9.2.1 *Let there exist a very good solution-measure of Eq.* (9.4) *with* $N \equiv 0$ *and* $\theta \in \tilde{\mathcal{S}}^2_\sigma$ *under the additional assumption that the* σ*-algebra* $\mathcal{F}$ *is separable. Then, it exists without this assumption.*

The proof of Lemma 9.2.1 uses the following strengthening of Theorem 2.55 of Lebedev (1996) or Lemma 2 of Lebedev (1983).

Lemma 9.2.2 *Let* $(E, \mathcal{E})$ *be a measurable space and* f *be a* $\mathcal{P}(\mathbf{F})\otimes\mathcal{E}$*- or* $\mathcal{O}(\mathbf{F})\otimes\mathcal{E}$*-measurable function on* $\Omega \times \mathsf{R}_+ \times E$ *taking values in a separable metric space* S*. Then, there exists a separable* σ*-algebra* $\mathcal{G} \subset \mathcal{F}$ *such that*

(a) f *is* $\mathcal{P}(\mathbf{G}) \otimes \mathcal{E}$*- or* $\mathcal{O}(\mathbf{G}) \otimes \mathcal{E}$*-measurable respectively, where* $\mathbf{G} = (\mathcal{G} \cap \mathcal{F}_t)_{t\in\mathsf{R}_+}$*;*

(b) *any* $\mathbf{G}$*-martingale is an* $\mathbf{F}$*-martingale [that is, the stochastic basis* $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ *is a very good extension of* $(\Omega, \mathcal{F}, \mathbf{G}, \mathrm{P})$*].*

PROOF. (a) is proved quite similarly to Theorem 2.55 of Lebedev (1996) or Lemma 2 of Lebedev (1983). Now, let $\mathcal{G}_{(0)}$ be a separable σ-algebra with which instead of $\mathcal{G}$ (a) is satisfied, $\mathcal{A} = (A_n)_{n\in\mathsf{N}}$ be a countable algebra generating the σ-algebra $\mathcal{G}_{(0)}$, and $\mathcal{G}_{(1)}$ be the σ-algebra generated by $\mathcal{G}_{(0)}$ and right-continuous martingales $(\mathrm{P}(A_n|\mathcal{F}_t))_{t\in\mathsf{R}_+}$ for $A_n \in \mathcal{A}$. Since the σ-algebra $\mathcal{G}_{(0)}$ is separable and the martingales $(\mathrm{P}(A_n|\mathcal{F}_t))$ are determined completely by their values at rational t, the σ-algebra $\mathcal{G}_{(1)}$ is also separable. Now, let $\mathcal{G}_{(2)}$ be the σ-algebra obtained from $\mathcal{G}_{(1)}$ similar to $\mathcal{G}_{(1)}$ from $\mathcal{G}_{(0)}$, and so on, and $\mathcal{G} = \bigvee_{n=0}^{\infty} \mathcal{G}_{(n)}$. Property (a) is obviously preserved under extension of the σ-algebra $\mathcal{G}$, and since $\mathcal{G}_{(0)} \subset \mathcal{G}$ it holds for the given $\mathcal{G}$. Now it is sufficient to prove that any bounded $\mathbf{G}$-martingale is an $\mathbf{F}$-martingale. Let M be an arbitrary right-continuous bounded $\mathbf{G}$-martingale. Then there exists P-a.s. $M_\infty = \lim_{t\to\infty} M_t$ and the random variable M_∞ can be chosen bounded and $\mathcal{G}$-measurable. To show that M is an $\mathbf{F}$-martingale, it suffices to verify that for every $t \in \mathsf{R}_+$ the random variable $\mathrm{E}(M_\infty|\mathcal{F}_t)$ is $\mathcal{G}$-measurable upto P-null sets. Let $\mathcal{H}$ be the set of bounded $\mathcal{G}$-measurable random variables X for which $\mathrm{E}(X|\mathcal{F}_t)$ are $\mathcal{G}$-measurable upto P-null sets for all $t \in \mathsf{R}_+$. Then $\mathcal{H}$ is obviously linear, closed under the uniform and the bounded monotone convergences and contains all random variables of the form 1_A for $A \in \mathcal{G}_{(n)}$, $n = 0, 1, 2, \ldots$. Let $\mathcal{M}$ consist of such random variables. Since $\mathcal{G}_{(m)} \subset \mathcal{G}_{(n)}$ for $m \leq n$, the set $\mathcal{M}$ is closed under multiplication, generating also the σ-algebra $\mathcal{G}$. Hence, by the monotone class theorem $\mathcal{H}$ contains all $\mathcal{G}$-measurable variables which gives the required result. ∎

PROOF OF LEMMA 9.2.1. First of all, we apply Lemma 9.2.2 to a strictly positive $\mathcal{P}(\mathbf{F}) \otimes \mathcal{E}$-measurable function V, belonging to $L^{1,2}(\theta)$, to the $\mathcal{P}(\mathbf{F})$-measurable processes c, B, X^1, $\langle X^i, X^i \rangle$ for $i \geq 2$,

$$\int_E 1_{C_j \cap \Gamma_k} \rho^1(du) \cdot X^1$$

and

$$\int_E 1_{C_j \cap \Gamma_k} \rho^i(du) \cdot \langle X^i, X^i \rangle$$

for $i \geq 2$, $j, k \in \mathsf{N}$, where $C_1 = \{V > 1\}$ and for $j \geq 2$ $C_j = \{2^{-j+1} < V \leq 2^{-j+2}\}$, and $\Gamma_k \in \mathcal{E}$ constitute an algebra generating the σ-algebra $\mathcal{E}$, to the $\mathcal{O}(\mathbf{F})$-measurable processes X^i for $i \geq 2$, and also to the $\mathcal{P}(\mathbf{F}) \otimes \check{\mathcal{F}} \otimes \mathcal{E}$-measurable function h. Let $\mathcal{G} \subset \mathcal{F}$ be a separable σ-algebra such that all enumerated functions are adapted properly to the filtration $\mathbf{G} = (\mathcal{G}_t)_{t \in \mathsf{R}_+}$ with $\mathcal{G}_t = \mathcal{G} \cap \mathcal{F}_t$ and $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$ is a very good extension of $(\Omega, \mathcal{F}, \mathbf{G}, \mathrm{P})$. Then the function V is $\mathcal{P}(\mathbf{G}) \otimes \mathcal{E})$-measurable so that Eq. (9.4) keeps its sense under passage from $\mathbf{F}$ to $\mathbf{G}$, and the decomposition (9.3) for $\varphi \in \hat{L}^2(\theta, \mathbf{G})$ remains such also with respect to $\mathbf{G}$ so that the conditions of the theorem also keep their sense with respect to $\mathbf{G}$.

By hypothesis, there exists a probability measure $\bar{\mathrm{P}}$ on $(\bar{\Omega}, \bar{\mathcal{G}})$ with $\bar{\mathcal{G}} = \mathcal{G} \otimes \check{\mathcal{F}}$ which is a very good weak solution of Eq. (9.4) with $N \equiv 0$ with respect to $\mathbf{G}$, and θ admits an extension to $(\bar{\Omega} \times \mathsf{R}_+ \times E, \mathcal{P}(\bar{\mathbf{G}}) \otimes \mathcal{E})$ as a σ-finite $L^2(\bar{\Omega}, \bar{\mathcal{G}}, \bar{\mathrm{P}})$-valued random measure. This means that the equality

$$X = h(X) * \theta, \tag{9.5}$$

where the stochastic integral process is taken on the filtration $\bar{\mathbf{G}}$ constructed from $\mathbf{G}$ similarly to $\bar{\mathbf{F}}$ from $\mathbf{F}$, is valid up to $\bar{\mathrm{P}}$-indistinguishability and the measure $\bar{\mathrm{P}}$ admits the factorization

$$\bar{\mathrm{P}}(d\omega \times d\check{\omega}) = \mathrm{P}(d\omega)\, \mathrm{Q}(\omega, d\check{\omega}) \tag{9.6}$$

with a regular transition measure Q from $(\Omega, \mathcal{G})$ to $(\check{\Omega}, \check{\mathcal{F}})$, for every $t \in \mathsf{R}_+$ and $F \in \check{\mathcal{F}}_t$ the function $\mathrm{Q}(\cdot, F)$ being $\mathcal{G}_t$-measurable up to P-null sets from $\mathcal{G}$.

Now we construct the measure $\bar{\mathrm{P}}$ on $(\bar{\Omega}, \bar{\mathcal{F}})$ with the factorization (9.6) and with the measure P on $\mathcal{F}$ instead of $\mathcal{G}$. It remains be to proved that the measure θ admits an extension to $(\bar{\Omega} \times \mathsf{R}_+ \times E, \mathcal{P}(\bar{\mathbf{F}}) \otimes \mathcal{E})$ and that the stochastic integral process on $\bar{\mathbf{G}}$ in the right-hand member of (9.5) remains such on $\bar{\mathbf{F}}$. First of all, for $\varphi \in \hat{L}^2(\theta, \mathbf{F})$ by Theorem 1 of Lebedev (1995), we have the decomposition (9.3), where the series converges unconditionally in $\mathcal{S}^2(\mathbf{F})$ and, in particular, for $\varphi \in \hat{L}^2(\theta, \mathbf{G})$ in $\mathcal{S}^2(\mathbf{G})$. For a $\mathcal{P}(\bar{\mathbf{F}}) \otimes \mathcal{E}$-measurable function φ on $\bar{\Omega} \times \mathsf{R}_+ \times E$ with $|\varphi| \leq V$ we define $\varphi * \theta$ by formula (9.3) and prove that the series in (9.3) converges unconditionally in $\mathcal{S}^2(\bar{\mathbf{F}})$. At first, we can show that by regularity (the very good property) of the passage from $\mathbf{F}$ to $\bar{\mathbf{F}}$ the measure

θ admits an extension to $\bar{\mathbf{F}}$ and, in particular, to $\bar{\mathbf{G}}$ as a σ-finite $L^2(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathrm{P}})$-valued random measure, as such it is extended uniquely, and $V \in L^{1,2}(\bar{\mathbf{F}})$. By applying Theorem 1 of Lebedev (1995) on $\bar{\mathbf{F}}$ for φ with $|\varphi| \le V$, we obtain the decomposition

$$\varphi * \theta = \sum_{i=1}^{\infty} \int_E \varphi\, \rho_s^i(\omega, du) \cdot X^i + \sum_{j=1}^{\infty} \int_E \varphi\, \tilde{\rho}_s^j(\omega, du) \cdot \tilde{X}^j, \tag{9.7}$$

where $\tilde{X}^j$ are martingales on $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbf{F}}, \bar{\mathrm{P}})$ orthogonal one to another and to X^i for $i \ge 2$, and $\tilde{\rho}^j$ are the corresponding transition measures, both series converging unconditionally in $\mathcal{S}^2(\bar{\mathbf{F}})$. But the first series in the right-hand member of (9.3) gives already an extension of the measure θ to $\bar{\mathbf{F}}$ and by its uniqueness the second series in (9.7) is equal to 0 identically up to $\bar{\mathrm{P}}$-indistinguishability. Besides that, the integral in the right-hand member of (9.5) has the same decomposition of the form (9.3) on $\bar{\mathbf{G}}$ and $\bar{\mathbf{F}}$. The lemma has been proved. ■

So let $N \equiv 0$ and the σ-algebra $\mathcal{F}$ be separable. Let us introduce $\mathcal{P}(\bar{\mathbf{F}}) \otimes \mathcal{E}$-measurable d-vector functions $h_{(n)}$ on $\bar{\Omega} \times E$ by the formula

$$h_{(n)t}(\omega, \breve{\omega}) = h_t(\omega, \breve{\omega}^{t_n}),$$

where $t_n = k/n$ if $k/n < t \le (k+1)/n$, and $t_n = 0$ if $t = 0$. Then analogous to Lemma 9.26 of Lebedev (1996) or the corresponding fragment of the proof of Theorem 1 of Lebedev (1983), for each $n \in \mathsf{N}$ the equation

$$X = h_{(n)} * \theta$$

has a unique solution-process $X_{(n)}$ on the original stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$. Now let

$$\begin{aligned} X^v_{(n)} &= \int_E h_{(n)}(X_{(n)})\, \rho^1(du) \cdot X^1, \\ X^c_{(n)} &= \textstyle\sum_{i=1}^\infty \int_E h_{(n)}(X_{(n)})\, \rho^{3i-1}(du) \cdot X^{3i-1}, \\ X^q_{(n)} &= \textstyle\sum_{i=1}^\infty \int_E h_{(n)}(X_{(n)})\, \rho^{3i}(du) \cdot X^{3i}, \\ X^j_{(n)} &= \textstyle\sum_{i=1}^\infty \int_E h_{(n)}(X_{(n)})\, \rho^{3i+1}(du) \cdot X^{3i+1}. \end{aligned}$$

Then analogous to Lemma 9.28 of Lebedev (1996) or the corresponding fragment of the proof of Theorem 1 of Lebedev (1983), Condition (1) of the theorem ensures the tightness of the sequence of distributions on $D_{[0,\infty[}(\mathsf{R}^{4d})$ with the Skorokhod J_1-topology of processes $R_{(n)} = (X^v_{(n)}, X^c_{(n)}, X^q_{(n)}, X^j_{(n)})$, and the condition of uniform convergence of the corresponding series on a J_1-compact set ensures also the tight majorization of jumps for the sequence $(R_{(n)})_{n \in \mathsf{N}}$.

Now let η be a random element of some compact metric space generating the σ-algebra $\mathcal{F}$ and now we apply the generalized Skorokhod theorem for the weak-strong convergence of probability measures [Theorem 5.13 with account of Remark 5.15 of Lebedev (1996), or Lemma 4 of Lebedev (1983)] to some subsequence extracted from the sequence $(\eta, R_{(n)})$. Let $(\eta^{(0)}, R)$ be the corresponding limiting random element on a probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathrm{P}})$, $\mathcal{G}_t$ for $t \in \mathsf{R}_+$ be the σ-algebra generated by random variables $\varphi(\eta^{(0)})$ for which $\varphi(\eta)$ are $\mathcal{F}_t$-measurable, $\bar{\mathcal{G}}_t$ be the intersection for $s > t$ of σ-algebras generated by $\mathcal{G}_s$ and R_r at $0 \le r \le s$, $\mathbf{G} = (\mathcal{G}_t)_{t \in \mathsf{R}_+}$ and $\bar{\mathbf{G}} = (\bar{\mathcal{G}}_t)_{t \in \mathsf{R}_+}$. Let $R = (X^v, X^c, X^q, X^j)$ and $X = X^v + X^c + X^q + X^j$. We must prove that any $\mathbf{G}$-martingale is a $\bar{\mathbf{G}}$-martingale [i.e., the stochastic basis $(\tilde{\Omega}, \tilde{\mathcal{F}}, \bar{\mathbf{G}}, \tilde{\mathrm{P}})$ is a very good extension for $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbf{G}, \tilde{\mathrm{P}})$] and

$$X = h(\eta^{(0)}, X) * \theta(\eta^{(0)})$$

or,

$$X = \sum_{i=1}^{\infty} \int_E h(\eta^{(0)}, X)\, \rho^i(\eta^{(0)}, du) \cdot X^i(\eta^{(0)}), \tag{9.8}$$

the series converging unconditionally in $\mathcal{S}^2(\bar{\mathbf{G}})$ under the proper $\mathbf{G}$-localization.

The $\mathbf{G}$-regularity in this sense of the stochastic basis $(\tilde{\Omega}, \tilde{\mathcal{F}}, \bar{\mathbf{G}}, \tilde{\mathrm{P}})$ can be proved as in Theorem 9.20 of Lebedev (1996) or in Theorem 1 of Lebedev (1983) without any changes. We prove analogously [which is also analogous to limit relations for stochastic integrals in § 5 of Chapter IX of Jacod and Shiryaev (1987)] that

$$X^v = \int_E h(\eta^{(0)}, X)\, \rho^1(\eta^{(0)}, du) \cdot X^1(\eta^{(0)}),$$

that X^c, X^q and X^j are locally square-integrable martingales on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \bar{\mathbf{G}}, \tilde{\mathrm{P}})$ and that for all $i \in \mathsf{N}$ and $k \in \mathsf{N}$

$$\begin{aligned}&\langle X^c, X^{3i-1}(\eta^{(0)})\rangle\\ &\quad= \int_E h(\eta^{(0)}, X)\, \rho^{3i-1}(\eta^{(0)}, du) \cdot \langle X^{3i-1}(\eta^{(0)}), X^{3i-1}(\eta^{(0)})\rangle,\end{aligned}$$

$$\begin{aligned}&\langle X^q, X^{3i}(\eta^{(0)})\rangle\\ &\quad= \int_E h(\eta^{(0)}, X)\, \rho^{3i}(\eta^{(0)}, du) \cdot \langle X^{3i}(\eta^{(0)}), X^{3i}(\eta^{(0)})\rangle,\end{aligned}$$

$$\begin{aligned}&\langle X^j, X^{3i+1}(\eta^{(0)})\rangle\\ &\quad= \int_E h(\eta^{(0)}, X)\, \rho^{3i+1}(\eta^{(0)}, du) \cdot \langle X^{3i+1}(\eta^{(0)}), X^{3i+1}(\eta^{(0)})\rangle,\end{aligned}$$

$\langle X^c, X^k(\eta^{(0)})\rangle \equiv 0$, $\langle X^q, X^k(\eta^{(0)})\rangle \equiv 0$ and $\langle X^j, X^k(\eta^{(0)})\rangle \equiv 0$, respectively, for $k \ne 3i-1$, $k \ne 3i$ and $k \ne 3i+1$ with some $i \in \mathsf{N}$. Hence, we conclude that

$$\begin{aligned}X^c &= \textstyle\sum_{i=1}^{\infty} \int_E h(\eta^{(0)}, X)\, \rho^{3i-1}(\eta^{(0)}, du) \cdot X^{3i-1}(\eta^{(0)}) + \tilde{X}^c,\\ X^q &= \textstyle\sum_{i=1}^{\infty} \int_E h(\eta^{(0)}, X)\, \rho^{3i}(\eta^{(0)}, du) \cdot X^{3i}(\eta^{(0)}) + \tilde{X}^q,\\ X^j &= \textstyle\sum_{i=1}^{\infty} \int_E h(\eta^{(0)}, X)\, \rho^{3i+1}(\eta^{(0)}, du) \cdot X^{3i+1}(\eta^{(0)}) + \tilde{X}^j,\end{aligned}$$

where $\tilde{X}^c$, $\tilde{X}^q$ and $\tilde{X}^j$ are locally square-integrable martingales on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \bar{\mathbf{G}}, \tilde{\mathrm{P}})$ orthogonal to every $X^i(\eta^{(0)})$ for $i \geq 2$, all series converging unconditionally in $\mathcal{S}^2(\bar{\mathbf{G}})$ under the proper $\mathbf{G}$-localization. From the continuity of functions

$$\sum_{i=1}^{\infty} \left| \int_E h\, \rho^{3i-1}(du) \right|^2 \frac{d\langle X^i, X^i \rangle}{dB}$$

and

$$\sum_{i=1}^{\infty} \left| \int_E h\, \rho^{3i}(du) \right|^2 \frac{d\langle X^i, X^i \rangle}{dB}$$

in $\check{\omega} \in \check{\Omega}$ for the U-topology following from Condition (2) of the theorem, we obtain that $\tilde{X}^c \equiv 0$ and $\tilde{X}^q \equiv 0$. Moreover, increments of $\langle X^j, X^j \rangle$ are concentrated on the union of graphs of $\mathbf{G}$-predictable stopping times exhausting jumps of $\langle X^{3i+1}, X^{3i+1} \rangle$ for all $n \in \mathbb{N}$ but $\tilde{X}^j$ has no jumps on this set, and hence $\tilde{X}^j \equiv 0$. Thus, (9.8) holds as required. So, we have constructed the solution-process X of Eq. (9.4) on the very good extension $(\tilde{\Omega}, \tilde{\mathcal{F}}, \bar{\mathbf{G}}, \tilde{\mathrm{P}})$ of the stochastic basis $(\tilde{\Omega}, \tilde{\mathcal{F}}, \mathbf{G}, \tilde{\mathrm{P}})$ isomorphic to the original $(\Omega, \mathcal{F}, \mathbf{F}, \mathrm{P})$, which is equivalent analogously to Proposition 9.16 of Lebedev (1996) or Theorem 2.18 of Jacod and Mémin (1981) to the existence of a very good solution-measure of Eq. (9.4). ■

References

1. Bichteler, K. (1981). Stochastic integration and L^p-theory of semimartingales, *Annals of Probability*, **9**, 49–89.

2. Bichteler, K. and Jacod, J. (1983). Random measures and stochastic integration, *Lecture Notes in Control and Information Sciences*, **49**, 1–18.

3. Emery, M. (1979). Une topologie sur l'espace des semimartingales, *Lecture Notes in Mathematics*, **721**, 260–280.

4. Jacod, J. and Mémin, J. (1981). Weak and strong solutions of stochastic differential equations: existence and stability, *Lecture Notes in Mathematics*, **851**, 169–212.

5. Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*, Berlin: Springer-Verlag.

6. Lebedev, V. A. (1983). On the existence of weak solutions for stochastic differential equations with driving martingales and random measures, *Stochastics*, **9**, 37–76.

7. Lebedev, V. A. (1995). Behavior of random measures under a change of filtration, *Theory of Probability and its Applications*, **40**, 645–652.

8. Lebedev, V. A. (1996). *Martingales, Convergence of Probability Measures and Stochastic Equations*, Moscow: MAI Publishing House (in Russian).

10

Tightness of Stochastic Families Arising From Randomization Procedures

Mikhail Lifshits and Michel Weber

St. Petersburg State University, St. Petersburg, Russia
Université Louis-Pasteur, Strasbourg, France

Abstract: We consider the laws of Gaussian random elements arising from randomization procedures in ergodic theory and real analysis. We find sufficient conditions for the tightness of the corresponding families in the spaces $C[0,1]$ and $L^p[0,1]$ and demonstrate some crucial situations where tightness does not take place.

Keywords and phrases: Gaussian random functions, measures in functional spaces, randomization procedure, tightness

10.1 Introduction

Let $(X, \mathcal{A}, \mu)$ be a measure space with $\mu(X) = 1$, endowed with an ergodic measure-preserving transformation $T : X \to X$. Let also $\{\xi_j,\ j \in \mathbf{N}\}$ be a standard Gaussian i.i.d. sequence defined on another probability space $(\Omega, \mathcal{B}, P)$. To each element $f \in L^p(\mu)$, we associate the following sequence of $L^p(\mu)$-valued random elements defined on Ω:

$$\forall J \geq 1,\ \forall(\omega, x) \in \Omega \times X,\ F^0_{J,f}(\omega, x) = \frac{1}{\sqrt{J}} \sum_{j \leq J} \xi_j(\omega) f \circ T^j(x).$$

One should regard this object as a randomized version of classical averages, $J^{-1} \sum_{j \leq J} f \circ T^j(x)$. The similar elements were introduced by Stein (1961, Theorem 1) in the study of the continuity principle for ergodic transformations where they played a key role in the probabilistic proof of his main result (to be precise, Stein used Rademacher sequence as ξ_j). More recently, by combining this randomization technique with the theory of Gaussian processes, Bourgain

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

(1988, Propositions 1 and 2) discovered a remarkable and useful entropy criterion for the study of the almost everywhere convergence in ergodic theory and related problems of real analysis. He considered more general randomizations

$$F_{J,f}(\omega,x)=\frac{1}{\sqrt{J}}\sum_{j\le J}\xi_j(\omega)(S_jf)(x),$$

related to different sequences of linear operators S_j.

The tightness properties of the laws of the family $\{F_{J,f}, J\in\mathbf{N}\}$ led to various refinements of Bourgain's entropy criterion in Weber (1994, 1996). We refer the interested reader to the recent monograph of Weber (1998) containing detailed exposition of applications of $F_{J,f}$.

Although the role of Gaussian elements $\{F_{J,f}\}$ in ergodic theory is now self-evident, they constitute a remarkable class of Gaussian random functions independently of this kind of applications. In our opinion, the study of their behavior may successfully contribute to the general theory of Gaussian random functions. In Lifshits and Weber (1998), we adopted a "J-trajectory approach" (with fixed $x\in X$ and integer variable J) and studied the oscillations of the corresponding random sequences.

Inspired with several most interesting examples in Bourgain (1988) where the sequences of shift operators S_j appear in the definition of $F_{J,f}$, we consider in this work the families of random elements

$$F_{J,f,\lambda}(x)=J^{-1/2}\sum_{j=1}^{J}f(x+\lambda_{J,j})\xi_j$$

in the space $L^p[0,1]$ (or $C[0,1]$) for $f\in L^p[0,1]$ (or $C[0,1]$ according to the context), $\lambda\in\Lambda$, where Λ is the class of all triangular arrays taking values in $[0,1]$, $J\in N$ and ξ_j a standard Gaussian i.i.d. sequence. We investigate the tightness of the laws of these elements in the corresponding spaces (in what follows, we identify the tightness of the family of random elements of vector space with the tightness of the family of the corresponding laws).

There are two types of arrays of special interest coming from Bourgain (1988): the sequences $\lambda_{J,j}=\lambda_j$ and the array corresponding to randomized Riemann sum $\lambda_{J,j}=j/J$.

We consider $[0,1]$ as a circle equipped with the structure of additive group. This factorization enables to treat $f(x+\lambda_{J,j})$ properly for all $\lambda_{J,j}$ and x.

Denote $\|\cdot\|_p$ the L^p-norm for $f\in L^p$ and let $\|\cdot\|$ denote L^p-norm or $C[0,1]$-norm according to the context. Define for $f\in L^p$ the modulus of continuity

$$\omega_f(u)=\sup_{0\le h\le u}\|f(\cdot+h)-f(\cdot)\|_p.$$

The modulus of continuity of the function $f\in C[0,1]$ coincides with that of the space $L^\infty[0,1]$.

10.2 Sufficient Condition of Tightness in $C[0,1]$

We prove the following result.

Proposition 10.2.1 *Let f be a continuous 1-periodic function. Assume that its modulus of continuity $\omega_f(\cdot)$ satisfies*

$$\int_0 \omega_f(u)/\left(u\sqrt{|\log u|}\right) du < \infty.$$

Let ξ_j be a standard Gaussian i.i.d. sequence. Then the family of the processes

$$\Phi_f = \left\{ F_J(x) = J^{-1/2}\sum_{j=1}^{J} \xi_j f(x+\lambda_{J,j}),\ J \in N, \lambda \in \Lambda \right\}$$

is tight in the space $C[0,1]$.

Remark 10.2.1 Moreover, if $\lambda_{J,j} = \lambda_j \to 0$ and $J\to\infty$, this family converges to the law of the degenerated process $\xi_1 f(\cdot)$.

PROOF OF THE PROPOSITION 10.2.1. For all $J \in N,\ x,y \in [0,1]$, we have

$$\begin{aligned}\mathbf{Var}_J(x,y) &= \mathbf{Var}(F_J(x)-F_J(y)) = J^{-1}\sum_{j=1}^{J}[f(x+\lambda_{J,j})-f(y+\lambda_{J,j})]^2 \\ &\le \sup_j [f(x+\lambda_{J,j})-f(y+\lambda_{J,j})]^2 \le \omega_f(|x-y|)^2.\end{aligned}$$

Hence,

$$\mathbf{Var}_J(x,y) \le \omega_f(|x-y|)^2.$$

We deduce now the estimate for Dudley integral [see Lifshits (1995, Section 15) for the definition; one could also use Fernique integral] which will be uniform over J and λ. Indeed, for each $r > 0$ and each J the intervals of length $\omega_f^{-1}(r)$ form a covering of $[0,1]$ by the sets of diameter not exceeding r with respect to the metric generated by the process F_J. Since the number of intervals is $1/\omega_f^{-1}(r)$, the Dudley integral $\psi_J(R)$ admits the estimate

$$\psi_J(R) \le \int_0^R \sqrt{\log(1/\omega_f^{-1}(r))}dr = \int_0^R \sqrt{|\log \omega_f^{-1}(r)|}dr.$$

By the variable change $r = \omega_f(u)$ and integration by parts, we obtain

$$\begin{aligned}\psi_J(R) &\le \int_0^{\omega_f^{-1}(R)} \sqrt{|\log(u)|}d\omega_f(u) \\ &= \omega_f(u)\sqrt{|\log(u)|}\ \Big|_0^{\omega_f^{-1}(R)} + \int_0^{\omega_f^{-1}(R)} \frac{\omega_f(u)}{2u\sqrt{|\log(u)|}}du.\end{aligned}$$

Moreover, the integral term is dominating, since the function ω_f is monotone and we have for each u

$$\omega_f(u)\sqrt{|\log(u)|} \leq 2\int_u^{u^{1/2}} \frac{\omega_f(v)}{v\sqrt{|\log(v)|}}dv.$$

Letting $u = \omega_f^{-1}(R)$, we obtain

$$\psi_J(R) \leq 2\int_0^{\omega_f^{-1}(R)^{1/2}} \frac{\omega_f(u)}{u\sqrt{|\log(u)|}}du \to 0, \quad (R \to 0).$$

Since the obtained estimate of $\psi_J(R)$ is J-uniform, the tightness easily follows via classical estimates of modulus of continuity using $\psi_J(R)$; see, for example, Lifshits (1995, Theorem 15.1, p. 216). ■

10.3 Continuous Generalization

We can transform the parameter λ from the statement of Proposition 10.2.1 in a continuous object. Let $\mathcal{M}$ denote the class of probabilistic measures on $[0,1]$. For each measure $\mu \in \mathcal{M}$, let W_μ denote the white noise with variance μ, and define a random function

$$F_{f,\mu}(x) = \int_0^1 f(x+\lambda)W_\mu(d\lambda).$$

Then the family

$$\Phi_f^{\mathcal{M}} = \{F_{f,\mu}, \mu \in \mathcal{M}\}$$

is tight in $C[0,1]$ since, for each $\mu \in \mathcal{M}$ and $x, y \in [0,1]$,

$$\begin{aligned}\mathbf{Var}\Big(F_{f,\mu}(x) - F_{f,\mu}(y)\Big) &= \mathbf{Var}\int_0^1 \Big(f(x+\lambda) - f(y+\lambda)\Big)W_\mu(d\lambda)\\ &= \int_0^1 |f(x+\lambda) - f(y+\lambda)|^2\mu(d\lambda) \leq \omega_f(|x-y|)^2,\end{aligned}$$

and the arguments of Proposition 10.2.1 work without further changes. Recall that Φ_f is a part of $\Phi_f^{\mathcal{M}}$, corresponding to the measures $\mu = J^{-1}\sum_{j=1}^J \delta_{\lambda_{J,j}}$ for integer J and $\lambda \in \Lambda$.

10.4 An Example of Non-Tightness in $C[0,1]$

We start our construction with the following definition.

Definition 10.4.1 Let $\{d_j\}$ and $\{\lambda_j\}$ be two positive sequences. We call a continuous 1-periodic function f on $(-\infty, \infty)$ a function with (d, λ)-complete system of values if for each $J \in N$ and for each sequence $\{\varepsilon_j\} \in \{-1; +1\}^J$ there exists $x \in [0,1]$ such that $f(x + \lambda_j) = d_j \varepsilon_j, 1 \le j \le J$.

This property means that all possible combinations of values appear simultaneously.

Proposition 10.4.1 *Let $\{\lambda_j\}$ be arbitrary sequence and the sequence $\{d_j\}$ be such that*

$$\limsup_{J \to \infty} J^{-1/2} \sum_{j=1}^{J} d_j = \infty.$$

Let f be a function with (d, λ)-complete system of values. Let ξ_j be a standard Gaussian i.i.d. sequence. Then the set of processes

$$\Phi_{f,\lambda} = \left\{ F_J(x) = J^{-1/2} \sum_{j=1}^{J} f(x + \lambda_j) \xi_j, \quad J \ge 1 \right\}$$

is not tight in $C[0,1]$.

PROOF OF PROPOSITION 10.4.1. For all $J \in N$, $\omega \in \Omega$, we have

$$\begin{aligned} \|F_J(\cdot)\| &= \sup_{x \in [0,1]} J^{-1/2} \left| \sum_{j=1}^{J} f(x + \lambda_j) \xi_j \right| \\ &\ge J^{-1/2} \left| \sum_{j=1}^{J} f(x_* + \lambda_j) \xi_j \right|, \end{aligned}$$

where $x_* = x_*(\omega)$ is such that $f(x_* + \lambda_j) = d_j \; sign(\xi_j)$. Moreover,

$$J^{-1/2} \left| \sum_{j=1}^{J} f(x_* + \lambda_j) \xi_j \right| = J^{-1/2} \sum_{j=1}^{J} d_j |\xi_j|$$

and hence

$$\mathbf{E} \|F_J(\cdot)\| \ge J^{-1/2} \sum_{j=1}^{J} d_j \mathbf{E} |\xi_j| = (2/\pi J)^{1/2} \sum_{j=1}^{J} d_j.$$

It follows that

$$\limsup_{J\to\infty} \mathbf{E}\|F_J(\cdot)\| = \infty.$$

■

The latter property is incompatible with tightness of Gaussian measures.

Proposition 10.4.2 *Let $\{\lambda_j = 2\cdot 8^{-j}\}$ and assume that the sequence $\{d_j\}$ satisfies*

$$\lim_{j\to\infty} d_j = 0.$$

Then there exists a function f with (d,λ)-complete system of values.

Corollary 10.4.1 *Let $d_j = j^{-1/3}$. Then, $\limsup_{J\to\infty} J^{-1/2}\sum_{j=1}^J d_j = \infty$ and $\lim_{j\to\infty} d_j = 0$. Assumptions of Propositions 10.4.1 and 10.4.2 are verified. Thus, we obtain a family F_j which is not tight.*

PROOF OF PROPOSITION 10.4.2. We base our construction on the continuous function $g : R^1 \to [-1,1]$ defined as follows. Let

(a) $g(t) = 0$ on $(-\infty, 1/8] \cup [4/8, 5/8] \cup [1,\infty)$;

(b) $g(t) = 1$ on $[2/8, 3/8]$ and $g(t) = -1$ over $[6/8, 7/8]$;

(c) g is linear on each of the remainder intervals $[1/8, 2/8]$, $[3/8, 4/8]$, $[5/8, 6/8]$, and $[7/8, 1]$.

For each $x \in R^1$, consider the expansion

$$x = z_x + \sum_{n=1}^{\infty} c_n(x) 8^{-n}, \qquad z_x \in Z,\ c_n(x) \in [0..7]\ .$$

Denote $|x| = 0$ for integer x and

$$|x| = \sup\{n : c_n(x) > 0\} \le \infty$$

for non-integer x.

Introduce a set of Cantor type

$$X = \Big\{x \in [0,1)\ :\ |x| < \infty,\ c_n(x) \in \{0\} \cup \{4\}\ \forall n\Big\}.$$

All possible combinations of values will appear on the elements of X.

Finally, define the function f on $[0,1)$ by

$$f(y) = \sum_{n=0}^{\infty} d_n \sum_{x\in X, |x|\le n} g_{x,n}(y)$$

with

$$g_{x,n}(y) = g(8^n(y-x))$$

and extend f periodically on R^1.

It is easy to verify that the support $S_{x,n}$ of $g_{x,n}$ is

$$\cdot \quad \Big\{y : c_j(y) = c_j(x) \in \{0\} \cup \{4\}\ j = 1, ..., n;\ c_{n+1}(y) \notin \{0\} \cup \{4\}\Big\}.$$

It follows that the supports of functions $g_{x,n}$ with different n are disjoint. In particular, this observation confirms that the series which defines f, is uniformly convergent.

It is also easy to see that for all $y \in S_{x,n}$ we have

$$g_{x,n}(y) = +1 \quad \text{if} \quad c_{n+1}(y) = 2$$

and

$$g_{x,n}(y) = -1 \quad \text{if} \quad c_{n+1}(y) = 6.$$

It follows that for each $j \geq 1$ and for each $x \in X$ we have

$$f(x + \lambda_j) = d_j g_{x_j, j-1}(x + \lambda_j)$$

where $x_j = \sum_{n=1}^{j-1} c_n(x) 8^{-n}$. We also obtain

$$f(x + \lambda_j) = d_j \quad \text{if} \quad c_j(x) = 0,$$

$$f(x + \lambda_j) = -d_j \quad \text{if} \quad c_j(x) = 4.$$

This is sufficient to derive that 2^J points $\{x \in X, |x| \leq J\}$ provide all possible combinations $\varepsilon_j d_j$, $\varepsilon_j = \pm 1$ of the values $f(x + \lambda_j)$. The system of values of f is therefore (d, λ)-complete. ∎

Remark 10.4.1 It is interesting to compare the variety of the values $f(x+\lambda_j)$ with identity $f(x) = 0$ that holds for all $x \in X$.

10.5 Sufficient Condition for Tightness in $L^p[0,1]$

We consider now the tightness of the same families of random elements $F_{J,f,\lambda}(x)$ in the space $L^p[0,1]$, for $f \in L^p[0,1], \lambda_j \in \Lambda$. We start from some general conditions providing tightness. For $\alpha, \beta \in [0,1]$, write $\alpha \leq \beta$ or $\alpha < \beta$ if $\beta - \alpha \in [0, 1/2)$ or $\beta - \alpha \in (0, 1/2)$, respectively. In these cases, we understand $[\alpha, \beta]$ as $\{x : \alpha \leq x \leq \beta\}$ etc.

Let $\|\cdot\|$ or $\|\cdot\|_p$ denote L^p-norm and for $f \in L^p$ let us use the notion of L^p-modulus of continuity $\omega_f(u)$ as stated in the introduction.

Recall some basic results about the relative compactness of sets in $L^p[0,1]$.

Theorem 10.5.1 *Let $p \in [1, \infty)$ and $\mathcal{F}$ be a subset of $L^p[0,1]$. Then $\mathcal{F}$ is relatively compact if and only if it is uniformly bounded, that is,*

$$\sup_{\mathcal{F}} \|f\| < \infty,$$

and uniformly continuous, that is,

$$\lim_{u \to 0} \sup_{\mathcal{F}} \omega_f(u) = 0.$$

This L^p-version Arzela-Ascoli theorem [see the proof in Dunford and Schwartz (1958, p. 298)] will not be applied directly, but it is useful for better understanding of the following criterion of the tightness of the family of the measures.

Theorem 10.5.2 *Let $p \in [1, \infty)$. A family Φ of random functions with sample paths in L^p is tight if and only if*

$$\lim_{M \to \infty} \sup_{F \in \Phi} P\{\|F\| > M\} = 0$$

and for each $\varepsilon > 0$

$$\lim_{u \to 0} \sup_{F \in \Phi} P\{\omega_F(u) > \varepsilon\} = 0.$$

This criterion yields the following simplified Gaussian version.

Theorem 10.5.3 *Let $p \in [1, \infty)$. A family Φ of centered Gaussian random functions is tight in L^p if*

$$\sup_{F \in \Phi} \mathbf{E}\|F\|^p < \infty \tag{10.1}$$

and

$$\lim_{u \to 0} \sup_{F \in \Phi} \mathbf{E}\,\omega_F(u)^p = 0. \tag{10.2}$$

In what follows, we always verify (10.1) and (10.2). In our special case, for (10.1) we have the estimate

$$\begin{aligned}
\mathbf{E}\|F_{J,f,\lambda}\|^p &= \mathbf{E}\int_0^1 |F_{J,f,\lambda}(x)|^p dx \\
&= \int_0^1 J^{-p/2}\mathbf{E}|\sum_{j=1}^{J} f(x+\lambda_{J,j})\xi_j|^p dx \\
&= c_p J^{-p/2} \int_0^1 [\sum_{j=1}^{J} f(x+\lambda_{J,j})^2]^{p/2} dx
\end{aligned}$$

with $c_p = \mathbf{E}|\xi_1|^p$.

It is good enough for $p \geq 2$, since discrete Hölder inequality yields

$$\sum_{j=1}^{J} f(x+\lambda_{J,j})^2 \leq \left[\sum_{j=1}^{J} |f(x+\lambda_{J,j})|^p\right]^{2/p} \left[\sum_{j=1}^{J} 1\right]^{1-2/p}$$

which implies

$$\left[\sum_{j=1}^{J} f(x+\lambda_{J,j})^2\right]^{p/2} \leq J^{p/2-1} \sum_{j=1}^{J} |f(x+\lambda_{J,j})|^p$$

and

$$\mathbf{E}\|F_{J,f,\lambda}\|^p \leq c_p J^{-1} \int_0^1 \sum_{j=1}^{J} |f(x+\lambda_{J,j})|^p dx = c_p \|f\|^p. \tag{10.3}$$

The latter inequality serves as a powerful instrument of "closure", i.e., for the passage from the "nice" functions to arbitrary ones.

In the case $p \in [1,2)$, we still have a Hölder estimate

$$\mathbf{E}\|F_{J,f,\lambda}\|^p \leq [\mathbf{E}\|F_{J,f,\lambda}\|^2]^{p/2} \leq \|f\|_2^p \tag{10.4}$$

which is not always efficient, especially for $f \in L^p \backslash L^2$. However, in certain situations, it is also useful (see a counter-example for L^p below).

Remark 10.5.1 The interested reader may compare Theorem 10.5.3 with alternative tightness criteria for L^p-spaces, [cf. Baushev (1987), Nguyen, Tarieladze, and Chobanyan (1978), and Suquet (1998).]

10.6 Indicator Functions

We show now that indicator functions $f_a = \mathbf{1}_{[0,a)}$ generate tight families in L^p, $1 \leq p < \infty$. Closing procedure will enable to extend this result on the class of arbitrary functions f in L^p, $2 \leq p < \infty$, while for $1 \leq p < 2$ the general result is false.

Theorem 10.6.1 *The family of random functions*

$$\Phi = \{F_{J,f_a,\lambda}, \quad a \in [0,1), \lambda \in \Lambda, J \in N\}$$

is tight in each L^p, $\quad p \in [1, \infty)$.

PROOF OF THEOREM 10.6.1. In order to keep transparent notations for intervals, we consider only the case $0 \le a < 1/2$; there is no loss of generality. We fix and abandon during this proof the indices J, a, λ for F. Our estimates will be uniform over these parameters.

We apply Theorem 10.5.3. Let us estimate the moments and the modulus of continuity. For the moments, we already have by (10.3) and (10.4)

$$\mathbf{E}\|F\|^p \le c_q \ \|f\|_q^p, \quad q = \max\{2, p\}.$$

This bound is uniform over Φ. Now we pass to the modulus of continuity.

Take an integer $M \ge 5$, let $u = M^{-1}$. For each integer $k \in [0..M-1]$, let $t_k = k/M$ and $I_k = [t_k, t_k + 2u)$. Then for each $x \in I_k$, we have

$$\begin{aligned} F(x) - F(t_k) &= J^{-1/2} \left[\sum_{j \le J, a - \lambda_{J,j} \in (t_k, x]} \xi_j - \sum_{j \le J, -\lambda_{J,j} \in (t_k, x]} \xi_j \right] \\ &= J^{-1/2} \left[W_k^+(x) - W_k^-(x) \right]. \end{aligned}$$

Next,

$$\begin{aligned} &\sup_{x,y \in I_k} |F(x) - F(y)| \\ &\le \ J^{-1/2} \left[\sup_{x,y \in I_k} |W_k^+(x) - W_k^+(y)| + \sup_{x,y \in I_k} |W_k^-(x) - W_k^-(y)| \right]. \end{aligned}$$

The oscillations of the processes W_k^+, W_k^- are bounded by the numbers of terms in the corresponding sums,

$$N_k^+ = \# \{ j \le J, a - \lambda_{J,j} \in (t_k, t_{k+2}) \},$$

$$N_k^- = \# \{ j \le J, -\lambda_{J,j} \in (t_k, t_{k+2}) \}.$$

By evident reasons,

$$\max \left\{ \sum_{k=0}^{M-1} N_k^+; \sum_{k=0}^{M-1} N_k^- \right\} \le 2J.$$

Since the process W_k^+ is a composition of consecutive sums of standard Gaussian i.i.d. variables, we bound its oscillation with oscillation of a Wiener process W.

Remark 10.6.1 This idea could also work for non-Gaussian ξ_j with symmetric distribution and finite moments of all orders.

We have

$$\begin{aligned}
&\mathbf{E} \sup_{x,y\in I_k} |W_k^+(x) - W_k^+(y)|^p \\
&= \mathbf{E}\left(\sup_{x\in I_k} W_k^+(x) + \sup_{y\in I_k}(-W_k^+(y))\right)^p \\
&\leq 2^{p-1}\mathbf{E}\left(\sup_{x\in I_k} W_k^+(x)^p + \sup_{y\in I_k}(-W_k^+(y))^p\right) \\
&= 2^p\mathbf{E}(\sup_{x\in I_k} W_k^+(x))^p \leq 2^p\mathbf{E}(\sup_{0\leq z\leq N_k^+} W(z))^p \\
&= 2^p\mathbf{E}|W(N_k^+)|^p = 2^p c_p (N_k^+)^{p/2}.
\end{aligned}$$

Similarly,

$$\mathbf{E} \sup_{x,y\in I_k} |W_k^-(x) - W_k^-(y)|^p \leq 2^p c_p (N_k^-)^{p/2}.$$

Now we are able to calculate the mean oscillations of the functions F. For each $h \in [0, u]$, we have

$$\begin{aligned}
\|F(\cdot + h) - F(\cdot)\|^p &= \int_0^1 |F(x+h) - F(x)|^p dx \\
&= \sum_{k=0}^{M-1} \int_{[t_k, t_k+u)} |F(x+h) - F(x)|^p dx \\
&\leq u \sum_{k=0}^{M-1} \sup_{x,y\in I_k} |F(x) - F(y)|^p
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}\,\omega_F(u)^p &= \mathbf{E}\sup_{h\leq u} \|F(\cdot + h) - F(\cdot)\|^p \\
&\leq 2^p c_p J^{-p/2} u \sum_{k=0}^{M-1} [(N_k^+)^{p/2} + (N_k^-)^{p/2}].
\end{aligned}$$

If $p \geq 2$, we have

$$\sum_{k=0}^{M-1} [(N_k^+)^{p/2} + (N_k^-)^{p/2}] \leq \left(\sum_{k=0}^{M-1} (N_k^+)\right)^{p/2} + \left(\sum_{k=0}^{M-1} (N_k^-)\right)^{p/2} \leq 2(2J)^{p/2}$$

and

$$\mathbf{E}\,\omega_F(u)^p \leq 2^{3p/2+1} c_p u \quad \to \quad 0 \qquad (u\to 0)$$

uniformly over $F \in \Phi$. If $p \in [1, 2)$, the estimate is not so good but still sufficient:

$$\sum_{k=0}^{M-1} [(N_k^+)^{p/2} + (N_k^-)^{p/2}] \quad \leq \quad M\left[\left(\sum_{k=0}^{M-1} N_k^+/M\right)^{p/2} + \left(\sum_{k=0}^{M-1} N_k^-/M\right)^{p/2}\right]$$

$$\begin{aligned} &= M^{1-p/2}\left[\left(\sum_{k=0}^{M-1} N_k^+\right)^{p/2} + \left(\sum_{k=0}^{M-1} N_k^-\right)^{p/2}\right] \\ &\le 2u^{p/2-1}(2J)^{p/2} \end{aligned}$$

and

$$\mathbf{E}\,\omega_F(u)^p \le 2^{3p/2+1} c_p u^{p/2} \quad \to \quad 0 \qquad (u \to 0)$$

uniformly over $F \in \Phi$. Condition (10.2) is verified and application of Theorem 10.5.3 completes the proof. ■

Corollary 10.6.1 *Let f be a function in L^p, $p \ge 2$. Then the family of random functions*

$$\Phi_f = \{F_{J,f,\lambda}, \quad \lambda \in \Lambda, J \in N\}$$

is tight in L^p.

PROOF. First, let

$$f(x) = \mathbf{1}_{[t,t+a]}(x), \quad 0 \le t \le t+a \le 1.$$

Then $\Phi_f = \{F_{J,f_a,\lambda}(\cdot - t), \quad \lambda \in \Lambda, J \in N\}$ and the tightness follows from Theorem 10.6.1. Next, recall that tightness is a property stable with respect to linear operations on random vectors (linear combination of compact sets is a compact). Therefore, we obtain the result for any function of the type

$$g(x) = \sum_{k=0}^{M-1} g_k \mathbf{1}_{[k/M,(k+1)/M]}(x). \tag{10.5}$$

Let now f be an arbitrary function. Inequality (10.1) for Φ_f follows directly from closure inequality (10.3). Next, for arbitrary $\varepsilon > 0$, choose g of type (10.5) such that $\|f - g\| \le \varepsilon$. Then we have

$$\omega_{F_{J,f,\lambda}}(u) \le \omega_{F_{J,g,\lambda}}(u) + 2\|F_{J,f-g,\lambda}\|,$$

$$\begin{aligned} \mathbf{E}\,\omega_{F_{J,f,\lambda}}(u)^p &\le 2^{p-1}\left(\mathbf{E}\,\omega_{F_{J,g,\lambda}}(u)^p + 2^p\mathbf{E}\|F_{J,f-g,\lambda}\|^p\right) \\ &\le 2^{p-1}\left(\mathbf{E}\,\omega_{F_{J,g,\lambda}}(u)^p + 2^p\varepsilon^p\right) \end{aligned}$$

and hence

$$\limsup_{u\to 0}\sup_{J,\lambda} \mathbf{E}\,\omega_{F_{J,f,\lambda}}(u)^p \le 2^{p-1}\left(\limsup_{u\to 0}\sup_{J,\lambda} \mathbf{E}\,\omega_{F_{J,g,\lambda}}(u)^p + 2^p\varepsilon^p\right) = 2^{2p-1}\varepsilon^p.$$

Since ε could be chosen arbitrarily small, we obtain (10.2) for Φ_f.

Remark 10.6.2 For $p \in [1,2)$, this method only yields the tightness of Φ_f in L^p for $f \in L^2$. This is trivial, since for $f \in L^2$ we have obtained the tightness in L^2-topology which is stronger than L^p-topology.

Continuous generalization

We trace here a "continuous" generalization analogous to that described above for $C[0,1]$. Recall that $\mathcal{M}$ denotes the class of all probabilistic measures on $[0,1]$, W_μ is the white noise with variance $\mu \in \mathcal{M}$, and

$$F_{f,\mu}(x) = \int_0^1 f(x+\lambda) W_\mu(d\lambda).$$

The family $\Phi_f^{\mathcal{M}} = \{F_{f,\mu}, \mu \in \mathcal{M}\}$ contains Φ_f and is still tight in L^p, $p \geq 2$. The proof of tightness remains almost the same. In particular, for closure inequality $(p \geq 2)$ we have

$$\begin{aligned} \mathbf{E}\|F_{f,\mu}\|^p &= \int_0^1 \mathbf{E}|F_{f,\mu}(x)|^p dx = \int \left[\int_0^1 f(x+\lambda)^2 \mu(d\lambda)\right]^{p/2} dx \\ &\leq \int_0^1 \left[\int_0^1 |f(x+\lambda)|^p \mu(d\lambda)\right] dx = \|f\|_p^p. \end{aligned}$$

Estimating the modulus of continuity for the key case $f(x) = \mathbf{1}_{[0,a)}(x)$ and using old notations for M, u, t_k, I_k, one obtains the representation

$$F_{f,\mu}(x) = W_\mu[-x, -x+a)$$

and for $x \in I_k$

$$F_{f,\mu}(x) - F_{f,\mu}(t_k) = W_\mu[-x, -t_k) - W_\mu[-x+a, -t_k+a).$$

Observe that these expressions are the processes with independent increments of argument x. Next, the proof follows the old way but instead of N_k^+ and N_k^- one should control the variances $N_{k,\mu}^+ = \mu[-t_{k+2}, t_k)$ and $N_{k,\mu}^- = \mu[-t_{k+2}+a, t_k+a)$; the sum of expressions of each type is bounded by 2.

10.7 An Example of Non-Tightness in L^p, $p \in [1,2)$

Consider the tightness of the families of random elements

$$F_{J,f,\lambda}(x) = J^{-1/2} \sum_{j=1}^{J} f(x+\lambda_j)\xi_j$$

in the space $L^p[0,1]$, for $f \in L^p[0,1], \{\lambda_j\} \in [0,1]^\infty, J \in N$. We give a parametric series of examples of functions $f \in L^p[0,1]$, $p \in [1,2)$ and sequences λ such that the family

$$\Phi = \{F_{J,f,\lambda},\ J \in N\}$$

is not tight in $L^p[0,1]$.

Construction

Fix $p \in [1,2), q \in [p,2)$ and let for $m \in N$

$$M_m = 2^{m^2/q},\ l_m = m^{-2}2^{-m^2},\ k_m = m2^{m^2}.$$

Consider the function

$$f(x) = \sum_{m=1}^{\infty} M_m \mathbf{1}_{[l_{m+1},l_m)}(x)$$

and the sequence

$$\{\lambda_j\} = \{-l_1, -2l_1, ..., -k_1 l_1, ..., -l_m, ..., -k_m l_m, ...\}.$$

We have immediately

$$\|f\|_p^p \le \|f\|_q^q \le \sum_{m=1}^{\infty} M_m^q l_m = \sum_{m=1}^{\infty} m^{-2} < \infty$$

and hence $f \in L^q[0,1] \subset L^p[0,1]$.

Estimation

Consider the subsequence $J_n = \sum_{m=1}^{n} k_m$.

From now on, fix n and omit J_n, λ in the notation thus replacing $F_{J_n,g,\lambda}$ by simple F_g. We have

$$f = f_1 + f_2 + f_3,$$

where

$$f_1 = \sum_{m=1}^{n-1} M_m \mathbf{1}_{[l_{m+1},l_m)}, \quad f_2 = M_n \mathbf{1}_{[l_{n+1},l_n)}, \quad f_3 = \sum_{m=n+1}^{\infty} M_m \mathbf{1}_{[l_{m+1},l_m)}.$$

In fact, we wish to get rid of f_1, f_3. Towards this aim, write

$$\begin{aligned}\|f_1\|_2^2 &= \sum_{m=1}^{n-1} M_m^2(l_m - l_{m+1}) \le \sum_{m=1}^{n-1} M_m^2 l_m \\ &= \sum_{m=1}^{n-1} 2^{(2/q-1)m^2} m^{-2} \le 2^3(2-q)^{-1} 2^{(2/q-1)(n-1)^2}.\end{aligned}$$

By closure inequality (10.4), we have

$$\mathbf{E}\|F_{f_1}\|_p^p \le \|f_1\|_2^p \le 2^{3p/2}(2-q)^{-p/2} 2^{(p/q-p/2)(n-1)^2}.$$

On the other hand, let U denote the support of the function F_{f_3}. Then U is contained in the union of J_n intervals

$$[-\lambda_j; -\lambda_j + l_{n+1}], \quad 1 \le j \le J_n.$$

We obtain a bound for Lebesgue measure of U,

$$mes\,U \le J_n l_{n+1} \le 2k_n l_{n+1} \le 2^{1+n^2-(n+1)^2} \le 2^{-2n} \le (4n)^{-1}.$$

On the other hand, for the main term we have

$$\begin{aligned} F_{f_2}(x) &= J_n^{-1/2} \sum_{j=1}^{J_n} f_2(x+\lambda_j)\xi_j \\ &= J_n^{-1/2} M_n \sum_{j=1}^{J_{n-1}} \mathbf{1}_{[l_{n+1}-\lambda_j, l_n-\lambda_j)}(x)\xi_j \\ &\quad + J_n^{-1/2} M_n \sum_{k=1}^{k_n} \mathbf{1}_{[l_{n+1}+kl_n,(k+1)l_n)}(x)\xi_{J_{n-1}+k} \\ &= F_A(x) + F_B(x) \end{aligned}$$

where the functions F_A and F_B are symmetric and independent. Let

$$I = \bigcup_{k=1}^{k_n} [l_{n+1} + kl_n, (k+1)l_n).$$

Then

$$mes\,I = k_n(l_n - l_{n+1}) \ge k_n l_n/2 = (2n)^{-1}.$$

We start the key estimate with

$$\mathbf{E}\|F_f\|_p^p = \int_0^1 \mathbf{E}|F_{f_1+f_2+f_3}(x)|^p dx \ge \int_{I-U} \mathbf{E}|F_{f_1+f_2}(x)|^p dx.$$

By linearity,

$$\begin{aligned} |F_{f_2}| &= |F_{-f_1} + F_{f_1+f_2}| \le |F_{f_1}| + |F_{f_1+f_2}|, \\ |F_{f_2}|^p &\le 2^{p-1}(|F_{f_1}|^p + |F_{f_1+f_2}|^p), \\ |F_{f_1+f_2}|^p &\ge 2^{1-p}|F_{f_2}|^p - |F_{f_1}|^p \end{aligned}$$

and we obtain for the expectations

$$\begin{aligned} \mathbf{E}\|F_f\|_p^p &\ge \int_{I-U} 2^{1-p}\mathbf{E}|F_{f_2}(x)|^p dx - \mathbf{E}\|F_{f_1}\|_p^p \\ &\ge (mes\,I - mes\,U)2^{1-p} \inf_{x\in I} \mathbf{E}|F_A(x) + F_B(x)|^p - \mathbf{E}\|F_{f_1}\|_p^p \\ &\ge (4n)^{-1}2^{1-p} \inf_{x\in I} \mathbf{E}|F_B(x)|^p - \mathbf{E}\|F_{f_1}\|_p^p \\ &\ge (4n)^{-1}2^{1-p} J_n^{-p/2} M_n^p c_p - 2^{3p/2}(2-q)^{-p/2} 2^{(q/p-p/2)(n-1)^2} \\ &\ge (4n)^{-1}2^{1-p}(2k_n)^{-p/2} M_n^p c_p - 2^{3p/2}(2-q)^{-p/2} 2^{(q/p-p/2)(n-1)^2} \\ &\ge n^{-1-p/2} c_p 2^{-1-3p/2+(p/q-p/2)n^2} - 2^{3p/2}(2-q)^{-p/2} 2^{(q/p-p/2)(n-1)^2} \\ &\quad \to \infty \end{aligned}$$

when $n \to \infty$. Therefore, we have obtained

$$\limsup_{J\to\infty} \mathbf{E}\|F_{J,f,\lambda}\|_p^p \geq \lim_{n\to\infty} \mathbf{E}\|F_{J_n,f,\lambda}\|_p^p = \infty$$

and the tightness does not take place.

Acknowledgement. The work of the first author was supported by Russian Foundation of Basic Research and INTAS, Grants 96-01-672, 95-0099.

References

1. Baushev, A. N. (1987). On the weak convergence of Gaussian measures, *Theory of Probability and Its Applications*, **32**, 670–677.

2. Bourgain, J. (1988). Almost sure convergence and bounded entropy, *Israel Journal of Mathematics*, **63**, 79–95.

3. Dunford, N. and Schwartz, J. T. (1958). *Linear Operators*, Volume 1, New York: Interscience Publishers.

4. Lifshits, M. A. (1995). *Gaussian Random Functions*, Dordrecht: Kluwer Academic Publishers.

5. Lifshits, M. A. and Weber, M. (1998). Oscillations of Gaussian Stein's elements, In *Progress in Probability*, Vol. **43**, pp. 249–261, Basel: Birkhäuser.

6. Nguyen, Z. T., Tarieladze, V. I. and Chobanyan, S. A. (1978). On compactness of families of second order measures in a Banach space, *Theory of Probability and Its Applications*, **22**, 805–810.

7. Stein, E. M. (1961). On limits of sequences of operators, *Annals of Mathematics*, **74**, 140–170.

8. Suquet, Ch. (1998). Tightness in Schauder decomposable Banach spaces, *Proceedings of the St-Petersburg Mathematical Society*, **5**, 289–327.

9. Weber, M. (1993). GC sets, Stein's elements and matrix summation methods, *Prépublication IRMA 1993/027*, Strasbourg.

10. Weber, M. (1994). GB and GC sets in ergodic theory, *Proceedings of the IX Conference on Probability in Banach Spaces, Sandberg, August 1993, Denmark, Progress in Probability*, **35**, pp. 129–151, Basel: Birkhäuser.

11. Weber, M. (1996). The Stein randomization procedure, *Rendiconti di Matematica (Roma)*, **16**, 569–605.

12. Weber, M. (1998) Entropie métrique et convergence presque partout, *Series Travaux en Cours*, **58**, Paris: Hermann.

11

Long-Time Behavior of Multi-Particle Markovian Models

A. D. Manita

M. V. Lomonosov Moscow State University, Moscow, Russia

Abstract: We find convergence time to equilibrium for wide classes of large multi-particle Markovian systems. We show that if a "one-particle" state space is large, then the long-time behavior of the multi-particle Markovian system strongly depends on the type of stochastic evolution of a single particle.

Keywords and phrases: Convergence time to equilibrium, multi-particle Markov chains, nonreversible Markov chains, Monte Carlo Markov chains, queueing models

11.1 Introduction

From the standpoint of computer simulation of Markovian stochastic systems, it is important to estimate the number of steps which are needed to approach the stationary distribution. This problem is important for dynamic Monte Carlo methods, Metropolis–Hastings algorithms, simulated annealing and image analysis. The special feature of these simulations is that the state space of a simulated system is finite but very large. This is the reason why this problem attracts widespread attention of experts in statistical physics, applied statistics and theory of computer algorithms.

In the majority of Markovian models of networks and multi-component systems in statistical physics, besides the standard Markovian property, there is some additional structure. This structure is determined by features peculiar to the state space and to the transitions between states. Undoubtedly, this structure exerts qualitative effect on *convergence time to equilibrium*. The notion of convergence time to equilibrium [see Manita (1996, 1999)] is an mathematical

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

formalization of "the number of steps of an algorithm needed for approaching the stationary distribution".

By a *multi-particle Markov chain*, we mean a system consisting of $M(N)$ particles, each evolving on large "one-particle" state space $\mathsf{K}(N)$. We are interested mainly in the situation when both $M(N)$ and $|\mathsf{K}(N)|$ tend to ∞ as $N \to \infty$. In this setting, the term "particle" is a matter of convention. It is used only to invoke physical intuition. The role of particles can be played by messages in large communication network or by customers in a queueing system with large number of nodes. Here we consider the case when an interaction between different particles is absent. A presence of interaction generally creates additional mathematical difficulties. The present work should be considered as the first step in studying the convergence time to equilibrium for more general Markovian systems of particles with interaction.

11.2 Convergence Time to Equilibrium

We consider sequences $\mathcal{L}(N)$, $N = 1, 2, \ldots$, where $\mathcal{L}(N)$ is a finite Markov chain on state space $X(N)$. We assume that $|X(N)| \to \infty$ as $N \to \infty$. In this Section, we recall the notion of *convergence time to equilibrium (CTE)* for the sequence of Markov chains $\mathcal{L}(N)$.

Let P_N be a transition matrix of the Markov chain $\mathcal{L}(N)$. Let $\mu = (\mu_\alpha, \alpha \in X(N))$ denote an initial distribution of $\mathcal{L}(N)$. Then distribution of the chain at time t is equal to μP_N^t, where P_N^t is the t-th power of the matrix P_N.

We assume that each chain $\mathcal{L}(N)$ is irreducible and aperiodic. Hence, each chain $\mathcal{L}(N)$ is ergodic. Denote by $\pi^N = (\pi_\alpha^N, \alpha \in X(N))$ the stationary distribution of $\mathcal{L}(N)$.

The *variation distance* between two probability distributions ν and ρ is defined as follows:

$$\| \nu - \rho \| = \sup_{B \subset X(N)} |\nu(B) - \rho(B)| \equiv \frac{1}{2} \sum_{x \in X(N)} |\nu_x - \rho_x|.$$

In the sequel, we denote by $\mathcal{P}(X)$ the set of probability distributions on a finite set X.

Definition 11.2.1 [Manita (1999)] We say that a function $T(N)$ is a **convergence time to equilibrium** if for any function $\psi(N) \uparrow \infty$

$$\sup_\mu \| \mu P_N^{T(N)\psi(N)} - \pi^N \| \to 0, \quad N \to \infty, \tag{11.1}$$

where sup is taken over all initial distributions of the chain $\mathcal{L}(N)$: $\mu \in \mathcal{P}(X(N))$.

The convergence time to equilibrium $T(N)$ is called a <u>minimal</u> CTE if for any function $T'(N)$ satisfying (11.1), we have $T(N) = O(T'(N))$ as $N \to \infty$.

Remark 11.2.1 The minimal CTE is unique up to the following equivalence relation:

$$T_1(N) \sim T_2(N) \overset{\text{def}}{\Longleftrightarrow} 0 < \liminf_{N\to\infty} \frac{T_1(N)}{T_2(N)} \le \limsup_{N\to\infty} \frac{T_1(N)}{T_2(N)} < +\infty.$$

Note that for any fixed N, the function $\sup_\mu \| \mu P_N^t - \pi^N \|$ decreases monotonically in t. Hence, we have the following statement.

Proposition 11.2.1 *Let a function $\theta(N)$ be such that $\sup_\mu \| \mu P_N^{\theta(N)} - \pi^N \| \to 0$, as $N \to \infty$. Then, $\theta(N)$ is the CTE in the sense of Definition* 11.2.1.

The next proposition follows from Definition 11.2.1.

Proposition 11.2.2 *A function $T(N)$ is the minimal CTE iff condition* (11.1) *holds and for any function $\phi(N) \to \infty$*

$$\sup_\mu \| \mu P_N^{T(N)/\phi(N)} - \pi^N \| \not\to 0, \qquad N \to \infty.$$

Remark 11.2.2 To show that CTE $T(N)$ is minimal, it is sufficient to prove the following statement: there exist a sequence of initial distributions $\{\mu^N\}$, $\mu^N \in \mathcal{P}(X(N))$, and a sequence of sets of states A_N, $A_N \subset X(N)$, such that

1. for any function $\phi(N) \to \infty$,

$$\mu^N P_N^{T(N)/\phi(N)}(A_N) \to 0, \qquad N \to \infty;$$

2. $\pi^N(A_N) \to 1, \qquad N \to \infty.$

In [Manita (1996, 1999)] the CTE $T(N)$ was found for sequences of finite Markov chains $\{\mathcal{L}(N)\}$ that are truncations of some geometrically ergodic countable chain. Moreover, in Manita (1999) some *queueing applications* of the obtained results were considered. Other examples are given in Section 11.4 of the present paper (see Remark 11.4.2 and Examples 11.4.1–11.4.3).

11.3 Multi-Particle Markov Chains

Let K be a finite set, $\mathcal{K} = \{\xi(t), t \in \mathbf{Z}_+\}$ be an irreducible aperiodic Markov chain on K with transition matrix $R = (r_{ij})$. It is well known that such a Markov chain is ergodic. Denote by $\nu = (\nu_j, j \in \mathsf{K}) \in \mathcal{P}(\mathsf{K})$ the stationary distribution of the chain $\mathcal{K}$. We shall interpret the random variable $\xi(t)$ as a position of some particle at time t. We assume that this particle moves over the set K according to the law of the chain $\mathcal{K}$.

Now we are going to consider a system $\mathcal{L}(\mathcal{K}, M)$ consisting of M noninteracting particles moving according to the law of the "one-particle" Markov chain $\mathcal{K}$. This system of particles can be described in two different ways: in terms of *distinguishable* or *indistinguishable* particles.

Let $\xi^{(1)}(t), \ldots, \xi^{(M)}(t)$ be M independent copies of the Markov chain $\mathcal{K}$. Consider a new Markov chain $\boldsymbol{\xi}(t) = (\xi^{(1)}(t), \ldots, \xi^{(M)}(t))$ on state space

$$\mathsf{K}^M \equiv \{\mathbf{k} = (k_1, \ldots, k_M) : \ k_m \in \mathsf{K}, \ m = 1, \ldots, M\}.$$

By assumption, the one-particle chain $\xi(t)$ is ergodic. This implies that the chain $\boldsymbol{\xi}(t)$ is also ergodic.

The Markov chain $\boldsymbol{\xi}(t)$ describes the evolution of a system consisting of M noninteracting particles. Therewith particles are numbered, and $\xi^{(m)}(t)$ is a position of the particle m at time t. Such representation of the system of noninteracting particles will be called the representation in terms of *distinguishable* particles or briefly the $\boldsymbol{\xi}$-representation. Sometimes, we shall use the notation $\mathcal{L}_\xi(\mathcal{K}, M)$ for the chain $\{\boldsymbol{\xi}(t), t \in \mathbf{Z}_+\}$ to point out its dependence on $\mathcal{K}$ and M.

Sometimes, in the situation when particles are identical, it is more convenient to consider another state space. Let $n_j(t)$ be the number of particles of the process $\boldsymbol{\xi}(\cdot)$ situated at state j at time t. Consider the random sequence

$$\mathbf{n}(t) = (n_j(t), j \in \mathsf{K}), \quad t = 0, 1, 2, \ldots.$$

It is easy to check that $\mathbf{n}(t)$ is a Markov chain on the state space

$$\mathcal{N}(\mathsf{K}, M) \stackrel{\text{def}}{=} \Big\{\mathbf{y} = (y_i, i \in \mathsf{K}) \in \mathbf{Z}_+^{\mathsf{K}} : \ \sum_{i \in \mathsf{K}} y_i = M\Big\}.$$

Under such choice of state space, we are interested only in that how many particles are placed at a specified state. On contrast to the case of $\boldsymbol{\xi}$-representation, here particles are *indistinguishable.* Such representation of a system of noninteracting particles will be called the $\mathbf{n}$-representation. To point out the dependence of this construction on the one-particle chain $\mathcal{K}$ and the number of particles M, we shall denote sometimes the chain $\{\mathbf{n}(t), t \in \mathbf{Z}_+\}$ by $\mathcal{L}_{\mathrm{n}}(\mathcal{K}, M)$.

Below we consider systems consisting of many particles under the assumption that one-particle chains may be "large". More precisely, let $\{\mathcal{K}(N)\}$ be a sequence of finite one-particle chains with state spaces $\mathsf{K}(N)$, and $\{M(N)\}$ be some sequence of positive integers. The subject of our investigation is the system $\mathcal{L}(\mathcal{K}(N), M(N))$ consisting of $M(N)$ noninteracting particles moving according to the law of the chain $\mathcal{K}(N)$. We shall always assume that $M(N) \to \infty$ or $|\mathsf{K}(N)| \to \infty$ as $N \to \infty$. Our aim is to find the CTE for the sequence $\mathcal{L}(N) \stackrel{\text{def}}{=} \mathcal{L}(\mathcal{K}(N), M(N))$. It will be seen from Sections 11.4 and 11.5 that the form of this CTE depends on the nature of chosen sequence $\mathcal{K}(N)$. All results of this chapter hold for both ($\boldsymbol{\xi}$ and $\mathbf{n}$) representations.

11.4 H and S-Classes of One-Particle Chains

Let $\mathcal{K}(N)$ be a irreducible aperiodic finite Markov chain with state space $\mathsf{K}(N)$, transition matrix $R_N = (r_{ij}(N))_{i,j\in\mathsf{K}(N)}$ and stationary distribution $\nu^N = (\nu_j^N, j \in \mathsf{K}(N))$.

Let $h(N) \geq 0$ be a monotone function increasing to ∞.

Definition 11.4.1 We say that a sequence of Markov chains $\mathcal{K}(N)$ belongs to the H-class with function $h = h(N)$, and write $\{\mathcal{K}(N)\} \in H(h)$, if the following conditions hold:

1. There exist $C_2, \gamma_2 > 0$ such that

$$\sup_{\nu_0 \in \mathcal{P}(\mathsf{K}(N))} \| \nu_0 R_N^t - \nu^N \| \leq C_2^{h(N)} \exp(-\gamma_2 t) \qquad \forall t;$$

2. There exist constants $a_1, a_2 > 0$ and sequences of states $\{i_N\}$ and $\{j_N\}$, $i_N, j_N \in \mathsf{K}(N)$, such that $|r^t_{i_N j_N} - \nu_{j_N}| > a_1 > 0 \quad \forall t \leq a_2 h(N)$;

3. There exist constants $C_1, \gamma_1 > 0$, a sequence of states $\{k_N\}$, $k_N \in \mathsf{K}(N)$, and $N_0 \in \mathbf{N}$ such that uniformly in $N \geq N_0$

$$|r^t_{k_N k_N} - \nu_{k_N}| > C_1 \exp(-\gamma_1 t) \quad \forall t.$$

Condition 1 of Definition 11.4.1 is typical for sequences of Markov chains $\mathcal{K}(N)$ that are truncations [see Manita (1999)] of some countable geometrically ergodic Markov chain $\mathcal{K}$ which possesses a Liapunov function. Markov chains characterized by Condition 2 have "bounded" one-step transitions in the sense that it is impossible to do transition between "widely separated" states in a limited time. Examples are provided by random walks with bounded jumps. Verifying of the fulfilment of Condition 3 is a particular problem.

Let $s(N) \geq 0$ be a monotone function increasing to ∞.

Definition 11.4.2 We say that the sequence of Markov chains $\mathcal{K}(N)$ belongs to the S-class with function $s = s(N)$, and write $\{\mathcal{K}(N)\} \in S(s)$, if the following conditions hold:

1. There exist $C_2, \gamma_2, a_0 > 0$ such that for all $t \geq a_0 s(N)$

$$\sup_{\nu_0} \| \nu_0 R_N^t - \nu^N \| \leq C_2 \exp(-\gamma_2 t / s(N));$$

2. There exist constants $C_1, \gamma_1 > 0$, a sequence of states $\{k_N\}$, $k_N \in \mathsf{K}(N)$, and a sequence of sets $\{B_N\}$, $B_N \subset \mathsf{K}(N)$, such that for sufficiently large N

$$|\mathsf{P}\{\xi(t) \in B_N | \xi(0) = k_N\} - \nu(B_N)| \geq C_1 \exp\left(-\gamma_1 \frac{t}{s(N)}\right) \quad \forall t.$$

Remark 11.4.1 Condition 2 of Definition 11.4.2 is equivalent to the following condition: *there exist constants $C_1, \gamma_1 > 0$ and a sequence of initial states $\{k_N\}$, $k_N \in \mathsf{K}(N)$, such that for all sufficiently large N*

$$\|r^t_{k_N \cdot} - \nu^N\| \geq C_1 \exp\left(-\gamma_1 \frac{t}{s(N)}\right) \quad \forall t.$$

[Here $r^t_{k_N \cdot} \in \mathcal{P}(\mathsf{K}(N))$ is the distribution of $\mathcal{K}(N)$ at time t, provided that the chain is located at time 0 at the initial state k_N.]

Remark 11.4.2 If $\{\mathcal{K}(N)\} \in H(h)$, then the minimal convergence time to equilibrium for the sequence of one-particle chains $\mathcal{K}(N)$ is equal to $T_{\mathcal{K}}(N) = h(N)$. If $\{\mathcal{K}(N)\} \in S(s)$, then $T_{\mathcal{K}}(N) = s(N)$.

Example 11.4.1 Markov chain $\mathcal{K}(N)$ in this example is the discrete analogue of the queueing system $\mathsf{M}\,|\,\mathsf{M}\,|\,1\,|\,N$. Namely, consider a set $\mathsf{K}(N) = \{0, 1, \ldots, N\} \equiv [0, N] \cap \mathbf{Z}$ and a Markov chain $\mathcal{K}(N)$ with state space $\mathsf{K}(N)$ and the following transition probabilities: $r_{i,i+1}(N) = p$ for $0 \leq i < N$, $r_{i,i-1}(N) = q$, for $0 < i \leq N$, $r_{ii}(N) = r$, for $0 < i < N$, $r_{00}(N) = q + r$, $r_{NN}(N) = p + r$, $p + q + r = 1$. If $p \neq q$, then $\{\mathcal{K}(N)\} \in H(N)$.

Example 11.4.2 Let $\mathcal{K}(N)$ be the same as in Example 11.4.1, but $p = q = (1 - r)/2$. In this case, $\{\mathcal{K}(N)\} \in S(N^2)$.

Example 11.4.3 Consider the discrete circle $\mathbf{Z}_{2N+1} \equiv \mathbf{Z}/(2N + 1)\mathbf{Z}$ and the simple random walk on it. More precisely, as an one-particle chain $\mathcal{K}(N)$, we consider the random walk $\xi(t)$ on the set $\mathsf{K}(N) = \{0, \ldots, 2N\}$ with periodic boundary conditions and the following jump probabilities:

$$r_{i,j}(N) = \frac{1}{2} \qquad \text{for } j - i = \pm 1 \pmod{2N + 1}$$

and $r_{i,j}(N) = 0$ for any other pair (i, j). As in Example 11.4.2, one can show that $\{\mathcal{K}(N)\} \in S(N^2)$.

Details can be found in Manita (1997).

11.5 Minimal CTE for Multi-Particle Chains

Let $T_{\mathcal{K}}(N)$ denote the minimal CTE for the sequence of one-particle chains $\mathcal{K}(N)$. Consider the sequence of multi-particle chains $\mathcal{L}(N) \stackrel{\text{def}}{=} \mathcal{L}(\mathcal{K}(N), M(N))$. We shall deal with the situation when *at least* one of sequences $\{M(N)\}$ and $\{|\mathsf{K}(N)|\}$ tends to ∞ as $N \to \infty$. Let $T(N)$ denote the minimal CTE for the sequence $\mathcal{L}(N)$. Our main results about large multi-particle Markov chains are summarized in the following theorem.

Theorem

1. *If* $M(N) \equiv M = \text{const}$ *(i.e., the number of particles is fixed), then* $T(N) = T_{\mathcal{K}}(N)$.

2. *If* $\mathcal{K}(N) \equiv \mathcal{K}$ *(i.e., the one-particle chain is fixed), then* $T(N) = \log M(N)$.

3. *If* $\{\mathcal{K}(N)\}$ *belongs to the H-class, then* $T(N) = \max(T_{\mathcal{K}}(N), \log M(N))$.

4. *If* $\{\mathcal{K}(N)\}$ *belongs to the S-class, then* $T(N) = T_{\mathcal{K}}(N) \log M(N)$.

Statements of the theorem hold for both ($\boldsymbol{\xi}$ and $\mathbf{n}$) representations of the multi-particle chain $\mathcal{L}(\mathcal{K}(N), M(N))$. Comparing items 3 and 4 of the theorem with item 2, we see that in the situation when one-particle chain is large the nature of $\{\mathcal{K}(N)\}$ is important for the CTE $T(N)$.

Proofs are given in Section 11.6.

Remark 11.5.1 Item 2 of the above theorem generalizes Proposition 7.7 in Aldous and Diaconis (1987) wherein random walks on finite groups were studied and result similar to item 2 was obtained for the case when $\mathcal{K}$ is a random walk on a finite group. In our theorem, $\mathcal{K}$ is an *arbitrary* finite ergodic Markov chain. Methods of this chapter are different from methods used by Aldous and Diaconis (1987).

11.6 Proofs

Preliminary results

Denote by $\mathbf{P} = (\mathbf{p_{kl}})_{\mathbf{k},\mathbf{l}\in \mathsf{K}^M}$ the transition matrix of the chain $\mathcal{L}_\xi(\mathcal{K}, M)$ and by $\boldsymbol{\pi} \equiv \boldsymbol{\pi}(\mathsf{K}, M) \in \mathcal{P}(\mathsf{K}^M)$ its stationary distribution. It is easy to see that

$$\mathbf{p_{kl}} = \prod_{m=1}^{M} r_{k_m l_m}, \qquad \mathbf{k} = (k_1, \ldots, k_M), \quad \mathbf{l} = (l_1, \ldots, l_M),$$

$$\boldsymbol{\pi} = \underbrace{\nu \times \cdots \times \nu}_{M}, \qquad \text{i.e.,} \qquad \boldsymbol{\pi}_{\mathbf{k}} = \nu_{k_1} \cdots \nu_{k_M}. \tag{11.2}$$

Lemma 11.6.1 $\sup\limits_{\boldsymbol{\mu}\in\mathcal{P}(\mathsf{K}^M)} \|\boldsymbol{\mu}\mathbf{P}^t - \boldsymbol{\pi}\| \;\le\; M \sup\limits_{\nu_0\in\mathcal{P}(\mathsf{K})} \|\nu_0 R^t - \nu\|.$

PROOF. Let $(i_1, \ldots, i_M) \in \mathsf{K}^M$. Then

$$\begin{aligned}
&\sum_{j_1,\ldots,j_M} \left| \prod_{l=1}^{M} r^t_{i_l j_l} - \prod_{l=1}^{M} \nu_{j_l} \right| \\
&= \sum_{j_1,\ldots,j_M} \left| (r^t_{i_1 j_1} - \nu_{j_1}) \prod_{l=2}^{M} r^t_{i_l j_l} + \nu_{j_1} \left(\prod_{l=2}^{M} r^t_{i_l j_l} - \prod_{l=2}^{M} \nu_{j_l} \right) \right| \\
&\le \sum_{j_1} |r^t_{i_1 j_1} - \nu_{j_1}| + \sum_{j_2,\ldots,j_M} \left| \prod_{l=2}^{M} r^t_{i_l j_l} - \prod_{l=2}^{M} \nu_{j_l} \right| \\
&\le 2 \sup_{\nu_0} \|\nu_0 R^t - \nu\| + \sum_{j_2,\ldots,j_M} \left| \prod_{l=2}^{M} r^t_{i_l j_l} - \prod_{l=2}^{M} \nu_{j_l} \right|.
\end{aligned}$$

Using this line of reasoning, we obtain Lemma 11.6.1. ■

Let us introduce the map $\chi : \mathsf{K}^M \longrightarrow \mathcal{N}(\mathsf{K}, M)$,

$$\chi(\mathbf{k}) = (\chi_j(\mathbf{k}), j \in \mathsf{K}), \qquad \chi_j(\mathbf{k}) = \#\{m : \ k_m = j\},$$

where $\#A \equiv |A|$ is the cardinality of the finite set A. For any distribution $\rho \in \mathcal{P}(\mathsf{K}^M)$, denote $\rho^{\mathbf{n}} \stackrel{\text{def}}{=} \rho \circ \chi^{-1} \in \mathcal{P}(\mathcal{N}(\mathsf{K}, M))$.

Lemma 11.6.2 $\|\rho^{\mathbf{n}} - \eta^{\mathbf{n}}\| \le \|\rho - \eta\| \quad \forall \rho, \eta \in \mathcal{P}(\mathsf{K}^M).$

PROOF. It is easily noted that

$$\begin{aligned} \|\rho^{\mathbf{n}} - \eta^{\mathbf{n}}\| &= \sup_{D \in \mathcal{N}(\mathsf{K},M)} |\rho^{\mathbf{n}}(D) - \eta^{\mathbf{n}}(D)| \\ &= \sup_{D \in \mathcal{N}(\mathsf{K},M)} |\rho(\chi^{-1}D) - \eta(\chi^{-1}D)| \\ &\leq \sup_{B \in \mathsf{K}^M} |\rho(B) - \eta(B)| = \|\rho - \eta\|. \end{aligned}$$

■

It is easy to check that $\mathbf{n}(t) = (n_j(t), j \in \mathsf{K}) = \chi(\boldsymbol{\xi}(t))$ and $\boldsymbol{\pi}^{\mathbf{n}} = \boldsymbol{\pi} \circ \chi^{-1}$ is the stationary distribution of the Markov chain $\mathcal{L}_{\mathrm{n}}(\mathcal{K}, M)$. Moreover, $\boldsymbol{\pi}^{\mathbf{n}}$ has the polynomial form

$$\boldsymbol{\pi}^{\mathbf{n}}_{\mathbf{y}} = M! \prod_{i \in \mathsf{K}} \frac{(\nu_i)^{y_i}}{y_i!}. \tag{11.3}$$

Denote $\mathbf{e}_i = (0, \ldots, 0, \underset{i}{1}, 0, \ldots, 0) \in \mathbf{R}^{|\mathsf{K}|}$, $r^t_{iB} = \sum_{j \in B} r^t_{ij}$ and $n_B(t) = \sum_{j \in B} n_j(t)$.

Lemma 11.6.3 *If $\delta > \frac{1}{2}, \epsilon > 0$ then uniformly in $i \in \mathsf{K}$, $B \subset \mathsf{K}$ and $t \in \mathbf{N}$*

$$\mathsf{P}\left\{ |n_B(t) - M r^t_{iB}| \geq \epsilon M^\delta \,\middle|\, \mathbf{n}(0) = M\mathbf{e}_i \right\} \leq \frac{1}{4\epsilon^2 M^{2\delta - 1}}.$$

PROOF. At given initial condition, at time t each of M particles is at the set B with probability r^t_{iB} independently of other particles. Since $n_B(t) = \sum_{m=1}^{M} \mathbf{1}(\xi^{(m)}(t) \in B)$, we have that the conditional distribution of $n_B(t)$ as binomial: $\mathsf{P}\{n_B(t) = l \mid \mathbf{n}(0) = M\mathbf{e}_i\} = C^l_M (r^t_{iB})^l (1 - r^t_{iB})^{M-l}$, $l = 0, 1, \ldots, M$. Hence $\mathsf{E}\,(n_B(t)) = M\, r^t_{iB}$ and $\mathsf{Var}\,(n_B(t)) = M\, r^t_{iB}(1 - r^t_{iB}) \leq M/4$ uniformly in i, B and t. Now statement of the lemma easily follows from the Chebyshev inequality. ■

Remark 11.6.1 It follows from (11.3) the stationary distribution of n_j is binomial, namely, $\boldsymbol{\pi}^{\mathbf{n},N}\{n_j = r\} = C^r_M \nu^r_j (1 - \nu_j)^{M-r}$, $r = 0, 1, \ldots, M$. Hence, $\mathsf{E}_\pi n_B = M\nu(B)$, $\mathsf{Var}_\pi(n_B) = M\nu(B)(1 - \nu(B))$, where $n_B = \sum_{j \in B} n_j$. Using the Chebyshev inequality, we get the following uniform estimate:

$$\forall \kappa > \frac{1}{2}, \epsilon > 0 \qquad \boldsymbol{\pi}^{\mathbf{n}}\left(|n_B - M\nu(B)| > \epsilon M^\kappa\right) \leq \frac{1}{4\epsilon^2 M^{2\kappa - 1}}. \tag{11.4}$$

Let $\{\mathcal{K}(N)\}$ be a sequence of finite one-particle chains with state spaces $\mathsf{K}(N)$, and $\{M(N)\}$ be some sequence of positive integers. We shall always assume that $M(N) \to \infty$ or $|\mathsf{K}(N)| \to \infty$ as $N \to \infty$. Consider the system of $M(N)$ noninteracting particles moving according to the law of the chain $\mathcal{K}(N)$. As already noted in Section 11.3, this system of particles can be described in terms of *distinguishable* or *indistinguishable* particles. The following question arises: does the convergence time to equilibrium depend on the choice of the representation or not ?

Proposition 11.6.1

1. *Suppose that a function $T'_\xi(N)$ is the convergence time to equilibrium for the sequence of chains $\mathcal{L}_\xi(\mathcal{K}(N), M(N))$. Then, the function $T'_\xi(N)$ is the convergence time to equilibrium for the sequence of chains $\mathcal{L}_{\mathrm{n}}(\mathcal{K}(N), M(N))$.*

2. *Suppose that a function $T_{\mathrm{n}}(N)$ is the minimal convergence time to equilibrium for the sequence $\mathcal{L}_{\mathrm{n}}(\mathcal{K}(N), M(N))$ and is the convergence time to equilibrium for the sequence $\mathcal{L}_\xi(\mathcal{K}(N), M(N))$. Then, the function $T_{\mathrm{n}}(N)$ is the minimal convergence time to equilibrium for the sequence of Markov chains $\mathcal{L}_\xi(\mathcal{K}(N), M(N))$.*

Proposition 11.6.1 is an easy consequence of Definition 11.2.1 and Lemma 11.6.2.

General idea

Let us fix some notation. Let $\mathsf{K}(N)$ denote the state space of one-particle Markov chain $\mathcal{K}(N)$, $R_N = (r_{ij}(N))_{i,j\in\mathsf{K}(N)}$ be its transition matrix, and $\nu^N = (\nu_j^N,\ j \in \mathsf{K}(N))$ be its stationary distribution. Consider a sequence of multi-particle systems $\{\mathcal{L}(\mathcal{K}(N), M(N)),\ N \in \mathbf{N}\}$. In $\boldsymbol{\xi}$-representation: the state space of the multi-particle chain $\mathcal{L}_\xi(\mathcal{K}(N), M(N))$ is the set $X(N) = \mathsf{K}(N)^{M(N)}$, transition probabilities have the form $\mathbf{p_{kl}} \equiv \mathbf{p_{kl}}(N) = \prod_{m=1}^{M(N)} r_{k_m l_m}(N)$, the stationary distribution is

$$\pi^N \equiv (\boldsymbol{\pi}_{\mathbf{l}}, \mathbf{l} \in X(N)) = \underbrace{\nu^N \times \cdots \times \nu^N}_{M(N)}.$$

In $\mathbf{n}$-representation: the state space of the chain $\mathcal{L}_{\mathrm{n}}(\mathcal{K}(N), M(N))$ is the set $X_{\mathrm{n}}(N) = \mathcal{N}(\mathsf{K}(N), M(N))$, the stationary distribution defined in (11.3) will be denoted by $\boldsymbol{\pi}^{\mathbf{n},N}$, and the state of the chain at time t will be denoted by $\mathbf{n}^N(t)$.

Proof of each item of the main theorem will consist of the following steps.

Step 1: We prove that the function $T(N)$ is the CTE for the sequence $\mathcal{L}_\xi(\mathcal{K}(N), M(N))$. To do this, we shall use Lemma 11.6.1.

Step 2: We prove that the function $T(N)$ is the minimal CTE for the sequence $\mathcal{L}_{\mathrm{n}}(\mathcal{K}(N), M(N))$. By Proposition 11.2.2 and Remark 11.2.2 it is sufficient to show that there exist a sequence of initial states $\mathbf{y}_0^N \in X_{\mathrm{n}}(N)$ and a sequence of sets of states $A_N \subset X_{\mathrm{n}}(N)$ such that the following conditions hold:

- For any function of the form $t(N) = T(N)/\phi(N)$, where $\phi(N)$ is an arbitrary function tending to ∞, we have

$$\mathsf{P}\left\{\mathbf{n}^N(t(N)) \notin A_N |\ \mathbf{n}^N(0) = \mathbf{y}_0^N\right\} \to 1 \quad (N \to \infty). \tag{11.5}$$

- $$\pi^{\mathbf{n},N}(A_N) \to 1 \quad (N \to \infty). \tag{11.6}$$

Then it will follow from Proposition 11.6.1 that $T(N)$ is the minimal CTE for both ($\boldsymbol{\xi}$ and $\mathbf{n}$) representations of the multi-particle system $\mathcal{L}(\mathcal{K}(N), M(N))$.

The case when the number of particles is fixed

Here we prove item 1 of the main theorem. It is assumed that the one-particle chain $\mathcal{K}(N)$ depends on N but the number of particles M *is fixed*, i.e. $\mathcal{L}(N) = \mathcal{L}(\mathcal{K}(N), M)$. This case is rather simple.

Let us prove that $T(N) = T_{\mathcal{K}}(N)$ is the CTE for the multi-particle chain $\mathcal{L}_{\xi}(\mathcal{K}(N), M)$. By Lemma 11.6.1 for any function $\psi(N) \uparrow \infty$ we have

$$\sup_{\mu \in \mathcal{P}(X(N))} \|\mu P_N^{T(N)\psi(N)} - \pi^N\| \leq M \sup_{\nu_0 \in \mathcal{P}(\mathsf{K}(N))} \|\nu_0 R_N^{T_{\mathcal{K}}(N)\psi(N)} - \nu^N\|.$$

By definition of CTE the r.h.s. of the bound vanishes as $N \to \infty$. *Step* 1 is thus proved.

Let us show now that the function $T(N) = T_{\mathcal{K}}(N)$ is the minimal CTE for the multi-particle chain $\mathcal{L}_{\mathrm{n}}(\mathcal{K}(N), M)$. By assumption, $T_{\mathcal{K}}(N)$ is the minimal CTE for the sequence of chains $\mathcal{K}(N)$. The application of Proposition 11.2.2 yields that for any function $\phi(N) \to \infty$

$$\sup_{\nu_0} \| \nu_0 R_N^{T_{\mathcal{K}}(N)/\phi(N)} - \nu^N \| \not\longrightarrow 0, \qquad N \to \infty.$$

It immediately follows that there exist sequences $i_N \in \mathsf{K}(N)$ and $B_N \subset \mathsf{K}(N)$ such that

$$\left| \sum_{j \in B_N} r_{i_N j}^{T_{\mathcal{K}}(N)/\phi(N)}(N) - \nu^N(B_N) \right| \not\longrightarrow 0, \qquad N \to \infty. \tag{11.7}$$

Consider a sequence of states $\mathbf{y}_0^N = M\mathbf{e}_{i_N} \in X_{\mathrm{n}}(N)$ and a sequence of sets

$$A_N = \{\mathbf{y} : \sum_{i \in B_N} y_i = M\} \subset X_{\mathrm{n}}(N).$$

It is easy to see that $\pi^{\mathbf{n},N}(A_N) = \left(\nu^N(B_N)\right)^M$ and

$$\mathsf{P}\left\{\mathbf{n}^N(t) \in A_N | \mathbf{n}^N(0) = \mathbf{y}_0^N\right\} = \left(\sum_{j \in B_N} r_{i_N j}^t(N)\right)^M.$$

Recall that M is fixed. Hence, using (11.7) we get

$$\left|\mathsf{P}\left\{\mathbf{n}^N(T_{\mathcal{K}}(N)/\phi(N)) \in A_N | \mathbf{n}^N(0) = \mathbf{y}_0^N\right\} - \pi^{\mathbf{n},N}(A_N)\right| \not\longrightarrow 0, \qquad N \to \infty.$$

Step 2 thus proved. Now, item 1 of the theorem follows from Proposition 11.6.1.

The case when one-particle chain is fixed

Here we prove item 2 of the main theorem. The one-particle chain $\mathcal{K}$ is *fixed* but the number of particles M tends to infinity. By assumption, the finite Markov chain $\mathcal{K}$ is ergodic. It is well known [Karlin (1968)] that such Markov chain converges exponentially fast to its stationary distribution, i.e., there exist $C > 0$ and $\gamma > 0$ such that

$$\sup_k \|r^t_{k\cdot} - \nu\| \le C\exp(-\gamma t). \tag{11.8}$$

Moreover, with the exception of the trivial case which we exclude from consideration, this convergence can not be faster than the exponential one, viz., there exist $i \in \mathsf{K}$, $\alpha > 0$ and $t_0 > 0$ such that

$$\|r^t_{i\cdot} - \nu\| > \exp(-\alpha t) \qquad \forall t > t_0. \tag{11.9}$$

We show first that the function $T(M) = \log M$ is the CTE for the sequence of multi-particle chains $\mathcal{L}_\xi(\mathcal{K}, M)$. Indeed, taking into account (11.8) and Lemma 11.6.1, we obtain

$$\begin{aligned}\sup_{\mu\in\mathcal{P}(X(M))} \|\mu P_M^{T(M)\psi(M)} - \pi^M\| &\le M \sup_{\nu_0\in\mathcal{P}(\mathsf{K})} \|\nu_0 R^{\log M\psi(M)} - \nu\| \\ &\le C\exp(-\gamma\log M\psi(M) + \log M).\end{aligned}$$

Since $\psi(M) \uparrow \infty$ as $M \to \infty$, the r.h.s. of the bound tends to 0. *Step* 1 is completed.

Let us prove now that $T(M) = \log M$ is the minimal CTE for the sequence of multi-particle chains $\mathcal{L}_\mathrm{n}(\mathcal{K}, M)$. Consider the following set of states $A_M = \{\mathbf{y} : |\mathbf{y} - M\cdot\nu|_1 \le \frac{1}{2}M^{5/6}\} \subset X_\mathrm{n}(M)$, where $|\cdot|_1$ is the L_1-norm in $\mathbf{R}^{|\mathsf{K}|}$, $|\mathbf{y}|_1 \stackrel{\mathrm{def}}{=} \sum_{k=1}^{K} |y_k|$. If M is sufficiently large, then the set A_M is not empty. Let i be the same as in (11.9). Put $\mathbf{y}_0^M = M\mathbf{e}_i \in X_\mathrm{n}(M)$. Let us prove that for the sequences A_M and $\mathbf{y}_0^M$, conditions (11.5) and (11.6) hold.

Let us show first that uniformly in t

$$\mathsf{P}\left\{|\mathbf{n}(t) - M\cdot r^t_{i\cdot}|_1 < M^{2/3}\right\} \longrightarrow 1 \quad (M \to \infty). \tag{11.10}$$

To do this, let us note that

$$\{|\mathbf{n}(t) - M\cdot r^t_{i\cdot}|_1 < M^{2/3}\} \supset \bigcap_{j\in\mathsf{K}} \left\{|n_j(t) - M\cdot r^t_{ij}| < M^{2/3}/K\right\},$$

where $K = |\mathsf{K}|$. Recall that K is fixed and does not depend on M. Hence, it is sufficient to show that $\forall j \in \mathsf{K}\ \mathsf{P}\left\{|n_j(t) - M\cdot r^t_{ij}| \ge M^{2/3}/K\right\} \to 0$ as $M \to \infty$. This easily follows from Lemma 11.6.3.

It follows from (11.9) that

$$|M \cdot r^t_{i\cdot} - M \cdot \nu|_1 > M \cdot \exp(-\alpha t). \tag{11.11}$$

Fix any function $\phi(M)$ such that $\phi(M) \to \infty$ as $M \to \infty$. Put $t(M) = \log M/\phi(M)$. It follows from (11.10) and (11.11) that

$$\mathsf{P}\left\{|\mathbf{n}(t(M)) - M \cdot \nu|_1 > M \cdot \exp(-\alpha t(M)) - M^{2/3} \,|\, \mathbf{n}(0) = \mathbf{y}_0^M\right\} \longrightarrow 1.$$

Let us choose $M_0 > 2^6$ such that $\alpha/\phi(M) < 1/6$ for $M > M_0$. Then for $M > M_0$, we have

$$\begin{aligned} M \cdot \exp(-\alpha t(M)) - M^{2/3} &= M \cdot \exp(-(\alpha \log M)/\phi(M)) - M^{2/3} \\ &= M \cdot M^{-\alpha/\phi(M)} - M^{2/3} \\ &\geq M^{5/6} - M^{2/3} \geq \frac{1}{2} M^{5/6}. \end{aligned}$$

Now condition (11.5) easily follows. To prove condition (11.6), note that the event $B_M = \left\{|y_j - \nu_j M| \leq (2K)^{-1} M^{5/6} \; \forall j \in \mathsf{K}\right\}$ is imbedded into the event A_M. Let us show that the stationary probability of the negation of the event B_M tends to zero: $\boldsymbol{\pi}^{\mathbf{n},M}\left\{\overline{B_M}\right\} \to 0$. Indeed,

$$\begin{aligned} \boldsymbol{\pi}^{\mathbf{n},M}\left\{\overline{B_M}\right\} &= \boldsymbol{\pi}^{\mathbf{n},M}\left\{\bigcup_{j \in \mathsf{K}} \{|y_j - \nu_j M| > (2K)^{-1} M^{5/6}\}\right\} \\ &\leq \sum_{j \in \mathsf{K}} \boldsymbol{\pi}^{\mathbf{n},M}\left\{|y_j - \nu_j M| > (2K)^{-1} M^{5/6}\right\}. \end{aligned}$$

Let us show that each term in the r.h.s. tends to zero. Fix any j and consider $\boldsymbol{\pi}^{\mathbf{n},M}\left\{|y_j - \nu_j M| > (2K)^{-1} M^{5/6}\right\}$. It follows from (11.4) that this probability is bounded by $K^2 M^{-2/3}$ and thus tends to 0 as $M \to \infty$. Condition (11.6) is proved and *step* 2 is completed.

Statement 2 of the main theorem is thus proved.

H-class of one-particle chains

We prove here item 3 of the theorem. Now the one-particle Markov chain $\mathcal{K}(N)$ is growing and the number of particles $M(N)$ tends to infinity as $N \to \infty$. It is assumed that $\{\mathcal{K}(N)\} \in H(h)$, where $h(N) \geq 0$ is some monotone function increasing to ∞.

Let us show that[1] $T(N) = \log M(N) \vee h(N)$ is the CTE for the sequence of chains $\mathcal{L}_\xi(\mathcal{K}(N), M(N))$. It follows from Lemma 11.6.1 and Definition 11.4.1

[1] $a \vee b \stackrel{\text{def}}{=} \max(a, b)$.

that

$$\sup_{\mu\in\mathcal{P}(X(N))} \|\mu P_N^t - \pi^N\| \le M(N)\cdot C_2^{h(N)}\exp(-\gamma_2 t) = 2\exp\left(\log M(N) + h(N)\log C_2 - \gamma_2 t\right). \tag{11.12}$$

If we put $t = (\log M(N) \vee h(N))\cdot\psi(N)$, where $\psi(N)\uparrow\infty$, $(N\to\infty)$, then the r.h.s. of (11.12) will tend to zero. *Step* 1 is thus completed.

Let us prove that the function $T(N) = \log M(N)\vee h(N)$ is the minimal CTE for $\mathcal{L}_{\rm n}(\mathcal{K}(N), M(N))$. Let us introduce sequences of states $\{u_N\}$, $\{l_N\}$,

$$u_N = \begin{cases} i_N, & \text{if } \log M(N)\le h(N),\\ k_N, & \text{if } \log M(N) > h(N),\end{cases} \qquad l_N = \begin{cases} j_N, & \text{if } \log M(N)\le h(N),\\ k_N, & \text{if } \log M(N) > h(N),\end{cases}$$

and sequence of sets $A_N = \left\{\mathbf{y} : |y_{l_N} - M(N)\nu_{l_N}| \le b\cdot(M(N))^{5/6}\right\}$. Let us show that there exists $b>0$ such that, for the sequence of sets A_N and the sequence of initial states $\mathbf{y}_0^N = M(N)\mathbf{e}_{u_N}$, conditions (11.5)–(11.6) hold. First we test the validity of condition (11.5). It follows from Lemma 11.6.3 that

$$\mathsf{P}\left\{|n_{l_N}^N(t) - M(N)r_{u_N l_N}^t| < (M(N))^{2/3}\mid \mathbf{n}^N(0) = \mathbf{y}_0^N\right\} \longrightarrow 1 \tag{11.13}$$

uniformly in t as $N\to\infty$. Fix any function $\phi(N)$, $\lim_{N\to\infty}\phi(N) = \infty$, and put $t(N) = (\log M(N)\vee h(N))/\phi(N)$. Let us prove now that for sufficiently large N

$$|r_{u_N l_N}^{t(N)} - \nu_{l_N}| > c\cdot(M(N))^{-1/6}, \tag{11.14}$$

where $c>0$ does not depend on N. It is necessary to consider two cases. Let N be such that $\log M(N)\le h(N)$. In this case, $t(N) = h(N)/\phi(N)$. Since $\phi(N)\to\infty$, we have $\phi(N) > 1/a_2$ for sufficiently large N. Hence, by Condition 2 from Definition 11.4.1, for such N we get

$$|r_{u_N l_N}^{t(N)} - \nu_{l_N}| > a_1 > 0. \tag{11.15}$$

Let N be now such that $\log M(N) > h(N)$, i.e. $t(N) = \log M(N)/\phi(N)$. Using Condition 3 from Definition 11.4.1, we obtain

$$\begin{aligned} |r_{u_N l_N}^{t(N)} - \nu_{l_N}| &> C_1\exp\left(-\gamma_1\log M(N)/\phi(N)\right)\\ &= C_1\cdot(M(N))^{-\gamma_1/\phi(N)}\\ &\ge C_1\cdot(M(N))^{-1/6} \qquad \text{for large } N.\end{aligned} \tag{11.16}$$

Combining (11.15) and (11.16), we get (11.14). Let us show now that (11.13) and (11.14) imply the validity of (11.5). Indeed, the probability that the following inequality holds

$$|n_{l_N}^N(t(N)) - M(N)r_{u_N l_N}^{t(N)}| < (M(N))^{2/3}$$

tends to 1 as $N \to \infty$. On other hand, by (11.14), it follows that for large N

$$|M(N)r_{u_N l_N}^{t(N)} - M(N)\nu_{l_N}| > c \cdot (M(N))^{5/6}.$$

Choosing $b < c$, we have that

$$|n_{l_N}^N(t(N)) - M(N)\nu_{l_N}| > c \cdot (M(N))^{5/6} - (M(N))^{2/3} > b \cdot (M(N))^{5/6}$$

with probability which tends to 1 as $N \to \infty$. This proves (11.5).

Using (11.4), we can estimate $\boldsymbol{\pi}^{\mathbf{n},N}(\overline{A_N})$:

$$\boldsymbol{\pi}^{\mathbf{n},N}\left(|y_{l_N} - M(N)\nu_{l_N}| > b \cdot (M(N))^{5/6}\right) \leq \frac{1}{4b^2 M(N)^{2/3}} \longrightarrow 0 \quad (N \to \infty).$$

The validity of (11.6) is proved. *Step* 2 is completed, thus providing. This completes the proof of statement 3 of the theorem.

S-class of one-particle chains

We prove here item 4 of the theorem. Situation is similar to the previous case but now $\{\mathcal{K}(N)\} \in S(s)$, where $s(N) \geq 0$ is some monotone function increasing to ∞. Applying Lemma 11.6.1 and using Condition 1 of Definition 11.4.2, we obtain

$$\sup_{\mu \in \mathcal{P}(X(N))} \|\mu P_N^t - \pi^N\| \leq \exp\left(\log M(N) + \log C_2 - \gamma_2 \frac{t}{s(N)}\right).$$

From this estimate, it follows that the function $T(N) = s(N) \log M(N)$ is the CTE for the sequence of chains $\mathcal{L}_\xi(\mathcal{K}(N), M(N))$. *Step* 1 is completed.

To prove that $T(N) = s(N) \log M(N)$ is <u>minimal</u> CTE for $\mathcal{L}_{\mathrm{n}}(\mathcal{K}(N), M(N))$, it is sufficient to show that, for the sequence of initial states $\mathbf{y}_0^N = M(N)\mathbf{e}_{k_N}$ and the sequence of sets $A_N = \left\{\mathbf{y} : |\sum_{l \in B_N} y_l - M(N)\nu(B_N)| \leq \frac{C_1}{2}(M(N))^{5/6}\right\}$, statements (11.5) and (11.6) hold. Similarly to the previous case, statement (11.5) will follow from the next two statements:

$$\mathsf{P}\left\{\left|n_{B_N}^N(t) - M(N)r_{k_N B_N}^t\right| < (M(N))^{2/3} \mid \mathbf{n}^N(0) = \mathbf{y}_0^N\right\} \longrightarrow 1 \qquad (11.17)$$

uniformly in t as $N \to \infty$, and

$$|\mathsf{P}\{\xi(t(N)) \in B_N | \xi(0) = k_N\} - \nu(B_N)| > C_1(M(N))^{-1/6}, \qquad (11.18)$$

where the constant $C_1 > 0$ is the same as in Definition 11.4.2 and N is sufficiently large. Statement (11.17) follows from Lemma 11.6.3. To prove (11.18), let us use Condition 2 of Definition 11.4.2. We have

$$\begin{aligned} |\mathsf{P}\{\xi(t(N)) \in B_N | \xi(0) = k_N\} - \nu(B_N)| &\geq C_1 \exp(-\gamma_1 t(N)/s(N)) \\ &= C_1 \exp(-\gamma_1 \log M(N)/\phi(N)) \\ &\geq C_1(M(N))^{-1/6} \end{aligned}$$

for large N. Statement (11.6) follows from (11.4).

This completes *step* 2 and the proof of the theorem. ■

Acknowledgements. This work was supported by RFBR (grants 99-01-01140 and 98-01-00653) and French-Russian Liapunov Institute of Applied Mathematics and Informatics (project 6-97).

References

1. Manita, A. D. (1996). On equilibrium convergence time for large finite Markov chains, *Moscow University Mathematics Bulletin*, **51**, No. 6, 38–39.

2. Manita, A. D. (1999). Convergence time to equilibrium for large finite Markov chains, *Fundamental and Applied Mathematics*, **5**, No. 1.

3. Manita, A. D. (1997). Convergence time to equilibrium for multi-particle Markov chains, *Preprint* of French-Russian Institute, No. 3. Moscow University.

4. Aldous, D. and Diaconis, P. (1987). Strong uniform times and finite random walks, *Advances in Applied Mathematics*, **8**, 69–97.

5. Karlin, S. (1968). *A First Course in Stochastic Processes*, Academic Press, New York.

12

Applications of Infinite-Dimensional Gaussian Integrals

A. M. Nikulin

St. Petersburg State Technical University, St. Petersburg, Russia

Abstract: In this chapter, the difference between an absolute moment of any Gaussian measure on the Hilbert space and the same moment of its projection onto some finite-dimensional subspace is evaluated.

Keywords and phrases: Gaussian measure, Hilbert space, Banach space, infinite-dimensional Gaussian integral

Let $X = L_2$ be the separable Hilbert space and μ be a Gaussian measure on X. Suppose that the mean value of the measure μ is equal to zero: $\mathbf{a}_\mu = 0$.

The correlative operator $\mathbf{K}_\mu$ is a symmetric positive kernel operator: its eigenvectors form an orthogonal basis, its eigenvalues λ_k are positive and

$$\sum_{k=1}^{\infty} \lambda_k < \infty, \quad \lambda_k \geq, 0 \quad k \in I\!N.$$

In this case, it is natural to choose the following orthonormal basis $\{e_k\}_{k=1}^{\infty}$ of the space X that $e_k, k = 1, 2, ...$ are the eigenvectors of $\mathbf{K}_\mu$, which is enumerated in decreasing order of the corresponding eigenvalues:

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq \ldots.$$

Note that

$$\Lambda(n) = \sum_{k=n+1}^{\infty} \lambda_k \to 0, \quad n \to \infty.$$

In the case $\operatorname{supp} \mu = X$, our measure is a product measure. Consider the system of projectors

$$\{\pi_n\}_{n=1}^{\infty}, \quad \pi_n : X \to X_n = \operatorname{Span}\{e_1, e_2, \ldots, e_n\},$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

where

$$\pi_n(x) = \pi_n\left(\sum_{k=1}^{\infty} x_k e_k\right) = \sum_{k=1}^{n} x_k e_k, \quad x \in X, \quad x_k = (x, e_k)$$

and $(\cdot,\cdot)$ is the scalar product in X.

Let $\|\cdot\|$ be the Hilbert norm in X, $\|\cdot\|_n$ the semi-norm in X_n generated by π_n, $n \in \mathbb{N}$, and

$$h(x) = \|x\|^p, \quad h_n(x) = \|\pi_n(x)\|^p = \|x\|_n^p, \quad p \geq 1, \quad x \in X.$$

Theorem 12.1.1 *Under these conditions, we have*

$$\forall p \geq 1 \quad \exists C_{1,p,\mu} < \infty, \quad C_{2,p,\mu} < \infty$$

such that

$$C_{1,p,\mu}\Lambda(n) \leq \Delta_{n,p} = \left| \int_X h(x)\,d\mu(x) - \int_X h_n(x)\,d\mu(x) \right| \leq C_{2,p,\mu}\Lambda(n).$$

For the proof of this theorem, we need the following result [see Fernique (1970) and Ledoux and Talagrand (1991)]

Theorem [Fernique (1970)] *Let $(E, \mathcal{B})$ be a measurable vector space, X be a Gaussian vector with values in E, and N be a semi-norm in E. Then if the probability $\mathbf{P}\{N(X) < \infty\}$ is strongly positive, there exists $\varepsilon > 0$ such that*

$$\forall \alpha < \varepsilon : \quad \mathbf{E}\{\exp(\alpha N^2(X))\} < \infty.$$

Using this result we obtain

$$\forall p > 0 : \int_X \|x\|^p d\mu(x) < \infty.$$

Note that

$$\Delta_{n,p} = \int_X h(x)\,d\mu(x) - \int_X h_n(x)\,d\mu(x)$$

because

$$\forall n \in \mathbb{N}, \quad \forall p \geq 1 \quad \|x\|_n^p \leq \|x\|^p.$$

For the proof, it is necessary to prove some auxiliary results.

Lemma 12.1.1 $\forall x, y > 0, \quad \forall p \geq 1,$

$$\frac{1}{2} y(x+y)^{p-1} \leq (x+y)^p - x^p \leq py(x+y)^{p-1}.$$

The proof of Lemma 12.1.1 is evident;

$$\|x\|^2 = \left\|\sum_{n=1}^{\infty}(x, e_n)e_n\right\|^2 = \sum_{n=1}^{\infty} x_n^2.$$

With the notation

$$l_n(x) = \sum_{k=n+1}^{\infty} x_k^2,$$

we notice that

$$\begin{aligned} \int\limits_X l_n(x)\, d\mu(x) &= \int\limits_X \sum_{k=n+1}^{\infty} x_k^2 d\mu(x) = \sum_{k=n+1}^{\infty} \int\limits_X x_k^2 d\mu(x) \\ &= \sum_{k=n+1}^{\infty} (\mathbf{K}_\mu e_k, e_k) = \sum_{k=n+1}^{\infty} \lambda_k = \Lambda(n). \end{aligned}$$

Lemma 12.1.2 *For all $p > 0$, there exist $K_p < \infty$ and $C_p < \infty$ such that*

$$K_p\Lambda(n) \le \int\limits_X l_n(x)\|x\|^p d\mu(x) \le C_p\Lambda(n).$$

Proof of Lemma 12.1.2. Let us find the upper bound. Denote

$$\xi = \sum_{i=n+1}^{\infty} x_i^2 \qquad \text{and} \qquad \alpha = \sum_{i=1}^{n} x_i^2.$$

Then it is sufficient to show that

$$\mathbf{E}\left\{\xi(\alpha+\xi)^{\frac{p}{2}}\right\} \le \text{Const}\,\mathbf{E}\xi.$$

Let $m \in I\!N$, $p/2 \le m$. We decompose the mathematical expectation in the sum

$$\begin{aligned} \mathbf{E}\left\{\xi(\alpha+\xi)^{\frac{p}{2}}\right\} &= \mathbf{E}\left\{\xi(\alpha+\xi)^{\frac{p}{2}}\mathbf{1}_{[0,1]}(\alpha+\xi)\right\} + \mathbf{E}\left\{\xi(\alpha+\xi)^{\frac{p}{2}}\mathbf{1}_{]1,\infty[}(\alpha+\xi)\right\} \\ &\le \mathbf{E}\xi + \mathbf{E}\left\{\xi(\alpha+\xi)^m\right\} = I_1 + I_2. \end{aligned}$$

By the independence of α and ξ, there is

$$I_2 = \sum_{k=0}^{m} \mathbf{C}_m^k \mathbf{E}\xi^{k+1}\alpha^{m-k} = \sum_{k=0}^{m} \mathbf{C}_m^k \mathbf{E}\alpha^{m-k}\mathbf{E}\xi^{k+1}.$$

Because

$$\alpha \le \sum_{i=1}^{\infty} x_i^2 \quad \text{and} \quad \mathbf{E}\xi = \Lambda(n),$$

it is sufficient to show that there exists a constant independent of n and, probably, independent of the measure μ, but depending on m:

$$\exists \text{ Const } = \text{ Const}(m) < \infty : \quad \mathbf{E}\xi^k \leq \text{ Const } \mathbf{E}\xi.$$

Let $\gamma_i = x_i^2, \quad i \in I\!N$, and it is evident that γ_i are independent;

$$\mathbf{E}\xi^k = \mathbf{E}\left\{\sum_{i=n+1}^{\infty} \gamma_i\right\}^k = \sum_{i_1,\dots,i_k=n+1}^{\infty} \mathbf{E}\gamma_{i_1}\cdots\gamma_{i_k},$$

$$\mathbf{E}\gamma_{i_1}\cdots\gamma_{i_k} = \mathbf{E}\gamma_{s_1}^{m_1}\cdots\gamma_{s_l}^{m_l} = \mathbf{E}\gamma_{s_1}^{m_1}\mathbf{E}\gamma_{s_2}^{m_2}\cdots\mathbf{E}\gamma_{s_l}^{m_l},$$

for

$$s_1 \neq s_2 \neq \ldots \neq s_l, \quad \sum_{i=1}^{l} m_i = k.$$

Therefore, we can compute

$$\mathbf{E}\gamma_{i_1}\cdots\gamma_{i_k} = \prod_{j=1}^{l} \mathbf{E}x_{s_j}^{2m_j} = \prod_{j=1}^{l}(2m_j-1)!!\lambda_{s_j}^{m_j} = C_{m_1,\dots,m_l}\prod_{r=1}^{k}\lambda_{i_r}.$$

It is clear that

$$C_{m_1,\dots,m_l} \leq \{(2k)!\}^k = A_k \leq A_m.$$

Thus we obtain

$$\mathbf{E}\xi^k \leq A_m\left\{\sum_{i=n+1}^{\infty}\lambda_i\right\}^k < A_m\Lambda(n)\left\{\sum_{i=1}^{\infty}\lambda_i\right\}^{k-1} = \text{ Const } \mathbf{E}\xi,$$

where

$$\text{Const} = A_m\left\{\sum_{i=1}^{\infty}\lambda_i\right\}^{k-1}.$$

Now we can look for the lower bound. We have

$$\int_X l_n(x)\|x\|^p d\mu(x) = I_1 + I_2,$$

where

$$I_1 = \int_A l_n(x)\|x\|^p d\mu(x), \quad I_2 = \int_B l_n(x)\|x\|^p d\mu(x),$$

$$A = \{x \in X : 0 < \|x\| < 1\}, \quad B = \{x \in X : \|x\| \geq 1\}.$$

It is evident that

$$I_2 > \int_B l_n(x)\,d\mu(x) = \int_B \{l_n(x)^{\frac{1}{2}}\}^2 d\mu(x) \geq \left\{\int_B l_n(x)^{\frac{1}{2}}d\mu(x)\right\}^2.$$

We first look at I_1: By Cauchy-Schwartz inequality, we obtain

$$\left\{\int\limits_A l_n(x)^{\frac{1}{2}} d\mu(x)\right\}^2 \leq \int\limits_A l_n(x)\|x\|^p d\mu(x) \int\limits_A \|x\|^{-p} d\mu(x).$$

For estimating a lower bound of I_1, it is sufficient to prove that the second factor is finite. It is clear that, without loss of generality, we can suppose that $\lambda_j > 0$ for infinite numbers of j. In this case, for any $m > 0$

$$\mu(\|x\| \leq \varepsilon) \leq C(m)\varepsilon^m.$$

Thus, for $m > p + 1$

$$\begin{aligned}
\int\limits_A \|x\|^{-p} d\mu(x) &\leq \int\limits_X \|x\|^{-p} d\mu(x) = \sum_{k=0}^{\infty} \int\limits_{\frac{1}{k+1} \leq \|x\| < \frac{1}{k}} \|x\|^{-p} d\mu(x) \\
&\leq \sum_{k=0}^{\infty} (k+1)^p \mu(\|x\| < \frac{1}{k}) \\
&\leq C(m)(\sum_{k=1}^{\infty} (k+1)^p \frac{1}{k^m} + 1) < \infty.
\end{aligned}$$

Consequently,

$$\forall p > 0 \quad \int\limits_X \|x\|^{-p} d\mu(x) < \infty.$$

So we have proved that

$$\exists \hat{C}_p < \infty : \quad \int\limits_A l_n(x)\|x\|^p d\mu(x) \geq \hat{C}_p \left\{\int\limits_A l_n(x)^{\frac{1}{2}} d\mu(x)\right\}^2 .$$

Therefore,

$$\exists \hat{K}_p < \infty : \quad \int\limits_X l_n(x)\|x\|^p d\mu(x) \geq \hat{K}_p \left\{\int\limits_X l_n(x)^{\frac{1}{2}} d\mu(x)\right\}^2 ;$$

for example

$$\hat{K}_p = \min\left\{\frac{\hat{C}_p}{2}, \frac{1}{2}\right\} \quad (\text{since } a^2 + b^2 \geq \frac{1}{2}(a+b)^2).$$

Now let us find the lower bound for integral

$$J = \left\{\int\limits_X l_n(x)^{\frac{1}{2}} d\mu(x)\right\}^2 .$$

By the Hőlder inequality, we get

$$\begin{aligned}\int_X l_n(x)d\mu(x) &= \int_X l_n(x)^{\frac{1}{3}} l_n(x)^{\frac{2}{3}} d\mu(x)\\ &\le \left\{\int_X l_n(x)^{\frac{1}{2}}\, d\mu(x)\right\}^{\frac{2}{3}} \left\{\int_X l_n(x)^2 d\mu(x)\right\}^{\frac{1}{3}}.\end{aligned}$$

Thus

$$\left\{\int_X l_n(x)^{\frac{1}{2}} d\mu(x)\right\}^2 \ge \frac{\left\{\int_X l_n(x)\, d\mu(x)\right\}^3}{\int_X l_n(x)^2 d\mu(x)}.$$

The numerator of the fraction is equal to

$$\left\{\int_X l_n(x)\, d\mu(x)\right\}^3 = (\Lambda(n))^3.$$

Now if we consider

$$I = \int_X l_n(x)^2 d\mu(x),$$

then

$$\begin{aligned} I &= \int_X \left(\sum_{k=n+1}^{\infty} x_k^2\right)\left(\sum_{k=n+1}^{\infty} x_k^2\right) d\mu(x) = \sum_{k=n+1}^{\infty} \int_X x_k^2 \sum_{j=n+1}^{\infty} x_j^2 d\mu(x)\\ &= \sum_{k=n+1}^{\infty} \left\{\mathbf{E}x_k^4 + \int_X x_k^2 d\mu(x) \int_X \sum_{j=n+1, j\ne k}^{\infty} x_j^2 d\mu(x)\right\}\\ &= \sum_{k=n+1}^{\infty} \left\{\mathbf{E}x_k^4 + \lambda_k(\Lambda(n) - \lambda_k)\right\}, \quad \mathbf{E}x_k^2 = \lambda_k, \quad x_k \in \mathbf{N}(0, \lambda_k).\end{aligned}$$

It is known that $\mathbf{E}x_k^4 = 3\lambda_k^2$. Then

$$I = \sum_{k=n+1}^{\infty} \left\{3\lambda_k^2 + \lambda_k(\Lambda(n) - \lambda_k)\right\} = \sum_{k=n+1}^{\infty} \left\{2\lambda_k^2 + \Lambda(n)\lambda_k\right\}.$$

In other words,

$$I = \Lambda(n)^2 + 2\sum_{k=n+1}^{\infty} \lambda_k^2 \le 3\Lambda(n)^2,$$

from which we conclude

$$\left\{\int_X l_n(x)^{\frac{1}{2}} d\mu(x)\right\}^2 \ge \frac{1}{3}\Lambda(n),$$

so that

$$\int_X l_n(x)\|x\|^p d\mu(x) \geq K_p \Lambda(n).$$

Remark 12.1.1 In the proof of Lemma 12.1.2, a stronger bound was obtained

$$\mathbf{E}\xi^k \leq A_m\{\Lambda(n)\}^k, \quad k \leq m.$$

Remark 12.1.2 In fact, our constant Const = Const(m) is depending on the measure μ, since the covariate operator and its eigenvalues are defined only by the measure μ.

Remark 12.1.3 In this lemma, it was shown that if we have a sequence $\{\lambda_n\}_{n=1}^{\infty}$, such that

$$\forall n \in I\!N \quad \lambda_n > 0, \quad \sum_{n=1}^{\infty} \lambda_n < \infty$$

and a sequence $\{z_n\}_{n=1}^{\infty}$ i.i.d. standard Gaussian random variables

$$\forall n \in I\!N \quad z_n \in N(0,1),$$

we get

$$\forall n \in I\!N \quad \left\{\mathbf{E}\left\{\sum_{k=n}^{\infty} \lambda_k z_k^2\right\}^{\frac{1}{2}}\right\}^2 \geq \frac{1}{3}\sum_{k=n}^{\infty} \lambda_k.$$

We need the following result

Lemma 12.1.3 *With the same notations, we have*

$$\forall p \in [0,1] \quad \exists \tilde{K}_p < \infty, \quad \tilde{C}_p < \infty$$

such that

$$\tilde{K}_p \Lambda(n) \leq \int_X \frac{l_n(x)}{\|x\|^p} d\mu(x) \leq \tilde{C}_p \Lambda(n). \qquad (*)$$

Proof of Lemma 12.1.3. Let us find the upper bound. Therefore, we decompose the integral in $(*)$ into the sum of two integrals

$$\int_X \frac{l_n(x)}{\|x\|^p} d\mu(x) = I_1 + I_2,$$

where

$$I_1 = \int_A \frac{l_n(x)}{\|x\|^p} d\mu(x), \qquad I_2 = \int_B \frac{l_n(x)}{\|x\|^p} d\mu(x),$$

$$A = \{x \in X : 0 < \|x\| \leq 1\}, \quad B = \{x \in X : \|x\| > 1\}.$$

Consider each integral separately; by the Cauchy-Schwartz inequality, we obtain

$$I_1 \leq \left\{ \int_A l_n(x)^2 d\mu(x) \right\}^{\frac{1}{2}} \left\{ \int_A \|x\|^{-2p} d\mu(x) \right\}^{\frac{1}{2}} = J_1 J_2.$$

It must be noted that in Lemma 12.1.2 we have proved that

$$\forall p > 0 \quad \int_X \|x\|^{-p} d\mu(x) < \infty.$$

Therefore, $J_2 < \infty$. Using Remark 12.1.1, we obtain $\mathbf{E} l_n(x)^2 \leq \text{const}\{\Lambda(n)\}^2$, hence $J_1 \leq C_p \Lambda(n)$. It is evident that

$$I_2 \leq \int_B l_n(x)\, d\mu(x) < \int_X l_n(x)\, d\mu(x) = \Lambda(n),$$

so $I = I_1 + I_2 \leq C_p \Lambda(n)$ holds. Now we go on to find the lower bound of our integral. For that we decompose it into the sum of the same two integrals: $I = I_1 + I_2$. Consider each integral separately as

$$I_1 > \int_A l_n(x)\, d\mu(x) \geq \left\{ \int_A l_n(x)^{\frac{1}{2}} d\mu(x) \right\}^2.$$

Furthermore, we estimate I_2:

$$\left\{ \int_B l_n(x)^{\frac{1}{2}} d\mu(x) \right\}^2 \leq \int_B \frac{l_n(x)}{\|x\|^p} d\mu(x) \int_B \|x\|^p d\mu(x),$$

and therefore

$$I_2 = \int_B \frac{l_n(x)}{\|x\|^p} d\mu(x) \geq C_p^* \left\{ \int_B l_n(x)^{\frac{1}{2}} d\mu(x) \right\}^2, \quad C_p^* < \infty.$$

Thus

$$\int_X \frac{l_n(x)}{\|x\|^p} d\mu(x) \geq \text{const}_p \left\{ \int_X l_n(x)^{\frac{1}{2}} d\mu(x) \right\}^2, \quad \text{const}_p < \infty,$$

and by using the result of Lemma 12.1.2

$$\left\{ \int_X l_n(x)^{\frac{1}{2}} d\mu(x) \right\}^2 \geq C\Lambda(n), \quad C < \infty,$$

we conclude that there exists a constant $\tilde{K}_p < \infty$ such that

$$\int_X \frac{l_n(x)}{\|x\|^p} d\mu(x) \geq \tilde{K}_p \Lambda(n), \quad 0 < p \leq 1.$$

PROOF OF THEOREM 12.1.1. The special case $p = 2$

$$\Delta_{n,2} = \int_X \|x\|^2 d\mu(x) - \int_X \|x\|_n^2 d\mu(x) = \int_X l_n(x)\, d\mu(x) = \Lambda(n)$$

delivers the result of our theorem. Let $p \geq 1,\ p \neq 2$. Then

$$\|x\|^p - \|x\|_n^p = \frac{\left\{\sum\limits_{k=1}^{n} x_k^2 + \sum\limits_{k=n+1}^{\infty} x_k^2\right\}^p - \left\{\sum\limits_{k=1}^{n} x_k^2\right\}^p}{\|x\|^p + \|x\|_n^p}.$$

By using Lemma 12.1.1, let us estimate the numerator of this fraction:

$$\frac{1}{2} l_n(x)\|x\|^{2p-2} \leq \|x\|^{2p} - \|x\|_n^{2p} \leq p l_n(x)\|x\|^{2p-2}$$

and note that

$$\|x\|^p \leq \|x\|^p + \|x\|_n^p \leq 2\|x\|^p.$$

Therefore,

$$\frac{1}{4} l_n(x)\|x\|^{p-2} \leq \|x\|^p - \|x\|_n^p \leq p l_n(x)\|x\|^{p-2}, \quad p \geq 1.$$

By using Lemma 12.1.2 (when $p > 2$) or the Lemma 12.1.3 (when $1 \leq p < 2$), we obtain constants $C_{1,p} < \infty$ and $C_{2,p} < \infty$ so that

$$C_{1,p}\Lambda(n) \leq \Delta_{n,p} \leq C_{2,p}\Lambda(n),$$

which concludes the proof. ■

Now we shall give some applications. Consider the Brownian motion $\xi(t)$ in the interval $t \in [0, 1]$. Let μ be the distribution of ξ in the space $L_2[0, 1]$.

In this case, the correlative operator is the integral operator with the kernel [see Gikhman and Skorohod (1971)]

$$B(t, s) = \min(t, s).$$

Its eigenvectors are the functions [see Gikhman and Skorohod (1971)]

$$\{\varphi_n(t)\}_{n=0}^{\infty}, \quad \varphi_n(t) = \sqrt{2} \sin\left\{\left(n + \frac{1}{2}\right)\pi t\right\},$$

and its eigenvalues corresponding are

$$\{\lambda_n\}_{n=0}^{\infty}, \quad \lambda_n = \frac{1}{(n+\frac{1}{2})^2\pi^2}.$$

Thus, we can decompose our process in the orthogonal sum

$$\xi(\omega,t) = \sqrt{2}\sum_{n=0}^{\infty} a_n(\omega)\frac{\sin\left\{(n+\frac{1}{2})\pi t\right\}}{(n+\frac{1}{2})\pi},$$

where $\{a_n(\omega)\}_{n=0}^{\infty}$ is the sequence of i.i.d. standard Gaussian random variables. At any fixed time t, we have the convergence with probability 1. Furthermore, we have the uniform convergence in $[0,1]$ and the limit (continuous almost sure) is the Wiener process in $[0,1]$.

Let h be the norm in the space $L_2[0,1]$. By using Theorem 12.1.1, we obtain

$$C_{1,p}\Lambda(n) \le \int_X \left|\|x\|^p - \|x\|_n^p\right| d\mu(x) \le C_{2,p}\Lambda(n), \quad p \ge 1,$$

$$\|\cdot\| = \|\cdot\|_{L_2}, \quad C_{1,p} < \infty, \quad C_{2,p} < \infty,$$

$$\Lambda(n) = \sum_{k=n+1}^{\infty} \lambda_k = \sum_{k=n+1}^{\infty} \frac{1}{(n+\frac{1}{2})^2\pi^2} \sim \frac{\text{Const}}{n},$$

where μ is the Wiener measure in the space $L_2[0,1]$.

Remark 12.1.4 In fact, we can rewrite our inequality

$$C_{1,p}\Lambda(n) \le \mathbf{E}\|\xi\|^p - \mathbf{E}\|\xi_n\|^p \le C_{2,p}\Lambda(n), \quad p \ge 1$$

where

$$\xi_n(\omega,t) = \sqrt{2}\sum_{k=0}^{n} a_k(\omega)\frac{\sin\left\{(k+\frac{1}{2})\pi t\right\}}{(k+\frac{1}{2})\pi}.$$

Now we look at another example. Let $\zeta = \{\zeta(t), \quad t \in [0,1]\}$ be the random process connected with the Brownian motion

$$\zeta(t) = \xi(t) - t\xi(1),$$

(it is the Brownian bridge). Then, ζ is the Gaussian process and its correlative operator is the integral operator with the kernel

$$B_1(t,s) = \min(t,s) - ts.$$

Its eigenvectors are $\{\varphi_n(t)\}_{n=1}^{\infty}$ and its eigenvalues are $\{\lambda_n\}_{n=1}^{\infty}$ [see Gikhman and Skorohod (1971)]

$$\varphi_n(t) = \sqrt{2}\sin\{n\pi t\}, \quad \lambda_n = \frac{1}{n^2\pi^2} \quad n \in I\!N.$$

Thus, we can decompose our process into the orthogonal sum

$$\zeta(\omega, t) = \sqrt{2} \sum_{n=1}^{\infty} a_n(\omega) \frac{\sin\{n\pi t\}}{n\pi},$$

where $\{a_n(\omega)\}_{n=0}^{\infty}$ is the sequence of i.i.d. standard Gaussian random variables. By using Theorem 12.1.1, we obtain

$$C_{1,p}^* \Lambda(n) \leq \int_X |\|x\|^p - \|x\|_n^p| \, d\mu(x) \leq C_{2,p}^* \Lambda(n), \quad p \geq 1,$$

$$\|\cdot\| = \|\cdot\|_{L_2}, \quad C_{1,p}^* < \infty, \quad C_{2,p}^* < \infty,$$

$$\Lambda(n) = \sum_{k=n+1}^{\infty} \lambda_k = \sum_{k=n+1}^{\infty} \frac{1}{n^2 \pi^2} \sim \frac{\text{Const}^*}{n},$$

where μ is the measure generated by our process in $L_2[0,1]$.

Remark 12.1.5 In fact, we can rewrite our inequality

$$C_{1,p}^* \Lambda(n) \leq \mathbf{E}\|\zeta\|^p - \mathbf{E}\|\zeta_n\|^p \leq C_{2,p}^* \Lambda(n), \quad p \geq 1,$$

where

$$\zeta_n(\omega, t) = \sqrt{2} \sum_{k=1}^{n} a_k(\omega) \frac{\sin\{k\pi t\}}{k\pi}.$$

It can be proved that a change of the basis of the space X does not improve the order of decrease of the value $\Delta_{n,p}$. We plan to publish this result in a future article.

References

1. Fernique, X. (1970). Intégrabilité des vecteurs Gaussiens, *Acad. Sci. Paris, Series A*.

2. Gikhman, I. I. and Skorohod, A. V. (1971). *Theory of the Stochastic Processes*, Moscow. Nauka. (in Russian).

3. Ibragimov, I. A. (1983). Probability for getting into a ball of a small radius in a Banach space for a Gaussian vector, *LOMI Seminars* (in Russian).

4. Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces, (Isoperimetry & Processes)*, Springer-Verlag, Berlin.

5. Vakhania, N. N., Tarieladze, V. I. and Chobanyan, S. A. (1985). *The Probability Measures in Banach Spaces*, Moscow. Nauka. (in Russian).

13

On Maximum of Gaussian Non-Centered Fields Indexed on Smooth Manifolds

Vladimir Piterbarg and Sinisha Stamatovich

M. V. Lomonosov Moscow State University, Moscow, Russia
University of Montenegro, Montenegro, Yugoslavia

Abstract: The double sum method of evaluation of probabilities of large deviations for Gaussian processes with non-zero expectations is developed. Asymptotic behaviors of the tail of non-centered locally stationary Gaussian fields indexed on smooth manifolds are evaluated. In particular, smooth Gaussian fields on smooth manifolds are considered.

Keywords and phrases: Gaussian fields, large excursions, maximum tail distribution, exact asymptotics

13.1 Introduction

The double-sum method is one of the main tools in studying asymptotic behavior of maxima distribution of Gaussian processes and fields; for example, see Adler (1998), Piterbarg (1996), Fatalov and Piterbarg and (1995) and references therein. Until recently, only centered processes have been considered. It can be seen from Piterbarg (1996) and this Chapter that the investigation of non-centered Gaussian fields can be performed with similar techniques, which, however, are far from trivial. Furthermore, there are examples when the need for the asymptotic behaviour for non-centered fields arises. In Piterbarg and Tyurin (1993, 1999), statistical procedures have been introduced to test non-parametric hypotheses for multi-dimensional distributions. The asymptotic decision rules are based on tail distributions of maxima of Gaussian fields indexed on spheres or products of spheres. In order to estimate power of the procedures, one might have to have asymptotic behaviour of tail maxima distributions for non-centered Gaussian fields.

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

In this Chapter we extend the double sum method to study Gaussian processes with non-zero expectations. We evaluate asymptotic behavior of the tail of non-centered locally $(\alpha_{\mathbf{t}}, D_{\mathbf{t}})$-stationary Gaussian field indexed on smooth manifolds, as defined below. In particular, smooth Gaussian fields on smooth manifolds are considered.

13.2 Definitions, Auxiliary Results, Main Results

Let the collection $\alpha_1, ..., \alpha_k$ of positive numbers be given, as well as the collection $l_1, ..., l_k$ of positive integers such that $\sum_{i=1}^{k} l_i = n$. We set $l_0 = 0$. These two collections will be called a *structure*; see Piterbarg (1996). For any vector $\mathbf{t} = (t_1, ..., t_n)^\top$ its *structural module* is defined by

$$|\mathbf{t}|_\alpha = \sum_{i=1}^{k} \left(\sum_{j=E(i-1)+1}^{E(i)} t_j^2 \right)^{\frac{\alpha_i}{2}}, \tag{13.1}$$

where $E(i) = \sum_{j=0}^{i} l_j$, $j = 1, ..., k$. The structure defines a decomposition of the space R^n into the direct sum $\mathrm{R}^n = \bigoplus_{i=1}^{k} \mathrm{R}^{l_i}$, such that the restriction of the structural module on either of R^{l_i} is just Euclidean norm taken to the degree α_i, $i = 1, ..., k$, respectively. For $u > 0$, denote by G_u^i the homothety of the subspace R^{l_i} with the coefficient u^{-2/α_i}, $i = 1, ..., k$, respectively, and by g_u, the superposition of the homotheties, $g_u = \bigcirc_{i=1}^{k} G_u^i$. It is clear that for any $\mathbf{t} \in \mathrm{R}^n$,

$$|g_u \mathbf{t}|_\alpha = u^{-2} |\mathbf{t}|_\alpha. \tag{13.2}$$

Let $\chi(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, be a Gaussian field with continuous paths, the expected value and the covariance function are given by

$$\mathbf{E}\chi(\mathbf{t}) = -|\mathbf{t}|_\alpha, \qquad \mathbf{Cov}(\chi(\mathbf{t}), \chi(\mathbf{s})) = |\mathbf{t}|_\alpha + |\mathbf{s}|_\alpha - |\mathbf{t} - \mathbf{s}|_\alpha, \tag{13.3}$$

respectively. Thus, $\chi(\mathbf{t})$ can be represented as a sum of independent multi-parameter drifted fractional Brownian motions (Lévy–Shönberg fields) indexed on R^{l_i}, with parameters α_i.

To proceed, we need a generalization of the Pickands' constant. Define the function on measurable subsets of R^n,

$$H_\alpha(B) = \exp\left\{ \sup_{\mathbf{t} \in B} \chi(\mathbf{t}) \right\}. \tag{13.4}$$

Let D be a non-degenerate matrix $n \times n$, and throughout we make no notational difference between a matrix and the corresponding linear transformation. Next, for any $S > 0$, we denote by

$$[0, S]^k = \{\mathbf{t} : \ 0 \le t_i \le S, \ i = 1, ..., k, \ t_i = 0, \ i = k+1, .., n\},$$

a cube of dimension k generated by the first k coordinates in R^n. In Belyaev and Piterbarg (1972), it is proved that there exists a positive limit

$$0 < H_\alpha^{DR^k} := \lim_{S\to\infty} \frac{H_\alpha(D[0,S]^k)}{\mathrm{mes}_k(D[0,S]^k)} < \infty, \tag{13.5}$$

where $\mathrm{mes}_k(D[0,S]^k)$ denotes the k-dimensional Lebesgue measure of $D[0,S]^k$. We write shortly $H_\alpha^{(k)} = H_\alpha^{IR^k}$ with I being the unit matrix. The constant $H_\alpha = H_\alpha^{(n)}$ is the Pickands' constant. Denoting

$$\Psi(u) = \frac{1}{\sqrt{2\pi}} \int_u^\infty e^{-\frac{x^2}{2}} dx, \tag{13.6}$$

it is well known that

$$\Psi(u) = \frac{1}{\sqrt{2\pi}u} e^{-\frac{u^2}{2}} (1 + o(1)) \quad \text{as } u \to \infty. \tag{13.7}$$

Lemma 13.2.1 *Let, $X(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, be a Gaussian homogeneous centered field. Let for a non-degenerate matrix A and α-structure on R^n, the covariance function $r(\mathbf{t})$ of $X(\mathbf{t})$ satisfies*

$$r(\mathbf{t}) = 1 - |A\mathbf{t}|_\alpha + o(|A\mathbf{t}|_\alpha) \quad as \quad \mathbf{t} \to \mathbf{0}. \tag{13.8}$$

Then for any compact set $T \subset \mathrm{R}^n$ and any function $\theta(u)$ with $\theta(u) \to 1$ as $u \to \infty$,

$$\mathbf{P}\left\{ \sup_{\mathbf{t}\in g_u T} X(\mathbf{t}) > u\theta(u) \right\} = H_\alpha(AT)\Psi(u\theta(u))(1 + o(1)) \quad as\ u \to \infty. \tag{13.9}$$

Definition 13.2.1 Let an α-structure be given on R^n. We say that $X(\mathbf{t})$, $\mathbf{t} \in T \subset \mathrm{R}^n$, has a local $(\alpha, D_\mathbf{t})$-stationary structure, or $X(\mathbf{t})$ is locally $(\alpha, D_\mathbf{t})$-stationary, if for any $\varepsilon > 0$ there exists a positive $\delta(\varepsilon)$ such that for any $\mathbf{s} \in T$ one can find a non-degenerate matrix $D_\mathbf{s}$ such that the covariance function $r(\mathbf{t}_1, \mathbf{t}_2)$ of $X(\mathbf{t})$ satisfies

$$1 - (1+\varepsilon)|D_\mathbf{s}(\mathbf{t}_1 - \mathbf{t}_2)|_\alpha \le r(\mathbf{t}_1, \mathbf{t}_2) \le 1 - (1-\varepsilon)|D_\mathbf{s}(\mathbf{t}_1 - \mathbf{t}_2)|_\alpha \tag{13.10}$$

provided $||\mathbf{t}_1 - \mathbf{s}|| < \delta(\varepsilon)$ and $||\mathbf{t}_2 - \mathbf{s}|| < \delta(\varepsilon)$.

It is convenient to cite here four theorems which are in use, and are suitable for our purposes. Before that, we need some notations. Let L be a k-dimensional subspace of R^n, for fixed orthogonal coordinate systems in R^n and in L, let $(x_1, ..., x_k)^\top$ be the coordinate presentation of a point $\mathbf{x} \in L$, and $(x_1', ...x_n')^\top$ be its coordinate presentation in R^n. Denote by $M = M(L)$ the corresponding transition matrix,

$$(x_1', ...x_n')^\top = M(x_1, ..., x_k)^\top,$$

i.e., $M = (\partial x_i'/\partial x_j, , \; i = 1, ..., n\,, \, j = 1, ..., k)$.

Next, for a matrix G of size $n \times k$ we denote by $V(G)$, the square root of the sum of squares of all minors of order k. This invariant transforms the volume when the dimension of vectors is changed, i.e., $d\mathbf{t} = V(G)^{-1} dG\mathbf{t}$. Note that since both coordinate systems in L and R^n are orthogonal, $V(M) = 1$.

Theorem 13.2.1 [Piterbarg (1996, Theorem 7.1)] *Let $X(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, be a Gaussian homogeneous centered field such that for some α, $0 < \alpha \leq 2$ and a non-degenerated matrix D, its covariance function satisfies*

$$r(\mathbf{t}) = 1 - ||D\mathbf{t}||^\alpha + o(||D\mathbf{t}||^\alpha) \quad \text{as } \mathbf{t} \to 0, \tag{13.11}$$

Then for any k, $0 < k \leq n$, every subspace L of R^n *with* $\dim L = k$, *any Jordan set $A \subset L$, and every function $w(u)$ with $w(u)/u = o(1)$ as $u \to \infty$,*

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\mathbf{t}\in A} X(\mathbf{t}) > u + w(u)\right\} \\ &\quad = H_\alpha^{(k)} V(DM(L)) \mathrm{mes}_L(A) u^{\frac{2k}{\alpha}} \Psi(u + w(u))(1 + o(1)) \end{aligned} \tag{13.12}$$

as $u \to \infty$, provided

$$r(\mathbf{t} - \mathbf{s}) < 1 \quad \textit{for all } \mathbf{t}, \mathbf{s} \in \bar{A},\ \mathbf{t} \neq \mathbf{s}, \tag{13.13}$$

with $\bar{A}$ the closure of A.

Theorem 13.2.2 [Michaleva and Piterbarg (1996, Theorem 1)] *Let $X(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, be a Gaussian centered locally $(\alpha, D_\mathbf{t})$-stationary field, with $\alpha > 0$ and a continuous matrix function $D_\mathbf{t}$. Let $\mathcal{M} \subset \mathrm{R}^n$ be a smooth compact of dimension k, $0 < k \leq n$. Then for any c,*

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\mathbf{t}\in\mathcal{M}} X(\mathbf{t}) > u - c\right\} \\ &\quad = H_\alpha^{(k)} u^{\frac{2k}{\alpha}} \Psi(u - c) \int_{\mathcal{M}} V(D_\mathbf{t} M_\mathbf{t})\, d\mathbf{t} (1 + o(1)) \end{aligned} \tag{13.14}$$

as $u \to \infty$, where $M_\mathbf{t} = M(T_\mathbf{t})$ with $T_\mathbf{t}$ the tangent subspace taken to $\mathcal{M}$ at the point $\mathbf{t}$ and $d\mathbf{t}$ is an element of volume of $\mathcal{M}$.

Theorem 13.2.3 [The Borell-Sudakov-Tsirelson inequality] *Let $X(t)$, $t \in T$, be a measurable Gaussian process indexed on an arbitrary set T, and let numbers σ, m, a be defined by relations*

$$\sigma^2 = \sup_{t\in T} \mathbf{Var} X(t) < \infty, \quad m = \sup_{t\in T} \mathbf{E} X(t) < \infty,$$

and

$$\mathbf{P}\left\{\sup_{t\in T} X(t) - \mathbf{E}X(t) \geq a\right\} \leq \frac{1}{2}. \tag{13.15}$$

Then for any x,

$$\mathbf{P}\left\{\sup_{t\in T} X(t) >\right\} \le 2\Psi\left(\frac{x-m-a}{\sigma}\right). \tag{13.16}$$

Theorem 13.2.4 [Slepian inequality] *Let* $X(t)$, $Y(t)$, $t \in T$, *be separable Gaussian processes indexed on an arbitrary set* T, *and suppose that for all* $t, s \in T$,

$$\begin{aligned} \mathbf{Var}X(t) = \mathbf{Var}Y(t), & \qquad \mathbf{E}X(t) = \mathbf{E}Y(t), \\ & \text{and} \\ \mathbf{Cov}(X(t), X(s)) \quad \le & \quad \mathbf{Cov}(Y(t), Y(s)). \end{aligned} \tag{13.17}$$

Then for all x,

$$\mathbf{P}\left\{\sup_{t\in T} X(t) < x\right\} \le \mathbf{P}\left\{\sup_{t\in T} Y(t) < x\right\}. \tag{13.18}$$

We now turn to our main results.

Theorem 13.2.5 *Let* $X(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, *be a Gaussian locally* $(\alpha, D_{\mathbf{t}})$*-stationary field, with some* $\alpha > 0$ *and continuous matrix function* $D_{\mathbf{t}}$. *Let* $\mathcal{M} \subset \mathrm{R}^n$ *be a smooth* k*-dimensional compact,* $0 < k \le n$. *Let the expectation* $m(\mathbf{t}) = \mathbf{E}X(\mathbf{t})$ *be continuous on* $\mathcal{M}$ *and attains its maximum on* $\mathcal{M}$ *at the only point* $\mathbf{t}_0$, *with*

$$m(\mathbf{t}) = m(\mathbf{t}_0) - (\mathbf{t}-\mathbf{t}_0)B(\mathbf{t}-\mathbf{t}_0)^\top + O(||\mathbf{t}-\mathbf{t}_0||^{2+\beta}) \quad \text{as } \mathbf{t} \to \mathbf{t}_0, \tag{13.19}$$

for some $\beta > 0$ *and positive matrix* B. *Then,*

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\mathbf{t}\in\mathcal{M}} X(\mathbf{t}) > u\right\} \\ &= \frac{\pi^{k/2}}{\sqrt{\mathbf{det} M^\top BM}} V(D_{\mathbf{t}_0}M)\, H_\alpha^{(k)} u^{\frac{2k}{\alpha}-\frac{k}{2}} \Psi(u - m(\mathbf{t}_0))(1+o(1)) \end{aligned}$$

as $u \to \infty$, *where* $M = M(T_{\mathbf{t}_0})$ *and* $T_{\mathbf{t}_0}$ *is the tangent subspace to* $\mathcal{M}$ *taken at the point* $\mathbf{t}_0$.

Theorem 13.2.6 *Let* $\mathcal{M} \subset \mathrm{R}^n$ *be a smooth* k*-dimensional compact,* $0 < k \le n$. *Let* $X(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, *be a differentiable in square mean sense Gaussian field with* $\mathbf{Var}X(\mathbf{t}) = 1$ *for all* $\mathbf{t} \in \mathcal{M}$ *and* $r(t,s) < 1$ *for all* $\mathbf{t}, \mathbf{s} \in \mathcal{M}$, $\mathbf{t} \ne \mathbf{s}$. *Let the expectation* $m(\mathbf{t}) = \mathbf{E}X(\mathbf{t})$ *be same as in Theorem 13.2.5. Then,*

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\mathbf{t}\in\mathcal{M}} X(\mathbf{t}) > u\right\} \\ &= \frac{\sqrt{V(\frac{1}{2}A_{\mathbf{t}_0}M)}}{\sqrt{\mathbf{det} M^\top BM}} u^{\frac{k}{2}} \Psi(u - m(\mathbf{t}_0))(1+o(1)) \end{aligned}$$

as $u \to \infty$, with M as in Theorem 13.2.5 and $A_{\mathbf{t}_0}$ the covariance matrix of the orthogonal projection of the gradient vector of the field $X(\mathbf{t})$ in point $\mathbf{t}_0$ onto the tangent subspace to the $\mathcal{M}$ taken at the point $\mathbf{t}_0$.

13.3 Proofs

PROOF OF LEMMA 13.2.1. First, observe that if one changes g_u on $g_{u\theta(u)}$, the lemma immediately follows from Lemma 6.1 of Piterbarg (1996). Second, observe that we can write $g_u T = g_{u\theta(u)}(I_u T)$, where I_u is a linear transformation of R^n, which also is a superposition of homotheties of R^{k_i} with coefficients tending to 1 as $u \to \infty$. Thus I_u tends to identity, and $I_u T$ tends to T in Euclidean distance. Third, note that $H_\alpha(T)$ is continuous in T in the topology of the space of measurable subsets of a compact, say K, generated by Euclidean distance. To prove that, observe that χ is a.s. continuous and $H_\alpha(T) \le H_\alpha(K) < \infty$, for all $T \subset K$, and use the dominated convergence theorem. These observations imply the Lemma assertion. ∎

PROOF OF THEOREM 13.2.5. Let $T_{\mathbf{t}_0}$ be the tangent plane to $\mathcal{M}$ taken at the point $\mathbf{t}_0$. Let $\mathcal{M}_0$ be a neighbourhood of $\mathbf{t}_0$ in $\mathcal{M}$, so small that it can be one-to-one projected on $T_{\mathbf{t}_0}$. We denote by P the corresponding one-to-one projector so that $P\mathcal{M}_0$ is the image of $\mathcal{M}_0$. The field $X(\mathbf{t})$, $\mathbf{t} \in \mathcal{M}$, generates on $P\mathcal{M}_0$ a field $\tilde{X}(\tilde{\mathbf{t}}) = X(\mathbf{t})$, $\tilde{\mathbf{t}} = P\mathbf{t}$. It is clear that $\mathbf{E}\tilde{X}(\tilde{\mathbf{t}}) = m(\mathbf{t}) = m(P^{-1}\tilde{\mathbf{t}})$. We denote by $\tilde{r}(\tilde{\mathbf{t}}, \tilde{\mathbf{s}}) = r(\mathbf{t}, \mathbf{s})$ the covariance function of $\tilde{X}(\tilde{\mathbf{t}})$. Choose an arbitrary $\varepsilon \in (0, \frac{1}{2})$. Due to the local stationary structure, one can find $\delta_0 = \delta(\varepsilon) > 0$ such that for all $\tilde{\mathbf{t}}_1, \tilde{\mathbf{t}}_2 \in T_{\mathbf{t}_0} \cap S(\delta_0, \mathbf{t}_0)$, where $S(\delta_0, \mathbf{t}_0)$ is centered at $\mathbf{t}_0$ ball with radius δ_0, we have

$$\exp\left\{-(1+\varepsilon)||D_{\mathbf{t}_0}(\tilde{\mathbf{t}}_1 - \tilde{\mathbf{t}}_2)||^\alpha\right\} \le \tilde{r}(\tilde{\mathbf{t}}_1, \tilde{\mathbf{t}}_2) \le \exp\left\{-(1-\varepsilon)||D_{\mathbf{t}_0}(\tilde{\mathbf{t}}_1 - \tilde{\mathbf{t}}_2)||^\alpha\right\}. \tag{13.20}$$

We also can assume δ_0 to be so small that we could let $\mathcal{M}_0 = P^{-1}\left[T_{\mathbf{t}_0} \cap S(\delta_0, \mathbf{t}_0)\right]$ and think of $P\mathcal{M}_0$ as a ball in $T_{\mathbf{t}_0}$ centered at $\tilde{\mathbf{t}}_0 = \mathbf{t}_0$, with the same radius. Denote $\mathcal{M}_1 = \mathcal{M} \setminus \mathcal{M}_0$. Since $m(\mathbf{t})$ is continuous,

$$\sup_{\mathbf{t} \in \mathcal{M}_1} m(\mathbf{t}) = m(\mathbf{t}_0) - c_0,$$

with $c_0 > 0$. By Theorem 13.2.2, for $X_0(\mathbf{t}) = X(\mathbf{t}) - m(\mathbf{t})$ we have

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\mathbf{t} \in \mathcal{M}_1} X(\mathbf{t}) > u\right\} \\ &= \mathbf{P}\left\{\sup_{\mathbf{t} \in \mathcal{M}_1} X_0(\mathbf{t}) + m(\mathbf{t}) > u\right\} \end{aligned}$$

$$\begin{aligned}
&\leq \mathbf{P}\left\{\sup_{\mathbf{t}\in\mathcal{M}_1} X_0(\mathbf{t}) > u - m(\mathbf{t}_0) + c_0\right\} \\
&= H_\alpha^{(k)} u^{\frac{2k}{\alpha}} \Psi(u - m(\mathbf{t}_0) + c_0)(1 + o(1)) \int_{\mathcal{M}_1} V(D_\mathbf{t} M_\mathbf{t})\, d\mathbf{t} \\
&= o(\Psi(u - m(\mathbf{t}_0) + c_1)), \qquad (13.21)
\end{aligned}$$

for any c_1 with $0 < c_1 < c_0$.

Now turn to $\mathcal{M}_0$. Note that

$$\mathbf{P}\left\{\sup_{\mathbf{t}\in\mathcal{M}_0} X(\mathbf{t}) > u\right\} = \mathbf{P}\left\{\sup_{\tilde{\mathbf{t}}\in P\mathcal{M}_0} \tilde{X}(\tilde{\mathbf{t}}) > u\right\}. \qquad (13.22)$$

Introduce a Gaussian stationary centered field $X_H(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, with covariance function

$$r_H(\mathbf{t}) = \exp\{-(1 + 2\varepsilon)||D_{\mathbf{t}_0}\mathbf{t}||^\alpha\}.$$

From (13.22) by Slepian inequality,

$$\mathbf{P}\left\{\sup_{\tilde{\mathbf{t}}\in P\mathcal{M}_0} \tilde{X}(\tilde{\mathbf{t}}) > u\right\} \leq \mathbf{P}\left\{\sup_{\mathbf{t}\in P\mathcal{M}_0} X_H(\tilde{\mathbf{t}}) + \tilde{m}(\tilde{\mathbf{t}}) > u\right\}. \qquad (13.23)$$

It is clear that, without loss of generality, we can put the origin of R^n at the point $\mathbf{t}_0$, so that the tangent plane $T_{\mathbf{t}_0}$ is now a tangent subspace and $\mathbf{t}_0 = \tilde{\mathbf{t}}_0 = \mathbf{0}$. From this point on, we restrict ourselves to the k-dimensional subspace $T_{\mathbf{t}_0}$ and will drop the "tilde". Let now $S = S(\mathbf{0}, \delta)$ be a ball in $T_{\mathbf{t}_0}$ centered at zero with radius δ with $\delta = \delta(u) = u^{-1/2} \log^{1/2} u$, and this choice will be clear later on. For all sufficiently large u, we have $S \subset P\mathcal{M}_0$, and there exists a positive c_1 such that

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{\mathbf{v}\in S^c\cap P\mathcal{M}_0} X_H(\mathbf{v}) + \tilde{m}(\mathbf{v}) > u\right\} \\
&\leq \mathbf{P}\left\{\sup_{\mathbf{v}\in S^c\cap P\mathcal{M}_0} X_H(\mathbf{v}) > u - \tilde{m}(\mathbf{t}_0) + c_1\delta^2(u)\right\} \\
&\leq \mathbf{P}\left\{\sup_{\mathbf{v}\in P\mathcal{M}_0} X_H(\mathbf{v}) > u - \tilde{m}(\mathbf{t}_0) + c_1\delta^2(u)\right\}. \qquad (13.24)
\end{aligned}$$

Applying Theorem 13.2.1 to the latter probability and making elementary calculations we get

$$\mathbf{P}\left\{\sup_{\mathbf{v}\in S^c\cap P\mathcal{M}_0} X_H(\mathbf{v}) + \tilde{m}(\mathbf{v}) > u\right\} = o\left(\Psi(u - m(\mathbf{t}_0))\right) \quad \text{as } u \to \infty. \quad (13.25)$$

Turn now to the ball S. Let $\mathbf{v}_1 = (v_{11}, ..., v_{n1})$, ..., $\mathbf{v}_k = (v_{1k}, ..., v_{nk})$ be an orthonormal basis in $T_{\mathbf{t}_0}$ given in the coordinates of R^n. In the coordinate system, consider the cubes

$$\Delta_\mathbf{0} = u^{-2/\alpha}[0, T]^k,\ \Delta_\mathbf{l} = u^{-2/\alpha} \times_{\nu=1}^k\ [l_\nu T, (l_\nu + 1)T],$$

$$\mathbf{l} = (l_1, ..., l_k) \in Z^k, \ T > 0.$$

We have

$$\sum_{\mathbf{i}\in L} \mathbf{P}\{A_\mathbf{i}\} - \sum_{\mathbf{i},\mathbf{j}\in L',\mathbf{i}\neq\mathbf{j}} \mathbf{P}\{A_\mathbf{i}A_\mathbf{j}\} \leq \mathbf{P}\left\{\sup_{\mathbf{v}\in S} X_H(\mathbf{v}) + \tilde{m}(\mathbf{v}) > u\right\} \leq \sum_{\mathbf{i}\in L'} \mathbf{P}\{A_\mathbf{i}\}, \tag{13.26}$$

where $A_\mathbf{i} = \left\{\sup_{\mathbf{v}\in\Delta_\mathbf{i}} X_H(\mathbf{v}) + \tilde{m}(\mathbf{v}) > u\right\}$, L' is the set of multi-indexes $\mathbf{i}$ with $\Delta_\mathbf{i} \cap S \neq \emptyset$, and L is the set of multi-indexes $\mathbf{i}$ with $\Delta_\mathbf{i} \subset S$. Using (13.19), we have

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\mathbf{v}\in\Delta_\mathbf{i}} X_H(\mathbf{v}) + \tilde{m}(\mathbf{v}) > u\right\} \\ &\quad\leq \ \mathbf{P}\left\{\sup_{\mathbf{v}\in\Delta_\mathbf{l}} X_H(\mathbf{v}) + m(\mathbf{t}_0) - \min_{\mathbf{v}\in\Delta_\mathbf{i}} ||\sqrt{B}\mathbf{v}||^2 + w_1(u) > u\right\}. \end{aligned}$$

Here, $uw_1(u) \to 0$ as $u \to \infty$ because of the choice of $\delta(u)$ and the remainder in (13.19). By Lemma 13.2.1 and the equivalence

$$\mathbf{t} = \tilde{\mathbf{t}} + O(||\tilde{\mathbf{t}}||^2) \ \text{ as } \ \mathbf{t} \to \mathbf{0}$$

(recall that we have assumed $\mathbf{t}_0 = \tilde{\mathbf{t}}_0 = \mathbf{0}$), there exists a function $\gamma_1(u)$, with $\gamma_1(u) \to 0$ as $u \to \infty$, such that for all sufficiently large u and every $\mathbf{i} \in L'$,

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\mathbf{v}\in\Delta_\mathbf{l}} X_H(\mathbf{v}) + \tilde{m}(\mathbf{v}) > u\right\} \\ &\quad\leq \ (1+\gamma_1(u))H_\alpha\left((1+\varepsilon)^{1/\alpha} D_{\mathbf{t}_0}[0,T]^k\right) \\ &\qquad\times \ \Psi\left(u - m(\mathbf{t}_0) + \min_{\mathbf{v}\in\Delta_\mathbf{l}} ||\sqrt{B}\mathbf{v}||^2 + w_1(u)\right). \end{aligned} \tag{13.27}$$

Using similar arguments, we obtain, that there exists $\gamma_2(u)$ with $\gamma_2(u) \to 0$ as $u \to \infty$, such that for all sufficiently large u and every $\mathbf{i} \in L$,

$$\begin{aligned} &\mathbf{P}\left\{\sup_{\mathbf{v}\in\Delta_\mathbf{l}} X_H(\mathbf{v}) + \tilde{m}(\mathbf{v}) u\right\} \\ &\quad\geq \ (1-\gamma_2(u))H_\alpha\left((1+\varepsilon)^{1/\alpha} D_{\mathbf{t}_0}[0,T]^k\right) \\ &\qquad\times \ \Psi\left(u - m(\mathbf{t}_0) + \min_{\mathbf{v}\in\Delta_\mathbf{l}} ||\sqrt{B}\mathbf{v}||^2 + w_2(u)\right), \end{aligned} \tag{13.28}$$

where $uw_2(u) \to 0$ as $u \to \infty$.

Now, in accordance with (13.26), we sum right-hand parts of (13.27) and (13.28) over L' and L, respectively. Using (13.7), we get for all sufficiently large u

$$\sum_{\mathbf{i}\in L'} \Psi\left(u - m(\mathbf{t}_0) + \min_{\mathbf{v}\in\Delta_\mathbf{i}} ||\sqrt{B}\mathbf{v}||^2 + w_1(u)\right)$$
$$\leq (1+\gamma_1'(u))\Psi(u - m(\mathbf{t}_0))T^{-k}u^{2k/\alpha}$$
$$\times \sum_{\mathbf{i}\in L'} \exp\left\{-u \min_{\mathbf{v}\in\Delta_\mathbf{i}} ||\sqrt{B}\mathbf{v}||^2 + o\,(1/u)\right\} T^k u^{-2k/\alpha}, \quad (13.29)$$

where $\gamma_1'(u) \to 0$ as $u \to \infty$. Changing variables $\mathbf{w} = \sqrt{u}\mathbf{t}$ and using the dominated convergence, we get

$$\sum_{\mathbf{i}\in L'} \exp\left\{-u \min_{\mathbf{v}\in\Delta_\mathbf{i}} ||\sqrt{B}\mathbf{v}||^2 + o\,(1/u)\right\}$$
$$= T^{-k} \int_{T_{\mathbf{t}_0}} \exp\{-B\mathbf{w}, \mathbf{w}\} d\mathbf{w} u^{2k/\alpha - k/2}(1 + o(1)) \quad (13.30)$$

as $u \to \infty$. Note that dw means here k-dimensional volume unite in $T_{\mathbf{t}_0}$. Similarly,

$$\sum_{\mathbf{i}\in L} \exp\left\{-u \min_{\mathbf{v}\in\Delta_\mathbf{i}} ||\sqrt{B}\mathbf{v}||^2 + o\,(1/u)\right\}$$
$$= T^{-k} \int_{T_{\mathbf{t}_0}} \exp\{-B\mathbf{w}, \mathbf{w}\} d\mathbf{w} u^{2k/\alpha - k/2}(1 + o(1)) \quad (13.31)$$

as $u \to \infty$. In order to compute the integral $\int_{T_{\mathbf{t}_0}} \exp\{-B\mathbf{w}, \mathbf{w}\} d\mathbf{w}$, we note that $\mathbf{w} = M\mathbf{t}$, where $\mathbf{t}$ denotes the vector $\mathbf{w}$ presented in the orthogonal coordinate system of $T_{\mathbf{t}_0}$, and recall that in this case $V(M) = 1$. Hence,

$$\int_{T_{\mathbf{t}_0}} \exp\{-B\mathbf{w}, \mathbf{w}\} d\mathbf{w} = \int_{T_{\mathbf{t}_0}} \exp\{-BM\mathbf{t}, M\mathbf{t}\}\, d\mathbf{t}$$
$$= \frac{\pi^{k/2}}{\sqrt{\det(M^\top BM)}} =: e^*. \quad (13.32)$$

Thus, for all sufficiently large u,

$$\sum_{\mathbf{i}\in L'} \mathbf{P}\{A_\mathbf{i}\} \leq (1+\gamma_1''(u))H_\alpha\left((1+\varepsilon)^{1/\alpha} D_{\mathbf{t}_0}[0,t]^k\right) e^* T^{-k} u^{2k/\alpha - k/2} \Psi(u - m(\mathbf{t}_0)) \quad (13.33)$$

and

$$\sum_{\mathbf{i}\in L} \mathbf{P}\{A_\mathbf{i}\} \geq (1-\gamma_1''(u))H_\alpha\left((1+\varepsilon)^{1/\alpha} D_{\mathbf{t}_0}[0,t]^k\right) e^* T^{-k} u^{2k/\alpha - k/2} \Psi(u - m(\mathbf{t}_0)), \quad (13.34)$$

where $\gamma_1''(u) \to 0$ as $u \to \infty$.

Now we are in a position to analyze the double sum in the left-hand part of (13.26). We begin with the estimation of the probability

$$\mathbf{P}\left\{\sup_{\mathbf{t}\in\Delta_1} X_H(\mathbf{t})+\tilde{m}(\mathbf{t})>u,\ \sup_{\mathbf{t}\in\Delta_2} X_H(\mathbf{t})+\tilde{m}(\mathbf{t})>u\right\},$$

with

$$\Delta_1 = u^{-2/\alpha} \times_{\nu=1}^{k} [S_\nu^1, T_\nu^1], \quad S_\nu < T_\nu,\ \nu=1,...,k,$$
$$\Delta_2 = u^{-2/\alpha}\left(\mathbf{w} + \times_{\nu=1}^{k} [S_\nu^1, T_\nu^1]\right), \quad S_\nu^1 < T_\nu^1,\ \nu=1,...,k,$$

where $\mathbf{w}$, T_ν, S_ν are such that $\rho(\Delta_1,\Delta_2)>0$, with $\rho(\cdot,\cdot)$ being the Euclidean distance in R^k. Recall that $\Delta_i \cap S(\mathbf{0},\delta(u)) \neq \emptyset$, $i=1,2$. Estimation of this probability follow the proof of Lemma 6.3 of Piterbarg (1996), but since the expectation of the field varies, more details have to be discussed, and so we give complete computations. Denote

$$K_1 = \times_{\nu=1}^{k} [S_\nu, T_\nu],\ K_2 = \mathbf{w}+K_1,\ c(u) = \max_{\mathbf{t}\in\Delta_1\cup\Delta_2} \tilde{m}(\mathbf{t}),\ \theta(u) = 1-\frac{c(u)}{u}.$$

We have

$$\mathbf{P}\left\{\sup_{\mathbf{t}\in\Delta_1} X_H(\mathbf{t})+\tilde{m}(\mathbf{t})>u,\ \sup_{\mathbf{t}\in\Delta_2} X_H(\mathbf{t})+\tilde{m}(\mathbf{t})>u\right\}$$
$$\leq \mathbf{P}\left\{\sup_{\mathbf{t}\in\Delta_1} X_H(\mathbf{t})>u\theta(u),\ \sup_{\mathbf{t}\in\Delta_2} X_H(\mathbf{t})>u\theta(u)\right\}. \quad (13.35)$$

Introduce a scaled Gaussian homogeneous field $\xi(\mathbf{t}) = X_H((1+2\varepsilon)^{-1/\alpha} D_{\mathbf{t}_0}^{-1}\mathbf{t})$. Note that

$$\mathbf{P}\left\{\sup_{\mathbf{t}\in\Delta_1} X_H(\mathbf{t})>u\theta(u),\ \sup_{\mathbf{t}\in\Delta_2} X_H(\mathbf{t})>u\theta(u)\right\}$$
$$= \mathbf{P}\left\{\sup_{\mathbf{t}\in(1+\varepsilon)^{1/\alpha}D_{\mathbf{t}_0}K_1} \xi(\mathbf{t})>u\theta(u),\ \sup_{\mathbf{t}\in(1+\varepsilon)^{1/\alpha}D_{\mathbf{t}_0}K_2} \xi(\mathbf{t})>u\theta(u)\right\}. \quad (13.36)$$

We have for the covariance function of ξ,

$$r_\xi(\mathbf{t}) = 1-||\mathbf{t}||^\alpha + o(||\mathbf{t}||^\alpha) \qquad \text{as } \mathbf{t}\to\mathbf{0}.$$

Hence there exists ε_0, $\varepsilon_0>0$, such that for all $\mathbf{t}\in B(\varepsilon_0/5)=\{\mathbf{t}: ||\mathbf{t}||^\alpha<\varepsilon_0/5\}$,

$$1-2||\mathbf{t}||^\alpha \leq r_\xi(\mathbf{t}) \leq 1-\frac{1}{2}||\mathbf{t}||^\alpha. \quad (13.37)$$

Let u be as large as

$$K_1' = (1+2\varepsilon)^{1/\alpha} D_{\mathbf{t}_0} K_1 \subset B(\varepsilon_0/5) \quad \text{and} \quad K_2' = (1+2\varepsilon)^{1/\alpha} D_{\mathbf{t}_0} K_2 \subset B(\varepsilon_0/5).$$

We have for the field $Y(\mathbf{t},\mathbf{s}) = \xi(\mathbf{t}) + \xi(\mathbf{s})$,

$$\mathbf{P}\left\{\sup_{\mathbf{t}\in K_1'} \xi(\mathbf{t}) > u\theta(u), \sup_{\mathbf{t}\in K_2'} \xi(\mathbf{t}) > u\theta(u)\right\} \le \mathbf{P}\left\{\sup_{(\mathbf{t},\mathbf{s})\in K_1'\times K_2'} Y(\mathbf{t},\mathbf{s}) > 2u\theta(u)\right\}. \tag{13.38}$$

For all $\mathbf{t} \in K_1'$, $\mathbf{s} \in K_2'$, we have $||\mathbf{t}-\mathbf{s}||^\alpha \le 2||\mathbf{t}||^\alpha + 2||\mathbf{s}||^\alpha < \varepsilon_0$. Since $D_{\mathbf{t}_0}$ is non-degenerate, for some $\kappa > 0$ and all $\mathbf{t}$, $||D_{\mathbf{t}_0}\mathbf{t}|| \ge \kappa||\mathbf{t}||$. The variance of Y equals $\sigma_Y^2(\mathbf{t},\mathbf{s}) = 2 + 2r_\xi(\mathbf{t}-\mathbf{s})$, hence for all $\mathbf{t} \in K_1'$, $\mathbf{s} \in K_2'$, we have

$$4 - 4||\mathbf{t}-\mathbf{s}||^\alpha \le \sigma^2(\mathbf{t},\mathbf{s}) \le 4 - ||\mathbf{t}-\mathbf{s}||^\alpha. \tag{13.39}$$

This implies that

$$\inf_{(\mathbf{t},\mathbf{s})\in K_1'\times K_2'} \sigma^2(\mathbf{t},\mathbf{s}) \ge 4 - 4\varepsilon_0 > 2 \tag{13.40}$$

provided ε_0 is sufficiently small, and

$$\sup_{(\mathbf{t},\mathbf{s})\in K_1'\times K_2'} \sigma^2(\mathbf{t},\mathbf{s}) \le 4 - u^{-2}(1+2\varepsilon)\kappa^\alpha\rho^\alpha(K_1,K_2) =: h(u,K_1,K_2) \tag{13.41}$$

For the standardized field $Y^*(\mathbf{t},\mathbf{s}) = Y(\mathbf{t},\mathbf{s})/\sigma(\mathbf{t},\mathbf{s})$, we have

$$\begin{aligned}&\mathbf{P}\left\{\sup_{(\mathbf{t},\mathbf{s})\in K_1'\times K_2'} Y(\mathbf{t},\mathbf{s}) > 2u\theta(u)\right\}\\ &\le \mathbf{P}\left\{\sup_{(\mathbf{t},\mathbf{s})\in K_1'\times K_2'} Y^*(\mathbf{t},\mathbf{s}) > 2u\theta(u)h^{-1/2}(u,K_1,K_2)\right\}.\end{aligned} \tag{13.42}$$

Algebraic calculations give

$$\mathbf{E}(Y^*(\mathbf{t},\mathbf{s}) - Y^*(\mathbf{t}_1,\mathbf{s}_1))^2 \le 16(||\mathbf{t}-\mathbf{t}_1||^\alpha + ||\mathbf{s}-\mathbf{s}_1||^\alpha). \tag{13.43}$$

Let $\eta_1(\mathbf{t})$, $\eta_2(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, be two independent identically distributed homogeneous Gaussian fields with expectations equal zero and covariance functions equal

$$r^*(\mathbf{t}) = \exp(-32||\mathbf{t}||^a).$$

Gaussian field

$$\eta(\mathbf{t},\mathbf{s}) = \frac{1}{\sqrt{2}}\left(\eta_1(\mathbf{t}) + \eta_2(\mathbf{s})\right), \qquad (\mathbf{t},\mathbf{s}) \in \mathrm{R}^n \times \mathrm{R}^n.$$

is homogeneous, and its covariance function is

$$r^{**}(\mathbf{t},\mathbf{s}) = \frac{1}{2}\left(\exp(-32||\mathbf{t}||^a + \exp(-32||\mathbf{s}||^a)\right). \tag{13.44}$$

As far as for the covariance function $r^{***}(\mathbf{t},\mathbf{s};\mathbf{t}_1,\mathbf{s}_1)$ of the field Y^*, we have

$$r^{***}(\mathbf{t},\mathbf{s};\mathbf{t}_1,\mathbf{s}_1) \ge 1 - 8(||\mathbf{t}-\mathbf{t}_1||^a + ||\mathbf{s}-\mathbf{s}_1||^a) \tag{13.45}$$

for all $(\mathbf{t},\mathbf{s}),(\mathbf{t}_1,\mathbf{s}_1)\in K_1'\times K_2'$, for these $(\mathbf{t},\mathbf{s}),(\mathbf{t}_1,\mathbf{s}_1)$ we also have that

$$r^{***}(\mathbf{t},\mathbf{s};\mathbf{t}_1,\mathbf{s}_1)\geq r^{**}(\mathbf{t}-\mathbf{t}_1;\mathbf{s}-\mathbf{s}_1).$$

Thus by Slepian inequality,

$$\mathbf{P}\left\{\sup_{(\mathbf{t},\mathbf{s})\in K_1'\times K_2'} Y^*(\mathbf{t},\mathbf{s})>2u\theta(u)h^{-1/2}(u,K_1,K_2)\right\}$$
$$\leq \mathbf{P}\left\{\sup_{(\mathbf{t},\mathbf{s})\in K_1'\times K_2'} \eta(\mathbf{t},\mathbf{s})>2u\theta(u)h^{-1/2}(u,K_1,K_2)\right\}. \qquad (13.46)$$

Further, for sufficiently large u,

$$4u^2\theta^2(u)h^{-1}(u,K_1,K_2)\geq u^2\theta^2(u)+\frac{\kappa^\alpha}{5}\rho^\alpha(K_1,K_2). \qquad (13.47)$$

Using the last two relations, Lemma 13.2.1 and (13.7), we get

$$\mathbf{P}\left\{\sup_{\mathbf{t}\in\Delta_1} X_H(\mathbf{t})+\tilde{m}(\mathbf{t}_0+\mathbf{t})u,\ \sup_{\mathbf{t}\in\Delta_2} X_H(\mathbf{t})+\tilde{m}(\mathbf{t}_0+\mathbf{t})u\right\}$$
$$\leq C\Psi(u\theta(u))H_\alpha(16(D_{\mathbf{t}_0}K_1\times D_{\mathbf{t}_0}K_2))\exp\left(-\frac{\kappa^\alpha}{10}\rho^\alpha(K_1,K_2)\right)$$
$$\leq C_1\prod_{\nu=1}^{k}(T_\nu^1-S_\nu^1)\prod_{\nu=1}^{k}(T_\nu^2-S_\nu^2)\exp\left(-\frac{\kappa^\alpha}{10}\rho^\alpha(K_1,K_2)\right)\Psi(u\theta(u)), \qquad (13.48)$$

which holds for all sufficiently large u and a constant C_1, independent of u, K_1, K_2. In order to estimate $H_\alpha(16(D_{\mathbf{t}_0}K_1\times D_{\mathbf{t}_0}K_2))$, we use here Lemmas 6.4 and 6.2 from Pickands (1969).

Now turn to the double sum $\sum_{\mathbf{i},\mathbf{j}\in L'}\mathbf{P}(A_\mathbf{i}A_\mathbf{j})$. We break it into two sums. The first one, denoted by Σ_1, is the sum over all non-neighbouring cubes (that is, the distance between any two of them is positive), and the second one, denoted by Σ_2, is the sum over all neighbouring cubes. Denote

$$x_\mathbf{i}=\min_{\mathbf{t}\in\Delta_\mathbf{i}}||\sqrt{B}\mathbf{t}||,\quad \mathbf{i}\in L'.$$

Using (13.48), we get

$$\mathbf{P}(A_\mathbf{i}A_\mathbf{j})\quad\leq C^\alpha T^{2k}\exp\left(-\frac{\kappa^\alpha}{10}T^\alpha(\max_{1\leq\nu\leq k}|i_\nu-j_\nu|-1)^\alpha\right)\Psi(u\theta(u))=:\theta_{\mathbf{i},\mathbf{j}}, \qquad (13.49)$$

where $\theta(u)=1-c(u)$, $c(u)=\max\{\max_{\mathbf{t}\in\Delta_\mathbf{i}}\tilde{m}(\mathbf{t}_0+\mathbf{t}),\max_{\mathbf{t}\in\Delta_\mathbf{j}}\tilde{m}(\mathbf{t}_0+\mathbf{t})\}$. This estimation holds for all members of the first sum and all sufficiently large u. Using it and approximating the sum by an integral, we get

$$\Sigma_1\leq 2\sum_{\mathbf{i}\in L'}\ \sum_{\mathbf{j}\in L',\,\mathbf{i}\neq\mathbf{j},\,x_\mathbf{i}<x_\mathbf{j}}\theta_{\mathbf{i},\mathbf{j}}\leq c^*e^*T^k\exp\left(-\frac{\kappa^\alpha}{10}T^\alpha\right)u^{\frac{2k}{\alpha}-\frac{k}{2}}\Psi(u-m(\mathbf{t}_0)). \qquad (13.50)$$

Now consider Σ_2. We can assume that $\max_{\mathbf{t}\in\Delta_\mathbf{i}} \tilde{m}(\mathbf{t}_0+\mathbf{t}) > \max_{\mathbf{t}\in\Delta_\mathbf{j}} \tilde{m}(\mathbf{t}_0+\mathbf{t})$. Denote

$$\Delta'_\mathbf{i} = u^{\frac{2}{\alpha}} \left([i_1 T, i_1 T + \sqrt{T}] \times \times_{\nu=2}^{k} [i_\nu T, (i_n u + 1)T] \right) \quad \text{and} \quad \Delta''_\mathbf{i} = \Delta_\mathbf{i} \setminus \Delta'_\mathbf{i}.$$

Clearly,

$$\begin{aligned} \mathbf{P}\{A_\mathbf{i} A_\mathbf{j}\} \leq \mathbf{P}\left\{ \sup_{\Delta_\mathbf{i}} X_H(\mathbf{t}) > u\theta(u) \right\} & \\ + \mathbf{P}\left\{ \sup_{\Delta''_\mathbf{i}} X_H(\mathbf{t}) > u\theta(u), \sup_{\Delta_\mathbf{j}} X_H(\mathbf{t}) > u\theta(u) \right\}. & \quad (13.51) \end{aligned}$$

Using now Lemma 13.2.1, (13.51), (13.48) and approximating the sum by an integral, we get for all sufficiently large u

$$\begin{aligned} \Sigma_2 \leq\ & C_2^* T^{k-1/2} e^* u^{\frac{2k}{\alpha}-\frac{k}{2}} \Psi(u - m(\mathbf{t}_0)) \\ & + C_3^* T^k e^* u^{\frac{2k}{\alpha}-\frac{k}{2}} \exp\left\{ -\frac{\kappa^\alpha T^\alpha}{10} \right\} \Psi(u - m(\mathbf{t}_0)). \qquad (13.52) \end{aligned}$$

Taking into account (13.33), (13.34), (13.50) and (13.52), we get for all positive T

$$\begin{aligned} & \frac{H_\alpha\left((1+2\varepsilon) D_{\mathbf{t}_0}[0,t]^k\right)}{T^k} - -C_1^* T^k \exp\left\{ -\frac{m^\alpha T^\alpha}{10} \right\} - C_2^* T^{-1/2} \\ & \quad - C_3^* T^k u^{\frac{2k}{\alpha}-\frac{k}{2}} \exp\left\{ -\frac{m^\alpha T^\alpha}{10} \right\} \\ & \leq \liminf_{u\to\infty} \frac{\mathbf{P}\left\{ \sup_{\mathbf{t}\in S} X_H(\mathbf{t}) + \tilde{m}(\mathbf{t}_0 + \mathbf{t}) > u \right\}}{e^* u^{\frac{2k}{\alpha}-\frac{k}{2}} \Psi(u - m(\mathbf{t}_0))} \\ & \leq \limsup_{u\to\infty} \frac{\mathbf{P}\left\{ \sup_{\mathbf{t}\in S} X_H(\mathbf{t}) + \tilde{m}(\mathbf{t}_0 + \mathbf{t}) > u \right\}}{e^* u^{\frac{2k}{\alpha}-\frac{k}{2}} \Psi(u - m(\mathbf{t}_0))} \\ & \leq \frac{H_\alpha\left((1+2\varepsilon) D_{\mathbf{t}_0}[0,t]^k\right)}{T^k}. \qquad (13.53) \end{aligned}$$

Now, letting T go to infinity and using (13.25), we obtain

$$\begin{aligned} & \mathbf{P}\left\{ \sup_{\mathbf{t}\in S} X_H(\tilde{\mathbf{t}}) + \tilde{m}(\tilde{\mathbf{t}}) > u \right\} \\ & \quad = (1+2\varepsilon)^k e^* V(D_{\mathbf{t}_0} M) H_\alpha^{(k)} u^{\frac{2k}{\alpha}-\frac{k}{2}} \Psi(u - m(\mathbf{t}_0))(1 + o(1)) \\ & \qquad (13.54) \end{aligned}$$

as $u \to \infty$.

Let now $X_H^*(\mathbf{t})$, $\mathbf{t} \in \mathrm{R}^n$, be a homogeneous centered Gaussian field with the covariance function $r_H^*(\mathbf{t}) = \exp(-(1-2\varepsilon)||D_{\mathbf{t}_0}\mathbf{t}||^\alpha)$. From Theorem 13.2.4, we have

$$\mathbf{P}\left\{\sup_{\mathbf{t}\in S} \tilde{X}(\tilde{\mathbf{t}}) > u\right\} \geq \mathbf{P}\left\{\sup_{\mathbf{t}\in S} X_H^*(\tilde{\mathbf{t}}) + \tilde{m}(\tilde{\mathbf{t}}) > u\right\}. \tag{13.55}$$

Proceeding as above for the latter probability, we get

$$\mathbf{P}\left\{\sup_{\mathbf{t}\in\Sigma_\Pi} X_H^*(\tilde{\mathbf{t}}) + \tilde{m}(\tilde{\mathbf{t}} > u\right\}$$
$$= (1-2\varepsilon)^k e^* V(D_{\mathbf{t}_0}M) H_\alpha^{(k)} u^{\frac{2k}{\alpha}-\frac{k}{2}} \Psi(u - m(\mathbf{t}_0))(1+o(1)) \tag{13.56}$$

as $u \to \infty$.

Now we collect (13.21), (13.23), (13.54), (13.55) and (13.56), and get

$$\begin{aligned}(1-2\varepsilon)^k &\leq \liminf_{u\to\infty} \frac{\mathbf{P}\left\{\sup_{\mathbf{t}\in\mathcal{M}} X(\mathbf{t}) > u\right\}}{e^* V(D_{\mathbf{t}_0}M) H_\alpha^{(k)} u^{\frac{2k}{\alpha}-\frac{k}{2}} \Psi(u - m(\mathbf{t}_0))} \\ &\leq \limsup_{u\to\infty} \frac{\mathbf{P}\left\{\sup_{\mathbf{t}\in\mathcal{M}} X(\mathbf{t}) > u\right\}}{e^* V(D_{\mathbf{t}_0}M) H_\alpha^{(k)} u^{\frac{2k}{\alpha}-\frac{k}{2}} \Psi(u - m(\mathbf{t}_0))} \leq (1+2\varepsilon)^k.\end{aligned} \tag{13.57}$$

It follows from this the assertion of the Theorem. ∎

Proof of Theorem 13.2.6. Let $\tilde{X}(\tilde{\mathbf{t}})$ be the field as it is defined in the proof of Theorem 13.2.5. Using Tailor expansion, we get

$$\tilde{X}(\tilde{\mathbf{t}}) = X(\mathbf{t}) = X(\mathbf{t}_0) + (\mathrm{grad}X(\mathbf{t}_0))^\top(\mathbf{t}-\mathbf{t}_0) + o(||\mathbf{t}-\mathbf{t}_0||), \quad \mathbf{t}\to\mathbf{t}_0. \tag{13.58}$$

From here, it follows that

$$\tilde{X}(\tilde{\mathbf{t}}) - \tilde{X}(\tilde{\mathbf{t}}_0) = (\widetilde{\mathrm{grad}}X(\mathbf{t}_0))^\top(\tilde{\mathbf{t}}-\tilde{\mathbf{t}}_0) + o(||\tilde{\mathbf{t}}-\tilde{\mathbf{t}}_0||), \quad \tilde{\mathbf{t}}\to\tilde{\mathbf{t}}_0, \tag{13.59}$$

where $\widetilde{\mathrm{grad}}$ is the orthogonal projection of the gradient of the field X onto the tangent subspace $T_{\mathbf{t}_0}$ to the $\mathcal{M}$ at the point $\mathbf{t}_0$. From (13.59), it follows that

$$\tilde{r}(\tilde{\mathbf{t}}-\tilde{\mathbf{t}}_0) = 1 - \frac{1}{2}(\tilde{\mathbf{t}}-\tilde{\mathbf{t}}_0)^\top A_{\mathbf{t}_0}(\tilde{\mathbf{t}}-\tilde{\mathbf{t}}_0) + o(||\tilde{\mathbf{t}}-\tilde{\mathbf{t}}_0||), \quad \tilde{\mathbf{t}}\to\tilde{\mathbf{t}}_0, \tag{13.60}$$

where $A_{\mathbf{t}_0}$ is the covariance matrix of the vector $\widetilde{\mathrm{grad}}X(\mathbf{t}_0)$. Note that the matrix $\sqrt{A_{\mathbf{t}_0}/2}$ is just the matrix $D_{\mathbf{t}_0}$ from Theorem 13.2.5. Now the proof repeats up to all details of the proof of Theorem 13.2.5. ∎

Acknowledgements. V. Piterbarg was partially supported by DFG-RFFI grant 36RUS113/4ti67/7 and RFFI Grants 97-01-00648, 98-01-00524. He also used the hospitality of the University of Montenegro during Summer, 1997.

References

1. Adler, R. J. (1998). On excursion sets, tube formulae, and maxima of random fields, *Manuscript*, Technion, Israel and University of North Carolina.

2. Belyaev, Yu. and Piterbarg, V. (1972). Asymptotic behaviour of the mean number of A-points of excursions of Gaussian field beyong high level, In *Excursions of Random Fields* (Ed., Yu. Belyaev), pp. 62–88, Moscow University. (In Russian)

3. Fatalov, V. R. and Piterbarg, V. I. (1995). The Laplace method for probability measures in Banach spaces, *Russian Math. Surveys*, **50**, 1151–1239.

4. Mikhaleva, T. L. and Piterbarg, V. I. (1996). On the maximum distribution of a Gaussian field with constant variance on a smooth manifold, *Probability Theory and Its Applications*, **41**, 438–451.

5. Borell, C. (1976). The Brunn-Minkovski inequality in Gauss space, *Invent. Math.*, **30**, 207–216.

6. Pickands, J., III. (1969). Upcrossings probabilities for stationary Gaussian process, *Transactions of the American Mathematical Society*, **145**, 51–73.

7. Piterbarg, V. I. (1996). *Asymptotic Methods in the Theory of Gaussian Processes and Fields*, Translations of Mathematical Monographs, American Mathematical Society, Providence, Rhode Island.

8. Piterbarg, V. I. and Tyurin, Yu. (1993). Homogeneity testing of two samples: Gaussian field on a sphere, *Mathematical Methods of Statistics*, **2**, 147-164.

9. Piterbarg, V. I. and Tyurin, Yu. (1999). A multidimensional rank correlation: A Gaussian field on a torus, *Probability Theory and Its Applications* (to appear).

10. Slepian, D. (1962). The one-sided barrier problem for the Gaussian noise, *Bell System Technical Journal*, **41**, 463–501.

14

Typical Distributions: Infinite-Dimensional Approaches

A. V. Sudakov, V. N. Sudakov, and H. v. Weizsäcker

Steklov Mathematical Institute, St. Petersburg, Russia
Steklov Mathematical Institute, St. Petersburg, Russia
University of Kaiserslautern, Kaiserslautern, Germany

Abstract: Some approaches to possible infinite-dimensional versions of the phenomenon of existence of typical distributions for vector spaces of random variables are under discussion and comparison.

Keywords and phrases: Measure concentration, typical distributions

14.1 Results

The existence and the structure of typical distributions of linear functionals on high-dimensional vector spaces with second order probability measures [Sudakov (1978), Nagaev (1982), Makarova (1985) and Weizsäcker (1997)] as well as the existence of typical distributions for finite-dimensional vector spaces of high dimension of 2nd order random variables [Sudakov (1994)] are manifestations of a general concentration of measure phenomenon, which was intensively studied last years by Gromov, Milman, Talagrand and others. The majority of the known theorems about the existence of typical distributions deal with finite-dimensional vector spaces (measure spaces or spaces of random variables). The wish to obtain an extension of these results to the infinite-dimensional case seems to be quite natural, though the very existence of such an adequate extension is not evident and not trivial.

Let E be a finite-dimensional vector space of random variables X with finite variances. This space is endowed with the canonical Euclidean structure induced from $L^2(\Omega, \mathcal{F}, \boldsymbol{P})$, and hence, the notion of the standard Gaussian measure γ_E on E is well defined as well as the notion of rotation invariant distributions on E. Let $\mathcal{M} = \mathcal{M}(\boldsymbol{R})$ stand for the space of all probability measures on $\boldsymbol{R}$ with finite first moment. The Kantorovich-Rubinstein distance

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

$\kappa(\mu,\nu) = \int |F_\mu(u) - F_\nu(u)|\, du$ will be usually considered on $\mathcal{M}$. Note that $\mathcal{M}$ consists of all probability measures on $\boldsymbol{R}$ such that their κ-distance from some (and, hence, any) δ-measure is finite. For arbitrary separable metric space S, the space $\mathcal{M}(S)$ is defined in a similar way.

The existence of typical distributions phenomenon consist in existence, for every arbitrary small $\varepsilon > 0$ and every E of dimension $d = \dim E$ large enough (depending on ε only), a probability distribution $\boldsymbol{P}_E \in \mathcal{M}$ (depending on E) such that it is "$(1-\varepsilon)$-typical for distributions $\mathcal{L}(X) = \boldsymbol{P} \circ X^{-1}$", $X \in E$ being chosen at random according to a "natural" probability distribution m on E. For instance, m may be the image of γ_E under the homothety $E \ni X \longmapsto (\dim E)^{-1/2} X$ or the probability measure uniformly concentrated on the unit ball or on the unit sphere (on the surface of the ball) of E. Let m_E denote the uniform probability distribution on the sphere. It is known that for large dimension these three kinds of "natural" distributions are close each to other (for instance, in sense of Kantorovich-Rubinstein distance). Here, "$(1-\varepsilon)$-typical" means that m-measure of the set of all such random variables X from E, for which $\kappa(\boldsymbol{P}_E, \mathcal{L}(X)$ does not exceed ε, is $(1-\varepsilon)$-massive:

$$m(\, X \in E : \kappa(\boldsymbol{P}_E, \mathcal{L}(X)) \le \varepsilon \,) \ge 1 - \varepsilon.$$

The measure $\boldsymbol{P}_E$ can always be chosen from the set of all mixtures of centered Gaussian univariate distributions. Instead of the Euclidean structure on E generated by the measure $\boldsymbol{P}$, any stronger Euclidean norm can be used for definition of the class $\mathcal{MN}_E$ of "natural" (i.e. certain rotation-invariant with respect to this stronger norm) measures m on E with not worse (not larger) rate of increasing $d(\varepsilon)$ or, the same, with not smaller rate of increasing of typicalness in dimension d.

In other words, the phenomenon of existence of typical distributions means that the image $\mathcal{L}m = m \circ \mathcal{L}^{-1}$ of the measure m under the map $\mathcal{L}\colon E \ni X \longmapsto \mathcal{L}(X) \in (\mathcal{M}, \kappa)$ is sharply κ-concentrated close to some element P_E of $\mathcal{M}$, which is just typical distribution for elements of E.

Another form of manifestation of this phenomenon is the existence of "typical distributions of linear functionals", or "typical marginals". Given a vector space $F = E'$ of large dimension with 2nd order probability measure $\boldsymbol{P}$ on it, consider the (Euclidean) trace of $L^2(\boldsymbol{P})$-norm on the conjugate space $F' = E'' = E$. Then for large dimensions for "natural" (in the previous sense) distributions m on the Euclidean space E, the assertion about existence of a typical for linear functionals on E' distribution with respect to $\boldsymbol{P}$ holds similarly to the previous case.

Trivial reformulations of the given finite-dimensional assertions for the infinite-dimensional case are senseless or even wrong: m-typical "elements" of E turn out not to be random variables for seemingly reasonable m. The property of $(1-\varepsilon)$-typicalness of a distribution $\boldsymbol{P}_E$ for $\varepsilon = 0$ must mean something like "essentially all (in an appropriate sense) $X \in E$ have the same distribution $\boldsymbol{P}_E$", what seems to be wrong for whatever reasonable sense. Indeed, in the separable infinite-dimensional Hilbert space H (corresponding to an infinite-dimensional E) there exists no non-trivial rotation-invariant measure. "The standard Gaussian measure γ_H" (or "the Gaussian white noise") is only a "weak distribution" and not a countably additive measure. Its "typical sample elements" may only be considered as elements of a suitable extension (the completion in a suitable weaker norm) of H and do not belong to H. They cannot be interpreted as linear functionals on E' or even as elements of $L^0(\Omega, \mathcal{F}, \boldsymbol{P})$. It is impossible to define their "distributions" with respect to $\boldsymbol{P}$ as elements of $\mathcal{M}$. Thus, in the case of infinite-dimensional E, there is a problem to justify the term "essentially all $X \in E$" in order to obtain a meaningful extension of the finite-dimensional assertion.

One of the seemingly reasonable ways to formulate a reasonable infinite-dimensional version for discussion is to define the notion of "the global limit distribution," or "the limit distribution in large" for the space E of second order random variables as follows. For infinite-dimensional separable Hilbert space E, we denote by γ_E the standard Gaussian white noise considered as a Gaussian measure on a Hilbert (with a weaker norm) superspace $\widehat{E} \supset E$. We preserve the notation X for elements of $\widehat{H}$, too. One can establish that γ_E is rotation invariant with respect to the Hilbert norm in E. Let $\xi_1 < \xi_2 < \ldots$, $\sup_i \xi_i = \epsilon$, where ξ_i are finite measurable partitions of the probability measure space $(\widehat{E}, \gamma_E)$, and ε denotes the partition into points. It can be verified that every barycenter (the mean value) $c_{i,j} \in \widehat{E}$ of the conditional distribution $\gamma_E(\,\cdot\,|C_{i,j})$, where $C_{i,j}$ stands for the jth element of ξ_i, is, in fact, an element of E, i.e., $\|c_{i,j}\| < \infty$.

We denote by $m^{\xi_i}_{\gamma_E}$ the discrete probability measure on E concentrated at the normalized barycenters $c^0_{i,j} = \|c_{i,j}\|^{-1} c_{i,j}$, $j = 1, \ldots$, and such that $m^{\xi_i}_{\gamma_E}(\{c^0_{i,j}\}) = \gamma_E(C_{i,j})$, $j = 1, \ldots$. Similar to the above definition of $\mathcal{L}m$, we define the corresponding distribution $\mathcal{L}m^{\xi_i}_{\gamma_E} = m^{\xi_i}_{\gamma_E} \circ \mathcal{L}^{-1}$ on $\mathcal{M}$. For $X \in \widehat{E}$, let $C_i(X)$ denote the element of the partition ξ_i containing X, and $c^0_i(X)$ its normalized barycenter. The map $\widehat{E} \ni X \longmapsto \mathcal{L}(c^0_i(X)) \in \mathcal{M}$ is a step-function, and in the finite-dimensional case its distribution in $\mathcal{M}$ tends to $\mathcal{L}m$ as $i \to \infty$ (here m is the uniform distribution on the unit sphere of E). For large dimensions, the distribution $\mathcal{L}m$ is sharply concentrated close to a point $\boldsymbol{P}_E$ which is just typical for random variables from E.

If for an infinite-dimensional Hilbert space E of random variables there exists a distribution $\boldsymbol{P}_E \in \mathcal{M}$ such that for every sequence ξ_i of finite partitions of $(\widehat{E}, \gamma_E)$ the sequence $\mathcal{L}m^{\xi_i}_{\gamma_E}$ converges to the degenerate distribution concentrated at $\boldsymbol{P}_E$, we say that $\boldsymbol{P}_E$ is *the global limit distribution*, or *the limit in large distribution with respect to* γ_E *of elements of* E. The simplest example: the global distribution for any Gaussian subspace of L^2 is the standard Gaussian measure.

In contrast to the finite-dimensional case, such a limit distribution exists not for every infinite-dimensional space E. Still, the common point is that an arbitrary limit distribution, if it corresponds to any sequence of finite partitions, is always degenerate, i.e., the limit in measure of the sequence of step-maps is a constant map. It should be noted that another reasonable definition of such a limit distribution turned out to be equivalent to one given above [Sudakov (1977)].

Theorem 14.1.1 *If for any sequence $\{\xi_i\}$ of finite measurable partitions of the space $(\widehat{E}, \gamma_E)$ the sequence $\{\mathcal{L}m_{\gamma_E}^{\xi_i}\}$ of elements of $\mathcal{M}(\mathcal{M}(\boldsymbol{R}))$ tends to a limit, then this limit is a degenerate (delta-) distribution, i.e., there exists a global limit distribution of elements of E. The global limit distribution is always a mixture of centered Gaussian ones.*

The question about criterion of existence of such a limit "in large" distribution is open.

Another and, probably, more rich in content approach to the infinite-dimensional situation arose from efforts to find for the finite-dimensional case how the smallest possible dimension of the space E depends on ε. Of course, estimates that do not depend on dimension are of particular interest. It is convenient to characterize the degree of possible typicalness of a distribution for a vector space E with respect to a measure m by the value of the average κ-distance between distribution laws of two elements of E chosen independently according to m, i.e., by the value of the integral

$$I(m) = \int \int_{E \oplus E} \kappa(\mathcal{L}(X_1), \mathcal{L}(X_2))(m \otimes m)\, d(X_1, X_2).$$

Here $E \oplus E$ means the orthogonal sum of two copies of E, and X_1, X_2 is the notation of elements of these copies. Since in the infinite-dimensional case this expression is senseless for $m = m_E$, it is important to study its behavior in m for finite-dimensional spaces E. For centered Gaussian measure γ on E, let $q(\gamma)$ be the quadratic form on the conjugate space $F = E'$, which is the restriction of the $L^2(\gamma)$-norm. We also use the evident notation $\gamma(q)$; sometimes we shall write $I(q(\gamma))$ instead of $I(\gamma)$. We say that $\gamma_1 < \gamma_2$ if $\gamma_2 = \gamma_1 * \gamma_3$ for some Gaussian measure γ_3 (or, the same, $q(\gamma_2) = q(\gamma_1) + q(\gamma_3)$). One can easily verify that in the one-dimensional case $I(\gamma)$ is monotonic in γ (or in q). Also for arbitrary dimension $d = \dim E$, any positive perturbation of $q(z) = q_0(z) = \|z\|_F^2$ (the case of the standard Gaussian measure) or, what is the same, of homothetical image of γ_E mentioned above leads to increasing of $I(q)$. Since $I(q_0)$ can be well estimated, it would be very useful to prove the monotonicity in q property of the function $I(q)$. In Sudakov (1979), one of the co-authors of this paper proposed, in particular, a short sketch of a supposed proof of such a monotonicity, which turned out to be wrong. The counterexample constructed by another co-author is based on the deduced explicit formula for the differential $D(\bar{q})(q)$ of the function $I(q)$ at an arbitrary point $\bar{q}$ and can be described as follows.

Let $(\Omega, \mathcal{F}, \boldsymbol{P})$ be a probability space with $\Omega = \{\omega_1, \omega_2\}$ and $\boldsymbol{P}(\{\omega_1\}) = \boldsymbol{P}(\{\omega_2\}) = \frac{1}{2}$. Let $E = \{(x,y)\}$ be the Euclidean space of all functions on Ω, i.e., the space $\boldsymbol{R}^2$ with corresponding weight norm. Let e_x and e_y denote coordinate functionals on E. Let the Gaussian measure $\gamma(\bar{q}_\delta)$ be a univariate Gaussian measure on the straight line $\delta x - y = 0$, where δ is a small negative number. Let $\gamma(q_\varepsilon)$ be a univariate Gaussian measure specified by the quadratic form $q_\varepsilon(e_x, e_y) = \varepsilon(e_x - e_y)^2$, i.e., for small ε it is sharply concentrated on the straight line $x + y = 0$ near zero point.

Theorem 14.1.2 *If $-\delta > 0$ and $\varepsilon > 0$ are sufficiently small, then $I(\bar{q}_\delta + q_\varepsilon) - I(\bar{q}_\delta) < 0$.*

Differential of $I(q)$ can be described as follows. Let for positive quadratic form q the Laplace operator with respect to the Euclidean norm $q^{\frac{1}{2}}$ is denoted by Δ_q.

Proposition 14.1.1 *The differential of I at the point $q \in Q_+$ (Q_+ is the cone of positive quadratic forms) can be written in the form*

$$DI(\bar{q})(q) = \frac{1}{2}\int\int_{E\oplus E} \Delta_{q\oplus q}\kappa(\mathcal{L}(X_1), \mathcal{L}(X_2))\gamma(\bar{q}\oplus\bar{q})\, d(X_1, X_2).$$

Here $\bar{q}\oplus\bar{q}$ denotes the quadratic form on $(E\oplus E)' = E'\oplus E'$ such that $(\bar{q}\oplus\bar{q})(f_1, f_2) = \bar{q}(f_1) + \bar{q}(f_2)$, $\gamma(\bar{q}\oplus\bar{q}) = \gamma(\bar{q})\otimes\gamma(\bar{q})$.

Eventhough the result of Theorem 14.1.2 is negative for our purpose, some theorems, which definitely can be considered as infinite-dimensional versions of the theorem about typical distributions, can be formulated and proved. They enable us to obtain the finite-dimensional versions with the best possible estimates of "the rate of convergence" in dimension and imply other interesting consequences. One of the possible directions of investigations is finding upper bounds for $I(q)$ in terms of the spectral radius of the correlation operator of $\gamma(q)$, i.e., dimension free estimates of $I(q)$ in terms of the first extremal value of q with respect to q_0.

A convenient tool for investigation of the problems under consideration is the solution of the isoperimetric problem for Gaussian measures [Sudakov and Tsirel′son (1974)]; see also Borell (1975). This chapter just gave the solution of such an isoperimetric problem for the infinite-dimensional case. Recall the main result of this chapter related to the infinite-dimensional centered Gaussian measure γ on a vector space E ((E,γ) is always supposed to be a Lebesgue-Rokhlin space; this always holds for separable metric E and Borel γ). If a measurable functional $R\colon E \longrightarrow \boldsymbol{R}$ obeys the γ-Lipschitz condition

$$\forall e \in \mathcal{E}_\gamma\ \forall\varepsilon > 0\ \ \gamma\{\, X \in E : |R(f + \varepsilon e) - R(e)| > \varepsilon\,\} = 0,$$

then the distribution law of R relative to γ is sublaplacean, i.e., $F_R^{-1} - \Phi^{-1}$ is non-strictly decreasing function (here $\mathcal{E}_\gamma$ denotes the "ellipsoid of dispersion of γ", i.e., the unit ball of the reproduction kernel space H_γ). Note that here the dimension plays no role as well as whatever topology on E.

For an arbitrary (possibly, infinite-dimensional) subspace $E \subset L^2(\Omega, \mathcal{F}, \boldsymbol{P})$ with a centered Gaussian measure $\gamma = \gamma(p)$ (supposed in what follows to be "substandard", i.e., obeying the condition $p \leq p_0$), the degree of concentration of the measure $\Gamma = \gamma \circ (\boldsymbol{P} \circ X^{-1})^{-1} = \gamma \circ \mathcal{L}^{-1}$ on the space $\mathcal{M}$ can be measured by the degree of concentration of the distribution of $\kappa(P_1, P_2)$, where P_1, P_2 are two elements of $\mathcal{M}$ chosen independently according to Γ.

We say that $\mu \in \mathcal{M}$ is sublaplacean if $\mu = \gamma_0 \circ T^{-1}$, where γ_0 is the univariate standard Gaussian measure and T is a Lipschitz map $\boldsymbol{R} \to \boldsymbol{R}$ with the Lipschitz constant 1. For a sublaplacean measure μ, a constant c is called its (*admissible*) *shift* if, in the above representation of μ, T can be chosen such that $T(0) = c$. For every sublaplacean measure μ, the set of all its shifts is a segment, which is degenerate for $\mu = \gamma_0$.

Theorem 14.1.3 *There exists a constant C with the following property. Let $E \subset L^2(\Omega, \mathcal{F}, \boldsymbol{P})$ be an arbitrary closed subspace of random variables. Let $\gamma = \gamma(q)$ be an arbitrary centered Gaussian measure on E with covariance operator majorized by the unit operator (i.e., the eigenvalues of this covariance operator do not exceed 1, $q \leq q_0$). Let X_1 and X_2 be two elements of E chosen independently from the distribution γ. Then the distribution of the random variable $\kappa(\boldsymbol{P} \circ X_1^{-1}, \boldsymbol{P} \circ X_2^{-1})$ is sublaplacean with a shift less than or equal to C.*

The value of C is closely connected with some entropy-type properties of $\mathcal{M}$. Theorem 14.1.3 allows us to obtain an upper bound of the value $I(\gamma)$ over the class of substandard ($\gamma \leq \gamma_E$) Gaussian measures for a space E of arbitrary dimension, though does not permit us to come to the conclusion that for finite-dimensional E the maximum value of $I(\gamma)$ is attained just at the standard Gaussian measure $\gamma = \gamma_E$. Still, Theorem 14.1.2 does not exclude such a possibility, though closes one of the approaches to prove it.

As an application of this theorem, an estimate of the (random) κ-distance between times of sojourn for two independently chosen sample functions of a centered Gaussian random process in terms of maximal eigenvalue of its covariance operator can be obtained. (The time of sojourn is the image of a given measure on the parametric set by a sample function. The density of time of sojourn is often called "the local time".) Proofs and consequences are to be published in *Zapiski Nauchnykh Seminarov POMI.*

A different infinite-dimensional approach to the problem has been given in Weizsäcker (1997). Here, we explain how to translate some part of Weizsäcker (1997) into the present setting. That paper deals with a, not necessarily Gaussian, random linear functional resp. cylindrical measure, on a Hilbert space

$\mathcal{H}$. This corresponds to our γ which induces the canonical Gaussian cylindrical measure on its "reproducing kernel" or "Cameron-Martin" subspace $\mathcal{H} \subset L$. Our abstract measure $\boldsymbol{P}$ is replaced in Weizsäcker (1997) by a (not necessarily second order) measure on $\mathcal{H}$. In our setting, such a measure on $\mathcal{H}$ can be constructed as follows: Choose an orthonormal base (e_i) of the Cameron-Martin space $\mathcal{H}$ which is also an orthogonal system in L. Then the numbers $\sigma_i^2 = E_P(e_i^2)$ are the eigenvalues of the correlation operator $\mathcal{C}$ of γ. We denote by $\rho(\mathcal{C})$ the spectral radius $\max_i \sigma_i^2$ and by $tr(\mathcal{C})$ the trace $\sum_i \sigma_i^2$. Since the operator $\mathcal{C}$ is of trace class, the series $S(\omega) = \sum e_i(\omega)e_i$ defines a square integrable random vector S on Ω with values in $\mathcal{H}$. The image measure $\boldsymbol{P} \circ S^{-1}$ is a second order measure which can take the role of the measure P in Weizsäcker (1997). We need the following notation: For every positive number a, the symbol $\mathcal{N}(a)$ denotes the centered Gaussian distribution on the real axis $\boldsymbol{R}$ with variance a; for every measure q on the positive real axis $\boldsymbol{R}_+$, the symbol $q \times \mathcal{N}$ denotes the measure on $\boldsymbol{R}_+ \times \boldsymbol{R}$ given by

$$q \times \mathcal{N}(A \times B) = \int_A \mathcal{N}(a)(B)\; q(da).$$

Theorem 14.1.4 *Let d be any metric on the space $\mathcal{P}(\boldsymbol{R}_+ \times \boldsymbol{R})$ which induces convergence in law. Then there is a function $\varphi: \boldsymbol{R}_+^3 \longrightarrow \boldsymbol{R}_+$ such that $\varphi(r,t,\varepsilon) \to 0$ as $r \to 0$ for all t, ε with the following property: For all γ and $(\Omega, \mathcal{F}, \boldsymbol{P})$ as in Theorem 14.1.1, there is a square integrable random variable σ on $(\Omega, \mathcal{F}, \boldsymbol{P})$ such that*

$$\gamma\{\, f : d(\mathcal{L}_P(\sigma^2, f), \mathcal{L}_P(\sigma^2) \times \mathcal{N}) > \varepsilon\,\} \le \varphi(\rho(\mathcal{C}), tr(\mathcal{C}), \varepsilon).$$

Remark 14.1.1 The random variable σ is the $\mathcal{H}$-norm of the vector S in the above construction. In particular, $E(\sigma^2) = \mathrm{tr}(\mathcal{C})$.

Thus, for γ most f, the law of f under P is close to the mixed normal $E_P(\mathcal{N}(\sigma^2))$ where the degree of closeness is determined alone by the pair $(\rho(\mathcal{C}), \mathrm{tr}(\mathcal{C}))$. Note that, as indicated above, the theorem has an extension to certain non Gaussian measures which have near Gaussian average marginals.

Acknowledgement. Research was supported by grants RFBI-DFG 96-01-00096, RFBR 96-01-00853, 96-15-96199.

References

1. Sudakov, V. N. (1978). Typical distributions of linear functionals on the spaces of high dimension, *Dokl. Akad. Nauk SSSR*, **243**, 195–269; English translation (1978), *Soviet Mathematical Doklady*, **19**, 1578–1582.

2. Nagaev, S. V. (1982). On distribution of linear functionals in finite-dimensional spaces of large dimension, *Dokl. Akad. Nauk SSSR*, **263**, 295–296; English translation in *Soviet Mathematical Doklady.*

3. Makarova, S. B. (1985). Mean distance for the occupation times of Gaussian processes, *Zap. Nauchn. Semin. LOMI*, **142**, 98–108; English translation in *Journal of Mathematical Sciences.*

4. Weizsäcker, H. v. (1997). Sudakov's typical marginals, random linear functionals and a conditional central limit theorem, *Probability Theory and Related Fields*, **107**, 313–324.

5. Sudakov, V. N. (1994). On distributions of random variables chosen at random, *Zap. Nauchn. Semin. POMI*, **216**, 153–160; English translation in *Journal of Mathematical Sciences.*

6. Sudakov, V. N. (1979). Remark on convergence in a limit theorem for distributions of random functional over a process, *Zap. Nauchn. Semin. LOMI*, **85**, 197–206; English translation in *Journal of Mathematical Sciences.*

7. Sudakov, V. N. (1980). On limit "in large" distributions of linear functionals over a random second order process, *Zap. Nauchn. Semin. LOMI*, **97**, 199–202; English translation in *Journal of Mathematical Sciences.*

8. Sudakov, V. N. and Tsirel′son, B. S. (1974). Extremal properties of half-spaces for spherically invariant measures, *Zap. Nauchn. Semin. LOMI*, **41**, 14–24; English translation in *Journal of Mathematical Sciences.*

9. Borell, C. (1975). The Brunn-Minkowski inequality in Gaussian space, *Inventionnes Mathematicae*, **30**, 207–216.

10. Sudakov, V. N. (1977). On limit distributions of random functionals over a process, *Zap. Nauchn. Semin. LOMI*, **72**, 154–160; English translation in *Journal of Mathematical Sciences.*

PART IV

Weak and Strong Limit Theorems

15

A Local Limit Theorem for Stationary Processes in the Domain of Attraction of a Normal Distribution

Jon Aaronson and Manfred Denker

Tel Aviv University, Tel Aviv, Israel
Universität Göttingen, Göttingen, Germany

Abstract: In this chapter, we prove local limit theorems for Gibbs-Markov processes in the domain of attraction of normal distributions.

Keywords and phrases: Local limit theorem, domain of attraction of normal distribution

15.1 Introduction

It is well known that a random variable X belongs to the domain of attraction of a normal distribution DA(2) if its characteristic function satisfies

$$\log E\exp[itX] = it\gamma - \frac{1}{2}t^2 L(1/|t|) \tag{15.1}$$

for some slowly varying function $L : \mathbf{R}_+ \to \mathbf{R}_+$ which is bounded below and some constant $\gamma \in \mathbf{R}$; see Ibragimov and Linnik (1971).

The normal (or classical) domain of attraction NDA(2) consists of the class L_2, and is characterized by the boundedness of the slowly varying function L in (15.1). Here, we consider the 'non-normal' domain of attraction DA(2) \ NDA(2).

The function L is unbounded and is determined (up to asymptotic equivalence) by the tails of the distribution of X which satisfy

$$1 - G(x) = P(X \geq x) \sim c_1 x^{-2} l(x),$$

$$G(-x) = P(X \leq -x) \sim c_2 x^{-2} l(x) \quad x \to \infty \tag{15.2}$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

for some constants $c_1, c_2 \geq 0, c_1 + c_2 = 1$ and some slowly varying function l, which in turn determines L by

$$L(x) = \int_{-x}^{x} u^2 dP_X(u). \tag{15.3}$$

It follows from (15.3) that

$$l(x) = o(L(x)) \tag{15.4}$$

as $x \to \infty$.

Let $X_1, X_2, \ldots$ be a stationary process of independent random variables with $X_k \in \mathrm{DA}(p) \quad (0 < p \leq 2)$.

The local limit theorem (LLT) for the partial sums $S_n := X_1 + \ldots + X_n$ is well known, which is that there exist constants $A_n, B_n \in \mathbf{R}$, $\mathbf{B_n} \to +\infty$ such that $\forall\ \kappa \in \mathbf{R}$ and $I \subset \mathbf{R}$ (an interval),

$$B_n P(S_n - k_n \in I) \to |I| g(\kappa) \text{ as } \frac{k_n - A_n}{B_n} \to \kappa,$$

where g is a p-stable density on $\mathbf{R}$. Extensions of the LLT to Markov chains are well known; for example, see Aaronson and Denker (1998) for a more detailed discussion.

Aaronson and Denker (1998) have established LLT's for Gibbs-Markov functionals (definition below) in the non-normal stable case ($p < 2$).

In the normal case ($p = 2$), such extensions are only known when $X_k \in \mathrm{NDA}(2)$; see Aaronson and Denker (1998), Rousseau-Egele (1983), Guivarc'h and Hardy (1988), and Morita (1994).

Here, in this Chapter, we prove the LLT for Gibbs-Markov functionals $X_1, X_2, \ldots$ in the case when $X_1 \in \mathrm{DA}(2) \setminus \mathrm{NDA}(2)$.

15.2 Gibbs-Markov Processes and Functionals

Definition 15.2.1 A mixing stationary process $\{Z_n : n \in \mathbf{N}\}$ is called *Gibbs-Markov* if its state space E is at most countable and if

- (*Markov property*)

$$P(Z_1 = a, Z_2 = b) > 0 \quad \text{and} \quad P(Z_1 = b, Z_2 = c) > 0$$
$$\Longrightarrow P(Z_1 = a, Z_2 = b, Z_3 = c) > 0$$

  for all $a, b, c \in E$ and

$$\inf\{ \sum_{b \in E; P(Z_1 = a, Z_2 = b) > 0} P(Z_1 = b) : a \in E\} > 0.$$

- (*Gibbs property*) There exist constants $M > 0$ and $0 < r < 1$ such that

$$\left|\frac{P(Z_1 = a_1, ..., Z_n = a_n | Z_{n+1} = b_1, ..., Z_{n+k} = b_k)}{P(Z_1 = a_1, ..., Z_n = a_n | Z_{n+1} = c_1, ..., Z_{n+k} = c_k)} - 1\right| \le Mr^{-\min\{l: c_l \neq b_l\}}$$

for all $a_i, b_j, c_j \in E$, $1 \le i \le n$, $1 \le j \le k$ and all $n, k \ge 1$.

Remark 15.2.1

1. Recall that a process $\mathbf{Z} = \{\mathbf{Z_n} : \mathbf{n} \ge \mathbf{1}\}$ is called *mixing* if for all square-integrable functions $f, g \in L_2(\mathbf{Z})$ one has

$$Ef(\mathbf{Z})\mathbf{g}(\mathbf{Z_n}, \mathbf{Z_{n+1}}, ...) \to \mathbf{Ef}(\mathbf{Z})\mathbf{Eg}(\mathbf{Z}),$$

where $L_q(\mathbf{Z})$ ($q \in \mathbf{N} \cup \{\infty\}$) is the space of functions $g : E^{\mathbf{N}} \to \mathbf{R}$ which are q-integrable with respect to the distribution of $\mathbf{Z}$.

2. The coordinate process on $E^{\mathbf{N}}$ of a probability preserving, mixing Gibbs-Markov map [as in Aaronson and Denker (1998)] is a Gibbs-Markov process in the sense of Definition 15.2.1. Conversely, the shift of a Gibbs-Markov process (equipped with its mixing, shift-invariant distribution on $E^{\mathbf{N}}$) is a probability preserving, mixing Gibbs-Markov map.

Definition 15.2.2 A function $f : E^{\mathbf{N}} \to \mathbf{R}$ is uniformly Lipschitz on states ($f \in Lip$) if

$$D(f) := \sup_{a \in E, x, y \in [a]} r^{\min\{l: x_l \neq y_l\}} |f(x) - f(y)| < \infty,$$

where $[a] = \{(x_1, x_2, ...) \in E^{\mathbf{N}} : x_1 = a\}$.

Definition 15.2.3 A stationary process $\{X_n : n \in \mathbf{N}\}$ is called a *Gibbs-Markov functional* if there exists a Gibbs-Markov process $\mathbf{Z} = \{\mathbf{Z_n} : \mathbf{n} \in \mathbf{N}\}$ and a function $f \in Lip$ such that

$$X_n = f(Z_n, Z_{n+1}, ...).$$

The Frobenius-Perron operators $P^n : L_1(\mathbf{Z}) \to \mathbf{L_1}(\mathbf{Z})$ are defined by

$$EP^n f(Z_1, Z_2, ...) g(Z_1, Z_2, ...) = Ef(Z_1, Z_2, ...) g(Z_{n+1}, Z_{n+2}, ...), \qquad (15.5)$$

and the characteristic function operator for the function $\varphi : E^{\mathbf{N}} \to \mathbf{R}$ by

$$P_t f = P(f \exp[it\varphi]). \qquad (15.6)$$

Aaronson and Denker (1998) have shown that when $\varphi \in Lip$, P_t acts on $\mathcal{L} := L_\infty(\mathbf{Z}) \cap \mathbf{Lip}$ equipped with the norm $\|f\|_{\mathcal{L}} = \|f\|_\infty + D(f)$. As an operator on $\mathcal{L}$, P_t has a unique eigenvalue of maximal modulus $\lambda(t)$ for $|t| < \epsilon$ and a decomposition

$$P_t^n f = \lambda(t)^n g(t) Ef(\mathbf{Z}) + \mathbf{Q_t^n f} \qquad (|\mathbf{t}| < \epsilon), \tag{15.7}$$

where the spectral radius of Q_t is uniformly bounded by some $\theta < 1$ and where $g(t)$ is the normalized eigenfunction for $\lambda(t)$. P_t is called the characteristic function operator, since

$$P_t^n 1 = P^n e^{itS_n} = \lambda(t)^n g(t) + Q_t^n 1,$$

where $S_n = X_1 + ... + X_n$.

15.3 Local Limit Theorems

In this section, we assume that $\{X_n : n \geq 1\}$ is a Gibbs-Markov functional with $X_1 = f(\mathbf{Z}) \in \mathrm{DA}(\mathbf{2})$, but $EX_1^2 = \infty$. Let the operator $P_t : \mathcal{L} \to \mathcal{L}$, $\lambda(t)$ and $g(t)$ be as defined (15.5)–(15.7) for $|t| < \epsilon$ and for $\phi = f$. Moreover, let G denote the distribution function of X_1 and l and L the associated slowly varying functions as defined in (15.2) and (15.3).

Theorem 15.3.1

$$\log \lambda(t) = it\gamma - \frac{1}{2}|t|^2 L(|t|^{-1})(1 + o(1)) \tag{15.8}$$

as $t \to 0$, where the constant $\gamma \in \mathbf{R}$ is defined by

$$\gamma = \int_{-\infty}^{\infty} xG(dx). \tag{15.9}$$

Remark 15.3.1 Theorem 15.3.1 may fail in the 'classical' case where $Ef(\mathbf{Z}) = \mathbf{0}$ and $Ef(\mathbf{Z})^{\mathbf{2}} < \infty$. Indeed, suppose $\phi \in \mathcal{L}$, then also $f := \phi \circ T - \phi \in \mathcal{L}$ (here T denotes the shift on $E^{\mathbf{N}}$). As can be easily checked,

$$P_t(e^{it\phi}) = e^{it\phi},$$

whence $\lambda(t) = 1$; see Aaronson and Denker (1998). On the other hand, Aaronson and Denker (1998) have indicated how to prove Theorem 15.3.1 in case $f \in Lip$, $Ef(\mathbf{Z}) = \mathbf{0}$, $Ef(\mathbf{Z})^{\mathbf{2}} < \infty$, and not of form $f = \phi \circ T - \phi$.

Remark 15.3.2 As a corollary, we obtain that under the conditions of Theorem 15.3.1,

$$|\log \lambda(t) - \log E \exp[itX_1]| = o(|t|^2 L(1/|t|)) \quad \text{as } t \to 0.$$

Lemma 15.3.1

$$E(|1 - e^{itX_1}|) = O(|t|)$$

as $t \to 0$.

PROOF. This estimate follows from the expansion of $E\exp[itX_1]$; see Theorem 2.6.5 of Ibragimov and Linnik (1971). ∎

PROOF OF THEOREM 15.3.1. Let $\tilde{g}_t = g(t)/Eg(t)(\mathbf{Z})$ denote the eigenfunction of P_t with eigenvalue $\lambda(t)$ satisfying $E\tilde{g}_t(\mathbf{Z}) = \mathbf{1}$; then by (15.5)

$$\lambda(t) = \lambda(t)E\tilde{g}_t(\mathbf{Z}) = \mathbf{E}\lambda(\mathbf{t})\tilde{\mathbf{g}}_\mathbf{t}(\mathbf{Z}) = \mathbf{EP}[\tilde{\mathbf{g}}_\mathbf{t}\mathbf{e}^{\mathbf{it}\phi}](\mathbf{Z}) = \mathbf{E}\tilde{\mathbf{g}}_\mathbf{t}(\mathbf{Z})\mathbf{e}^{\mathbf{itX_1}}. \quad (15.10)$$

By Theorem 4.1 of Aaronson and Denker (1998), and by Lemma 3.4,

$$\|\tilde{g}_t - 1\|_\infty = O(|t|) \text{ as } t \to 0.$$

Denote by $\mathcal{F}_0$ the σ-algebra generated by X_1 and let $\hat{g}_t \circ X_1 = E(\tilde{g}_t(\mathbf{Z})|\mathcal{F}_0)$; then by (15.10)

$$\lambda(t) = E\hat{g}_t(X_1)\exp[itX_1] = \int_{-\infty}^{\infty} \hat{g}_t(x)\exp[itx]G(dx), \quad (15.11)$$

$$\|\hat{g}_t - 1\|_{L_\infty(G)} \le \|\tilde{g}_t - 1\|_\infty = O(|t|) \text{ as } t \to 0, \quad (15.12)$$

and

$$\int_{-\infty}^{\infty} \hat{g}_t(x)\ G(dx) = 1 \quad \forall\ t \in \mathbf{R}.$$

It follows from (15.12) that for $|t|$ small enough, Re $\hat{g}_t \ge 0$. Write

$$\hat{g}_t = g_t^r + ig_t^+ - ig_t^-$$

where $g_t^\pm := \max\{\pm\text{Im } \hat{g}_t, 0\} \ge 0$ and $g_t^r = \text{Re } \hat{g}_t \ge 0$.

For $* = r, +, -$, we fix $g_t = g_t^*$. Then $dG_t := g_t dG$ is a (positive) measure on $\mathbf{R}$. Note that by (15.12)

$$\lim_{t\to 0}\sup_{x\in\mathbf{R}} |g_t(x) - K| = 0$$

where $K = K_* = 1$ for $* = r$ and $K = 0$ otherwise.

Define distribution functions G^j, G_t^j $(j = 1, 2)$ on $\mathbf{R}_+$ by

$$G_t^1(x) := G_t(x) - G_t(0),\ G_t^2(x) := G_t(0) - G_t(-x),$$

$$G^1(x) := G(x) - G(0), \text{ and } G^2(x) := G(0) - G(-x).$$

We have that

$$G_t^j(\infty) - G_t^j(x) = \frac{h_j(x)}{x^2}g_j(t, x), \quad (15.13)$$

where

$$h_j(x) := \begin{cases} x^2(1-G(x)) = (c_1+o(1))l(x) & \text{if } j=1 \\ x^2G(-x) = (c_2+o(1))l(x) & \text{if } j=2 \end{cases}$$

as $x \to \infty$, and

$$g_1(t,x) := \frac{\int_x^\infty g_t(u)\ G(du)}{\int_x^\infty G(du)}, \quad g_2(t,x) := \frac{\int_{-\infty}^{-x} g_t(u)\ G(du)}{\int_{-\infty}^{-x} G(du)}.$$

It follows from (15.12) again that $\sup_{x\in\mathbf{R}} |g_j(t,x) - K| \to 0$ as $t \to 0$.

We need the following calculations. First note that

$$\begin{aligned} &\int_{\mathbf{R}} (1+itx-e^{itx})\ G_t(dx) \\ &= \int_0^\infty (1+itx-e^{itx})\ G_t^1(dx) + \int_0^\infty (1-itx-e^{-itx})\ G_t^2(dx), \end{aligned}$$

and secondly that integration by parts (for $j=1,2$) yields

$$\begin{aligned} &\int_0^\infty (1-(-1)^j itx - \exp[-(-1)^j itx])G_t^j(dx) \\ &= -[(G_t^j(\infty)-G_t^j(x))(1-(-1)^j itx - \exp[-(-1)^j itx])]_0^\infty \\ &\quad + \int_0^\infty (G_t^j(\infty)-G_t^j(x))((-1)^j it \exp[-(-1)^j itx] - (-1)^j it)dx \\ &= i(-1)^j t \int_0^\infty \Big(\exp[-i(-1)^j tx] - 1\Big) g_j(t,x)\frac{h_j(x)}{x^2}dx. \end{aligned}$$

We split the last integral into three parts:

$$\begin{aligned} & t\int_{|t|^{-1}}^\infty \Big(\exp[-i(-1)^j tx] - 1\Big) g_j(t,x)\frac{h_j(x)}{x^2}dx \\ &+ t\int_0^{|t|^{-1}} \Big(\exp[-i(-1)^j tx] - 1 + i(-1)^j tx\Big) g_j(t,x)\frac{h_j(x)}{x^2}dx \\ &- t\int_0^{|t|^{-1}} i(-1)^j tx g_j(t,x)\frac{h_j(x)}{x^2}dx. \end{aligned}$$

For the first integral, we obtain using (15.4)

$$\begin{aligned} & t\int_{|t|^{-1}}^\infty \Big(\exp[-i(-1)^j tx] - 1\Big) g_j(t,x)\frac{h_j(x)}{x^2}dx \\ &= \operatorname{sgn}(t)\int_1^\infty \Big(\exp[-i(-1)^j y \operatorname{sgn}(t)] - 1\Big) g_j(t,y/|t|)\frac{h_j(y/|t|)}{(y/|t|)^2}dy \\ &= O\Big(\int_1^\infty \frac{t^2}{y^2} h_j(y/|t|)dy\Big) \\ &= O\Big(\int_1^\infty \frac{t^2}{y^2} l(y/|t|)dy\Big) \\ &= O\Big(t^2 l(1/|t|)\Big) = o\Big(t^2 L(1/|t|)\Big). \end{aligned}$$

Since l is slowly varying,

$$|t|\int_0^{|t|^{-1}} l(x)dx = O\Big(l(|t|^{-1})\Big).$$

From this and (15.4), we obtain for the second integral that

$$\begin{aligned} t\int_0^{|t|^{-1}} &\Big(\exp[-i(-1)^j tx] - 1 + i(-1)^j tx\Big) g_j(t,x)\frac{h_j(x)}{x^2}dx \\ &= O\Big(t^3\int_0^{|t|^{-1}} h_j(x)dx\Big) \\ &= O\Big(t^2 l(1/|t|)\Big) = o\Big(t^2 L(1/|t|)\Big). \end{aligned}$$

The third integral, multiplied by $i(-1)^j$, is equal to

$$\begin{aligned} &t^2\int_0^{|t|^{-1}} x g_j(t,x)\frac{h_j(x)}{x^2}dx \\ =\ & t^2\int_0^{|t|^{-1}} x(G_t^j(\infty) - G_t^j(x))dx \\ =\ & \frac{t^2}{2}[(G_t^j(\infty) - G_t^j(x))x^2]_0^{|t|^{-1}} + \frac{t^2}{2}\int_0^{|t|^{-1}} x^2 G_t^j(dx) \\ =\ & \frac{t^2}{2}\int_0^{|t|^{-1}} x^2 G_t^j(dx)\ +\ o(t^2 L(1/|t|)) \\ =\ & \begin{cases} (K+o(1))\frac{t^2}{2}\int_0^{|t|^{-1}} x^2 G(dx)\ +\ o(t^2L(1/|t|)) & j=1 \\ (K+o(1))\frac{t^2}{2}\int_{-|t|^{-1}}^0 x^2 G(dx)\ +\ o(t^2L(1/|t|)) & j=2 \end{cases} \\ =\ & (K+o(1))\frac{t^2}{2}L(1/|t|), \end{aligned}$$

where we used (15.2), (15.12) and (15.13). Finally, note that by (15.12)

$$\gamma_t := \int_{\mathbf{R}} x\hat{g}_t(x)G(dx) = \gamma + O(|t|) \quad \text{as } t \to 0$$

and, since G is not in the normal domain of attraction, we have $t^2 = o(t^2L(1/|t|))$.

The proof of Theorem 15.3.1 is completed by using (15.11) and the previous estimates:

$$\begin{aligned} &\log\lambda(t) - it\gamma \sim \lambda(t) - 1 - it\gamma \\ =\ & \int_{-\infty}^{\infty}\Big(e^{itx} - 1 - itx\Big)\hat{g}_t(x)G(dx) + o(t^2L(1/|t|)) \\ =\ & \int_{-\infty}^{\infty}\Big(e^{itx} - 1 - itx\Big)(g_t^r(x) + ig_t^+(x) - ig_t^-(x))G(dx) + o(t^2L(1/|t|)) \\ =\ & \frac{t^2}{2}\int_{-|t|^{-1}}^{|t|^{-1}} x^2\hat{g}_t^r(x)G(dx) + o(t^2L(1/|t|)) \\ =\ & t^2L(1/|t|)(1+o(1)). \qquad \blacksquare \end{aligned}$$

Let

$$nL(B_n) = B_n^2, \quad A_n = \gamma n. \tag{15.14}$$

The following corollaries contain the local and central limit theorems. Their proofs are straightforward using Theorem 15.3.1; for example, see corresponding statements in Aaronson and Denker (1998). We write, as before,

$$S_n = X_1 + X_2 + ... + X_n,$$

and denote by ϕ the density of the standard normal distribution.

Corollary 15.3.1 [Conditional lattice local limit theorem] *Suppose that X_1 is aperiodic.*

Let A_n and B_n be as defined in (15.14), and suppose that $k_n \in \mathbf{Z}$, $\frac{\mathbf{k_n - A_n}}{\mathbf{B_n}} \to \kappa \in \mathbf{R}$ as $n \to \infty$, then

$$\|B_n P^n(1_{[S_n=k_n]}) - \phi(\kappa)\|_\infty \to 0 \quad \text{as } n \to \infty,$$

and, in particular,

$$B_n E 1_{[S_n=k_n]} \to \phi(\kappa) \quad \text{as } n \to \infty.$$

Corollary 15.3.2 [Conditional non-lattice local limit theorem] *Suppose that X_1 is aperiodic.*

Let A_n and B_n be as defined in (15.14), let $I \subset \mathbf{R}$ be an interval, and suppose that $k_n \in \mathbf{Z}$, $\frac{\mathbf{k_n - A_n}}{\mathbf{B_n}} \to \kappa \in \mathbf{R}$ as $n \to \infty$, then

$$B_n P^n(1_{[S_n \in k_n + I]}) \to |I|\phi(\kappa) \quad \text{as } n \to \infty,$$

where $|I|$ is the length of I, and in particular,

$$B_n E 1_{[S_n \in k_n + I]} \to |I|\phi(\kappa) \quad \text{as } n \to \infty.$$

Corollary 15.3.3 [Distributional limit theorem] *Let A_n and B_n be as defined in (15.14). Then,*

$$\frac{S_n - A_n}{B_n}$$

is asymptotically standard normal.

Acknowledgement. This research was supported by a grant from G.I.F., the German-Israel Foundation for Scientific Research and Development.

References

1. Aaronson, J. and Denker, M. (1998). Local limit theorems for Gibbs-Markov maps, *Preprint.* http://www.math.tau.ac.il/~aaro

2. Guivarc'h, Y. and Hardy, J. (1988). Théorèmes limites pour une classe de chaînes de Markov et applications aux difféomorphismes d'Anosov, *Ann. Inst. H. Poincaré*, **24**, 73–98.

3. Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables* (Ed., J. F. C. Kingman), Groningen, Netherlands: Wolters-Noordhoff.

4. Morita, T. (1994). Local limit theorems and density of periodic points of Lasota-Yorke transformations, *Journal of the Mathematical Society of Japan*, **46**, 309–343.

5. Nagaev, S. V. (1957). Some limit theorems for stationary Markov chains, *Theory of Probability and Applications*, **2**, 378–406.

6. Rousseau-Egele, J. (1983). Un theorème de la limite locale pour une classe de transformations dilatantes et monotones par morceaux, *Annals of Probability*, **11**, 772–788.

16

On the Maximal Excursion Over Increasing Runs

Andrei Frolov, Alexander Martikainen, and Josef Steinebach

St. Petersburg State University, St. Petersburg, Russia
St. Petersburg State University, St. Petersburg, Russia
Philipps-Universität Marburg, Marburg, Germany

Abstract: Let $\{(X_i, Y_i)\}$ be a sequence of i.i.d. random vectors with $P(Y_1 = y) = 0$ for all y. Put $M_n(j) = \max_{0 \le k \le n-j}(X_{k+1} + \cdots + X_{k+j})I_{k,j}$, where $I_{k,j} = I\{Y_{k+1} \le \cdots \le Y_{k+j}\}$, $I\{\,\cdot\,\}$ denotes the indicator function of the event in brackets. If, for example, $X_i = Y_i$, $i \ge 1$, and X_i denotes the gain in the i-th repetition of a game of chance, then $M_n(j)$ is the maximal gain over increasing runs of length j. We investigate the asymptotic behaviour of $M_n(j)$, $j = j_n \le L_n$, where L_n is the length of the longest increasing run in $Y_1, \ldots, Y_n$. We show that the asymptotics of $M_n(j)$ crucially depend on the growth rate of j, and they vary from strong non-invariance as in the Erdős–Rényi law of large numbers to strong invariance as in the Csörgő–Révész strong approximation laws.

Keywords and phrases: Increasing run, head run, monotone block, increments of sums, Erdős–Rényi laws, strong approximation laws, strong limit theorems

16.1 Introduction

Let (X, Y), (X_1, Y_1), $(X_2, Y_2), \ldots$ be a sequence of i.i.d. random vectors satisfying $P(Y = y) = 0$ for all y. Put

$$S_k = \sum_{i=1}^{k} X_i\ , \quad S_0 = 0\ .$$

Let $M_n(j)$ be the "maximal excursion" of $\{S_k\}$ over subintervals of length j on which $\{Y_i\}$ increases, that is,

$$M_n(j) = \max_{0 \le k \le n-j} (S_{k+[j]} - S_k) I\{Y_{k+1} \le \cdots \le Y_{k+[j]}\}\ , \tag{16.1}$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

where $1 \le j \le n$, $I\{\,\cdot\,\}$ denotes the indicator function of the event in brackets, and $[\,\cdot\,]$ is the integer part function.

The first question arising here is how long "increasing runs" may be for which the indicator in (16.1) equals one. The length of the longest increasing run in the first n observations of Y-values, that is,

$$L_n = \max\{k \mid Y_{i+1} \le \cdots \le Y_{i+k} \quad \text{for some } i,\ 0 \le i \le n-k\}\ ,$$

has been investigated by several authors. Pittel (1981), for instance, has proved that

$$\lim_{n\to\infty} \frac{L_n}{k(n)} = 1 \quad a.s.,$$

where $k(n) = \log n / \log\log n$. A more precise result on the asymptotics of L_n has been given by Révész (1983) and Novak (1992).

Theorem 16.1.1 *If* $l = l_n$ *is the solution of* $l^l e^{-l}(2\pi l)^{1/2} = n$, *then*

$$\limsup_{n\to\infty}\,(L_n - l_n) = 0 \quad a.s., \tag{16.2}$$

$$\liminf_{n\to\infty}\,(L_n - l_n) = -2 \quad a.s.; \tag{16.3}$$

see, for example, Novak (1992, Corollary 2.2).

One can check that

$$l_n = \frac{\log n - \frac{1}{2}\log(2\pi e)}{\log\log n\ - \log\log\log n - 1 + o(1)} - \frac{1}{2} \quad (n \to \infty) \tag{16.4}$$

[Novak (1992, Remark 2.4)]. Hence, $l_n \sim k(n)$ as $n \to \infty$.

Results on the length of the longest increasing run in case of discrete distributions have recently been obtained by Csáki and Földes (1996). For asymptotics of the length of the longest increasing run in R^d, see Frolov and Martikainen (1998).

It is interesting to investigate the growth rates of $M_n(aL_n)$ and $M_n(al_n)$, $a = a_n \in (0,1]$. The above setting includes some important special cases. For example, if $X = 1$ a.s., then $M_n(L_n) = L_n$ a.s., and precise limiting results are described above. Another special case of interest arises when $X_i = Y_i$ for all i. Then, $\{X_i\}$ can be interpreted as a sequence of "gains" of a player in a game of chance, and the random walk $\{S_k\}$ describes the player's fortune. So, $M_n(j)$ gives the "maximal gain" of a player over increasing runs of length j. It turns out that some surprising phenomena can be observed. For instance, the maximal gain of a player is not always attained over increasing runs of maximal length L_n. Indeed, the optimal length may depend on the full underlying distribution.

Similar phenomena have been observed for the maximal (unrestricted) gain (say) U_n over subintervals of length $j = j(n)$, i.e.

$$U_n(j) = \max_{0 \le k \le n-j} (S_{k+j} - S_k).$$

The asymptotic behaviour of $U_n(j)$ depends on the growth rate of $j(n)$. Erdős and Rényi (1970) have investigated the case $j = [c \log n]$. When $j/\log n \to \infty$, the asymptotics of $U_n(j)$ have been studied by Csörgő and Révész (1981). The case $j/\log n \to 0$ has been dealt with by Mason (1989).

For the sake of comparison, we briefly describe these results here. Assume that

(i) X is non-degenerate, $0 \le EX < \infty$;
(ii) $t_0 = \sup\{t : \varphi(t) = Ee^{tX} < \infty\} > 0$.

Denote

$$F(x) = P(X < x)\ ,\ \omega = \text{ ess sup} X\ ,$$

$$\zeta(z) = \sup\{zt - \log \varphi(t) : t \ge 0,\ \varphi(t) < \infty\},$$

$$\gamma(x) = \sup\{z : \zeta(z) \le x\}.$$

The function $\gamma(x)$ is a concave, continuous and non-decreasing function on $[0, \infty)$ with $\gamma(x) \to \omega$ as $x \to \infty$ [Mason (1989)].

Put

$$c_0 = 1/\Big(\int_0^{t_0} tm'(t)dt\Big)\ , \tag{16.5}$$

where $m(t) = \varphi'(t)/\varphi(t)$, $0 \le t < t_0$. It is known that $c_0 = 1/(At_0 - \log \varphi(t_0))$ if $t_0 < \infty$ and $A = \lim_{t \uparrow t_0} m(t) < \infty$. Furthermore, $c_0 = -1/\log P(X = A)$ if $t_0 = \infty$, $A < \infty$ and $P(X = A) > 0$ (in this case, $A = \omega$). In all other cases, $c_0 = 0$ [see Deheuvels, Devroye and Lynch (1986)].

The following result has been proved by Erdős and Rényi (1970) for $c > c_0$ and by Deheuvels and Devroye (1987) for $0 < c \le c_0$. See also Theorem 2.4.3 of Csörgő and Révész (1981).

Theorem 16.1.2 *Put $j = [c \log n]$. Then, almost surely*

$$\lim_{n\to\infty} \frac{U_n(j)}{j} = \begin{cases} \gamma(\frac{1}{c})\ , & \text{if } c > c_0\ , \\ A + \frac{1}{t_0}\Big(\frac{1}{c} - \frac{1}{c_0}\Big)\ , & \text{if } 0 < c \le c_0\ . \end{cases}$$

Mason (1989) has extended the Erdős–Rényi law of large numbers as follows.

Theorem 16.1.3 *Assume that* $j = j(n)$ *and* $d = (\log n)/j \to \infty$ *. Then,*

$$\limsup_{n\to\infty} \frac{U_n(j)}{j\gamma(d)} = 1 \quad a.s. \tag{16.6}$$

Furthermore, one can replace $\limsup$ *by* $\lim$ *if* $\omega < \infty$ *or if*

$$\lim_{x\to\infty} \frac{\gamma(-\log(1-F(x)))}{x} = 1 . \tag{16.7}$$

Remark 16.1.1 *Mason (1989) has proved that (16.7) is equivalent to*

$$\lim_{n\to\infty} \frac{\max_{1\le k\le n} X_k}{\gamma(\log n)} = 1 \qquad a.s.,$$

which, in turn, is equivalent to (16.6) with $j = 1$.

The case $j/\log n \to \infty$ has been investigated by Csörgő and Révész (1981). The following one-sided generalization of their results is due to Frolov (1998).

Theorem 16.1.4 *Let* $EX = 0$ *and* $EX^2 = 1$. *Assume that* $j = j(n) \le n$ *and* $j \sim h_n$ *for a non-decreasing sequence* $\{h_n\}$ *such that* $\{n/h_n\}$ *is also non-decreasing. (Note that* j*'s are integers, which usually do not satisfy the monotonicity assumptions on* h_n*.)*

If either, for some $t > 0$, $0 < \beta \le 1$,

$$\int_0^\infty \exp\{tx^\beta\} dF(x) < \infty \qquad and \qquad j/(\log n)^{2/\beta - 1} \to \infty ,$$

or, for some $p > 2$,

$$\int_0^\infty x^p dF(x) < \infty , \quad \log n \int_{-\infty}^{-n} x^2 dF(x) \to 0 ,$$

$$and \quad \liminf_{n\to\infty} j n^{-2/p} \log n > 0 ,$$

then

$$\limsup_{n\to\infty} \frac{U_n(j)}{(2j(\log(n/j) + \log\log n))^{1/2}} = 1 \quad a.s.$$

If additionally $\log\log n = o(\log(n/j))$, *then* $\limsup$ *can be replaced by* $\lim$.

Our attention focuses on the limiting behaviour of the maximal gain $M_n(j)$ over increasing runs of length $j = j_n$. Such statistics play a role in various contexts. For example, if $\{S_n\}$ describes the log–return process of a certain portfolio, the investor might be interested in "maximal increasing draw–downs" of $\{S_n\}$ in order to estimate the risk of his investment [see, for example, Binswanger and Embrechts (1994)]. Similar questions arise in connection with molecular sequence comparisons or in the change–analysis of engineering systems [Dembo and Karlin (1991)]. In any case, a.s. limiting relations of the oscillations of $\{S_n\}$ provide strong measures of the randomness of the underlying sequences $\{X_i\}$ or $\{(X_i, Y_i)\}$. For corresponding results on the maximal gain over (so–called) "head runs", see Frolov, Martikainen and Steinebach (1998).

Let us now describe the asymptotics of $M_n(j)$ which crucially depend on the growth rate of $j = j_n = a_n l_n$, $a_n \in (0,1)$, l_n as in (16.4). It turns out that there are essentially three different cases:

If $(1-a)\log\log n \to \infty$, Theorem 16.2.1 shows that the a.s. asymptotics depend on the underlying distribution, a similar phenomenon as in Theorem 16.1.3 above. It also demonstrates that the maximal gain over increasing runs is not necessarily attained during increasing runs of maximal length. It may happen that the maximal gain is attained over increasing runs of length aL_n (or equivalently al_n), with a coefficient $a \in (0,1)$ depending on the distribution of X. For example, $a = 1/2$ if X has a standard normal distribution, and $a = (\alpha - 1)/\alpha$ if X has a Weibull(α, λ) distribution with parameter $\alpha > 1$. Another surprising phenomenon appears when the random variable X is bounded. By Theorem 16.2.1, the maximal gain over increasing runs of length al_n (or aL_n) a.s. increases faster whenever $a < 1$ is closer to 1. Nevertheless, by Theorem 16.2.2, the maximal gain $M_n(a_n l_n)$ suddenly can have a smaller growth rate if $a_n \to 1$ faster than in Theorem 16.2.1. The case when j does not depend on n is also included in Theorem 16.2.1. Then, $M_n(j)$ typically increases a.s. as $j\gamma(\log n/j)$.

If $(1-a)\log\log n \to B$ with $0 < B < \infty$, then $M_n(l_n)$ is a.s. proportional to al_n for some coefficient a analogous to the limiting constant in Theorem 16.1.2. In this sense, we obtain an Erdős–Rényi (1970) type analogue of Theorem 16.1.2, but it is important to keep in mind that, though in both Theorems 16.1.2 and 16.2.2 the maxima are normed by the lengths of increments, these lengths have different order of growth, which is $\log n$ in Theorem 16.1.2 and $\log n/\log\log n$ in Theorem 16.2.2. It is worthwhile mentioning that the distribution of X can be uniquely determined by the limit function of Theorem 16.2.2 and therefore a.s. by the maxima $M_n(j)$.

If $a = a_n \to 1$ fast enough, i.e. if $(1-a)\log\log n \to 0$, the asymptotics of $M_n(j)$ do not depend on the distribution of X, provided $EX = 0$ and $EX^2 = 1$. Theorems 16.2.3 and 16.2.4 give universal norming sequences in this case. This type of behaviour is similar to the Csörgő–Révész results of Theorem 16.1.4.

16.2 Results

Given a real sequence $\{a_n\}$ with $a_n \in (0,1)$, put

$$i(n) = [a_n l_n], \; c(n) = \frac{1-a_n}{i(n)} \log n,$$

$$b_n = i(n)\, \gamma(\, c(n)\,)\ .$$

Note that, if $i(n) \to \infty$, then

$$b_n \sim \frac{a_n \log n}{\log\log n}\, \gamma\Big(\frac{1-a_n}{a_n}\, \log\log n\Big) \quad \text{as} \quad n \to \infty.$$

Recall that γ is concave, continuous and nondecreasing with $\gamma(\infty) = \omega$.

We assume in the sequel that the length of a run is at least 1, i.e. $i(n) \geq 1$.

Theorem 16.2.1 *Assume that* $\{i(n)\}$ *is non-decreasing, and*

$$(1-a_n)\, \log\log n \to \infty \quad \textit{as} \quad n \to \infty\ .$$

Then,

$$\limsup_{n\to\infty} \frac{M_n(a_n l_n)}{b_n} = 1 \quad a.s. \tag{16.8}$$

Moreover, one can replace lim sup *in (16.8) by* lim *if* $\omega < \infty$ *or if (16.7) holds.*

According to Mason (1989), (16.7) holds for the normal, geometric, Poisson and Weibull distributions. He also gave an example of a distribution for which (16.7) fails, but (16.6) still holds.

Examples

(1) **Normal distribution**: $X \sim N(\alpha, \sigma^2)$, $\alpha \geq 0$, $\sigma^2 > 0$. Here,

$$\varphi(t) = \exp\{\alpha t + \frac{\sigma^2 t^2}{2}\}\,, \quad \zeta(z) = \frac{(z-\alpha)^2}{2\sigma^2}\,, \quad \gamma(x) = \sqrt{2\sigma^2 x} + \alpha\,,$$

$$b_n \sim \sqrt{2\sigma^2 a_n (1-a_n)}\, \frac{\log n}{\sqrt{\log\log n}}\ .$$

It follows that b_n has maximal growth rate if $a_n = 1/2$. Hence, the maximal gain over increasing runs is attained in runs of length $l_n/2$.

(2) **Weibull(α,λ) distribution**: $P(X \geq x) = \exp\{-\lambda x^\alpha\}$, $x \geq 0$, $\lambda > 0$, $\alpha > 1$. It follows from (16.7) that

$$\gamma(x) \sim \left(\frac{x}{\lambda}\right)^{1/\alpha} \quad \text{as} \quad x \to \infty\,.$$

Hence,

$$b_n \sim \left(\frac{1}{\lambda}\right)^{1/\alpha} a_n^{1-1/\alpha}(1-a_n)^{1/\alpha}\log n\,(\,\log\log n\,)^{-1+1/\alpha}\,.$$

The maximal gain over increasing runs is attained in the case of $a_n = (\alpha-1)/\alpha$.

(3) **Exponential distribution**: $X \sim E(\lambda)$, $\lambda > 0$. Here,

$$\varphi(t) = \frac{\lambda}{\lambda - t}\,, t < \lambda\,, \quad \gamma(x) \sim \frac{x}{\lambda} \quad as \quad x \to \infty\,, \quad b_n \sim \frac{1-a_n}{\lambda}\log n\,.$$

Here, it seems that the maximal gain is attained when $a_n \to 0$ and $b_n \sim \log n/\lambda$. Moreover, since Theorem 16.2.1 admits the case $i(n) = 1$, the maximal gain is already attained in a single game.

Remark 16.2.1 *One can replace $M_n(a_n l_n)$ by $M_n(a_n L_n)$ in Theorem 16.2.1.*

Next, we study the case $a_n \to 1$ in more detail.

Theorem 16.2.2 *Let $\{a_n\}$ be a sequence of real numbers such that $a_n \to 1$, and $(1-a_n)\log\log n \to B$ as $n \to \infty$, $0 < B < \infty$. Put $B_0 = 1/c_0$, where c_0 is as in (16.5). Then, almost surely*

$$\lim_{n\to\infty} \frac{M_n(a_n l_n)}{l_n} = \begin{cases} \gamma(B)\,, & \text{if } B < B_0\,, \\ A + \frac{1}{t_0}(B - B_0)\,, & \text{if } B \geq B_0\,. \end{cases}$$

From here on, assumptions (i) and (ii) are not used any longer.

Theorem 16.2.3 *Assume that $EX = 0$, $EX^2 = 1$ and, for some $0 < \beta \leq 1$, $t > 0$,*

$$\int_0^\infty \exp\{tx^\beta\}dF(x) < \infty.$$

Let $\{a_n\}$ be a sequence of real numbers such that $a_n \to 1$, and

$$(1-a_n)(\log n)^{2(1-\beta)/(2-\beta)}(\,\log\log n\,)^{\beta/(2-\beta)} \to 0\,.$$

Then,

$$\limsup_{n\to\infty} \frac{M_n(a_n l_n)}{d_n} \leq 1 \quad a.s.\,, \tag{16.9}$$

where $d_n = (2l_n(l_n - i(n) + 1)\log l_n)^{1/2}$.

If, additionally, $\liminf(l_n - i(n)) > 1$, *then*

$$\liminf_{n\to\infty} \frac{M_n(a_n l_n)}{b_n} \geq 1 \quad a.s. , \tag{16.10}$$

where $b_n = (2l_n(l_n - i(n) - 1)\log l_n)^{1/2}$.

Corollary 16.2.1 *If the assumptions of Theorem 16.2.3 are satisfied, and* $l_n - i(n) \to \infty$, *then*

$$\lim_{n\to\infty} \frac{M_n(a_n l_n)}{b_n} = 1 \quad a.s.$$

Theorem 16.2.4 *Assume that* $EX = 0$, $EX^2 = 1$, *and*

$$n^p P(X > n) \to 0 \ \ for\ some\ p > 4 , \tag{16.11}$$

$$\int_{|x|>n} x^2 dF(x) = o(\frac{1}{\log n}) .$$

Let $\{a_n\}$ *be a sequence of real numbers such that* $a_n \to 1$.

If $\limsup(l_n - i(n)) < (p-2)/2 - 1$, *then (16.9) holds.*

If, additionally, $\liminf(l_n - i(n)) > 1$, *then (16.10) holds.*

16.3 Proofs

We first prove the following lemma.

Lemma 16.3.1 *For any* $x > 0$,

$$P(S_k I\{Y_1 \leq \cdots \leq Y_k\} > x) = \frac{1}{k!} P(S_k > x) .$$

PROOF. By independence and identical distribution of $(X_1, Y_1), \ldots, (X_k, Y_k)$ together with the continuity of the distribution of Y's we have, for any $x > 0$,

$$\begin{aligned} P(S_k > x) &= \sum_{\pi\ perm.} P(X_1 + \cdots + X_k > x,\ Y_{\pi(1)} < \cdots < Y_{\pi(k)}) \\ &= \sum_{\pi\ perm.} P(X_{\pi(1)} + \cdots + X_{\pi(k)} > x,\ Y_{\pi(1)} < \cdots < Y_{\pi(k)}) \\ &= k! P(X_1 + \cdots + X_k > x,\ Y_1 < \cdots < Y_k) \\ &= k! P(S_k I\{Y_1 \leq \cdots \leq Y_k\} > x) . \end{aligned}$$

This proves Lemma 16.3.1. ∎

For sake of brevity, we put $l = l_n$, $a = a_n$.

PROOF OF THEOREM 16.2.1. In the estimations below, C_i will always denote a positive constant.

Assume first that $\omega = \infty$.

Let

$$N_{ij0} = \{n : 2^j \le n < 2^{j+1},\ i(n) = i,\ a \le \beta\}, \quad 0 < \beta < 1\,,$$

$$N_{ijr} = \left\{n : 2^j \le n < 2^{j+1},\ i(n) = i,\ \frac{r}{\log\log 2^j} \le a < \frac{r+1}{\log\log 2^j}\right\},$$

for $\log\log 2^j > r \ge \beta \log\log 2^j$.

Put $n_{ijr} = \min\{n : n \in N_{ijr}\}$. For fixed $\varepsilon > 0$, we have

$$\begin{aligned} P_{ij0} &= P(\max_{n\in N_{ij0}} M_n(al)/b_n \ge 1+\varepsilon) \\ &\le P(\max_{0\le k\le 2^{j+1}-i} (S_{k+i} - S_k) I\{Y_{k+1} \le \cdots \le Y_{k+i}\} \\ &\qquad \ge (1+\varepsilon)\, i\, \gamma((1-\beta)\log n_{ij0}/i)\,) \\ &\le 2^{j+1} P(\, S_i I\{Y_1 \le \cdots \le Y_i\} \ge (1+\varepsilon)\, i\, \gamma((1-\beta)\log n_{ij0}/i)\,) \\ &= \frac{2^{j+1}}{i!} P(\, S_i \ge (1+\varepsilon)\, i\, \gamma((1-\beta)\log n_{ij0}/i)\,)\,. \end{aligned}$$

By Lemma 2.3 in Mason (1989), we get

$$\begin{aligned} P_{ij0} &\le 2^{j+1} \exp\{-(1+\varepsilon)(1-\beta)\log n_{ij0}\} \\ &\le 2\exp\{\,\log 2^j\,(1-(1-\beta)(1+\varepsilon))\} \le 2^{1-\varepsilon j/2}, \end{aligned} \tag{16.12}$$

if β is chosen small enough to satisfy $\beta < \varepsilon/2(1+\varepsilon)$.

Now, we turn to the case of positive r. Put $A_{ijr} = \max_{n\in N_{ijr}} a_n$, $a_{ijr} = \min_{n\in N_{ijr}} a_n$.

As before, we obtain

$$\begin{aligned} P_{ijr} &= P(\max_{n\in N_{ijr}} M_n(al)/b_n \ge 1+\varepsilon) \\ &\le \frac{2^{j+1}}{i!} P(\, S_i \ge (1+\varepsilon)\, i\, \gamma((1-A_{ijr})\log n_{ijr}/i)\,)\,. \end{aligned}$$

By Stirling's formula and Lemma 2.3 in Mason (1989), we get

$$P_{ijr} \le C_0 2^j \exp\{-(i+1/2)\log i + i - (1+\varepsilon)(1-A_{ijr})\log n_{ijr}\}\,.$$

Recall $k(n) = \log n/\log\log n$. By (16.4),

$$i \ge a_{ijr} k(2^j)(1 + \log\ \log\log 2^j\ /\ \log\log 2^j\,)$$

for large j. Since $a_{ijr} \geq \beta$, we have

$$\begin{aligned} i \log i &\geq a_{ijr} k(2^j)(1 + \log \log\log 2^j \,/\, \log\log 2^j\,) \\ &\qquad \times (\log \beta + \log\log 2^j - \log \log\log 2^j\,) \\ &\geq C_1 a_{ijr} k(2^j) + a_{ijr} \log 2^j\,(1 - (\log \log\log 2^j \,/\, \log\log 2^j\,)^2)\,. \end{aligned}$$

The last inequality together with $i \leq 2A_{ijr}k(2^j)$ and $n_{ijr} \geq 2^j$ implies

$$P_{ijr} \leq C_0 \exp\{\log 2^j\,(1 - a_{ijr} - (1+\varepsilon)(1 - A_{ijr})) + 3k(2^j)\}$$

for large j. By the definition of N_{ijr}, $A_{ijr} - a_{ijr} \leq 1/\log\log 2^j$. Note that $1 - A_{ijr} \geq 5/(\varepsilon \log\log 2^j\,)$ for large j, by the assumption of Theorem 16.2.1, and therefore

$$\begin{aligned} P_{ijr} &\leq C_0 \exp\{\log 2^j\,(A_{ijr} - a_{ijr} - \varepsilon)(1 - A_{ijr})) + 3k(2^j)\} \\ &\leq C_0 \exp\{-k(2^j)\} \end{aligned}$$

for large j.

Since the right-hand sides of this estimate and (16.12) do not depend on i, we have

$$\begin{aligned} P_j &= P(\max_{2^j \leq n < 2^{j+1}} M_n(al)/b_n \geq 1 + \varepsilon) \\ &\leq 2m_j 2^{-\varepsilon j/2} + C_0 m_j \sum_{\beta \log\log 2^j \leq r < \log\log 2^j} \exp\{-k(2^j)\}\,, \end{aligned}$$

where $m_j = \#\{i : N_{ij} \neq \emptyset\}$, $N_{ij} = \bigcup_r N_{ijr}$. Note that $m_j \leq al_{2^{j+1}} \leq a(j+1)\log 2$. This implies that the series $\sum_j P_j$ converges. By the Borel–Cantelli lemma, we conclude that

$$\limsup_{n\to\infty} \frac{M_n(al)}{b_n} \leq 1 \quad \text{a.s.} \tag{16.13}$$

Now, put $i = i(n)$. For $\varepsilon > 0$, we have

$$\begin{aligned} Q(n) &= P(M_n(al) < (1+\varepsilon)^{-3} b_n) \\ &\leq P\Big(\bigcap_{m=0}^{n/i-1} \Big\{(S_{(m+1)i} - S_{mi}) I\{Y_{mi+1} \leq \cdots \leq Y_{(m+1)i}\} < (1+\varepsilon)^{-3} b_n\Big\}\Big) \\ &= (P(S_i I\{Y_1 \leq \cdots \leq Y_i\} < (1+\varepsilon)^{-3} b_n))^{n/i-1} \\ &\leq \exp\{-C_2 \frac{n}{i\,i!}\, P(S_i \geq (1+\varepsilon)^{-3} b_n)\} \\ &= \exp\{-C_2 \frac{n}{i\,i!}\, P(S_i \geq (1+\varepsilon)^{-3} i\gamma(\,c(n)\,))\} \\ &\leq \exp\{-C_2 \frac{n}{i\,i!}\, (P(X_1 \geq (1+\varepsilon)^{-3} \gamma(\,c(n)\,)))^i\}\,. \end{aligned} \tag{16.14}$$

Here, we have used Lemma 16.3.1 together with the inequalities $1 - x \le e^{-x}$ and $P(S_m \ge mx) \ge (P(X_1 \ge x))^m$.

By the method used in Lemma 2.5 of Mason (1989), we can construct a sequence $\{n_r\}$ of natural numbers such that

$$(P(X_1 \ge (1+\varepsilon)^{-3}\gamma(c(n_r))))^{\log n_r/c(n_r)} \ge \exp\{-(1+\varepsilon)^{-1}\log n_r\} \quad (16.15)$$

and $n_r \ge r$ for all large r.

By Stirling's formula and the definition of l_n,

$$\begin{aligned} \frac{n}{i\,i!} &\ge C_3 l^{-1} n \exp\{-(i+\frac{1}{2})\log i + i\} \\ &\ge C_4 l^{-1} n \exp\{-(al+\frac{1}{2})\log l + al\} \\ &\ge C_4 l^{-3/2} n^{1-a} \exp\{a(\log n - l\log l + l)\} \\ &\ge C_4 l^{-3/2} n^{1-a} \exp\{a \log l/2\} \ge C_4 l^{-3/2} n^{1-a}\,. \end{aligned}$$

On combining (16.14) and (16.15), we have

$$\begin{aligned} Q(n_r) &\le \exp\{-C_4 l_{n_r}^{-3/2} n_r^{(1-(1+\varepsilon)^{-1})(1-a_{n_r})}\} \\ &\le \exp\{-C_4 l_{n_r}^{-3/2} \exp\{k(n_r)\}\} \le \exp\{-C_4 k(n_r)^{-3/2} \exp\{k(n_r)\}\} \end{aligned}$$

for large r. Hence, the series $\sum_r Q(n_r)$ converges. By the Borel–Cantelli lemma, it follows that

$$\limsup_{n\to\infty} \frac{M_n(al)}{b_n} \ge 1 \quad \text{a.s.}$$

Taking (16.13) into account, we arrive at (16.8).

Assume now that (16.7) holds. Replacing $(1+\varepsilon)^{-3}$ in (16.14) by $(1+\varepsilon)^{-2}$, we have

$$\begin{aligned} R(n) &= P(M_n(al) < (1+\varepsilon)^{-2} b_n) \\ &\le \exp\{-C_5 \frac{n}{i\,i!} (P(X_1 \ge (1+\varepsilon)^{-2}\gamma(\,c(n)\,)))^i\}\,. \end{aligned}$$

From the proof of Theorem in Mason (1989, p. 264), we adopt the inequality

$$P(X_1 \ge (1+\varepsilon)^{-2}\gamma(c(n))) \ge \exp\{-(1+\varepsilon)^{-1}\,c(n)\,\}$$

which, in combination with Stirling's formula and the definition of l_n again, implies convergence of the series $\sum_n R(n)$. From the Borel–Cantelli lemma, we conclude that

$$\liminf_{n\to\infty} \frac{M_n(al)}{b_n} \ge 1 \quad a.s.$$

This completes the proof of Theorem 16.2.1 for the case $\omega = \infty$.

In the case of $\omega < \infty$, choose $0 < \delta_1 < 1$. As in (16.14), we obtain

$$\begin{aligned} Q(n) &= P(M_n(al) < \delta_1 \omega i) \le \exp\left\{-C_6 \frac{n}{i\,i!} P(S_i \ge \delta_1 \omega i)\right\} \\ &\le \exp\left\{-C_6 \frac{n}{i\,i!} (P(X_1 \ge \delta_1 \omega))^i\right\} . \end{aligned}$$

Similar estimations as before then show that the series $\sum_n Q(n)$ converges. By Borel–Cantelli lemma, we have

$$\liminf_{n\to\infty} \frac{M_n(al)}{l_n} \ge a\omega \quad a.s.$$

This inequality, in combination with the definitions of $M_n(al)$ and ω, completes the proof of Theorem 16.2.1. ∎

In the proof of our next result, we make use of the following theorem. Put $\gamma = \gamma(B)$, and let $t^* = t^*(\gamma)$ be the solution of the equation $m(t^*) = \gamma$.

Theorem 16.3.1 [Petrov (1965)]. *For any* $\varepsilon > 0$,

$$P(S_n \ge n\gamma) \sim \frac{\psi(t^*)}{\sqrt{n}} \exp\{-nB\}$$

uniformly for $\gamma \in [\varepsilon, \min\{A-\varepsilon, 1/\varepsilon\}]$, *where* $\psi(t^*)$ *is a finite positive constant depending only on* t^* *and the distribution of* X_1.

For nonlattice distributions, $\psi(t^*) = 1/(t^*\sigma(t^*)\sqrt{2\pi})$, *while for lattice distributions with span* H, $\psi(t^*) = H/(\{1-e^{-Ht^*}\}\sigma(t^*)\sqrt{2\pi})$, *where* $\sigma(t) = m'(t)$.

PROOF OF THEOREM 16.2.2. For sake of brevity, we put $i = i(n) = [al]$.

Assume first that $B < B_0$. Making use of Theorem 16.3.1, Stirling's formula and the properties of $\gamma(x)$, we have

$$\begin{aligned} R(n) &= P(M_n(i) \ge (1+\varepsilon)\gamma(B)i) \le \frac{n}{i!} P(S_i \ge (1+\varepsilon)\gamma(B)i) \\ &\le C_7 \exp\left\{(l+\frac{1}{2})\log l - l - (i+\frac{1}{2})\log i + i - i(1+\delta_2)B\right\} \\ &\le C_7 \exp\{-\delta_3 al\} \end{aligned}$$

for any $\varepsilon > 0$, some positive δ_2, δ_3, and all large n.

Put $n_j = \max\{n : i(n) = j\}$. Then the series $\sum_j R(n_j)$ converges. By Borel–Cantelli lemma,

$$\limsup_{j\to\infty} \frac{M_{n_j}(a_{n_j} l_{n_j})}{a_{n_j} l_{n_j}} \le \gamma(B) \quad \text{a.s.}$$

Since $a \to 1$ and $M_n(i) \le M_{n_j}(a_{n_j} l_{n_j})$ for n such that $i(n) = j$, we conclude that

$$\limsup_{n\to\infty} \frac{M_n(i)}{l} \le \gamma(B) \quad \text{a.s.}$$

On the other hand, by Theorem 16.3.1, Stirling's formula and the properties of $\gamma(x)$, we also have

$$\begin{aligned}
Q(n) &= P(M_n(i) < (1-\varepsilon)\gamma(B)al) \\
&\le \exp\left\{-\frac{n}{al\, i!} P(S_i \ge (1-\varepsilon)\gamma(B)al)\right\} \\
&\le \exp\left\{-C_8 \exp\left\{(l+\frac{1}{2})\log l - l - (i+\frac{3}{2})\log i + i - i(1-\delta_4)B\right\}\right\} \\
&\le \exp\{-C_8 \exp\{\delta_5 al\}\} \le n^{-2}
\end{aligned}$$

for all large n. Hence, the series $\sum_n Q(n)$ converges and, by Borel–Cantelli lemma,

$$\liminf_{n\to\infty} \frac{M_n(i)}{l} \ge \gamma(B) \quad \text{a.s.},$$

which completes the proof in the first case.

We now turn to the case $B \ge B_0$. Put

$$L = A + \frac{1}{t_0}(B - B_0)\,.$$

Since $B_0 = 1/c_0$, $B_0 < \infty$ only if $A < \infty$.

Assume first that $A < \infty$ and $t_0 = \infty$. Then $A = \omega$, $L = A$ and we need only prove that

$$\lim_{n\to\infty} \frac{M_n(i)}{l} \ge L \quad \text{a.s.} \tag{16.16}$$

To do so, we use the same arguments as in the proof of Theorem 16.2.1 and get

$$\begin{aligned}
Q(n) &= P(M_n(i) < (L-\varepsilon)l) \\
&\le \exp\left\{-C_9 \exp\{(1 - a + C_{10}\frac{1}{\log\log n})\log n\right\}
\end{aligned}$$

for all large n. Since $(1-a)\ \log\log n\ \to B$, we have

$$(1 - a + C_{10}\frac{1}{\log\log n})\log n = C_{11}\frac{\log n}{\log\log n} \ge 2\ \log\log n$$

for all large n. This yields

$$Q(n) \le n^{-2}$$

for all large n, and the series $\sum_n Q(n)$ converges. Then Borel–Cantelli lemma gives (16.16).

Finally, assume that $A < \infty$ and $t_0 < \infty$. Then, $\omega = \infty$ and $c_0 = 1/(At_0 - \log\varphi(t_0))$. Hence, $B_0 = At_0 - \log\varphi(t_0)$.

We have

$$\begin{aligned}
R(n) &= P(M_n(i) \ge (L+\varepsilon)al) \le \frac{n}{i!} P(S_i \ge (L+\varepsilon)al) \\
&\le \frac{n}{i!} \exp\{i \log \varphi(t_0) - (L+\varepsilon)al\} \\
&\le C_{12} \exp\left\{(l+\frac{1}{2}) \log l - l - (i+\frac{1}{2}) \log i + i + i \log \varphi(t_0) - (L+\varepsilon)al\right\} \\
&\le C_{12} \exp\{-\delta_6 al\}
\end{aligned}$$

for all large n. Here, we have used Markov's inequality and Stirling's formula. Applying the same arguments as in the case $B < B_0$, we get

$$\limsup_{n\to\infty} \frac{M_n(i)}{l} \le L \quad \text{a.s.}$$

Put $\lambda = L - \varepsilon$, $\mu = L + \varepsilon$. From Deheuvels and Devroye (1987, p. 1376), we adopt the following inequality. For any small $\delta > 0$, there exists a positive constant v such that

$$P(S_k \ge k\lambda) \ge v k^{-1/2} \exp\{-\frac{\mu B_0 k}{A-\delta}\}$$

for all large n.

Then, we have

$$\begin{aligned}
&Q(n) \\
&= P(M_n(i) < (1-\varepsilon)Lal) \le \exp\left\{-\frac{n}{al\, i!} P(S_i \ge (1-\varepsilon)Lal)\right\} \\
&\le \exp\left\{-C_{13} \exp\left\{(l+\frac{1}{2}) \log l - l - (i+\frac{3}{2}) \log i + i - \frac{1}{2} \log i - \frac{\mu B_0 i}{A-\delta}\right\}\right\} \\
&\le \exp\left\{-C_{13} \exp\left\{\delta_7 l(B - \frac{\mu B_0}{A-\delta}\right\}\right\}
\end{aligned}$$

for all large n. Putting $c = 1/B$, we get

$$B - \frac{\mu B_0}{A-\delta} = \frac{1}{c} - \frac{1}{c_0} \frac{\mu}{A-\delta} .$$

In Deheuvels and Devroye (1987, p. 1377), it is proved that for any small $\varepsilon > 0$ there exists a small $\delta > 0$ such that

$$\frac{1}{c} - \frac{1}{c_0} \frac{\mu}{A-\delta} > 0 .$$

So,

$$Q(n) \le \exp\{-\exp\{\delta_8 l\}\}$$

for all large n. Applying the same arguments as in the case $B < B_0$, we complete the proof of Theorem 16.2.2. ■

Lemma 16.3.2 *Assume* $(1-a)\log\log n \to 0$. *Let* d_n *and* b_n *be as defined in Theorem 16.2.3. If, for some* $r > 1$,

$$\log P(S_i \geq Dd_n) \sim -\frac{D^2 d_n^2}{2l} \tag{16.17}$$

for any $0 < D < r$, *then the conclusion of Theorem 16.2.3 holds true.*

PROOF OF LEMMA 16.3.2. Take $\varepsilon > 0$ such that $(1+\varepsilon)^{1/2} < r$. We have for large n

$$\begin{aligned} R(n) &= P(M_n(i) \geq (1+2\varepsilon)^{1/2} d_n) \leq \frac{n}{i!} P(S_i \geq (1+2\varepsilon)^{1/2} d_n) \\ &\leq C_{14} \exp\left\{ (l+\frac{1}{2})\log l - l - (i+\frac{1}{2})\log i + i - \frac{(1+\varepsilon)d_n^2}{2l} \right\} \\ &\leq C_{14} \exp\left\{-(1+\varepsilon\tau_n)\log l\right\} , \end{aligned}$$

where $\tau_n = l - i + 1 \geq 1$. Here, we have used Stirling's formula together with (16.17).

Fix $q > 1$. Setting $N_{jk} = \{n : i(n) = j\ ,\ , q^k \leq \tau_n < q^{k+1}\}$, $n_{jk} = \max\{n : n \in N_{jk}\}$, $j \geq 2$, $k \geq 0$, we conclude that the series $\sum_{j,k} R(n_{jk})$ converges. By Borel–Cantelli lemma,

$$\limsup_{j\to\infty} \frac{M_{n_{jk}}(a_{n_{jk}} l_{n_{jk}})}{d_{n_{jk}}} \leq 1 \quad \text{a.s.},$$

uniformly over k. Since $d_n \geq d_{n_{jk}}/q$ and $M_n(a_n l) \leq M_{n_{jk}}(a_{n_{jk}} l_{n_{jk}})$ for $n \in N_{jk}$, all large j, all k, we conclude that

$$\limsup_{n\to\infty} \frac{M_n(i)}{d_n} \leq q \quad \text{a.s.} .$$

Since $q > 1$ is arbitrary otherwise, this yields (16.9).

On the other hand, making use of Stirling's formula and (16.17) again, we get for large n

$$\begin{aligned} Q(n) &= P(M_n(i) < (1-\varepsilon)b_n) \leq \exp\left\{-\frac{n}{al\, i!} P(S_i \geq (1-\varepsilon)b_n)\right\} \\ &\leq \exp\left\{-C_{15}\exp\left\{(l+\frac{1}{2})\log l - l - (i+\frac{3}{2})\log i + i - \frac{(1-2\varepsilon)b_n^2}{2l}\right\}\right\} \\ &\leq \exp\{-\delta_9 l^{\varepsilon\tau_n}\} , \end{aligned}$$

where $\tau_n = l - i - 1$.

For N_{jk} defined above, we put now $n_j = \min\{n : n \in N_{jk}\}$. Then the series $\sum_{j,k} Q(n_{jk})$ converges, and Borel–Cantelli lemma implies

$$\liminf_{j\to\infty} \frac{M_{n_{jk}}(a_{n_{jk}} l_{n_{jk}})}{d_{n_{jk}}} \geq 1 \quad \text{a.s.}$$

uniformly over k. In the same way as before, we get

$$\liminf_{n\to\infty} \frac{M_n(i)}{b_n} \geq 1 \quad \text{a.s.}$$

This completes the proof of Lemma 16.3.2. ∎

PROOF OF THEOREMS 16.2.3 AND 16.2.4. Theorems 16.2.3 and 16.2.4 are immediate from Lemma 16.3.2 and the following results on large [Frolov (1998)] and moderate [Amosova (1979)] deviations. ∎

Theorem 16.3.2 *If the assumptions of Theorem 16.2.3 are satisfied, then, for any $\delta > 0$ and any sequence $\{x_n\}$ with $x_n = o(n^{\beta/(4-2\beta)})$, the inequalities*

$$\exp\{-(1+\delta)x_n^2/2\} \leq P(S_n \geq x_n\sqrt{n}) \leq \exp\{-(1-\delta)x_n^2/2\}$$

hold for all large n.

PROOF. If $\beta = 1$, see Feller's (1969) result. If $\beta < 1$, it follows from Lemma 2 in §3 of Chapter VIII and relations (3.28), (3.30) in Petrov (1975, p. 241). ∎

Theorem 16.3.3 *If the assumptions of Theorem 16.2.4 are satisfied, then for any $c < \sqrt{p-2}$,*

$$P(S_n > x\sqrt{n}) \sim 1 - \Phi(x)$$

uniformly over $0 \leq x \leq c\sqrt{\log n}$. Here, $\Phi(x)$ denotes the standard normal distribution function.

Acknowledgement. This research was started when the first two authors visited Marburg. They gratefully acknowledge partial support by a special research grant from the University of Marburg, which made this exchange possible, and by grant 96–01–00547 from RFFI.

References

1. Amosova, N. N. (1979). On the probability of moderate deviations for sums of independent random variables, *Theory of Probability and Its Applications*, **24**, 858–865.

2. Binswanger, K. and Embrechts, P. (1994). Longest runs in coin tossing, *Insur. Math. Econom.*, **15**, 139–149.

3. Csáki, E. and Földes, A. (1996). On the length of the longest monotone block, *Studia Sci. Math. Hung.*, **31**, 35–46.

4. Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*, New York: Academic Press.

5. Deheuvels, P. and Devroye, L. (1987). Limit laws of Erdős–Rényi–Shepp type, *Annals of Probability*, **15**, 1363–1386.

6. Deheuvels, P., Devroye, L. and Lynch, J. (1986). Exact convergence rate in the limit theorems of Erdős–Rényi and Shepp, *Annals of Probability*, **14**, 209–223.

7. Dembo, A. and Karlin, S. (1991). Strong limit theorems of empirical functionals for large exceedances of partial sums of i.i.d. variables, *Annals of Probability*, **19**, 1737–1755.

8. Erdős, P. and Rényi, A. (1970). On a new law of large numbers, *J. Analyse Math.*, **23**, 103–111.

9. Feller, W. (1969). Limit theorems for probabilities of large deviations, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **14**, 21–26.

10. Frolov, A. N. (1998). On one-sided strong laws for large increments of sums, *Statistics & Probability Letters*, **37**, 155–165.

11. Frolov, A. N. and Martikainen, A. (1998). On the length of the longest increasing run in R^d, *Statistics & Probability Letters* (to appear).

12. Frolov, A. N., Martikainen, A. and Steinebach, J. (1998). On the maximal gain over head runs, *Studia Sci. Math. Hung.*, **34**, 165–181.

13. Mason, D. M. (1989). An extended version of the Erdős–Rényi strong law of large numbers, *Annals of Probability*, **17**, 257–265.

14. Novak, S. Yu. (1992). Longest runs in a sequence of m–dependent random variables. *Probability Theory and Related Fields*, **91**, 269–281.

15. Petrov, V. V. (1965). On the probabilities of large deviations of sums of independent random variables, *Theory of Probability and Its Applications*, **10**, 287–298.

16. Petrov, V. V. (1975). *Sums of Independent Random Variables*, New York: Springer-Verlag.

17. Pittel, B. G. (1981). Limiting behaviour of a process of runs, *Annals of Probability*, **9**, 119–129.

18. Révész, P. (1983). Three problems on the lengths of increasing runs, *Stochastic Processes and their Applications*, **15**, 169–179.

17

Almost Sure Behaviour of Partial Maxima Sequences of Some m-Dependent Stationary Sequences

George Haiman and Lhassan Habach

Université de Sciences et Technologies de Lille, Villeneuve d'Ascq, France

Abstract: For some m-dependent stationary sequences, one can construct an i.i.d. sequence, with same marginal distribution, such that a.s. from a random range on the distribution tail, the corresponding partial maxima sequences coincide. In this chapter, we give a short presentation of the proofs of these results.

Keywords and phrases: m-dependent, stationary, extremes, partial maxima

17.1 Introduction

Let $\{X_n\}$ be a stationary m-dependent sequence of random variables (for any $t \geq 1$, $\sigma(\ldots, X_t)$ and $\sigma(X_{t+m+1}, \ldots)$ are independent).

Suppose that $F(x) = P(X_1 \leq x)$ is continuous and there exist $\beta > 0$ and $k > 0$ such that for all $2 \leq i \leq m+1$ we have

$$\limsup_{u \to \omega} \left\{ \sup_{u < v < \varphi(u)} P\{X_1 > u \mid X_i = v\}(-\log(1 - F(u)))^{2+\beta} \right\} < \infty, \qquad (17.1)$$

where $\omega = \text{Max}\{x; F(x) < 1\}$ and $\varphi(u)$ is the solution of the equation $1 - F(\varphi(u)) = (1 - F(u))^{1+k}$.

The following Theorem 17.1.1 has been proved in Haiman (1987) [see also Haiman (1992) and Haiman *et al.* (1998)].

Theorem 17.1.1 *Let $\{X_n\}$ satisfy condition (17.1). Then, one can construct, on the probability space on which $\{X_n\}$ is defined, enlarged by independent*

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

factors, an i.i.d. sequence $\{\hat{X}_n\}_{n\geq 1}$*, such that* $X_1 \stackrel{\mathcal{L}}{=} \hat{X}_1$ *and a.s. there exists a random range* N *such that for* $n \geq N$ *we have*

$$\mathrm{Max}(X_1, \ldots, X_n) = \mathrm{Max}(\hat{X}_1, \ldots, \hat{X}_n). \tag{17.2}$$

Several examples of sequences satisfying the following condition that is stronger than (17.1): there exist $\gamma > 0$ and $k > 0$ such that for all $2 \leq i \leq m+1$, we have

$$\limsup_{u \to \omega} \left\{ \sup_{u<v<\varphi(u)} P\{X_1 > u \mid X_i = v\}(P\{X_1 > u\})^{-\gamma} \right\} < \infty, \tag{17.3}$$

as given in Haiman *et al.* (1998).

Among these, for $m = 1$, examples of the form $X_n = f(U_n, U_{n+1}), n \geq 1$, where $\{U_n\}_{n\geq 1}$ is a sequence of i.i.d. uniformly on $[0, 1]$ distributed random variables, are $X_n = U_n + U_{n+1}$, $X_n = U_n \times U_{n+1}$ and $X_n = \inf(U_n, U_{n+1})$. For sequences $\{X_n\}$ which do not satisfy (17.1), such as $X_n = \mathrm{Max}(U_n, U_{n+1})$, a similar result to Theorem 17.1.1 result was obtained in the above cited paper. Another example of 1-dependent sequence satisfying (17.3) [see Haiman (1999)] is

$$X_n = \max_{n-1\leq t<n} (W(t) - W(t+1)), n \geq 1,$$

where $W(t)$ is the Wiener process.

Applied to this example, (17.2) means that for $n \geq N$ we have

$$\mathop{\mathrm{Max}}_{0\leq t\leq n} (W(t) - W(t+1)) = \mathrm{Max}(\hat{X}_1, \ldots, \hat{X}_n)$$

and the i.i.d. sequence $\{\hat{X}_n\}$ satisfies

$$P\{X_1 < x\} = P\{\hat{X}_1 < x\} = 1 - \varphi x - 2\psi + \varphi x \psi - \varphi^2 + \psi^2,$$

with $\varphi = \varphi(x) = 1/\sqrt{2\pi}e^{-x^2/2}$ and $\psi = \psi(x) = 1 - \int_{-\infty}^{x} \varphi(u)du$. In Haiman and Habach (1999), the following Theorem 17.1.2 has been proved which completes Theorem 17.1.1.

Theorem 17.1.2 *The random variable* N *of Theorem 17.1.1 satisfies*

$$P(N > s) = O((\log s)^{-\beta/(4+\beta)}), s \geq 2. \tag{17.4}$$

It may be easily seen that (17.3) implies (17.1) for any $\beta > 0$. Thus, we have the following corollary.

Corollary 17.1.1 *If* $\{X_n\}$ *satisfies (17.3), then for any* $0 < \epsilon < 1$ *we have*

$$P(N > s) = o((\log s)^{\epsilon-1}), \qquad s \geq 2. \tag{17.5}$$

The proof of Theorem 17.1.2, quite long and complicated, is closely connected to the method of construction of the i.i.d. sequence $\{\hat{X}_n\}$.

In this Chapter we give a short presentation of the proofs of these results, leaving aside auxiliary technical aspects which may be found in the above cited papers.

17.2 Proof of Theorem 17.1.2

Let $r_0 < \omega$ be fixed and consider the sequence $\{(T_n, R_n)\}_{n\geq 1}$ of record times T_n and record values R_n of $\{X_n\}$, defined with respect to the initial threshold r_0, as

$$T_1 = \inf\{k \geq 1\,;\, X_k > r_0\}, R_1 = X_{T_1}$$

and for $n \geq 1$,

$$T_{n+1} = \inf\{k > T_n\,;\, X_k > R_n\}, R_{n+1} = X_{T_{n+1}}. \tag{17.6}$$

By the hypotheses, a.s. for any $n \geq 1$, T_n is finite. Let $\{(\hat{T}_n, \hat{R}_n)\}_{n\geq 1}$ be a sequence of random vectors, taking values in $\boldsymbol{N} \times \boldsymbol{R}$ and having the same distribution as the records defined with respect to r_0 of an i.i.d. sequence, with same marginal distribution as $\{X_n\}$.

It is well known that $\{(\hat{T}_n, \hat{R}_n)\}_{n\geq 1}$ form a Markov chain such that for any integers $1 \leq t_1 < t_2 < \cdots < t_{n+1}$ and any $r_0 < r_1 < \cdots < r_{n+1} < \omega$ we have

$$\begin{aligned}
&P\left\{\hat{T}_{n+1} = t_{n+1}, \hat{R}_{n+1} > r_{n+1} \mid \hat{T}_1 = t_1, \hat{R}_1 = r_1, \ldots, \hat{T}_n = t_n, \hat{R}_n = r_n\right\} \\
&= P\left\{\hat{T}_{n+1} = t_{n+1}, \hat{R}_{n+1} > r_{n+1} \mid \hat{R}_n = r_n\right\} \\
&= (F(r_n))^{t_{n+1}-t_n-1}(1 - F(r_{n+1}))\,.
\end{aligned} \tag{17.7}$$

The construction in Theorem 17.1.1 easily follows from the following Theorem 17.2.1.

Theorem 17.2.1 *Let $\{X_n\}$ satisfy condition (17.1). Then one can construct, on the probability space on which $\{X_n\}$ is defined, enlarged by independent factors, a sequence $\{(\hat{T}_n, \hat{R}_n)\}$, satisfying (17.7), such that a.s. there exists a random range ν such that for $n \geq \nu$, we have*

$$\begin{aligned}
\hat{T}_{n+1} &= \inf\{k > \hat{T}_n\,;\, X_k > \hat{R}_n\} \\
&= \inf\{k > \hat{T}_n + m + 1\,;\, X_k > \hat{R}_n\}
\end{aligned} \tag{17.8}$$

and

$$\hat{R}_{n+1} = X_{\hat{T}_{n+1}}.$$

Indeed, observe that, if (17.8) is satisfied, then there exist integers ν' and Q such that a.s. for $n \geq \nu'$ we have

$$\hat{T}_n = T_{n-Q} \quad \text{and} \quad \hat{R}_n = R_{n-Q}\,.$$

Thus, since the T_n's are the instants when the M_n's change and the R_n's the corresponding values, it is not difficult to construct the final i.i.d. sequence $\{\hat{X}_n\}$ the records of which, with respect to r_0, are $\{(\hat{T}_n, \hat{R}_n)\}$.
The first step of the proof of Theorem 17.1.2 is the following Lemma 17.2.1.

Lemma 17.2.1 *The random variable ν in Theorem 17.2.1 satisfies*

$$P\{\nu > t\} = O(t^{-\beta/2}), t > 0. \tag{17.9}$$

PROOF. The construction in Theorem 17.2.1 is such that if C_n denotes the event $C_n = (\hat{T}_{n+1} \neq \inf\{k > \hat{T}_n, X_k > \hat{R}_n\}$ or $\hat{R}_{n+1} \neq X_{\hat{T}_{n+1}})$, we have

$$P(C_n) = O(n^{-1-\beta/2})$$

and then the statement of the theorem follows by the first Borel Cantelli Lemma.
But then, we also have

$$P(\nu > t) \leq P(\bigcup_{n \geq t} C_n) = O(t^{-\beta/2})\,.$$

■

The next step involves proving the following Lemma 17.2.2.

Lemma 17.2.2 *We have for $s \geq 2$,*

$$P\{\hat{T}_\nu > s\} = O((\log s))^{-\beta/(4+\beta)}\,. \tag{17.10}$$

PROOF. We first establish, for any n, the formula

$$\begin{aligned} E(\log \hat{T}_n) &= n - c + \log\left(\frac{F(r_0)}{1-F(r_0)}\right) + \frac{1-F(r_0)}{F(r_0)} + \frac{d}{2(F(r_0))^2} + O(1/2^n) \\ &= :A(n, r_0)\,, \end{aligned} \tag{17.11}$$

with $c =$ Euler constant, $|d| \leq 1$, and $|O(x)| \leq K(r_0)\,|x|$.
From (17.11) and the Markov inequality, we get

$$P\{\hat{T}_n > s\} \leq \frac{A(n, r_0)}{\log s}, s \geq 2\,. \tag{17.12}$$

The proof of (17.11) is based on the corresponding result obtained in Pfeifer (1984) for the classical record times sequence $\{L_n\}$ of an i.i.d. sequence $\{Y_n\}$.

$\{L_n\}$ are defined as $L_1 = 1$, $L_{k+1} = \inf\{t > L_k, Y_t > Y_{L_k}\}, k \geq 1$ and Pfeifer's result is the formula

$$E(\log L_n) = n - c + O(E(1/L_n)), \text{ where } E(1/L_n) = O(n^2/2^n).$$

The latter formula was improved in Nevzorov (1995) [see also Nevzorov and Balakrishnan (1998)], who obtained

$$E(1/L_n) = \frac{1}{2^n} + O(1/3^n)\,.$$

Notice that the L_n's are different from the record times defined with respect to an initial threshold. However, if $\{\tau_n\}_{n\geq 1}$ is the record time sequence of $\{Y_n\}$ with respect to some initial threshold u_0 (i.e. $\tau_1 = \inf\{t \geq 1\,;\, Y_t > u_0\}$ and $\tau_{k+1} = \inf\{k > \tau_k\,;\, Y_k > Y_{\tau_k}\}$), then there exist η_0 and q such that for $n \geq \eta_0$ we have $L_n = \tau_{n-q}$.

Going back to the proof of (17.10), for any $t \geq 1$ we have

$$P(\hat{T}_\nu > s) \leq P(\hat{T}_\nu > s, \nu \leq t) + P(\nu > t)\,,$$

where, by (17.9), there is a constant K_1 such that $P(\nu > t) \leq K_1 t^{-\beta/2}$. Next,

$$\begin{aligned} P(\hat{T}_\nu > s, \nu \leq t) &= \sum_{n=1}^{t} P(\hat{T}_n > s, \nu = n) \\ &\leq \sum_{n=1}^{t} P(\hat{T}_n > s). \end{aligned} \tag{17.13}$$

By (17.12), there is a constant K_2 such that the last sum in (17.13) is majorized by $K_2 t^2/\log s$.

Thus, for any $t \geq 1$,

$$P(\hat{T}_\nu > s) \leq K_1 t^{-\beta/2} + K_2 t^2/\log s\,, \qquad s \geq 2. \tag{17.14}$$

Taking $t = [(\log s)^{1/(2+\beta/2)}]$, (where [ ] stands for integer part), there is a constant K such that the right hand term in the above inequality is majorized by $K(\log s)^{-\beta/(4+\beta)}$. ■

In order to prove (17.4) we now assume, without loss of generality, that the X_n's (thus the $\hat{X}_n$'s) are uniformly on $[0, 1]$ distributed.

We are now in position to prove (17.4). Let

$$L = \inf\{k \geq 0\,;\, X_{\hat{T}_\nu + k} > M_{\hat{T}_\nu}\}. \tag{17.15}$$

Let $0 < \alpha < 1$ be a fixed constant and $\{u_n\}_{n\geq 1}$, $0 < u_n < 1$, an increasing sequence, $\lim_{n\to\infty} u_n = 1$. We then have, for any $s \geq 1$,

$$\begin{aligned} P(N > s) &\leq P\{\hat{T}_\nu + L > s\} \qquad (17.16)\\ &\leq P\{(\hat{T}_\nu + L > s) \cap (M_{[s^\alpha]} < u_{[s^\alpha]})\} + P\{M_{[s^\alpha]} \geq u_{[s^\alpha]}\}. \end{aligned}$$

Next,

$$P\{(\hat{T}_\nu + L > s) \cap (M_{[s^\alpha]} < u_{[s^\alpha]})\} \leq P(A) + P\{\hat{T}_\nu > [s^\alpha]\}\,, \tag{17.17}$$

where

$$A = (\hat{T}_\nu \leq [s^\alpha]) \cap (\hat{T}_\nu + L > s) \cap (M_{[s^\alpha]} < u_{[s^\alpha]})\,.$$

By (17.10), we have

$$P\{\hat{T}_\nu > [s^\alpha]\} = O((\log s)^{-\beta/(4+\beta)}), s \geq 2\,. \tag{17.18}$$

Next,

$$\begin{aligned} A &\subset (\hat{T}_\nu \leq [s^\alpha]) \cap (L > s - [s^\alpha]) \cap (M_{[s^\alpha]} < u_{[s^\alpha]}) \\ &\subset \{\hat{T}_\nu \leq [s^\alpha]\} \cap \{\mathrm{Max}(X_{\hat{T}_\nu+1}, \dots, X_{\hat{T}+s-[s^\alpha]}) < u_{[s^\alpha]}\} \\ &\subset \bigcup_{\tau=1}^{[s^\alpha]} \{\mathrm{Max}(X_{\tau+1}, \dots, X_{\tau+s-[s^\alpha]}) < u_{[s^\alpha]}\}\,. \end{aligned}$$

Thus, by stationarity,

$$P(A) \leq s^\alpha \; P\{M_{s-[s^\alpha]} < u_{[s^\alpha]}\} \leq s^\alpha (u_{[s^\alpha]})^{(s-[s^\alpha])/2}. \tag{17.19}$$

Let

$$u_n = 1 - \frac{(\log n)^{-\beta/(4+\beta)}}{n}\,. \tag{17.20}$$

Then, (17.19) implies that there exists a $\gamma > 0$ such that

$$P(A) = O(s^\alpha \cdot e^{-s^\gamma}) = o((\log s))^{-\beta/(4+\beta)}). \tag{17.21}$$

In order to bound the term $P\{M_{[s^\alpha]} \geq u_{[s^\alpha]}\}$, we apply the next Lemma, which is a consequence of Haiman (1987, Theorem 1).

Lemma 17.2.3 *If the sequence* $\{u_n\}_{n\geq 1}$ *is such that* $\lim_{n\to\infty} n(1-u_n) = 0$, *then*

$$\lim_{n\to\infty} P\{M_n \geq u_n\}/n(1-u_n) = 1\,. \tag{17.22}$$

With $\{u_n\}$ defined in (17.20), we then obtain

$$P\{M_{[s^\alpha]} \geq u_{[s^\alpha]}\} = O\Big((\log s)^{-\beta/(4+\beta)}\Big)\,. \tag{17.23}$$

Combining (17.16), (17.17), (17.18), (17.21) and (17.23), we get (17.4).

References

1. Haiman, G. (1981). Valeurs extrêmales de suites stationnaires de variables aléatoires m-dépendantes, *Annals of the Institute of H. Poincaré*, **17**, 309–330.

2. Haiman, G. (1987). Etude des extrêmes d'une suite stationnaire m-dépendante avec une application relative aux accroissements du processus de Wiener, *Annals of the Institute of H. Poincaré*, **23**, 425–458.

3. Haiman, G. (1992). A strong invariance principle for the extremes of multivariate stationary m-dependent sequences, *Journal of Statistical Planning and Inference*, **32**, 147–163.

4. Haiman, G. (1999). Block records and maxima of the increments of the Wiener process, *Journal of Statistical Planning and Inference* (to appear).

5. Haiman, G. and Habach, L. (1999). On the a.s. behaviour of extremes of m-dependent sequences, *Preprint.*

6. Haiman, G., Mayeur, N., Nevzorov, V. and Puri, M. (1998). Records and 2-blocks records of 1-dependent stationary sequences under local dependence, *Annals of the Institute of Henri Poincaré*, **34**, 481–503.

7. Pfeifer, D. (1984). A note on moments of certain record statistics, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **66**, 293–296.

8. Nevzorov, V. B. (1995). Asymptotic distributions of records in nonstationary schemes, *Journal of Statistical Planning and Inference*, **45**, 261–273.

9. Nevzorov, V. B. and Balakrishnan, N. (1998). A record of records, In *Handbook of Statistics–16: Order Statistics: Theory and Methods* (Eds., N. Balakrishnan and C. R. Rao), North-Holland, Amsterdam.

18

On a Strong Limit Theorem for Sums of Independent Random Variables

Valentin V. Petrov

St. Petersburg State University, St. Petersburg, Russia

Abstract: This chapter examines the almost sure behaviour of sums of independent non-identically distributed random variables. An extension of some results of Chung and Erdös is obtained.

Keywords and phrases: Almost sure convergence, sums of independent random variables, Chung–Erdös theorem

18.1 Introduction and Results

The set of functions $\psi(x)$ that are positive and non-decreasing in the region $x > x_0$ for some x_0 and such that the series $\sum 1/(n\psi(n))$ converges (diverges) will be denoted by Ψ_c (respectively, Ψ_d).

Chung and Erdös (1947) proved that if $\{X_n\}$ is a sequence of independent random variables having a common distribution function with non-zero absolutely continuous component and if $EX_1 = 0$, $E|X_1|^5 < \infty$, then

$$\liminf_{n\to\infty} n^{1/2}\psi(n)|S_n| > 0 \quad a.s. \tag{18.1}$$

for every function $\psi \in \Psi_c$, but if $\psi \in \Psi_d$, then

$$\liminf_{n\to\infty} n^{1/2}\psi(n)|S_n| = 0 \quad a.s. \tag{18.2}$$

Here, $S_n = X_1 + X_2 + \cdots + X_n$. Some analogues of these results were obtained by Cote (1955) for sequences of independent non-identically distributed random variables under severe assumptions including uniformly bounded moments of the 5th order.

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

In Petrov (1978a), and Petrov (1978b) [see also Petrov (1995a, p. 224)], it was proved that if $\{X_n\}$ is a sequence of independent identically distributed random variables satisfying the Cramér condition

$$\limsup_{|t|\to\infty} |Ee^{itX_1}| < 1, \tag{C}$$

then

$$\lim_{n\to\infty} n^{1/2}\psi(n)|S_n| = \infty \quad a.s. \tag{18.3}$$

for every $\psi \in \Psi_c$; if the additional assumptions $EX_1 = 0$ and $EX_1^2 < \infty$ are satisfied, then (18.2) holds for every $\psi \in \Psi_d$.

In Petrov (1995b), sufficient conditions are given for relation (18.3) with $\psi \in \Psi_c$ in the case of independent non-identically distributed random variables. We shall be interested in sufficient conditions for (18.2) when $\psi \in \Psi_d$ also in the case of independent non-identically distributed summands.

Let $\{X_n, n = 1, 2, \ldots\}$ be a sequence of independent random variables with zero means and finite absolute third moments. We put

$$B_{n,k} = \sum_{j=k+1}^{n} EX_j^2, \quad D_{n,k} = \sum_{j=k+1}^{n} E|X_j|^3 \;\; (0 \le k \le n-1), \quad B_n = B_{n,0}.$$

Theorem 18.1.1 *Suppose that*

$$B_n \ge c_0 n \;\text{ for all sufficiently large }\; n, \tag{18.4}$$

$$D_{n,k} \le c_1(n-k) \;\text{ for all sufficiently large }\; n-k \;\text{ and }\; k, \tag{18.5}$$

where c_0 and c_1 are some positive constants. Moreover, we assume that

$$B_{n,k} \ge (1-\delta)B_{n-k} \tag{18.6}$$

for every $\delta > 0$ and all sufficiently large $n-k$ and k. Finally, we assume that the sequence $\{f_n(t), n = 1, 2, \ldots\}$, where $f_n(t) = Ee^{itX_n}$, contains a subsequence $\{f_{n_m}(t); m = 1, 2, \ldots\}$ with the following properties: (A) $|f_{n_m}(t)| \le C|t|^{-\alpha}$ for $|t| \ge R$ and some positive constants C, α and R $(m = 1, 2, \ldots)$, (B) if $r_N(n)$ is the number of elements of the subsequence $\{f_{n_m}(t)\}$ in the set $f_{N+1}(t), \ldots, f_{N+n}(t)$, then $r_N(n) \ge cn$ for all sufficiently large N and n, where c is a positive constant not depending on n and N.

Under the above-mentioned assumptions, relation (18.2) holds for every $\psi \in \Psi_d$.

It follows from Petrov (1995b) that conditions (A) and (B) with $N = 0$ are sufficient for relation (18.3) when $\psi \in \Psi_c$ even without any moment conditions. More general results related to the case $\psi \in \Psi_c$ can be found in Petrov (1998) for arbitrary sequences of random variables without the independence condition and any moment assumptions.

18.2 Proofs

In what follows, we assume that conditions of Theorem 18.1.1 are satisfied.

Lemma 18.2.1 *Let $a < b$, $a = o(B_{n,k}^{1/2})$ and $b = o(B_{n,k}^{1/2})$ as $n - k \to \infty$. Then,*

$$P(a \le S_n - S_k \le b) = \frac{b-a}{(2\pi B_{n,k})^{1/2}}(1 + o(1))$$

as $n - k \to \infty$.

This lemma of independent interest is an extension of a result of Shepp (1964) obtained for independent identically distributed random variables under weaker conditions. Its proof is lengthy and will be published elsewhere. Lemma 18.2.1 remains true if we replace condition (18.4) and (18.6) by the condition $B_{n,k} \ge c_0(n-k)$ for all sufficiently large $n - k$ and k.

Lemma 18.2.2 *Let c be an arbitrary positive constant. Then*

$$P(n^{1/2}\psi(n)|S_n| \le c \text{ i.o. }) = 1$$

for every function $\psi \in \Psi_d$.

PROOF. We put

$$\lambda_n = \frac{c}{n^{1/2}\psi(n)}, \quad D_n = [|S_n| \le \lambda_n], \quad E_n = [|S_n| > \lambda_n], \quad P_n = P(D_n).$$

Moreover, for any fixed sufficiently large integer N, we put

$$G_k = \bigcap_{r=N}^{k-1} E_r \;\; (k > N), \;\; Q_N = P_N, \;\; Q_k = P(D_k G_k) \;\; (k > N),$$

$$P_{N,N} = 1, \;\; P_{N,n} = P(D_n | D_N) \;\; (n > N),$$

$$P_{k,n} = P(D_n | D_k G_k) \;\; (N < k < n).$$

It is easy to show by induction that

$$\sum_{n=N}^{m} P_n = \sum_{k=N}^{m} Q_k \sum_{n=k}^{m} P_{k,n} \tag{18.7}$$

for every $m \ge N$.

We shall prove the following statement:
for arbitrary $\varepsilon > 0$, there exists an integer M such that

$$P_{k,n} \le (1+\varepsilon) P_{n-k} \tag{18.8}$$

for $k > M$ and $n - k > M$.

Let γ be an arbitrary positive number. We represent the interval $|x| \leq \lambda_n$ as the union of non-overlapping intervals I_j with length $l_j \leq \gamma\lambda_n$ for every j. Let T_j be an arbitrary interval with length $t_j \leq 2\lambda_n + \gamma\lambda_n$ that contains in the interval $[-\lambda_k - \lambda_n, \lambda_k + \lambda_n]$.

Taking into account Lemma 18.2.1 and condition (18.6), we get

$$\max_j P(S_n - S_k \in T_j) \leq \frac{2(1+\gamma)\lambda_{n-k}}{(2\pi B_{n-k})^{1/2}} \tag{18.9}$$

for all sufficiently large k and $n-k$. We put

$$H_j^{(k)} = [S_k \in I_j].$$

Obviously,

$$\bigcup_j H_j^{(k)} = [|S_k| \leq \lambda_k] = D_k, \quad \sum_j P(H_j^{(k)}) = P_k.$$

We have

$$\begin{aligned} P_{k,n} &= \frac{P(D_n D_k G_k)}{P(D_k G_k)} = \Big\{\sum_j P(G_k H_j^{(k)})\Big\}^{-1} \sum_j P(D_n G_k H_j^{(k)}) \\ &\leq \Big\{\sum_j P(G_k H_j^{(k)})\Big\}^{-1} \sum_j P(G_k H_j^{(k)} \cap [S_n - S_k \in T_j]) \end{aligned}$$

where T_j $(j = 1, 2, \ldots)$ is an interval of length $t_j \leq 2\lambda_n + \gamma\lambda_n$ containing in the interval $[-\lambda_k - \lambda_n, \lambda_k + \lambda_n]$. Since $X_1, X_2, \ldots$ is a sequence of independent random variables, we obtain

$$P(G_k H_j^{(k)} \cap [S_n - S_k \in T_j]) = P(G_k H_j^{(k)}) P(S_n - S_k \in T_j)$$

and

$$P_{k,n} \leq \max_j P(S_n - S_k \in T_j).$$

Making use of (18.9) and definition of λ_n, we get

$$P_{k,n} \leq \frac{2(1+\gamma)\lambda_{n-k}}{(2\pi B_{n-k})^{1/2}} \leq \frac{2(1+\gamma)c}{(2\pi B_{n-k}(n-k))^{1/2}\psi(n-k)} \tag{18.10}$$

for all sufficiently large k and $n-k$. It follows from Lemma 18.2.1 that

$$P_{k,n} \geq \frac{2(1-\gamma)c}{(2\pi B_{n-k}(n-k))^{1/2}\psi(n-k)} \tag{18.11}$$

for every positive fixed γ and for all sufficiently large $n-k$. Inequalities (18.10) and (18.11) imply (18.8).

Let c be an arbitrary fixed positive number. If N and L are sufficiently large numbers, we obtain from (18.7) and (18.8) that

$$\sum_{n=N}^{m} P_n = \sum_{k=N}^{m} Q_k \Big\{ \sum_{n=k}^{k+L} P_{k,n} + (1+\varepsilon) \sum_{n=k+L+1}^{m} P_{n-k} \Big\}$$
$$\leq \sum_{k=N}^{m} Q_k \Big\{ L + 1 + (1+\varepsilon) \sum_{r=L+1}^{m} P_r \Big\}.$$

Therefore,

$$\sum_{k=N}^{m} Q_k \geq \Big\{ L + 1 + (1+\varepsilon) \sum_{r=L+1}^{m} P_r \Big\}^{-1} \sum_{n=N}^{m} P_n. \tag{18.12}$$

Taking into account inequality (18.11) and conditions (18.5) and $\psi \in \Psi_d$, we have $\sum_{n=N}^{\infty} P_n = \infty$. Passing to the limit in (18.12) as $m \to \infty$, we get

$$\sum_{k=N}^{\infty} Q_k \geq \frac{1}{1+\varepsilon}. \tag{18.13}$$

Here, ε is an arbitrary positive number. The left-hand side of (18.13) does not depend on ε. Therefore, $\sum_{k=N}^{\infty} Q_k \geq 1$. Since

$$\sum_{k=N}^{\infty} Q_k = P(D_n \cup E_N D_{N+1} \cup E_N E_{N+1} D_{N+2} \cup \cdots) = P(\bigcup_{n=N}^{\infty} D_n),$$

we have

$$P(\bigcup_{n=N}^{\infty} D_n) = 1. \tag{18.14}$$

It follows from relations

$$\bigcup_{n=N}^{\infty} D_n \supset \bigcup_{n=N+1}^{\infty} D_n \supset \cdots$$

and

$$\bigcap_{N=1}^{\infty} \bigcup_{n=N}^{\infty} D_n = \limsup D_n$$

that there exists the limit

$$\lim_{N\to\infty} P(\bigcup_{n=N}^{\infty} D_n) = P(\limsup D_n).$$

Applying equality (18.14), we conclude that $P(\limsup D_n) = 1$. Lemma 18.2.2 is proved. ■

Relation (18.2) readily follows from Lemma 18.2.2.

Acknowledgement. This research was partially supported by Ministry of Common and Professional Education of Russia, grant N 97–M3–32.

References

1. Chung, K. L. and Erdös, P. (1947). On the lower limit of sums of independent random variables, *Annals of Mathematics*, **48**, 1003–1013.

2. Cote, L. J. (1955). On fluctuations of sums of random variables, *Proceedings of the American Mathematical Society*, **6**, 135–144.

3. Petrov, V. V. (1978a). On the lower limit of the modulus of sums of independent random variables, *Vestnik Leningrad Univ. Ser. Math.*, N 7, 49–54. English translation: *Vestnik Leningrad Univ. Math.*, **11** (1983), 147–152.

4. Petrov, V. V. (1978b). Remark on the lower limit for the modulus of sums of independent random variables, *Lithuanian Mathematical Journal*, **18**, 528–531.

5. Petrov, V. V. (1995a). *Limit Theorems of Probability Theory,* London, England: Oxford University Press.

6. Petrov, V. V. (1995b). On the almost sure behaviour of sums of random variables, *Statistics & Probability Letters*, **24**, 229–231.

7. Petrov, V. V. (1998). Almost sure behaviour of some random sequences, *Studia Sci. Math. Hungar.*, **34**, 345–348.

8. Shepp, L. A. (1964). A local limit theorem, *Annals of Mathematics & Statistics*, **35**, 419–423.

PART V

Large Deviation Probabilities

19

Development of Linnik's Work in His Investigation of the Probabilities of Large Deviation

A. Aleškevičienė, V. Statulevičius, and K. Padvelskis

Institute of Mathematics and Informatics, Vilnius, Lithuania
Institute of Mathematics and Informatics, Vilnius, Lithuania
Kaunas Vytautas Magnus University, Kaunas, Lithuania

Abstract: In this chapter, we present some large deviation results for the distribution of the normed sum of random variables related to a Markov chain with the ergodicity coefficient under conditions similar to the case of independent summands. The method of cumulants will be used to prove these results.

Keywords and phrases: Large deviation, ergodicity coefficient, Markov chain, cumulants, factorial cumulants, Cramer series, partition, correlation function, factorial moments

19.1 Reminiscences on Yu. V. Linnik (V. Statulevičius)

In the period of 1954–1957, I was a post-graduate student in the Ph.D. studies in Leningrad. The topic of my thesis was *Local limit theorems for nonhomogeneous Markov chains.* In the case when all transition probabilities $p_{ij}(n) \geq \lambda > 0$, the local limit theorem was proved by Yu. V. Linnik himself. I succeeded to prove it in a rather general case (when $\alpha^{(n)} B_n \to \infty, n \to \infty$, where $\alpha^{(n)}$ is the ergodicity coefficient), after introducing the so-called characteristic functions of transition.

His post-graduate students used to go to his place in the evening to discuss some mathematical problems and to report on the work done. His wife Lyudmila Pavlovna usually treated us to very tasty tea. One such evening, Yu. V. Linnik, being in a good mood, told about himself. He said he liked to solve only difficult problems. Suppose a known mathematician formulates a problem the solution of which is of great importance to mathematics or its applications. Yu. V. Linnik said: "One year passes, the problem is not solved; another year

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

passes, no solution. Then I set to its solution myself!" He said that in applying the theory of complex functions, he was ready to outdo anyone. Having chosen the method of solution, he used to go straight out, without looking around. He reported his successful results at a seminar. However, he worried about where to publish his results since they were about 80 pages long. His post-graduate students such as A. A. Zinger, V. V. Petrov, I. A. Ibragimov, and the author of these lines curtailed the paper to 10 to 20 pages. In brief, if some 'loops' appeared in Linnik's proof, sometimes we succeeded in curtailing them (and to curtail the length of the paper). He did not like undisciplined, ungifted students in Mathematics. Once I recall him asking a student during an exam whether Doob was not his relative. The student swore he was not and he never knew Doob. It was his humor—the famous book *Theory of Random Processes* by D. Doob had been published at that time. In Russian, the word 'doob' ('dub') has another meaning, namely; 'blockhead'. A dull, uneducated man was called 'doob'. I have already mentioned, academician Yu. V. Linnik was fond of solving delayed and difficult problems. So, if a post-graduate student got new good results, he used to send him for discussion with A. N. Kolmogorov or N. V. Smirnov (if it was a statistical result). This is how we got familiar with Moscow representatives of the school of probability theory and statistics which included A. N. Kolmogorov, N. V. Smirnov, B. V. Gnedenko, E. B. Dynkin, Yu. V. Prokhorov, R. L. Dobrushin, L. N. Bolshev, and others. If you were a success, he tried to 'settle you down on the track' and there was no way back; you could not help but work a lot.

Based on the results I had obtained, I was awarded a prize with a nice diploma 'Laureate of Leningrad University prize', while the newspaper *Evening Leningrad* published an article entitled "A young scientist can be greeted with a good beginning." I think each of his students can tell such interesting stories with deep gratitude to Yu. V. Linnik, and we are always proud to say that we are Yu. V. Linnik's students.

19.2 Theorems of Large Deviations for Sums of Random Variables Related to a Markov Chain

On a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, let there be given a Markov process $\xi_t(\omega)$ with the values from a measurable space $(\mathbf{X}_t, \mathcal{B}_t)$, $t = 1, \ldots, n$, with transition probabilities $P_{s,t}(x, B) = \mathbf{P}\{\xi_t \in B \mid \xi_s = x\}$ from the state $x \in \mathbf{X}_s$ at time s to a set of states $B \in \mathcal{B}_t$ at time t, $0 \le s < t \le n$, and the initial probability distribution $P_0(B) = \mathbf{P}(\xi_0 \in B)$, $B \in \mathcal{B}_0$. Let us introduce σ-algebras $\mathcal{F}_t = \sigma\{\xi_t\}$ and $\mathcal{F}_t^{t'} = \sigma\{\xi_s,\ t \le s \le t'\}$. We consider random variables

$$X_1, \ldots X_n,$$

related to a Markov chain ξ_t, $t = 0, 1, \ldots, n$, i.e., $X_t = g_t(\xi_t)$, $t = 1, \ldots, n$, where $g_t(x)$ is a real $\mathcal{B}_t$-measurable function defined on $\mathbf{X}_t$.

As a measure of dependence of random variables $X_1, \ldots, X_n$, we will use the ergodicity coefficient of the transition function $P_{s,t}(x, \mathcal{B})$ introduced by Dobrushin (1956) given by

$$\alpha_{s,t} = 1 - \sup_{\substack{x,y \in \mathbf{X}_s \\ \mathcal{B} \in \mathcal{B}_t}} |P_{s,t}(x, B) - P_{s,t}(y, B)|.$$

Denote $\alpha^{(n)} = \min_{1 \le t \le n} \alpha_{t-1,t}$. Let us introduce the following functions of α-mixing, φ-mixing, and ψ-mixing:

$$\alpha(s,t) = \sup_{A \in \mathcal{F}_0^s, \mathcal{B} \in \mathcal{F}_t^n} |\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)|,$$

$$\varphi(s,t) = \sup_{\substack{A \in \mathcal{F}_0^s, \mathcal{B} \in \mathcal{F}_t^n, \\ \mathbf{P}(A) > 0}} \left| \frac{\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)}{\mathbf{P}(A)} \right|,$$

$$\psi(s,t) = \sup_{\substack{A \in \mathcal{F}_0^s, \mathcal{B} \in \mathcal{F}_t^n, \\ \mathbf{P}(A) > 0, \mathbf{P}(B) > 0}} \left| \frac{\mathbf{P}(AB) - \mathbf{P}(A)\mathbf{P}(B)}{\mathbf{P}(A)\mathbf{P}(B)} \right|.$$

It is known that

$$\alpha(s,t) \le \varphi(s,t) \le \psi(s,t).$$

It is also easy to show that

$$\frac{1}{2}(1 - \alpha_{s,t}) \le \varphi(s,t) \le 1 - \alpha_{s,t}.$$

If we have inequalities connecting different mixing functions with the ergodicity coefficient $\alpha_{s,t}$, we can obtain, as consequences, some propositions proved by applying these functions. Suppose that $\mathbf{E}X_t = 0$ and $0 < \sigma_t^2 = \mathbf{E}X_t^2 < \infty$, $t = 1, \ldots, n$. Let

$$S_n = \sum_{t=1}^{n} X_t, \quad S_{k,l} = \sum_{t=k+1}^{l} X_t, \quad 0 \le k < l \le n,$$
$$B_n^2 = \mathbf{D}S_n, \quad Z_n = B_n^{-1} S_n.$$

We are interested in large deviation probabilities of the distribution $F_{Z_n}(x) = \mathbf{P}(Z_n < x)$ of the normed sum Z_n of random variables related to a Markov chain with the ergodicity coefficient α_n under conditions similar to the case of independent summands. We will use the method of cumulants in the investigation.

Statulevičius (1969) obtained optimal results in the case of bounded random variables X_t, i.e., when $|X_t| \le C_n$, $t = 1, \ldots, n$ with probability 1. To investigate the behavior of cumulants of the sum S_n with respect to B_n, we used the scheme of enlargement of summands X_t proposed by Dobrushin (1956). For unbounded X_t, we have found another scheme of enlargement of summands.

The case of a homogeneous chain is much simpler (because we can do without the enlargement of summands) and has been studied in detail by Saulis and Statulevičius (1989).

Denote by $\Phi(x)$ a (0,1) – normal distribution function, and by $\Gamma_k(\xi)$ a cumulant of order k of a random variable ξ given by

$$\Gamma_k(\xi) = \frac{1}{i^k}\frac{d^k}{dt^k}(\log f_\xi(t))\Big|_{t=0},$$

if $\mathbf{E}|\xi|^k < \infty$. Here, $f_\xi(t)$ is a characteristic function of the random variable ξ. Put

$$\mathcal{F}_{t-1,t+1} = \begin{cases} \mathcal{F}_{t-1} \otimes \mathcal{F}_{t+1}, & \text{if } t = 1, \ldots, n-1, \\ \mathcal{F}_{n-1}, & \text{if } t = n. \end{cases}$$

By C, $C(\gamma)$, we denote finite positive not always the same constants absolute or dependent only on γ, respectively.

Theorem 19.2.1 *If* $\alpha^{(n)} > 0$ *and with probability* 1

$$\mathbf{E}(|X_t|^k|\mathcal{F}_{t-1,t+1}) \le (k!)^{1+\gamma_1} H^k \sigma_t^2, \quad t = 1, \ldots, n, \quad k = 2, 3, \ldots,$$

for some $\gamma_1 \ge 0$ *and* $H > 0$, *then*

$$|\Gamma_k(Z_n)| \le (k!)^{1+\gamma_1} \left(\frac{C(\gamma_1) H \max\limits_{1\le t\le n} \sigma_t}{\alpha^{(n)} B_n} \right)^{k-2}, \qquad k = 3, 4, \ldots$$

and, for

$$\gamma = \gamma_1, \qquad \Delta_\gamma = \left(\frac{C(\gamma)\, \alpha^{(n)} B_n}{H \max\limits_{1\le t\le n} \sigma_t} \right)^{\frac{1}{1+2\gamma}}$$

in the interval

$$0 \le x < \Delta_\gamma$$

the following relations of large deviations

$$\begin{aligned} \frac{1 - F_{Z_n}(x)}{1 - \Phi(x)} &= \exp\{L_\gamma(x)\} \left(1 + \theta f(x) \frac{x+1}{\Delta_\gamma}\right), \\ \frac{F_{Z_n}(-x)}{\Phi(-x)} &= \exp\{L_\gamma(-x)\}\left(1 + \theta f(x) \frac{x+1}{\Delta_\gamma}\right) \end{aligned} \qquad (19.1)$$

hold. Here,

$$L_\gamma(x) = \sum_{3\le k<p} \lambda_k x^k, \qquad p = \begin{cases} \dfrac{1}{\gamma} + 2, & \textit{if } \gamma > 0, \\ \infty, & \textit{if } \gamma = 0. \end{cases}$$

The coefficients λ_k *are expressed by cumulants of the random variable* Z_n *and are the same as those of the Cramer series, and*

$$f(x) = \frac{60\left(1 + 10\,\Delta_\gamma^2 \exp\left\{-\left(1 - \frac{x}{\Delta_\gamma}\right)\sqrt{\Delta_\gamma}\right\}\right)}{1 - \frac{x}{\Delta_\gamma}}, \qquad |\theta| \le 1.$$

Theorem 19.2.2 *If* $\alpha^{(n)} > 0$ *and* [1]

$$\widetilde{L}_{k,n} = \frac{\sum_{t=1}^{n} \mathbf{E}^*(|X_t|^k \mid \mathcal{F}_{t-1,t+1})}{{\alpha^{(n)}}^{k-1} B_n^k} \le \frac{(k!)^{1+\gamma_2}}{\Delta_n^{k-2}}, \qquad k = 3,\, 4,\, \ldots$$

for some $\gamma_2 \ge 0$ *and* $\Delta_n \ge e$, *then* [2]

$$|\Gamma_k(Z_n)| \le \begin{cases} k!\left(\dfrac{C\log\Delta_n}{\Delta_n}\right)^{k-2}, & k = 3,\, 4,\, \ldots, \qquad \textit{if } \gamma_2 = 0 \\ \dfrac{k!}{\left(C(\gamma_2)\,\Delta_n^{\frac{1}{1+2\gamma_2}}\right)^{k-2}}, & k = 3,\, 4,\, \ldots,\, 2\left[\Delta_n^{\frac{2}{1+2\gamma_2}}\right], \qquad \textit{if } \gamma_2 > 0 \end{cases}$$

and, for

$$\gamma = \gamma_2, \qquad \Delta_\gamma = \overline{\Delta} = \begin{cases} \dfrac{C\Delta_n}{\log\Delta_n}, & \textit{if } \gamma = 0 \\ C(\gamma)\,\Delta_n^{\frac{1}{1+2\gamma}}, & \textit{if } \gamma > 0 \end{cases}$$

in the interval

$$0 \le x < \Delta_\gamma,$$

the relations of large deviations in (19.1) hold, where

$$L_\gamma(x) = \sum_{3 \le k < p} \lambda_k x^k + \theta\left(\frac{x}{\Delta_\gamma}\right)^3,$$

$$p = \begin{cases} \min\left\{\dfrac{1}{\gamma} + 2,\, 2\left[\Delta_n^{\frac{2}{1+2\gamma}}\right]\right\}, & \textit{if } \gamma > 0, \\ \infty, & \textit{if } \gamma = 0. \end{cases}$$

---

[1] $\mathbf{E}^*(\xi \mid \widehat{\mathcal{F}}) = \operatorname{ess\,sup} \mathbf{E}(\xi \mid \widehat{\mathcal{F}}) = \sup_{A \in \widehat{\mathcal{F}},\, \mathbf{P}(A) > 0} \mathbf{E}(\xi \mid \widehat{\mathcal{F}})$.

[2] $[x]$ denotes the integral part of the number x.

Remark. The multiplier $|\log \Delta_n|$ is essential. As $\gamma_2 = 0$, the estimate for the k-th order cumulant of the random variable Z_n is unimprovable with an accuracy up to a constant [see Saulis and Statulevičius (1989)], i.e., when investigating extremely large deviations $|x| = \varepsilon\,\Delta_n$, as $\gamma_2 = 0$, additional we need conditions on individual properties of the summands X_t, $t = 1, \dots, n$, not just on mean characteristics $\widetilde{L}_{k,n}$.

Below, we present some main steps of the proof.

In order to investigate the behavior of $\Gamma_k(S_n)$ with respect to B_n, we need the lower estimate for $\mathbf{D}S_n$. Since the random variables X_t $t = 1, \dots, n$, are correlated, the variance of sum S_n is not necessarily equal to the sum of variances of the summands. Therefore, we present S_n by the sum of enlarged summands so that B_n^2 will be of the same order as the sum of variances of these summands.

Let

$$r_i = i\left[\frac{1}{\alpha^{(n)}} + 1\right], \qquad i = 0,\, 1,\, 2, \dots,\, N-1,$$

where the number N is defined by the inequalities

$$r_{N-1} < n \;\le\; r_{N-1} + \left[\tfrac{1}{\alpha^{(n)}} + 1\right].$$

Let

$$W_0^{(1)} = W_0^{(2)} = 0,$$

$$W_i^{(1)} = \mathbf{E}(X_1 + \cdots + X_{r_i} \mid \mathcal{F}_{r_i}),$$

and

$$W_i^{(2)} = \mathbf{E}(|X_{r_{i-1}+1}| + \cdots + |X_{r_i}| \mid \mathcal{F}_{r_i}), \quad i = 1, \dots, N.$$

We determine random variables $\overline{W}_i$, $i = 0,\, 1, \dots,\, N$, as

$$\overline{W}_i = \min\{|W_i^{(1)}|, W_i^{(2)}\}\operatorname{sign} W_i^{(1)}.$$

We can the represent the sum S_n

$$S_n = \sum_{i=1}^{N} \sum_{t=r_{i-1}+1}^{r_i} X_t = \sum_{i=1}^{N} S_{r_{i-1}, r_i}.$$

When defining new random variables

$$\overline{Y}_i \;=\; \overline{W}_{i-1} + S_{r_{i-1}, r_i} - \overline{W}_i, \qquad i = 1, \dots, N-1,$$

$$\overline{Y}_N \;=\; \overline{W}_{N-1} + S_{r_{N-1}, r_N},$$

$$Y_i \;=\; \overline{Y}_i - \mathbf{E}\overline{Y}_i, \qquad i = 1, \dots, N,$$

we obtain

$$S_n = \sum_{i=1}^{N} \overline{Y}_i = \sum_{i=1}^{N} Y_i. \tag{19.2}$$

The random variables $Y_1, \ldots, Y_N$ possess the following: properties

(A) $|\mathbf{E}Y_kY_l| \le 2(1-\beta_{k,l})^{\frac{1}{2}}\mathbf{E}^{\frac{1}{2}}Y_k^2\mathbf{E}^{\frac{1}{2}}Y_l^2, \quad 1 \le k \le l \le N,$

where $0 < \beta_{k,l} < 1$, $1 \le k < l \le N$, and $1 - \beta_{k,l} \le \exp\{-(l-k-1)\}$;

(B) $\frac{1}{16} \min_{1\le t\le n} \sigma_t^2 \le \mathbf{D}Y_i \le \frac{36}{\alpha^{(n)2}} \max_{1\le t\le n} \sigma_t^2, \quad i = 1, \ldots, N;$

(C) $\frac{1}{24} \sum_{i=1}^{N} \mathbf{D}Y_i \le B_n^2 \le \frac{43}{5} \sum_{i=1}^{N} \mathbf{D}Y_i.$

The random variables Y_i are $\mathcal{F}_{r_{i-1}}^{r_i}$-measurable ($i = 1, \ldots, N$). Suppose $\widetilde{\mathcal{F}}_i = \mathcal{F}_{r_{i-1}}^{r_i-1}$ ($i = 1, \ldots, N$) and $\widetilde{\mathcal{F}}_0 = \mathcal{F}_0$. Next we evaluate the moments, centered moments, and k-th order cumulants of enlarged summands Y_i, $i = 1, \ldots, N$.

Lemma 19.2.1

(i) If the estimate

$$\mathbf{E}(|X_t|^k|\mathcal{F}_{t-1,t+1}) \le (k!)^{1+\gamma_1} H^k \sigma_t^2, \qquad t = 1, \ldots, n, \quad k = 2,\, 3 \ldots$$

holds with probability 1 *for some* $\gamma_1 \ge 0$ *and* $H > 0$, *then with probability* 1

$$\mathbf{E}(|Y_i|^k|\widetilde{\mathcal{F}}_{i-1}) \le (k!)^{1+\gamma_1} \left(\frac{10H}{\alpha^{(n)}}\right)^k \max_{1\le t\le n} \sigma_t^2 \tag{19.3}$$

for all $i = 1, \ldots, N$ *and* $k = 2,\, 3 \ldots.$

(ii) If

$$\widetilde{L}_{k,n} = \frac{\sum_{t=1}^{n} \mathbf{E}^*(|X_t|^k|\mathcal{F}_{t-1,t+1})}{\alpha^{(n)k-1} B_n^k} \le \frac{(k!)^{1+\gamma_2}}{\Delta_n^{k-2}}, \qquad k = 3,\, 4, \ldots$$

for some $\gamma_2 \ge 0$ *and* $\Delta_n \ge e$, *then, as* $\gamma_2 = 0$,

$$\mathbf{E}^*(|Y_i|^k|\widetilde{\mathcal{F}}_{i-1}) \le \frac{29}{9}\, k! \left(\frac{40 B_n \log \Delta_n}{\Delta_n}\right)^k \tag{19.4}$$

for all $i = 1, \ldots, N$ *and* $k = 2, 3, \ldots$, *and, as* $\gamma_2 > 0$,

$$\mathbf{E}^*(|Y_i|^k|\widetilde{\mathcal{F}}_{i-1}) \le k! \left(20\, \widetilde{C}(\gamma_2)\, B_n \Delta_n^{-\frac{1}{1+2\gamma_2}}\right)^k \tag{19.5}$$

for all $i = 1, \ldots, N$ *and* $k = 2, \ldots, s$, $s = 2\left[\Delta_n^{\frac{2}{1+2\gamma_2}}\right]$. *Here,*

$$\widetilde{C}(\gamma_2) = \max\left\{5^{\widehat{1+\gamma_2}}\,2^{\gamma_2},\ e\,(\widehat{1+\gamma_2})\left(\frac{(1+2\gamma_2)^2}{e\,\gamma_2}\right)^{1+2\gamma_2}\right\} \tag{19.6}$$

and $\widehat{x}$ *is the minimal integer* $\geq x$.

Let $\mathcal{N} = \{1, \ldots, n\}$ and $I = \{t_1, \ldots, t_k \,|\, t_j \in \mathcal{N},\ 1 \leq j \leq k\}$. If for some $k > 2$, $\mathbf{E}|X_t|^k < \infty$, $t = 1, \ldots$, then the centered moment $\widehat{\mathbf{E}}X_I$ is determined in the following way

$$\widehat{\mathbf{E}}X_I = \widehat{\mathbf{E}}X_{t_1}\ldots X_{t_k} = \mathbf{E}X_{t_1}\,\widehat{X_{t_2}\ldots \widehat{X_{t_{k-1}}\,\widehat{X_{t_k}}}}\,.$$

The sign $\frown$ over random variables means that it is centered by its own mathematical expectation $\widehat{\xi} = \xi - \mathbf{E}\xi$.

In the case of independent random variables $X_{t_1}, \ldots, X_{t_k}$, $\widehat{\mathbf{E}}X_{t_1}\ldots X_{t_k}$ differs from zero only as $t_1 = \cdots = t_k$.

If $X_t = g_t(\xi_t)$, $(t = 1, \ldots, n)$ are random variables related to a Markov chain ξ_t, then one can easily notice that

$$\begin{aligned}\widehat{\mathbf{E}}X_I &= \widehat{\mathbf{E}}X_{t_1}\ldots X_{t_k} \\ &= \underbrace{\int\!\!\int\cdots\int}_{k} g_{t_1}(x_1)\,P_{t_1}(dx_1) \\ &\qquad \times \prod_{j=2}^{k} g_{t_j}(x_j)\left(P_{t_{j-1},t_j}(x_{j-1},\,dx_j) - P_{t_j}(dx_j)\right).\end{aligned}$$

Here, $P_t(B) = \mathbf{P}(\xi_t \in B)$.

We call an unordered collection of non-overlapping, non-empty sets $\{I_1, \ldots, I_\nu\}$ $(1 \leq \nu \leq k)$ such that $\bigcup_{p=1}^{\nu} I_p = I$ as a partition.

Assume $\{I_1', I_2', \ldots, I_\nu'\}$ $(1 \leq \nu \leq k)$ to be a set of all ν block partitions of the set $I' = \{t_1, \ldots, t_k \,|\, t_j \in \mathcal{N}\}$, $t_1 \leq \cdots \leq t_k$, i.e. $I_p' = \left\{t_1^{(p)}, \ldots, t_{k_p}^{(p)}\right\}$, $t_1^{(p)} \leq \cdots \leq t_{k_p}^{(p)}$, $1 \leq p \leq \nu$, $k_1 + \cdots + k_\nu = k$.

Further, let $\mathcal{L} = \{l_1, \ldots, l_r\}$, $l_1 < \cdots < l_r$, be the set of indices of the k-set I'. Suppose that $\{\mathcal{L}_1, \ldots, \mathcal{L}_\nu\}$ is a partition of the set $\mathcal{L}$, corresponding to the partition $\{I_1', \ldots, I_\nu'\}$ of the set I', i.e., $\mathcal{L}_p = \left\{l_1^{(p)}, \ldots, l_{r_p}^{(p)}\right\}$ is the set of indices of the k_p-set I_p', $l_1^{(p)} < \cdots < l_{r_p}^{(p)}$, $1 \leq p \leq \nu$, $r_1 + \cdots + r_\nu \geq r$.

Let $(m_1, \ldots, m_r)$ be the index vector of the set $\mathcal{L}$, $m_1 + \cdots + m_r = k$ and $(m_1^{(p)}, \ldots, m_{r_p}^{(p)})$ be the index vector of the set I_p', generated by the set $\mathcal{L}_p$, $m_1^{(p)} + \cdots + m_{r_p}^{(p)} = k_p$, $1 \le p \le \nu$.

Now, let us consider the estimation of centered moments $\widehat{\mathbf{E}} Y_{I_p'}$, $1 \le p \le \nu \le k$.

Lemma 19.2.2

(i) If, with probability 1, *the estimate*

$$\mathbf{E}(|X_t|^k | \mathcal{F}_{t-1,t+1}) \le (k!)^{1+\gamma_1} H^k \sigma_t^2, \quad t = 1, \ldots, n, \quad k = 2, 3, \ldots$$

holds for some $\gamma_1 \ge 0$ *and* $H > 0$, *then*

$$|\widehat{\mathbf{E}} Y_{I_p'}| \le \frac{1}{2} \prod_{j=1}^{r_p} \left(m_j^{(p)}!\right)^{1+\gamma_1} \left(\frac{40 \cdot 2^{\gamma_1} H}{\alpha^{(n)}}\right)^{k_p - \chi_{I_p}(t_1) - \chi_{I_p}(t_2)}$$

$$\times \prod_{j=1}^{r_p - 1} \left(1 - \beta_{l_j^{(p)}, l_{j+1}^{(p)}}\right)^{\frac{1}{2}} \mathbf{E}^{\frac{1}{2}\chi_{I_p}(t_1)} Y_{t_1}^2 \, \mathbf{E}^{\frac{1}{2}\chi_{I_p}(t_2)} Y_{t_2}^2 \left(\max_{1 \le t \le n} \sigma_t\right)^{r_p}, \tag{19.7}$$

where

$$\chi_B(t) = \begin{cases} 0, & \text{if } t \notin B, \\ 1, & \text{if } t \in B. \end{cases}$$

(ii) If

$$\widetilde{L}_{k,n} = \frac{\sum_{t=1}^{n} \mathbf{E}^*(|X_t|^k | \mathcal{F}_{t-1,t+1})}{{\alpha^{(n)}}^{k-1} B_n^k} \le \frac{(k!)^{1+\gamma_2}}{\Delta_n^{k-2}}, \qquad k = 3, 4, \ldots$$

for some $\gamma_2 \ge 0$ *and* $\Delta_n \ge e$, *then, as* $\gamma_2 = 0$,

$$|\widehat{\mathbf{E}} Y_{I_p'}| \le \frac{1}{2} \left(\frac{9}{5}\right)^{r_p} \prod_{j=1}^{r_p} \overline{m}_j^{(p)}! \left(\frac{160 B_n \log \Delta_n}{\Delta_n}\right)^{k_p - \chi_{I_p}(t_1) - \chi_{I_p}(t_2)}$$

$$\times \prod_{j=1}^{r_p - 1} \left(1 - \beta_{l_j^{(p)}, l_{j+1}^{(p)}}\right)^{\frac{1}{2}} \mathbf{E}^{\frac{1}{2}\chi_{I_p}(t_1)} Y_{t_1}^2 \, \mathbf{E}^{\frac{1}{2}\chi_{I_p}(t_2)} Y_{t_2}^2, \tag{19.8}$$

and as $\gamma_2 > 0$,

$$|\widehat{\mathbf{E}} Y_{I_p'}| \le \frac{1}{2} \prod_{j=1}^{r_p} \overline{m}_j^{(p)}! \left(80\, \widetilde{C}(\gamma_2)\, B_n\, \Delta_n^{-\frac{1}{1+2\gamma_2}}\right)^{k_p - \chi_{I_p}(t_1) - \chi_{I_p}(t_2)}$$

$$\times \prod_{j=1}^{r_p - 1} \left(1 - \beta_{l_j^{(p)}, l_{j+1}^{(p)}}\right)^{\frac{1}{2}} \mathbf{E}^{\frac{1}{2}\chi_{I_p}(t_1)} Y_{t_1}^2 \, \mathbf{E}^{\frac{1}{2}\chi_{I_p}(t_2)} Y_{t_2}^2 \tag{19.9}$$

for $k_p = 2, 3, \ldots, 2\left[\Delta_n^{\frac{2}{1+2\gamma_2}}\right]$. *Here,* $\widetilde{C}(\gamma_2)$ *is as defined in (19.6), and*

$$\overline{m}_j^{(p)} = \begin{cases} m_j^{(p)} - \chi_{I_p}(t_j), & \text{if } j = 1, 2, \\ m_j^{(p)}, & \text{if } j = 3, 4, \ldots, r_p. \end{cases}$$

For some $k \geq 2$, let $\mathbf{E}|X_t|^k < \infty$, $t = 1, \ldots$. Then, $\Gamma(X_{t_1}, \ldots, X_{t_k})$ is a correlation function (or a simple k-th order cumulant) of random variables $X_{t_1}, X_{t_2}, \ldots, X_{t_k}$ defined as follows:

$$\Gamma(X_{t_1}, \ldots, X_{t_k}) = \frac{1}{i^k} \frac{\partial^k}{\partial u_1 \ldots \partial u_k} \log \mathbf{E} \exp\left\{ i \sum_{j=1}^{k} u_j X_{t_j} \right\}\Big|_{u_1 = \cdots = u_k = 0}.$$

Obviously, $\Gamma(X_{t_1}, \ldots, X_{t_k})$ is a symmetric function of its arguments. The k-th order cumulant of the sum S_n is defined by

$$\Gamma_k(S_n) = \sum_{1 \leq t_1, \ldots, t_k \leq n} \Gamma(X_{t_1}, \ldots, X_{t_k}). \tag{19.10}$$

Note that $\Gamma(X_{t_1}, \ldots, X_{t_k})$ can be expressed through the function $\widehat{\mathbf{E}} X_{t_1} \ldots X_{t_k}$ by the following formula [see Statulevičius (1969) and Saulis and Statulevičius (1989)]:

$$\Gamma(X_{t_1}, \ldots, X_{t_k}) = \sum_{\nu=1}^{k} (-1)^{\nu-1} \sum_{\bigcup_{p=1}^{\nu} I_p = I} N_\nu(I_1, \ldots, I_\nu) \prod_{p=1}^{\nu} \widehat{\mathbf{E}} X_{I_p}, \tag{19.11}$$

where $\sum_{\bigcup_{p=1}^{\nu} I_p = I}$ stands for the summation over all the ν -block partitions $\{I_1, \ldots, I_\nu\}$ of the set I; the integers $N_\nu(I_1, \ldots, I_\nu)$,

$$0 \leq N_\nu(I_1, \ldots, I_\nu) \leq (\nu - 1)!,$$

depend only on the set $\{I_1, \ldots, I_\nu\}$, and moreover, if $N_\nu(I_1, \ldots, I_\nu) > 0$, then

$$\sum_{p=1}^{\nu} \max_{t_i^{(p)}, t_j^{(p)} \in I_p} \left(t_j^{(p)} - t_i^{(p)} \right) \geq \max_{1 \leq i, j \leq k} (t_j - t_i).$$

From (19.10), (19.11) and (19.2), we get

$$\Gamma_k(S_n) = \sum_{\nu=1}^{k} (-1)^{\nu-1} \sum_{\bigcup_{p=1}^{\nu} I_p = I} N_\nu(I_1, \ldots, I_\nu) \times \sum_{1 \leq t_1, \ldots, t_k \leq N} \prod_{p=1}^{\nu} \widehat{\mathbf{E}} Y_{t_1^{(p)}} \ldots Y_{t_{k_p}^{(p)}}. \tag{19.12}$$

Next, we need the following formula, valid for any nonnegative symmetric function $f(a_1, \ldots, a_s)$, $a_i \in \mathcal{N}$, $i = 1, \ldots, s$ [see Saulis and Statulevičius (1989)]:

$$\begin{aligned} &\sum_{1 \le a_1, \ldots, a_s \le n} f(a_1, \ldots, a_s) \\ &= \sum_{r=1}^{s} \sum_{m_1 + \cdots + m_r = s} \frac{s!}{m_1! \ldots m_r!} \\ &\times \sum_{1 \le a_1 < \cdots < a_r \le n} f(\underbrace{a_1, \ldots, a_1}_{m_1}, \ldots, \underbrace{a_r, \ldots, a_r}_{m_r}) \end{aligned} \tag{19.13}$$

and its consequence

$$\sum_{1 \le a_1, \ldots, a_s \le n} f(a_1, \ldots, a_s) \le s! \sum_{1 \le a_1 \le \cdots \le a_s \le n} f(a_1, \ldots, a_s). \tag{19.14}$$

It ought to be noted that for $f \equiv 1$, (19.13) takes on the form

$$n^s = \sum_{r=1}^{s} \sum_{m_1 + \cdots + m_r = s} \frac{s!}{m_1! \ldots m_r!} \binom{n}{r}.$$

Making use of (19.14), we obtain from (19.12)

$$|\Gamma_k(S_n)| \le k! \sum_{\nu=1}^{k} \sum_{\bigcup_{p=1}^{\nu} I_p = I} N_\nu(I_1, \ldots, I_\nu) \sum_{1 \le t_1 \le \cdots \le t_k \le N} \prod_{p=1}^{\nu} |\widehat{\mathbf{E}} Y_{I_p}| \tag{19.15}$$

and using (19.13), we have from (19.12)

$$\begin{aligned} |\Gamma_k(S_n)| \le & \sum_{r=1}^{k} \sum_{m_1 + \cdots + m_r = k} \frac{k!}{m_1! \ldots m_r!} \\ & \times \sum_{\nu=1}^{k} \sum_{\bigcup_{p=1}^{\nu} I_p = I} N_\nu(I_1, \ldots, I_\nu) \sum_{1 \le l_1 < \cdots < l_r \le N} \prod_{p=1}^{\nu} |\widehat{\mathbf{E}} Y_{I_p}|. \end{aligned} \tag{19.16}$$

Formulas (19.15) and (19.16) are the basic ones needed to prove the following lemma.

Lemma 19.2.3

(i) If, with probability 1,

$$\mathbf{E}(|X_t|^k|\mathcal{F}_{t-1,t+1}) \le (k!)^{1+\gamma_1} H^k \max_{1\le t\le n} \sigma_t^2, \qquad t=1,\dots,n, \quad k=2,\,3,\,\dots$$

for some $\gamma_1 \ge 0$ *and* $H > 0$, *then*

$$|\Gamma_k(S_n)| \le 171\,(k!)^{1+\gamma_1} \left(\frac{568 \cdot 2^{\gamma_1} H \max_{1\le t\le n} \sigma_t}{\alpha^{(n)}} \right)^{k-2} B_n^2 \tag{19.17}$$

for all $k = 3,\,4,\,\dots$.

(ii) If

$$\widetilde{L}_{k,n} = \frac{\sum_{t=1}^{n} \mathbf{E}^*(|X_t|^k|\mathcal{F}_{t-1,t+1})}{\alpha^{(n)k-1} B_n^k} \le \frac{(k!)^{1+\gamma_2}}{\Delta_n^{k-2}}, \qquad k=3,\,4,\,\dots$$

for some $\gamma_2 \ge 0$ *and* $\Delta_n \ge e$, *then, as* $\gamma_2 = 0$,

$$|\Gamma_k(S_n)| \le 171\,k! \left(\frac{4090 \log \Delta_n}{\Delta_n} \right)^{k-2} B_n^k \tag{19.18}$$

for all $k = 3,\,4,\,\dots$, *and as* $\gamma_2 > 0$,

$$|\Gamma_k(S_n)| \le 171\,k! \left(1136\,\widetilde{C}(\gamma_2)\,\Delta_n^{-\frac{1}{1+2\gamma_2}} \right)^{k-2} B_n^k \tag{19.19}$$

for all $k = 2,\,3,\,\dots,\,2\left[\Delta_n^{\frac{2}{1+2\gamma_2}}\right]$. *Here,* $\widetilde{C}(\gamma_2)$ *is defined earlier in (19.6).*

To prove the basic results, we also need the following lemma.

Assume that there exist constants $\gamma \ge 0$ and $\Delta > 0$ such that the random variable ξ satisfies the condition

$$|\Gamma_k(\xi)| \le \frac{(k!)^{1+\gamma}}{\Delta^{k-2}}, \qquad k=3,\,4,\,\dots \tag{19.20}$$

Denote

$$\Delta_\gamma = c_\gamma\,\Delta^{\frac{1}{1+2\gamma}}, \qquad c_\gamma = \tfrac{1}{6}\left(\tfrac{\sqrt{2}}{6}\right)^{\frac{1}{1+2\gamma}}.$$

Lemma 19.2.4 [Saulis and Statulevičius (1989)] *If condition (19.20) is fulfilled for the random variable* ξ *with* $\mathbf{E}\xi = 0$ *and* $\mathbf{E}\xi^2 = 1$, *then in the interval*

$$0 \le x < \Delta_\gamma$$

the following relations of large deviations

$$\frac{\mathbf{P}(\xi > x)}{1-\Phi(x)} = \exp\{L_\gamma(x)\}\left(1+\theta f(x)\,\frac{x+1}{\Delta_\gamma}\right),$$

$$\frac{\mathbf{P}(\xi < -x)}{\Phi(-x)} = \exp\{L_\gamma(-x)\}\left(1+\theta f(x)\,\frac{x+1}{\Delta_\gamma}\right)$$

hold. Here,

$$f(x) = \frac{60\left(1+10\,\Delta_\gamma^2\exp\left\{-\left(1-\frac{x}{\Delta_\gamma}\right)\sqrt{\Delta_\gamma}\right\}\right)}{1-\frac{x}{\Delta_\gamma}}, \qquad |\theta| \le 1,$$

$$L_\gamma(x) = \sum_{3\le k<p} \lambda_k x^k, \qquad p = \begin{cases} \dfrac{1}{\gamma}+2, & \text{if } \gamma > 0, \\ \infty, & \text{if } \gamma = 0. \end{cases}$$

The coefficients λ_k *are expressed through the cumulants of the random variable* ξ *and are the same as those of the Cramer series. For the coefficients* λ_k *, the estimate*

$$|\lambda_k| \le \frac{2}{k}\left(\frac{16}{\Delta}\right)^{k-2}((k+1)!)^\gamma, \qquad k = 3,\, 4,\, \ldots .$$

holds. As $\gamma > 0$*, we obtain Linnik zones.*

Theorems 19.2.1, 19.2.2 are proved by direct application of the results[3] of Lemmas 19.2.1–19.2.4 and direct calculation of Δ_γ.

Proof of Theorem 19.2.1. From (19.17), as $\gamma_1 \ge 0$, we get

$$|\Gamma_k(Z_n)| \le (k!)^{1+\gamma_1}\left(\frac{C(\gamma_1)H\max\limits_{1\le t\le n}\sigma_t}{\alpha^{(n)}B_n}\right)^{k-2}, \qquad k = 3,\, 4,\, \ldots .$$

By applying Lemma 19.2.4, as $\gamma = \gamma_1$ and

$$\Delta_\gamma = \left(\frac{C(\gamma)\,\alpha^{(n)}B_n}{H\max\limits_{1\le t\le n}\sigma_t}\right)^{\frac{1}{1+2\gamma}},$$

we obtain the proposition of Theorem 19.2.1. ■

[3]N. N. Amosova showed that condition (19.20) was also necessary with an accuracy up to the constant c_γ [see *Lithuanian Mathematical Journal,* (1999)].

PROOF OF THEOREM 19.2.2. From (19.18) and (19.19), as $\Delta_n \geq e$, we have

$$|\Gamma_k(Z_n)| \leq \begin{cases} k!\left(\dfrac{C\log\Delta_n}{\Delta_n}\right)^{k-2}, & k=3,4,\ldots, \quad \text{if } \gamma_2=0,\\[2ex] \dfrac{k!}{\left(C(\gamma_2)\,\Delta_n^{\frac{1}{1+2\gamma_2}}\right)^{k-2}}, & k=3,=4,\ldots,2\left[\Delta_n^{\frac{2}{1+2\gamma_2}}\right], \quad \text{if } \gamma_2>0. \end{cases}$$

By applying Lemma 19.2.4, as $\xi = Z_n$, $\gamma = \gamma_2$ and

$$\Delta_\gamma = \begin{cases} \dfrac{C\Delta_n}{\log\Delta_n}, & \text{if } \gamma=0,\\[2ex] C(\gamma)\,\Delta_n^{\frac{1}{1+2\gamma}}, & \text{if } \gamma>0, \end{cases}$$

we obtain the proposition of Theorem 19.2.2. ■

19.3 Non-Gaussian Approximation

If the distribution of a random variable X_n (as $n \to \infty$) converges to the Poisson distribution, then in the asymptotic analysis (the rate of convergence, asymptotic expansions, behaviour of the probabilities of large deviations) one has to employ factorial moments and factorial cumulants $\widetilde{\Gamma}_k(X_n)$, which are defined as coefficients of the expansion

$$\log \mathbf{E}e^{itX_n} = \sum_{k=1}^{s} \frac{\widetilde{\Gamma}_k(X_n)}{k!} z_1^k(t) + o(|t|^s)$$

if $\mathbf{E}|X_t|^s < \infty$, where $z_1(it) = e^{it} - 1$.

In the normal approximation, we took $z_1(it) = it$ and obtained simple cumulants $\Gamma_k(X_n)$. In the approximation by the Poisson law, general lemmas of large deviations, if we have the estimates for $\widetilde{\Gamma}_k(X_n)$, were proved by Aleškevičienė and Statulevičius (1995).

We can offer the general principle of choice of special cumulants for each approximation as follows: if we wish to approximate the distribution of random variable X_n by the distribution F with the characteristic function $f(t)$, then we expand $\log \mathbf{E}e^{itX_n}$ in a neighbourhood of the point

$$z_1(it) = \frac{d\log f(t)}{d(it)}\Big|_0^t.$$

Thus, for example, if the approximation law is the normal one, then $f(t) = e^{-\frac{t^2}{2}}$ and $z_1(it) = it$; if it is the Poisson law, then $z_2(it) = \lambda(e^{it} - 1)$; and if it is

χ^2_m with m degrees of freedom, then $z_1(it) = m\frac{2it}{1-2it}$ as far as $f(t) = f_{\chi^2_m}(t) = \frac{1}{(1-2it)^{\frac{m}{2}}}$.

Instead of multipliers λ and m, we can take 1, in which case λ and m will be included in the expression of coefficients which we declare as the corresponding cumulants. In the case of χ^2_m approximation, we denote such cumulants by $\overline{\Gamma}_k(X_n)$ [see Aleškevičienė and Statulevičius (1997)].

Let

$$X_n = \sum_{i=1}^{m} \left(\frac{S_n^{(i)}}{\sqrt{n}} \right)^2,$$

where $S_n^{(i)} = \xi_{(i-1)n} + \cdots + \xi_{in}$, $i = 1, \dots, m$, and $\xi_1, \xi_2, \dots$ is a sequence of independent identically distributed random variables with $\mathbf{E}\xi_1 = 0$, $\mathbf{E}\xi_1^2 = 1$, satisfying the Cramer condition

$$|\mathbf{E}\xi_1^2| \le k! H^{k-2} \tag{19.21}$$

for all $k \ge 3$. Obviously, $X_n \xrightarrow{d} \chi^2_m$.

Theorem 19.3.1 *If (19.21) is fulfilled, then in the interval* $1 < x < c_0 n^{\frac{1}{3}}$, *the relation*

$$\mathbf{P}(X_n \ge x) = \mathbf{P}(\chi^2_m \ge x) \left(1 + \theta C_0 \frac{x^3}{n} \right)$$

holds, where c_0 *and* C_0 *depend only on* m *and* H *with* $|\theta| \le 1$.

The main idea of the proof is the following. The summands in (19.2) are independent, and therefore it suffices to estimate only $\overline{\Gamma}(Y^2)$, where $Y = \frac{S_n^{(1)}}{\sqrt{n}}$. These cumulants are expressed by simple cumulants $\Gamma_\nu(Y^2)$, $\nu = 1, \dots, k$, as follows:

$$2^k \overline{\Gamma}_k(Y^2) = k! \sum_{\nu=2}^{k} (-1)^{k-\nu} 2^{k-\nu} \binom{k-1}{\nu-1} \frac{\Gamma_\nu(Y^2)}{\nu!} + (-1)^{k-1} 2^{k-1} (k-1)! k.$$

The cumulants $\Gamma_\nu(Y^2)$ are expressed through the moments $\mathbf{E}Y^{2j}$, $j = 1, \dots, \nu$, while the latter can be inversely expressed through the cumulants $\Gamma_l(Y)$, $l = 1, \dots, 2\nu$, of the random variable Y (which are trivially calculated and estimated)

$$\Gamma_l(Y) = \frac{\Gamma_l(\xi_l)}{(\sqrt{n})^{l-2}} = \theta \frac{k! H^{l-2}}{(\sqrt{n})^{l-2}}, \quad |\theta| \le 1.$$

The majority of summands in (19.4) is reduced after transformations, and we finally find that

$$\left| \overline{\Gamma}_k(Y^2) \right| \le \frac{(k!)^2 H_1^k}{n^{k-1}} + \frac{(k!) H_2^k}{n^{\frac{k}{3}}},$$

where H_1 and H_2 depend only on H. Hence, we obtain the assertion of Theorem 19.3.1. ∎

These ideas are also applied in the investigation of the probabilities of large deviations for Pearson's χ^2 statistics

$$\chi^2 = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i}.$$

Let a hypothetical distribution $\mathbf{P}(A) = \mathbf{P}(\xi \in A)$ be completely determined, $A_1, \ldots, A_r$ be a partitioning of the space of values of the random variable ξ into a finite number of parts (say r) without common points, $p_i = P(A_i)$ and ν_i be the number of incidences of the sample $\xi_1, \ldots, \xi_n$ values in the set A_i, $i = 1, \ldots, r$. If the assumption on the fact that the sample corresponds to the distribution $\mathbf{P}(A)$ is true, then, as it is known by Pearson's theorem, the distribution of statistic χ^2 tends to the χ^2-distribution with $r-1$ degrees of freedom. In this case, it is possible to show that

$$\chi^2 = \sum_{k=1}^{r-1} Y_k^2,$$

where

$$Y_k^2 = \frac{n\pi_k\pi_{k+1}}{p_{k+1}} \left[\left(\frac{N_{k+2}}{n} - \Pi_{k+2} \right) \frac{1}{\pi_{k+1}} + \left(\frac{\mathcal{N}_{k+1}}{n} - \Pi_{k+1} \right) \frac{1}{\pi_k} \right]^2,$$

$$\pi_k = p_1 + \cdots + p_k, \quad \Pi_{k+1} = p_{k+1} + \cdots + p_r,$$

$$\mathcal{N}_{k+1} = \nu_{k+1} + \cdots + \nu_r,$$

$$\mathbf{E}Y_iY_j = 0, \quad i \neq j, \quad \mathbf{E}Y_i^2 = 1, \quad i, j = 1, \ldots, r-1.$$

Random variables $Y_1, \ldots, Y_r$ are dependent though noncorrelated and, in order to estimate $\Gamma_k(\chi^2)$, we have to use mixed cumulants [see Saulis and Statulevičius (1989)].

References

1. Aleškevičienė, A. and Statulevičius, V. (1995). Large deviations in power zones in the approximation by the Poisson law, *Uspekhi Mat. Nauk*, **50**, 63–82.

2. Aleškevičienė, A. and Statulevičius, V. (1997). Probabilities of large deviations in the approximation by χ^2-law, *Lithuanian Mathematical Journal*, **37**.

3. Dobrushin, R. L. (1956). Central limit theorem for nonhomogeneous Markov chains. I, II. *Probability Theory and Its Applications*, **1**, No. 1, 72–89; No. 4, 365–425.

4. Statulevičius, V. (1969). Limit theorems for sums of random variables related to a Markov chain. I, II, III, *Liet. Mat. Rink. (Lithuanian Mathematical Journal)*, **9**, No. 2, 345–362; No. 3, 635–679; , **10**, No. 1, 161–169.

5. Saulis, L. and Statulevičius, V. (1989). *Limit Theorems for Large Deviations*, Dodrecht, The Netherlands: Kluwer Academic Publishers.

6. Tumanyan, S. Kh. (1956). Asymptotic distribution of the χ^2 criterion when the number of observations and number of groups increase simultaneously, *Theory of Probability and Its Applications*, **1**, 99–116.

20

Lower Bounds on Large Deviation Probabilities for Sums of Independent Random Variables

S. V. Nagaev

Sobolev Institute of Mathematics, Novosibirsk, Russia

Abstract: In this chapter, we discuss some lower bounds on the probability $\mathbf{P}(S > x)$, where $S = \sum_{j=1}^{n} X_j$ with X_j's being independent random variables with zero means and finite third moments.

Keywords and phrases: Large deviation, law of iterated logarithm, Mills's function

20.1 Introduction. Statement of Results

Let $X_1, X_2, ...X_n$ be independent random variables with zero expectations and finite third moments, and let $S = \sum X_j$. Here and in the sequel $\sum$ denotes $\sum_j^n$, the case $n = \infty$ is not excluded.

We shall deal with lower bounds on the probability $\mathbf{P}(S > x)$. Denote $\sigma_j^2 = EX_j^2$, $B^2 = \sum \sigma_j^2$, $\beta_j = E|X_j|^3$, and $C = \sum \beta_j$. Let $L = C/B^3$ be a Lyapunov ratio.

In what follows, we omit the index j in definitions and statements relative to every $1 \le j \le n$, i.e. write σ^2 instead of σ_j^2, X instead of X_j, and so forth. We use the notations $a \vee b := \max(a, b)$, $a \wedge b := \min(a, b)$.

Let $\Phi(x)$ be a standard normal law. Denote $\Phi_1(x)$ the Mills function $\sqrt{2\pi}\Big(1 - \Phi(x)\Big)e^{x^2/2}$. Let $\Psi(x) = x\Phi_1(x)$.

The first lower bound on the probability $\mathbf{P}(S > x)$ was obtained by Kolmogoroff (1929) for uniformly bounded random variables $|X_j| < M$ in connection with proving the law of the iterated logarithm. In our notation, the Kolmogoroff inequality looks as follows: *if* $x^2 > 512$ *and* $a = xM/B \le 1/256$, *then*

$$\mathbf{P}(S > xB) > \exp\Big\{-\frac{x^2}{2}(1 + \epsilon)\Big\},$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

where $\epsilon = \max\Big\{64\sqrt{a},\ 32\sqrt{\ln x^2}/x\Big\}$. The result stated below is due to Lenart (1968).

If $|X_k| < M$ *and* $0 < xM/B < 1/12$, *then*

$$\mathbf{P}(S > xB) = \exp\Big\{-\frac{x^2}{2}Q(x)\Big\}\Big(1 - \Phi(x) + \tau(M/B)e^{-x^2/2}\Big), \tag{20.1}$$

where $Q(x) = \sum_1^\infty q_k x^k,\ q_1 < M/3B,\ q_k < \frac{1}{8}(12M/B)^k,\ k = \overline{2,\infty},\ |\tau| < 7.465$ [see also Petrov (1972, p. 308), in this connection].

This result coincides in form with that by Feller (1943) and differs from the latter in somewhat lesser values of constants. It is easy to derive from (20.1) the inequality

$$\mathbf{P}(S > Bx) \geq \Big(1 - \Phi(x)\Big)\exp\Big\{-(\gamma)\frac{Mx^3}{B}\Big\}\Big(1 - 7.465\sqrt{2\pi}\frac{Mx}{B\Psi(x)}\Big), \tag{20.2}$$

where $1 < x < \gamma B/M,\ c(\gamma) = 1/6 + 9\gamma/(1-12\gamma),\ \gamma < 1/12$.

Estimating the right hand side of (20.2) in terms of ϵ which is present in Kolmogoroff inequality, we arrive after rather complicated calculations at the inequality

$$\mathbf{P}(S > Bx) > \frac{\exp\Big\{-(1 + 545 \cdot 10^{-5}\epsilon)x^2/2\Big\}}{\sqrt{2\pi}},$$

where $16\sqrt{2} < x < B/256M$.

Compare now the bound (20.2) with that from the recent paper [Rozovskii (1997)]

$$\mathbf{P}(S > Bx) > \exp\Big\{-(1 + 50\alpha)\frac{x^2}{2}\Big\} \tag{20.3}$$

which holds under the condition

$$\frac{1}{\alpha} < x < \frac{\alpha M}{B},\ 0 < \alpha < 10^{-2}.$$

Under this condition, the inequality (20.2) leads to the bound which is much sharper than (20.3)

$$\mathbf{P}(S > Bx) > \frac{\exp\Big\{-(1 + 0.099\alpha)x^2/2\Big\}}{\sqrt{2\pi}}.$$

Thus the bound (20.2) is sharper than (20.3).

Up to now, we dealt with lower bounds for uniformly bounded summands. Proceed to the case that this restriction does not hold.

Under the Bernstein condition

$$\mathbf{E}|X_j|^k \leq \frac{k!}{2}\sigma_j^2 c^{k-2},\ j = 1 \div n, \quad k = 3, 4...$$

one can derive a lower bound from the paper of Statulevicius (1966) [see in this connection Arkhangel'skii (1989)]. The essential advance was obtained by Arkhangel'skii (1989) who replaced the condition of the uniform boundedness of summands by the condition $\sup_j \mathbf{E}|X_j|^{2+\delta}/\sigma_j^2 < \infty,\ \delta > 0$.

Theorem 20.1.1 *Let the conditions*

$$0 < x < (1-4\gamma)\Big(\frac{\gamma}{L} \wedge \frac{\alpha B}{\max_j \sigma_j}\Big),\ \gamma < \frac{1}{16},\ e\alpha^2 < \frac{\gamma}{2}. \tag{20.4}$$

hold.

Then

$$\begin{aligned} &\mathbf{P}(S > Bx) \\ &\quad > \Big(1-\Phi(x)\Big)\exp\Big\{-\Big(c_1(\gamma)L + c_2(\gamma,\alpha)\sum \sigma_j^3/B^3\Big)x^3\Big\} \\ &\qquad \times\left(1 - \frac{\Big(c_3(\gamma)L + c_4(\alpha)\sum \sigma_j^3/B^3\Big)x}{\Psi(x)}\right), \end{aligned} \tag{20.5}$$

where $c_1(\gamma) \le 1.2/(1-4\gamma)^3$, $c_2(\gamma,\alpha) \le 2\alpha/(1-4\gamma)^3$, $c_3(\gamma) \le 9.79 + 76.26\gamma$, $c_4(\alpha) \le 106\alpha$.

The close result have been stated in Nagaev (1979), but without specifying the values of constants.

Corollary 20.1.1 *If the conditions (20.4) hold, then*

$$\mathbf{P}(S > Bx) > \Big(1-\Phi(x)\Big)\exp\Big\{-c_1(\alpha,\gamma)Lx^3\Big\}\left(1 - \frac{c_2(\alpha,\gamma)Lx}{\Psi(x)}\right), \tag{20.6}$$

where

$$c_1(\alpha,\gamma) \le \frac{(1.2+2\alpha)}{(1-4\gamma)^3},\ c_2(\alpha,\gamma) \le 9.79 + 76.26\gamma + 106\alpha. \tag{20.7}$$

Putting in (20.4) $\gamma = 1/20,\ \alpha = 1/20$ and taking into account that $\Psi(1.7) > 0.805$, we obtain the following result.

Corollary 20.1.2 *If* $1.7 < x < \Big(1/25\Big)(1/L \wedge (B/\max_j \sigma_j)\Big)$, *then*

$$\mathbf{P}(S > Bx) > \Big(1-\Phi(x)\Big)e^{-2.55Lx^3}(1 - 23.4Lx) > 0.06\Big(1-\Phi(x)\Big)e^{-2.55Lx^3}. \tag{20.8}$$

If the random variables X_i's are identically distributed, then

$$\max_j \frac{\sigma_j}{B} = \frac{1}{\sqrt{n}} \leq C/B^3 = \frac{\beta}{\sigma^3\sqrt{n}},$$

where σ^2 and β are, respectively, the second and third absolute moments of X_i.

If $|X_i| < M,\ i = \overline{1,n}$, then

$$\frac{C}{B^3} < \frac{M}{B},\ \max_j \frac{\sigma_j}{B} < \frac{M}{B}.$$

Thus, one can exclude $B/\max_j \sigma_j$ from condition (20.4) in both the above-mentioned cases. The following bounds are then obtained as a result.

Corollary 20.1.3 *If X_i's are identically distributed and*

$$1.7 < x \leq \frac{1}{25}\sqrt{n}\frac{\sigma^3}{\beta}, \tag{20.9}$$

then

$$\mathbf{P}(S > Bx) > \Big(1 - \Phi(x)\Big)\exp\Big\{-\frac{(2.35\beta/\sigma^3 + 0.2)x^3}{\sqrt{n}}\Big\}\left(1 - \frac{(16.88\beta/\sigma^3 + 6.58)x}{\sqrt{n}}\right). \tag{20.10}$$

Corollary 20.1.4 *If $|X_j| < M$ and*

$$1.7 < x \leq \frac{1}{25}\frac{B}{M}, \tag{20.11}$$

then

$$\mathbf{P}(S > xB) > \Big(1 - \Phi(x)\Big)\exp\Big\{-2.35\frac{Mx^3}{B}\Big\}\Big(1 - 23.4\frac{Mx}{B}\Big). \tag{20.12}$$

Not very complicated calculations show that for $1.7 < x < (2/25)M/B$, the inequality (20.2) implies slightly sharper bound than in Corollary 20.1.4

$$\mathbf{P}(S > xB) > \Big(1 - \Phi(x)\Big)\exp\Big\{-1.291\frac{Mx^3}{B}\Big\}\Big(1 - 23.2\frac{Mx}{B}\Big).$$

It is possible to exclude $B/\max \sigma_j$ from condition (20.4) in the general case as well, however at the cost of considerable increase in constants, viz. the following bound holds.

Theorem 20.1.2 *If* $3/2 < x < \gamma/L$, *and* $\gamma < 1/144$, *then*

$$\mathbf{P}(S > Bx) > \Big(1 - \Phi(x)\Big) \exp\Big\{-c_1(\gamma)Lx^3\Big\}(1 - c_2(\gamma)Lx),$$

where

$$c_1(\gamma) \le \frac{23 + 53\sqrt{\gamma}}{(1 - 28\gamma)^3} + \frac{14}{\sqrt{\gamma}}, \; c_2(\gamma) \le 60 + 1630\gamma + 647\sqrt{\gamma} + \frac{0.63}{\sqrt{\gamma}}. \qquad (20.13)$$

Corollary 20.1.5 *If* $3/2 < x < \gamma/L$, *and* $\gamma < 1/144$, *then*

$$\mathbf{P}(S > Bx) > \Big(1 - \Phi(x)\Big) \exp\Big\{-215Lx^3\Big\}(1 - 133Lx). \qquad (20.14)$$

It is better to apply the above stated lower bounds combining them with

$$\mathbf{P}(S > Bx) \ge \frac{1}{2}\sum_j \mathbf{P}(X_j > 2Bx), \qquad (20.15)$$

which holds at least for $x > 2$ [see Nagaev (1979, p. 759)].

If the distributions of X_j's have heavy tails, then the bounds suggested in this Chapter are sharper than the bound (20.15) only for rather small x.

Assume for simplicity that X_j's are identically distributed, and

$$\mathbf{P}(X > x) > \frac{c}{x^t}, \; x > 1, \; t > 3.$$

If $x > b\sqrt{\ln n}$, where $b > \sigma(t-2)^{1/2}$, then for sufficiently large n

$$n\mathbf{P}(X > 2\sigma x\sqrt{n}) > 2(1 - \Phi(x)).$$

Therefore, for $x > b\sqrt{\ln n}$, the bound (20.15) turns out to be shaper than the bounds (20.10), if n is sufficiently large.

Consider now the specific example. Let the density $p(x)$ be defined by

$$p(x) = \begin{cases} 0, & |x| < 1, \\ 2/|x|^5, & |x| \ge 1. \end{cases}$$

Compare the bounds (20.10) and (20.15) for $x = 3$ in the case when X_j's are identically distributed with density $p(x)$. In the case under review, $\sigma^2 = 2$, $\beta = 4$.

For $\gamma = 1/16$, the bound (20.10) is applicable, if $n \ge 2^{13} = 8192$. The inequality (20.10), for $n = 2^{13}$, takes the form

$$\mathbf{P}(S > 3\sqrt{2n}) > 0.57e^{-1.92}(1 - \Phi(3)) > 0.156(1 - \Phi(3)) > 21 \cdot 10^{-5}.$$

On the other hand, using the inequality (20.15), we get, for $n \ge 2^{13}$, the bound

$$\mathbf{P}(S > 3\sqrt{2n}) > \frac{n}{2}\mathbf{P}(X > 6\sqrt{2n}) > \frac{1}{3^4 2^{20}} \sim 11 \cdot 10^{-9}.$$

We observe that, for $x = 3,\ n = 2^{13}$, the bound (20.10) is 10^4 times sharper than (20.15). Note that in the example considered,

$$x < \sigma((t-2)\ln n)^{1/2}.$$

Look what the Berry-Esseen bound

$$\left|\Phi(3) - \mathbf{P}(S > 3\sqrt{2n})\right| < c_0\sqrt{\frac{2}{n}},$$

where constant $c_0 = 0.7655$, does yield in this case.

Putting $n = 2^{13}$, we obtain the trivial bound

$$\mathbf{P}(S > 3\sqrt{2n}) > 1 - \Phi(3)) - 0.01197 > -0.01062.$$

Nontrivial bound takes place only for $n > 643 \cdot 10^3$. Our reasonings show once more what important part constants play in practical calculations. At first glance, the bound 3/64 in the condition (20.9) is not too small, but nevertheless the latter is responsible for the inequality (20.10) being valid only for large n.

It should be remarked that (20.5) and (20.13) do not cover all possible cases. For example, let

$$\mathbf{P}(X = 1) = \mathbf{P}(X = -1) = p,\ \mathbf{P}(X = 0) = 1 - 2p.$$

Evidently, in this case $L = \frac{1}{\sqrt{2pn}}$. Therefore, the inequality (20.5) is applicable only if $x < \frac{3}{64}\sqrt{2pn}$, if $x > \sqrt{2pn} > 1$. On the other hand, the inequality (20.13) implies the trivial bound $\mathbf{P}(S > x) \geq 0$.

Now we state a bound which supplements the above-mentioned inequalities.

Theorem 20.1.3 *Let X_j's be identically distributed, X being symmetric, and $p = \mathbf{P}(X \geq b) < 1/2,\ b > 0$. Then,*

$$\mathbf{P}(S > x) > \frac{1}{2}\sum_{x/b<k\leq n} C_n^k p^k (1-2p)^{n-k}.$$

Let X^s be the symmetrization of X. By using the symmetrization inequality, we get the following bound.

Corollary 20.1.6 *If X_j's are identically distributed, then*

$$\mathbf{P}(|S| > x) > \frac{1}{4}\sum_{2x/b<k\leq n} C_n^k p^k (1-2p)^{n-k},$$

where $p = \mathbf{P}(X^s \geq b) < 1/2$.

In the sequel, we need the following definitions and notations. Define the truncation $X(y)$ of X by the equality

$$X(y) = \begin{cases} X, & X \le y, \\ 0, & X > y. \end{cases}$$

Put $r(h,y) = \mathbf{E}e^{hX(y)}$, $a(h,y) = \mathbf{E}X(y)e^{hX(y)}$, $\sigma^2(h,y) = \mathbf{E}X^2(y)e^{hX(y)}$, and $\beta(h,y) = \mathbf{E}|X(y)|^3 e^{hX(y)}/r(h,y)$.

Denote $a(y) = a(0,y)$, $\sigma^2(y) = \sigma^2(0,y)$, $\beta(y) = \beta(0,y)$. Let $A(y) = \sum a_j(y)$, $B^2(y) = \sum \sigma^2(y)$. Put $m(h,y) = a(h,y)/r(h,y) = \frac{\partial}{\partial h} r(h,y)$, $M(h,y) = \sum m_j(h,y)$, $b^2(h,y) = \sigma^2(h,y)/r(h,y) - m^2(h,y) \equiv \frac{\partial^2}{\partial h^2} \ln r(h,y)$. Note that $M(h,y)$ does not increase as function of h since

$$\frac{\partial}{\partial h} M(h,y) = \sum b_j^2(h,y) > 0.$$

On the other hand, $M(0,y) = A(y) \le 0$. Therefore, the equation with respect to h

$$M(h,y) = x$$

has (for y fixed) the unique solution which we denote by $h(x,y)$.

Put $F(x) = \mathbf{P}(X < x)$, $B^2(h,y) = \sum_j b_j^2(h,y)$. Let

$$C(h,y) = \sum \mathbf{E} \left| \frac{e^{hX_j(y)} X_j^3(y)}{r_j(h,y)} - m_j(h,y) \right|^3.$$

20.2 Auxiliary Results

In this section we state, without proof, several lemmas which we need in Sections 20.3 and 20.4.

Lemma 20.2.1 *If $0 < h \le 1/y$, then*

$$0 < r(h,y) - 1 - a(y)h < \frac{eh^2\sigma^2(y)}{2}. \tag{20.16}$$

Put $\beta_+ = E\{X^3; X > 0\}$, $\beta_- = E\{X^3; X < 0\}$.

Lemma 20.2.2 *If $0 < h \le 1/y$, then*

$$a(h,y) \ge a(y), \tag{20.17}$$

$$\frac{\beta_- h^2}{2} < a(h,y) - a(y) - \sigma^2(y)h < \beta_+ \frac{(e-2)}{y^2} \wedge \frac{eh^2}{2}. \tag{20.18}$$

Lemma 20.2.3 *Let* $0 < h < 1/y$,

$$y > (\alpha^{-1} \max_j \sigma_j) \vee \frac{C}{\gamma B^2}, \ e\alpha^2 < \gamma. \tag{20.19}$$

Then, for every $\gamma < 1$,

$$M(h, y) > \Big((1 - 2\gamma)h - \frac{2\gamma}{y}\Big)B^2. \tag{20.20}$$

Lemma 20.2.4 *Let the conditions, (20.19) hold. Then, for every* $\gamma \le 1/4$,

$$h(x, y) < \frac{x + 2\gamma B^2/y}{(1 - 2\gamma)B^2} \le \frac{1}{y} \tag{20.21}$$

if

$$x \le \frac{(1 - 4\gamma)B^2}{y}. \tag{20.22}$$

Put $Q(h, y) = \sum(r_j(h, y) - 1)$.

Lemma 20.2.5 *If* $h < 1/y$, *then*

$$A(y)h + \frac{B^2(y)h^2}{2} - \frac{Ch^3}{6} < Q(h, y) < \frac{eB^2h^2}{2}. \tag{20.23}$$

Lemma 20.2.6 *Let* $0 < \gamma < 1/16$,

$$y = \frac{(1 - 4\gamma)B^2}{x}, \tag{20.24}$$

$$x < (1 - 4\gamma)B^2\Big(\frac{\gamma B^2}{C} \wedge \frac{\alpha}{\max_j \sigma_j}\Big), \tag{20.25}$$

$$e\alpha^2 < \gamma. \tag{20.26}$$

Then, for $h = h(x, y)$,

$$Q(h, y) - hx > -\frac{x^2}{2B^2} - 1.2C\Big(\frac{x}{(1 - 4\gamma)B^2}\Big)^3. \tag{20.27}$$

Lemma 20.2.7 *Let* $0 < h < 1/y$ *and the conditions in (20.19) hold. Then, for* $\alpha^2 < 1$,

$$h(x, y) > \frac{(1 - \alpha^2)x}{(1 + e\gamma/2)B^2}. \tag{20.28}$$

Lemma 20.2.8 *If the conditions in (20.19) and (20.24) hold, then for* $h < 1/y$

$$\frac{1}{y} \le \frac{1 + e\gamma/2}{(1 - 4\gamma)(1 - \alpha^2)}h(x, y). \tag{20.29}$$

Lemma 20.2.9 *If* $0 < h \leq 1/y$, *then*

$$-\frac{\sigma^2}{y} < a(h,y) < \sigma^2\Big(eh \wedge \frac{(e-1)}{y}\Big). \tag{20.30}$$

Lemma 20.2.10 *Let* $0 < h \leq 1/y$ *and the conditions*

$$y > \frac{\max\limits_j \sigma_j}{\alpha}, \ \alpha^2 < \frac{1}{16e} \tag{20.31}$$

hold. Then,

$$\max_j |m_j(h,y)| < 1.76\alpha\sigma. \tag{20.32}$$

Lemma 20.2.11 *Under conditions of Lemma 20.2.10,*

$$C(h,y) < 2.782C + 19\alpha \sum \sigma_j^3 \tag{20.33}$$

if $\gamma < 1/16$.

Lemma 20.2.12 *If* $h \leq 1/y$, *then*

$$(e-1)\beta y > \sigma^2(h,y) - \sigma^2(y) > -\beta h. \tag{20.34}$$

Lemma 20.2.13 *Let* $0 < h \leq 1/y$ *and the conditions in (20.19) hold with* $\gamma < 1/16$. *Then,*

$$B^2(h,y) > \frac{32}{33}(1 - 2\gamma - 3.2\alpha^2)B^2. \tag{20.35}$$

Lemma 20.2.14 *Let* $0 < h < 1/y$ *and the conditions in (20.19) hold with* $\gamma < 1/16$. *Then,*

$$|hB(h,y) - \frac{M(h,y)}{B(y)}| < \frac{3.32\alpha \sum \sigma_j^3 h}{By} + 1.686\frac{C}{By^2}. \tag{20.36}$$

Lemma 20.2.15 *For every* $0 < u_1 \leq u_2$,

$$1 < \frac{\Phi_1(u_1)}{\Phi_1(u_2)} < \exp\Big\{\frac{u_2 - u_1}{u_1^2\Phi_1(u_1)}\Big\}. \tag{20.37}$$

Lemma 20.2.16 *For every* $0 < u_1 < u_2$,

$$\frac{1 - \Phi(u_2)}{1 - \Phi(u_1)} > \exp\left\{\frac{(u_1 - u_2)u_2}{\Psi(u_1)}\right\}. \tag{20.38}$$

Define

$$\lambda(x) = (1 - ax)e^{-cx^3}(1 - \Phi(x)),$$

where a and c are some constants.

Lemma 20.2.17 *If $a > 0$, then the function $\lambda(x)$ is convex for $0 < x \le 1/a$.*

Lemma 20.2.18 *For every $0 < x < 1$,*

$$e^{-x-x^2/(1-x)} < 1 - x. \tag{20.39}$$

Lemma 20.2.19 *If the condition $\max_j \sigma_j < B_\epsilon$ holds, then for every $h > 0$*

$$\begin{aligned}&\left|h\int_0^\infty e^{-hx}\Big(\mathbf{P}(S<x)-\Phi(x/B)\Big)dx\right| \\ &\quad < \frac{L(\epsilon)}{6\sqrt{2\pi}\Big(1-0.27L^{2/3}(\epsilon)\Big)^{3/2}} + \frac{2.473C^2h}{\pi B^5} \\ &\quad + \frac{L^{2/3}}{\pi}\Big(0.607e^{-0.82L^{-2/3}} + 0.304e^{-1.64L^{-2/3}}\Big) + L^2e^{-.0225L^{-2}},\end{aligned} \tag{20.40}$$

where $L(\epsilon) = C/B_\epsilon^3$, $B_\epsilon^2 = (1-\epsilon^2)B^2$.

20.3 Proof of Theorem 20.1.1

Since $X_j(y) \le X_j$,

$$\mathbf{P}(S > x) \ge \mathbf{P}(S(y) > x), \tag{20.41}$$

where $S(y) = \sum X_j(y)$. Let $G(x;y) = \mathbf{P}(S(y) < x)$. It is not hard to show that

$$\mathbf{P}(S(y) > x) = R(h;y)\int_x^\infty e^{-hu}G_h(du),$$

where $R(h,y) = \mathbf{E}e^{hS(y)} = \Pi r_j(h,y)$, $G_h(du) = e^{hu}G(du;y)/R(h;y)$. Putting $\overline{G}_h(u) = G_h(u+x)$, we have

$$\mathbf{P}(S(y) > x) = R(h;y)e^{-hx}\int_0^\infty e^{-hu}d\overline{G}_h(u). \tag{20.42}$$

The distribution function G_h is the convolution of the distribution functions $F_j(u;h,y)$, $j = \overline{1,n}$, where $F(du;h,y) = e^{hu}F(du;y)/r(h,y)$. It is easily seen that

$$\int_{-\infty}^y uF(du;y) = m(h,y), \int_{-\infty}^y \Big(u - m(h,y)\Big)^2 F(du;y) = \sigma^2(h,y).$$

Hence, the expectation $M(h,y) = \sum m_j(h,y)$ and the variance $B^2(h,y) = \sum b_j^2(h,y)$ correspond to the distribution function G_h.

In what follows, we assume that y satisfies (20.24). Put now $h = h(x, y)$. Then according to the definition, $x = M(h, y)$ (see Section 20.1).

Without loss of generality, one may assume that

$$x \geq 1.7. \tag{20.43}$$

In fact, according to Berry-Esseen bound

$$\mathbf{P}(S > x) > (1 - \Phi(x))\left(1 - \frac{c_0 L}{1 - \Phi(x)}\right), \tag{20.44}$$

where $c_0 < 0.7915$ [see Shiganov (1982)]. On the other hand, $c_3(\gamma)$ in (20.5) can not be less than 9.79. Therefore, the bound (20.44) is sharper than (20.5) provided $(1/\sqrt{2\pi})e^{-x^2/2} > c_0/9.79$, i.e. $x < 1.7$.

Consider the identity

$$\begin{aligned}\int_0^\infty e^{-hu} d\overline{G}_h(u) &= \int_0^\infty e^{-hu} d\Phi(u/B(h,y)) + h\int_0^\infty e^{-hu} r_h(u)du - r_h(0) \\ &= I_1 + hI_2 - r_h(0), \end{aligned} \tag{20.45}$$

where $r_h(u) = \overline{G}_h(u) - \Phi(u/B(h, y))$. It is not hard to show that

$$I_1 = \frac{1}{\sqrt{2\pi}} e^{h^2 B^2(h,y)/2} \int_{hB(h,y)}^\infty e^{-h^2/2} du = \frac{\Phi_1(hB(h,y))}{\sqrt{2\pi}}. \tag{20.46}$$

Let $\Delta(h, y) = hB(h, y) - M(h, y)/B(y)$. Suppose that $\Delta(h, y) > 0$. In view of Lemma 20.2.14 and formula (20.24)

$$\frac{\Delta(h,y)B(y)}{x} < \frac{x\left(1.686C + 3.32\alpha\sum\sigma_j^3\right)}{(1-4\gamma)^2 B^4} < \frac{x\left((1.686 + 20\gamma)C + 5.91\alpha\sum\sigma_j^3\right)}{B^4}.$$

Letting $u_1 = x/B(y)$, and $u_2 = hB(h, y)$ in Lemma 20.2.15 and applying the previous bound, we have

$$\begin{aligned}\sqrt{2\pi} I_1 &= \Phi_1(hB(h,y)) \\ &> \Phi_1(x/B(y)) \exp\left\{-\frac{\left((1.686 + 20\gamma)C + 5.91\alpha\sum\sigma_j^3\right)x}{B^4 \Psi(x/B)}\right\}. \end{aligned} \tag{20.47}$$

We took into account here that $\Psi(x/B) < \Psi(x/B(y))$. Let now $u_1 = x/B$, and $u_2 = x/B(y)$. Substituting these values into inequality (20.37), we obtain

$$\Phi_1(x/B(y)) > \Phi_1(x/B) \exp\left\{-\frac{1 - B/B(y)}{\Psi(x/B)}\right\}.$$

By using identity (20.24) and the inequality

$$M(h,y) > \Big((1-2\gamma)h - \frac{2\gamma}{y}\Big)B^2,$$

we get

$$\frac{B}{B(y)} - 1 = \frac{B^2 - B^2(y)}{B(y)(B+B(y))} < \frac{C}{2yB^2(y)} < \frac{Cx}{2(1-\gamma)(1-4\gamma)B^4}.$$

Thus,

$$\Phi_1(x/B(y)) > \Phi_1(x/B)\bigg(1 - \frac{(1+6.76\gamma)Cx}{2\Psi(x/B)B^4}\bigg). \tag{20.48}$$

It follows then from (20.47) and (20.48) that

$$\sqrt{2\pi}I_1 > \Phi_1(x/B)\exp\bigg\{-\frac{\Big((2.186+23.38\gamma)C + 5.91\alpha\sum\sigma_j^3\Big)x}{B^4\Psi(x/B)}\bigg\}. \tag{20.49}$$

If $\Delta(h,y) < 0$, then $\Phi_1(hB(h,y) > \Phi_1(x/B(y))$. Hence, by using (20.48), we obtain that for $\Delta(h,y) < 0$

$$\sqrt{2\pi}I_1 > \Phi_1(x/B)\bigg(1 - \frac{(1+6.76\gamma)Cx}{2\Psi(x/B)B^4}\bigg). \tag{20.50}$$

We now proceed to estimate I_2. For this purpose, we apply Lemma 20.2.19 to the sum of independent random variable $\overline{G}_h$. It is not hard to see that Lyapunov ratio $L(h,y)$ corresponding to $\overline{G}_h$ is equal to $C(h,y)/B^3(h,y)$. It follows from Lemma 20.2.11 that for $\alpha^2 < 1/16e$,

$$C(h,y) < 5.67C. \tag{20.51}$$

Put $B_1^2(h,y) = \inf_j\Big(B^2(h,y) - b_j^2(h,y)\Big)$. By Lemma 20.2.13

$$B_1^2(h,y) > 0.777\Big(B^2 - \sup_j \sigma_j^2\Big).$$

Note that in view of (20.4), $x/B < 3\alpha(B/4\max\sigma_j)$. Hence, by (20.43)

$$\bigg(\frac{\max\sigma_j}{B}\bigg)^2 < \bigg(\frac{3}{4x}\bigg)^2\alpha^2 < 0.008. \tag{20.52}$$

Thus,

$$B_1^2(h,y) > 0.769B^2. \tag{20.53}$$

Combining (20.51) and (20.53), we have

$$\frac{C^{2/3}(h,y)}{B_1^2(h,y)} < 4.15L^{2/3}. \tag{20.54}$$

By (20.4) and (20.43)

$$L < \frac{3}{64x} < 0.037. \tag{20.55}$$

Combining (20.54) and (20.55), we obtain as a result of rather simple calculations

$$R_1 := \frac{L_1(h,y)}{6\sqrt{2\pi}\Big(1 - 0.27L_1^{2/3}(h,y)\Big)^{3/2}} < 0.081L_1(h,y), \tag{20.56}$$

where $L_1(h,y) = C(h,y)/B_1^3(h,y)$.

By (20.35) and (20.52),

$$B_1^2(h,y) > 0.96B^2(1 - 2\gamma - 3.2\alpha^2). \tag{20.57}$$

It is not hard to show that

$$\frac{1}{(1 - 2\gamma - 3.2\alpha^2)^{3/2}} < \frac{1}{(1 - 3.178\gamma)^{3/2}} < 1 + 6.3\gamma < 1.394. \tag{20.58}$$

It follows from (20.33), (20.57) and (20.58) that

$$L_1(h,y) < \frac{(2.96 + 18.7\gamma)C + 28.2\alpha\sum\sigma_j^3}{B^3}. \tag{20.59}$$

Comparing (20.56) and (20.59), we obtain

$$R_1 < \frac{(0.24 + 1.515\gamma)C + 2.285\alpha\sum\sigma_j^3}{B^3}. \tag{20.60}$$

By (20.4), (20.21) and (20.24), $h(x,y) < x/(1-4\gamma)B^2 < \gamma B^2/C < B^2/16C$. Hence, taking into account (20.33) and (20.35), we conclude that

$$\frac{hC^2(h,y)}{B^5(h,y)} < 1.08\frac{2.782^2C + 38\cdot 2.782\alpha + 361\alpha^2\sum\sigma_j^3}{16(1 - 3.178\gamma)^{5/2}B^3}.$$

Therefore,

$$\frac{2.473hC^2(h,y)}{\pi B^5(h,y)} < \frac{(0.412 + 4.89\gamma)C + 14.91\alpha\sum\sigma_j^3}{B^3}. \tag{20.61}$$

Substitute now $L(h,y)$ and $L_1(h,y)$ into the right-hand side of (20.40) in place of L and $L(\epsilon)$, respectively. The first two summands are estimated via (20.60) and (20.61). As to the other three, calculations show that one may ignore them. As a result, we are led to the bound

$$h|I_2| < \frac{(0.652 + 6.1\gamma)C + 17.195\alpha\sum\sigma_j^3}{B^3}. \tag{20.62}$$

Further, according to the Berry–Esseen bound

$$\sup_u |r_h(u)| < \frac{c_0 C(h,y)}{B^3(h,y)}. \tag{20.63}$$

Applying now Lemmas 20.2.11 and 20.2.13, we get

$$|r_h(0)| < \frac{1.08c_0(2.782C + 19\alpha \sum \sigma_j^3)}{(1-2\gamma-3.5\alpha^2)^{3/2}B^3} < \frac{(2.38+15\gamma)C + 22.65\alpha \sum \sigma_j^3}{B^3}. \tag{20.64}$$

Incidentally, one can use in inequality (20.63) hI_2 as well since $h|I_2| < \sup_u |r(u)|$.

Up to now, everyone proceeded in this manner, beginning with fundamental work of Kramer (1938). However, one loses an accuracy in doing so. It is sufficient to compare the bounds (20.62) and (20.64) to be sure.

One can prove with the aid of Lemma 20.2.1 that for $e\alpha^2 < \gamma$, and $\gamma < 1/16$

$$|r(h,y) - 1| < \frac{1}{32}.$$

Hence, by using Lemma 20.2.18 and the inequality $1 + x > e^{x - x^2/2}$, where $0 \le x \le 1/2$, we get

$$R(h,y) > \exp\left\{Q(h,y) - \frac{32}{31}\sum\Big(r_j(h,y) - 1\Big)^2\right\}, \tag{20.65}$$

(see the definition of $Q(h,y)$ before Lemma 20.2.5). By (20.16) and (20.29)

$$|r(h,y) - 1| < \frac{eh^2\sigma^2}{2} \vee \frac{\sigma^2 h}{y} \le \frac{e\sigma^2}{2y^2}.$$

Hence, by using (20.16) and the conditions (20.24) and (20.25), we have

$$\begin{aligned} \sum\Big(r_j(h,y) - 1\Big)^2 &< \Big(\frac{e}{2}\Big)^2 y^{-4} \max_l \sigma_l \sum \sigma_j^3 < \frac{x^4 \max_l \sigma_l \sum \sigma_j^3}{(1-4\gamma)^4 B^8}\Big(\frac{e}{2}\Big)^2 \\ &< \frac{1.85\alpha x^3 \sum \sigma_j^3}{(1-4\gamma)^3 B^6}. \end{aligned} \tag{20.66}$$

It follows from (20.27), (20.65) and (20.66) that

$$e^{-hx} R(h,y) > \exp\Big\{-\frac{x^2}{2B^2} - \frac{(1.2C + 1.91\alpha \sum \sigma_j^3)x^3}{(1-4\gamma)^3 B^6}\Big\}. \tag{20.67}$$

Combining (20.41), (20.42), (20.45), (20.49), (20.50), (20.62), (20.64) and (20.67), we obtain

$$\mathbf{P}(S > x) \;>\; \left[\Big((1 - \Phi(x/B)\Big)\exp\left\{-\frac{((2.186 + 23.38\gamma)C + 5.91\alpha \sum \sigma_j^3)x}{B^4 \Psi(x/B)}\right\}\right.$$

$$- \frac{e^{-x^2/2B^2}}{B^3}\left((3.032+21.1\gamma)C+39.845\alpha\sum\sigma_j^3\right)\Bigg]$$
$$\times\ \exp\left\{-\frac{(1.2C+1.98\alpha\sum\sigma_j^3)x^3}{(1-4\gamma)^3B^6}\right\}.$$

Hence, taking into account that

$$e^{-x^2/2B^2}=\sqrt{2\pi}\frac{(1-\Phi(x/B))x}{\Psi(x/B)B},$$

we are led to the desired result. ■

20.4 Proof Theorem 20.1.2

Let the numbers γ and $\alpha>0$ be such that $\gamma=\alpha^2$. Assume that the inequality

$$x<\frac{\gamma B^4}{C} \tag{20.68}$$

holds.

Define

$$N(x)=\left\{j:\sigma_j<\frac{\alpha^2B^2}{2x}\right\}.$$

Put $S_1=\sum' X_j$, and $S_2=\sum'' X_j$. Here and in what follows, $\sum'=\sum\limits_{j\in N(x)}$ and $\sum''=\sum\limits_{j\notin N(x)}$. Let $B_1^2=\sum'\sigma_j^2$, $B_2^2=\sum''\sigma_j^2$, and $C_1=\sum'\beta_j$. Then

$$B_2^2<\frac{2xC}{\alpha B^2}<2\frac{\gamma B^2}{\alpha}=2\sqrt{\gamma}B^2. \tag{20.69}$$

If $\gamma<1/36$, then $B_2^2<B^2/3$, that is,

$$B_1^2>\frac{2B^2}{3}. \tag{20.70}$$

Put $\gamma_1=3.5\gamma$. Clearly,

$$\frac{(1-4\gamma_1)\alpha B_1^2}{\max\limits_{j\in N(x)}\sigma_j}\geq\frac{4(1-14\gamma)}{3}>x$$

if $\gamma\leq 1/56$. Further, for $\gamma_1<5/56$,

$$\frac{\gamma B^4}{C}<\left(\frac{3}{2}\right)^2\frac{\gamma_1B_1^4}{3.5C_1}<\frac{(1-4\gamma_1)\gamma_1B_1^4}{C_1}.$$

Thus, for $\gamma < 1/56 (\gamma_1 < 1/16)$, the condition

$$x < (1 - 4\gamma_1)B_1^2\Big(\gamma_1 B_1^2/C_1 \wedge \alpha/ \max_{j\in N(x)} \sigma_j\Big)$$

holds, i.e. condition (20.4) with $\gamma = \gamma_1$ is fulfilled for S_1 and x which satisfies condition (20.68). Applying now Corollary 20.1.1 to S_1, we conclude that for $B_1 < x < \omega = (1 - 4\gamma_1)B_1^2\Big(\gamma_1 B_1^2/C_1 \wedge \alpha/ \max_{j\in N(x)} \sigma_j\Big)$

$$\mathbf{P}(S_1 > x) > (1 - \Phi(x/B_1)) \exp\Big\{-\frac{c_1(\alpha,\gamma_1)C_1x^3}{B_1^3}\Big\}\left(1 - \frac{c_2(\alpha,\gamma_1)C_1x}{B_1^4\Psi(1)}\right) \equiv f(x),$$

where $c_j(\cdot,\cdot)$ are defined by (20.7). Hence, denoting $\mathbf{P}(S_i \le x) = G_i(x)$, $i = 1, 2$, we get

$$\mathbf{P}(S > x) = \int_{-\infty}^{\infty} \Big(G_1(x-u)\Big) dG_2(u) > \int_{u_1}^{u_2} f(x-u)dG_2(u) + f(x)(1 - G_2(u_2)), \tag{20.71}$$

where $u_1 = x - \omega$, and $u_2 = x - B_1$. According to Lemma 20.2.17, the function $f(x-u)$ is convex in u for $u_1 \le u < x$. Therefore, by the Young inequality

$$\int_{u_1}^{u_2} f(x-u)dG_2(u) > pf(x-q), \tag{20.72}$$

where $p = G_2(u_2) - G_2(u_1)$, and $q = \int_{u_1}^{u_2} u dG_2(u)/p$. If $3B/2 < x < \omega/2$, then $|u_1| > \omega/2 > 3B/2$ and by (20.69)

$$G_2(u_1) < \frac{B_2^2}{u_1^2} < \frac{8}{9}\frac{xC}{\alpha B^4}. \tag{20.73}$$

Estimate now the quantity q. If $x > B$, then by (20.69)

$$-\int_{u_1}^{u_2} u dG_2(u) < \frac{B_2^2}{u_2} < \frac{3B_2^2}{x} < \frac{6C}{\sqrt{\gamma}B^2}. \tag{20.74}$$

Further, $p > 1 - B_2^2/u_1^2 \wedge u_2^2$. In view of (20.69), for $3B/2 < x < \omega/2$

$$\frac{B_2^2}{u_1^2} < \frac{4}{9}\sqrt{\gamma}, \quad \frac{B_2^2}{u_2^2} < \frac{4B_2^2}{B^2} < 4\sqrt{\gamma}.$$

Thus,

$$p > 1 - 4\sqrt{\gamma} > 0.465 \tag{20.75}$$

provided $3B/2 < x < \omega/2$, and $\gamma < 1/56$. It follows from (20.74) and (20.75) that

$$-q < \frac{12.91C}{\sqrt{\gamma}B^2} = \eta. \tag{20.76}$$

By using Lemma 20.2.16 and the bound in (20.76), we conclude that for $3B/2 < x < \omega/2$, and $\gamma < 1/56$

$$f(x-q) > \Big(1-\Phi(x/B_1)\Big)\exp\Big\{-\frac{1}{\Psi(3/2)}\Big(\frac{x\eta}{B_1^2}+\frac{\eta^2}{B_1^2}\Big)-\frac{c_1(\alpha,\gamma_1)(x+\eta)^3C}{B_1^6}\Big\}$$
$$\times\Big(1-\frac{c_2(\alpha,\gamma_1)(x+\eta)C}{B_1^4\Psi(3/2)}\Big). \tag{20.77}$$

In view of condition (20.68), for $\gamma < 1/56$

$$\frac{Cx}{\sqrt{\gamma}B^4} < \sqrt{\gamma} < \frac{1}{2\sqrt{14}}.$$

Hence $L/\sqrt{\gamma} < 1/3\sqrt{14}$, since $x > 3B/2$. Therefore,

$$\frac{\eta}{B} < 1.16,\ \frac{\eta^2}{B^2} < 1.16\frac{\eta}{B} < \frac{14.98C}{\sqrt{\gamma}B^3}. \tag{20.78}$$

By (20.78) and (20.70),

$$\frac{x\eta+\eta^2}{B_1^2} < 63.76\frac{Cx}{\sqrt{\gamma}B^4} < 27.89\frac{Cx^3}{\sqrt{\gamma}B^6} \tag{20.79}$$

and

$$x+\eta < x+1.16B < 1.774x. \tag{20.80}$$

In view of (20.70), (20.79) and (20.80), it follows from (20.77) that

$$f(x-q) > \Big(1-\Phi(x/B_1)\Big)\exp\Big\{-\Big(\frac{36.07}{\sqrt{\gamma}}+18.85c_1(\alpha,\gamma_1)\Big)\frac{Cx^3}{B^6}\Big\}$$
$$\times\Big(1-\frac{4c_2(\alpha,\gamma_1)Cx}{\Psi(1)B^4}\Big). \tag{20.81}$$

By using (20.69), (20.70) and Lemma 20.2.16, we conclude that

$$1-\Phi(x/B_1) \geq \Big(1-\Phi(x/B)\Big)\exp\Big\{-\frac{B_2^2x^2}{B_1^2B(B_1+B_2)\Psi(3/2)}\Big\}$$
$$> \Big(1-\Phi(x/B)\Big)\exp\Big\{-\frac{3\sqrt{3}Cx^3}{\sqrt{2\gamma}B^6\Psi(3/2)}\Big\}, \tag{20.82}$$

provided that $3B/2 < x < \gamma B^4/2C$, and $\gamma < 1/36$. Combining the bounds (20.71), (20.73), (20.81) and (20.82), we obtain that for $3B/2 < x < \gamma B^4/2C$, and $\gamma < 1/56$

$$\mathbf{P}(S > x) > \Big(1-\Phi(x/B)\Big)\exp\Big\{-\frac{c_1'(\alpha,\gamma)Cx^3}{B^4}\Big\}\Big(1-\frac{c_2'(\alpha,\gamma)Cx}{B^4}\Big),$$

where

$$c_1'(\alpha,\gamma) = 18.85c_1(\alpha, 3.5\gamma) + \frac{3.68}{\sqrt{\gamma}\Psi(3/2)} + \frac{36.07}{\sqrt{\gamma}},$$

$$c_2'(\alpha,\gamma) = \frac{4c_2(\alpha, 3.5\gamma)}{\Psi(1)} + \frac{0.89}{\sqrt{\gamma}}, \ \alpha = \sqrt{\gamma}.$$

Now, substituting for $c_1(\cdot,\cdot)$ and $c_2(\cdot,\cdot)$ their expressions in (20.7), we get the desired inequality. ■

Acknowledgements. This research was supported by RFFI grant 96-01-01529.

References

1. Arkhangel'skii, A. N. (1989). Lower bounds for probabilities of large deviations of sums of independent random variables, *Teor. Veroyat. i primen.*, **34**, 625–655.

2. Cramer, H. (1938). Sur un nouveau theoreme-limite de la theorie des probabilites, *Actual. sci. et ind.*, **736**.

3. Feller, W. (1943). Generalization of a probability limit theorem of Cramer, *Transactions of the American Mathematical Society*, **54**, 361–372.

4. Kolmogoroff, A. N. (1929). Uber das Gesetz des iterierten Logarithmus, *Math. Ann.*, **101**, 126–135.

5. Lenart, C. (1968). On certain theorems of Berry and a limit theorem of Feller, *Mat. casopis*, **18**, 59–75.

6. Nagaev, S. V. (1979). Large deviations of sums of independent random variables, *Annals of Probability*, **7**, 745–789.

7. Nagaev, S. V. (1996). Some refinements of probability inequalities, *Vestnik Moskov. Univ. Series 1, Mat. Mekh.*, **6**, 64–66.

8. Petrov, V. V. (1972). Sums of independent random variables, *M. Nauka*

9. Rozovskii, L. V. (1997). A generalization of the Kolmogorov theorem on the law of the iterated logarithm, *Teor. Veroyat. i primen.*, **42**, 134–143.

10. Shiganov, I. S. (1982). Refinement of the upper bound of a constant in the reminder of central limit theorem, In *Stability Problems for Stochastic Models*, pp. 105-115, Vsesoyuz. Nauch.- Issled. Inst. Sistem. Issled., Moskow, 1982. (In Russian)

11. Statulevicius, V. A. (1966). On large deviations, *Z. Wahrscheinlichkeitstheor. verw. Geb.*, **6**, 133–144.

PART VI

Empirical Processes, Order Statistics, and Records

21

Characterization of Geometric Distribution Through Weak Records

Fazil A. Aliev

Ankara University, Ankara, Turkey

Abstract: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables (r.v.'s) taking on values $0, 1, \ldots$ with a distribution function F such that $F(n) < 1$ for any $n = 0, 1, \ldots$ and $EX_1 < \infty$. Let $X_{L(n)}$ be the n-th weak record value. In this chapter we show that X_1 has a geometric distribution iff $E(X_{L(n+2)} - X_{L(n)} \mid X_{L(n)} = i) = \alpha$ for some $n > 0, \alpha > 0$ and for all $i \geq 0$.

Keywords and phrases: Records, weak records, characterization of geometric distribution

21.1 Introduction

A lot of papers in the field of records are devoted to characterizations of distributions via records; for example, Ahsanullah (1995), Ahsanullah and Holland (1984), Aliev (1998), Arnold, Balakrishnan and Nagaraja (1998), Kirmani and Beg (1984), Korwar (1984), Nagaraja (1998), Nevzorov (1987), Nevzorov and Balakrishnan (1998), Stepanov (1994), and Vervaat (1973). Great interest in these records exists because they are widely available and they often provide a degree of mathematical accuracy.

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed random variables (r.v.'s) taking on values $0, 1, \ldots$ with a distribution function F such that $F(n) < 1$ for any $n = 0, 1, \ldots$ and $EX_1 < \infty$. Define the sequence of *weak record times* $L(n)$ and *weak record values* $X_{L(n)}$ as follows:

$$L(1) = 1, \quad L(n+1) = \min\left\{j > L(n) : X_j \geq X_{L(n)}\right\} \quad n = 1, 2, \ldots . \quad (21.1)$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

If we replace the sign $\geq$ by $>$ in (21.1), then we obtain record times and record values instead of weak record times and weak record values. Let $p_k = P\{X_1 = k\}$ and $\overline{F}(k) = 1 - F(k)$ $(k \geq 0)$.

It is known that

(1) X_1 has a distribution of the form

$$P\{X_1 \geq m\} = \left(\prod_{i=1}^{m}(\alpha + (i-1)\beta)\right)\left(\prod_{i=1}^{m}(1 + \alpha + i\beta)\right)^{-1}$$

for some $\alpha > 0$, $\beta \geq 0$, and $m = 1, 2, ...$, iff $E(X_{L(n+1)} - X_{L(n)} \mid X_{L(n)} = s) = \alpha + \beta s$, for all $s = 0, 1, ...$ $(n > 0)$; see Stepanov (1994). If $\beta = 0$, this result corresponds to the geometric distribution,

(2) If $\{A_i\}_{i=0}^{\infty}$ is any sequence of positive numbers such that $\frac{A_{i-1}}{(1+A_i)} < 1$ for all i and $\prod_{i=1}^{\infty} \frac{A_i}{(1+A_i)} = 0$, then X_1 has distribution of the form

$$P\{X_1 \geq m\} = \prod_{i=1}^{m} \frac{A_{i-1}}{(1 + A_i)}$$

for all $m = 1, 2, ...$ iff $E\left\{X_{L(n+1)} - X_{L(n)} \mid X_{L(n)} = s\right\} = A_s$ for all $s = 0, 1, ...$ $(n \geq 1)$; see Aliev (1998). In the case of $A_s = \alpha + \beta s$, this result implies the above stated result of Stepanov (1994).

In this paper, we first give a characterization of geometric distribution in terms of $E\{X_{L(n+2)} - X_{L(n)} \mid X_{L(n)} = s\}$ instead of $E\{X_{L(n+1)} - X_{L(n)} \mid X_{L(n)} = s\}$.

21.2 Characterization Theorem

Theorem 21.2.1 *A necessary and sufficient condition for a random variable X_1 to have a geometric distribution is that*

$$E\left\{X_{L(n+2)} - X_{L(n)} \mid X_{L(n)} = s\right\} = \alpha \tag{21.2}$$

for some $n \geq 1$, $\alpha > 0$ and all $s = 0, 1, \ldots$.

Proof. Let us consider the probability $P\left\{X_{L(n+2)} - X_{L(n)} = k, X_{L(n)} = s\right\}$ $(k, s \geq 0)$. We have

$$P\left\{X_{L(n+2)} - X_{L(n)} = k, X_{L(n)} = s\right\}$$

$$
\begin{aligned}
&= P\left\{X_{L(n+2)} = k+s, X_{L(n)} = s\right\} \\
&= \sum_{d=n}^{\infty} P\left\{X_{L(n+2)} = k+s, X_{L(n)} = s, L(n) = d\right\} \\
&= \sum_{d=n}^{\infty} \sum_{m=d+2}^{\infty} P\left\{X_{L(n+2)} = k+s, X_{L(n)} = s, L(n) = d, L(n+2) = m\right\} \\
&= \sum_{d=n}^{\infty} \sum_{m=d+2}^{\infty} \sum_{l=d+1}^{m-1} P\{X_m = k+s, X_d = s, L(n) = d, \\
&\qquad\qquad L(n+1) = l, L(n+2) = m\} \\
&= \sum_{d=n}^{\infty} \sum_{m=d+2}^{\infty} \sum_{l=d+1}^{m-1} \sum_{t=s}^{k+s} P\{X_m = k+s, X_d = s, X_{L(n+1)} = t, L(n) = d, \\
&\qquad\qquad L(n+1) = l, L(n+2) = m\} \\
&= \sum_{d=n}^{\infty} \sum_{m=d+2}^{\infty} \sum_{l=d+1}^{m-1} \sum_{t=s}^{k+s} P\{X_m = k+s, X_d = s, X_l = t, L(n) = d, \\
&\qquad\qquad L(n+1) = l, L(n+2) = m\}.
\end{aligned}
\tag{21.3}
$$

The probability under summation may be rewritten as

$$
\begin{aligned}
&P(X_m = k+s, X_d = s, X_l = t, L(n) = d, L(n+1) = l, L(n+2) = m) \\
&\quad = P(X_d = s, L(n) = d, X_{d+1} < s, ..., X_{l-1} < s, \\
&\qquad\qquad X_l = t, X_{l+1} < t, ..., X_{m-1} < t, X_m = k+s).
\end{aligned}
\tag{21.4}
$$

Note that the event $\{X_d = s, L(n) = d\}$ is defined only by the random variables $X_1, X_2, ..., X_d$ and, therefore, is independent of

$$\{X_{d+1} < s, ..., X_{l-1} < s, X_l = t, X_{l+1} < t, ..., X_{m-1} < t, X_m = k+s\},$$

and consequently we have from (21.4)

$$
\begin{aligned}
&P\{X_m = k+s, X_d = s, X_l = t, L(n) = d, L(n+1) = l, L(n+2) = m\} \\
&\quad = P\{X_d = s, L(n) = d\}\; P\{X_{d+1} < s, ..., X_{l-1} < s, \\
&\qquad\qquad X_l = t, X_{l+1} < t, ..., X_{m-1} < t, X_m = k+s\} \\
&\quad = P\{X_d = s, L(n) = d\}\; F^{l-d-1}(s)\; F^{m-l-1}(t)\; P(X_l = t)\; P(X_m = k+s) \\
&\quad = p_t p_{k+s} P\{X_d = s, L(n) = d\}\; F^{l-d-1}(s)\; F^{m-l-1}(t).
\end{aligned}
\tag{21.5}
$$

From (21.3) and (21.5), upon changing the order of summation for t, l and m, one can write

$$\begin{aligned} P&\left\{X_{L(n+2)}-X_{L(n)}=k, X_{L(n)}=s\right\} \\ &= \sum_{d=n}^{\infty}\sum_{m=d+2}^{\infty}\sum_{l=d+1}^{m-1}\sum_{t=s}^{k+s}\left\{p_t p_{k+s} P\left\{X_d=s, L(n)=d\right\}\; F^{l-d-1}(s)\; F^{m-l-1}(t)\right\} \\ &= p_{k+s}\sum_{d=n}^{\infty} P\left\{X_d=s, L(n)=d\right\}\sum_{t=s}^{k+s} p_t \sum_{l=d+1}^{\infty} F^{l-d-1}(s)\sum_{m=l+1}^{\infty} F^{m-l-1}(t). \end{aligned} \tag{21.6}$$

Using the obvious facts

$$\sum_{m=l+1}^{\infty} F^{m-l-1}(t)=\frac{1}{\overline{F}(t)},\qquad \sum_{l=d+1}^{\infty} F^{l-d-1}(s)=\frac{1}{\overline{F}(s)}$$

and

$$P(X_{L(n)}=s)=\sum_{d=n}^{\infty} P\left\{X_d=s, L(n)=d\right\}$$

in (21.6), we obtain

$$P\left\{X_{L(n+2)}-X_{L(n)}=k, X_{L(n)}=s\right\}=P(X_{L(n)}=s)\; p_{k+s}\,\frac{1}{\overline{F}(s)}\sum_{t=s}^{k+s}\frac{p_t}{\overline{F}(t)},$$

or, equivalently, we have for the conditional probability

$$P\{X_{L(n+2)}-X_{L(n)}=k \mid X_{L(n)}=s\}=p_{k+s}\,\frac{1}{\overline{F}(s)}\sum_{t=s}^{k+s}\frac{p_t}{\overline{F}(t)}. \tag{21.7}$$

Note that, since this probability does not depend on n, we may, without loss of any generality, assume that $n=1$.

From (21.7), the conditional expectation becomes

$$E\left\{X_{L(3)}-X_1 \mid X_1=s\right\}=\frac{1}{\overline{F}(s)}\cdot\sum_{k=0}^{\infty}\left(kp_{k+s}\sum_{l=s}^{k+s}\frac{p_l}{\overline{F}(l)}\right). \tag{21.8}$$

By changing the order of summation in (21.8) and taking $k+s=z$, one can write

$$\begin{aligned} E\left\{X_{L(3)}-X_1 \mid X_1=s\right\} &= \frac{1}{\overline{F}(s)}\sum_{l=s}^{\infty}\left(\frac{p_l}{\overline{F}(l)}\;\sum_{z=l}^{\infty}((z-s)\;p_z)\right) \\ &= \frac{1}{\overline{F}(s)}\sum_{l=s}^{\infty}\left(\frac{p_l}{\overline{F}(l)}\;\left(\sum_{z=l}^{\infty}(z\;p_z)-s\overline{F}(l)\right)\right) \\ &= \sum_{l=s}^{\infty}\sum_{z=l}^{\infty}\frac{zp_zp_l}{\overline{F}(s)\overline{F}(l)}-s. \end{aligned}$$

Therefore, the basic formula for the conditional expectation for future references is

$$E\left\{X_{L(3)} - X_1 \mid X_1 = s\right\} = \sum_{l=s}^{\infty}\sum_{z=l}^{\infty} \frac{z p_z p_l}{\overline{F}(s)\overline{F}(l)} - s. \tag{21.9}$$

NECESSITY. Let X_1 have a geometric distribution with $p_k = P(X_1 = k) = pq^k$, $k = 0, 1, ...,$ where $q = 1 - p$. Then, it is obvious that $\overline{F}(s) = q^s$ for all $s \geq 0$. Using the known formula that $\sum_{z=l}^{\infty}(zq^z) = \frac{lpq^l + q^{l+1}}{p^2}$ and (21.9), it may be easily seen that

$$\begin{aligned}
E\left\{X_{L(3)} - X_1 \mid X_1 = s\right\} &= \sum_{l=s}^{\infty}\sum_{z=l}^{\infty} \frac{zpq^z pq^l}{q^s q^l} - s \\
&= \frac{p^2}{q^s}\sum_{l=s}^{\infty}\sum_{z=l}^{\infty}(zq^z) - s \\
&= \frac{p^2}{q^s}\sum_{l=s}^{\infty}\frac{lpq^l + q^{l+1}}{p^2} - s \\
&= \frac{1}{q^s}\sum_{l=s}^{\infty}(lpq^l + q^{l+1}) - s \\
&= \frac{p}{q^s}\sum_{l=s}^{\infty}(lq^l) + \frac{1}{q^s}\sum_{l=s}^{\infty} q^{l+1} - s \\
&= \frac{p}{q^s}\frac{spq^s + q^{s+1}}{p^2} + \frac{q}{p} - s = \frac{2q}{p}
\end{aligned}$$

which proves the necessity part of the theorem.

SUFFICIENCY. Let condition (21.2) hold. Also using (21.9), we take the equality

$$\sum_{l=s}^{\infty}\sum_{z=l}^{\infty} \frac{z p_z p_l}{\overline{F}(s)\overline{F}(l)} - s = \alpha,$$

or, equivalently,

$$\begin{aligned}
B_s &\equiv \sum_{l=s}^{\infty}\sum_{z=l}^{\infty} \frac{z p_z p_l}{\overline{F}(s)\overline{F}(l)} \\
&= \alpha + s \qquad\qquad \text{for all } s \geq 0.
\end{aligned} \tag{21.10}$$

Rewriting (21.10) for $s = k$, we have

$$\begin{aligned}
\alpha + k &= B_k \equiv \sum_{l=k}^{\infty}\sum_{z=l}^{\infty} \frac{zp_z p_l}{\overline{F}(k)\overline{F}(l)} \\
&= \sum_{z=k}^{\infty} \frac{zp_z p_k}{\overline{F}(k)\overline{F}(k)} + \sum_{l=k+1}^{\infty}\sum_{z=l}^{\infty} \frac{zp_z p_l}{\overline{F}(k)\overline{F}(l)} \\
&= \frac{p_k}{\overline{F}^2(k)} \sum_{z=k}^{\infty}(zp_z) + \sum_{l=k+1}^{\infty}\sum_{z=l}^{\infty} \frac{zp_z p_l}{\overline{F}(k)\overline{F}(l)} \\
&= \frac{p_k}{\overline{F}^2(k)} \sum_{z=k}^{\infty}(zp_z) + \frac{\overline{F}(k+1)}{\overline{F}(k)} \sum_{l=k+1}^{\infty}\sum_{z=l}^{\infty} \frac{zp_z p_l}{\overline{F}(k+1)\overline{F}(l)} \\
&= \frac{p_k}{\overline{F}^2(k)} \sum_{z=k}^{\infty}(zp_z) + \frac{\overline{F}(k+1)}{\overline{F}(k)} B_{k+1}. \qquad (21.11)
\end{aligned}$$

By condition (21.10), $B_{k+1} = \alpha + k + 1$, and therefore, from (21.11) for all $k \geq 0$, one can write

$$\alpha + k = \frac{p_k}{\overline{F}^2(k)} \sum_{z=k}^{\infty}(zp_z) + \frac{\overline{F}(k+1)}{\overline{F}(k)} (\alpha + k + 1). \qquad (21.12)$$

Observing that $\overline{F}(k+1) = \overline{F}(k) - p_k$, (21.12) gives the identity

$$\alpha + k = \frac{p_k}{\overline{F}^2(k)} \sum_{z=k}^{\infty}(zp_z) + \frac{\overline{F}(k) - p_k}{\overline{F}(k)} (\alpha + k + 1),$$

or, equivalently,

$$\alpha + k = \frac{p_k}{\overline{F}^2(k)} \sum_{z=k}^{\infty}(zp_z) + \alpha + k + 1 - \frac{p_k}{\overline{F}(k)} (\alpha + k + 1).$$

From the last identity, we may write

$$p_k \left(\alpha + k + 1 - \frac{1}{\overline{F}(k)} \sum_{z=k}^{\infty}(zp_z) \right) = \overline{F}(k) \quad \text{for all } k \geq 0. \qquad (21.13)$$

Here, in the case of $k = 0$, we have

$$EX_1 + \frac{1 - p_0}{p_0} = \alpha,$$

or

$$\sum_{z=1}^{\infty}(zp_z) + \frac{1 - p_0}{p_0} = \alpha$$

and rewriting (21.13) in the form

$$p_k = \frac{1-(p_0+...+p_{k-1})}{\alpha+k+1} - \frac{1}{1-(p_0+...+p_{k-1})}\left(\alpha - \frac{1-p_0}{p_0} - \sum_{z=1}^{k-1}(zp_z)\right),$$

we have a recurrence relation for determining p_k for any given p_0. It is clear that the set of probabilities $p_0, p_1, p_2, ...$ must satisfy the conditions

$$p_0+p_1+p_2+p_3+... = 1 \text{ and } \sum_{z=1}^{\infty}(zp_z) + \frac{1-p_0}{p_0} = \alpha. \tag{21.14}$$

For proving that such a set of $p_0, p_1, p_2, ...$ exists and is unique, rewrite (21.13) in terms of $\overline{F}(k)$. Using the obvious equality

$$\begin{aligned}\frac{1}{\overline{F}(k)}\sum_{z=k}^{\infty}(zp_z) &= \frac{kp_k+(k+1)p_{k+1}+(k+2)p_{k+2}+...}{\overline{F}(k)} \\ &= (k-1)+\frac{\overline{F}(k)+\overline{F}(k+1)+\overline{F}(k+2)+\overline{F}(k+3)+...}{\overline{F}(k)}\end{aligned}$$

with (21.13) and the identity $p_k = \overline{F}(k) - \overline{F}(k+1)$, we get

$$\begin{aligned}&\left\{\overline{F}(k)-\overline{F}(k+1)\right\} \\ &\quad\times\left(\alpha+2-\frac{\overline{F}(k)+\overline{F}(k+1)+\overline{F}(k+2)+\overline{F}(k+3)+...}{\overline{F}(k)}\right) = \overline{F}(k).\end{aligned}$$

This equality may be equivalently changed to

$$\alpha\overline{F}(k) = \overline{F}(k+1)+\overline{F}(k+2)+\overline{F}(k+3)+...+\frac{\overline{F}(k)\overline{F}(k+1)}{\overline{F}(k)-\overline{F}(k+1)}. \tag{21.15}$$

Now using (21.15) for k and $k+1$ and subtracting, we take

$$\alpha\left(\overline{F}(k)-\overline{F}(k+1)\right)-\overline{F}(k+1)-\frac{\overline{F}(k)\overline{F}(k+1)}{\overline{F}(k)-\overline{F}(k+1)} = -\frac{\overline{F}(k+1)\overline{F}(k+2)}{\overline{F}(k+1)-\overline{F}(k+2)}.$$

Denoting $\beta_k = \overline{F}(k+1)\overline{F}(k)$ $(k \geq 0)$, noting that $\overline{F}(k+1) = \beta_k\overline{F}(k)$ and that $\overline{F}(k+2) = \beta_k\beta_{k+1}\overline{F}(k)$, we have the recurrence relation for β_k

$$\beta_{k+1} = 1+\frac{\beta_k(1-\beta_k)}{\alpha(1-\beta_k)^2-3\beta_k+2\beta_k^2}\ . \tag{21.16}$$

Consider the second part of condition (21.14) for β_k. Note that $\beta_0 = 1-p_0 = \overline{F}(1)$, $\beta_0\beta_1...\beta_k = \overline{F}(k+1)$ and $0 \leq \beta_k \leq 1$ for all k. The second part of (21.14) together with the fact that

$$\sum_{z=1}^{\infty}(zp_z) = EX_1 = \sum_{k=1}^{\infty}\overline{F}(k),$$

we have

$$\beta_0 + \beta_0\beta_1 + ... + \beta_0\beta_1...\beta_k + ... + \frac{\beta_0}{1-\beta_0} = \alpha. \tag{21.17}$$

In this step, first note that taking $\beta_0 = \frac{\alpha}{\alpha+2}$ in (21.16) we have $\beta_0 = \beta_1 = \beta_2 = ... = \frac{\alpha}{\alpha+2}$ which implies that $\overline{F}(k) = (\frac{\alpha}{\alpha+2})^{k-1}$. So, at least one solution for (21.16) and (21.17) (also satisfying (21.13), (21.14) and (21.15)) exists. This solution corresponds to the geometric distribution with $p_0 = p = \frac{2}{\alpha+2}$. Let us show that this solution is unique. Consider the real-valued function $f(x) = 1 + \frac{x(1-x)}{\alpha(1-x)^2-3x+2x^2}$ $(0 \le x \le 1)$ with two points of discontinuity. For all continuity points x of $f(x)$, we may write

$$f'(x) = \frac{\alpha(1-x)^2 + x^2}{(\alpha(1-x)^2 - 3x + 2x^2)^2} > 0.$$

Therefore, $f(x)$ is a monotonically increasing function in continuity intervals. Let x_1 and x_2 $(x_1 \le x_2)$ be the discontinuity points of $f(x)$. It may be verified then that these points are different, $x_1 \in (0,1)$ and $x_2 > 1$ for any $\alpha > 0$. Furthermore, $f(x) > 1$ for $0 < x < x_1$ and from (21.17) we may have $\beta_0 > 0$ and $\beta_0 < 1$. Eq. (21.16) may be written as $\beta_{k+1} = f(\beta_k)$ from which we have $\beta_0 > x_1$, and vice versa we have $\beta_1 > 1$, which contradicts with condition $0 \le \beta_1 \le 1$. By the same process, it may be seen that from the condition $0 < \beta_1 = f(\beta_0)$, we have $\beta_0 > 1-(\alpha+1)^{-1/2}$. Note that last point $1-(\alpha+1)^{-1/2}$ is the smaller one of the two roots of the equation $f(x) = 0$. For all x such that $1-(\alpha+1)^{-1/2} < x < 1$, $f(x)$ is strictly increasing function and (21.17) then becomes

$$\begin{aligned}\beta_0 &+ \beta_0\, f(\beta_0) + \beta_0\, f(\beta_0)\, f(f(\beta_0)) \\ &+ ... + \beta_0\, f(\beta_0) \dots f(...f(f(\beta_0))...) + ... + \frac{\beta_0}{1-\beta_0} = \alpha.\end{aligned} \tag{21.18}$$

Because $f(\beta_0)$ is a monotonically increasing function of β_0 and $\frac{\beta_0}{1-\beta_0}$ is also monotonically increasing expression of β_0 $(0 \le \beta_0 \le 1)$, we have the left hand side of (21.18) to be monotonically increasing expression of β_0. Therefore, for the constant right hand side of (21.18), we may have only one β_0 satisfying (21.18), which completes the proof of the Theorem. ■

References

1. Ahsanullah, M. (1995). *Record Statistics*, Commack, NY: Nova Science Publishers.

2. Ahsanullah, M. and Holland, B. (1984). Record values and the geometric distribution, *Statistische Helfe*, **25**, 319–327.

3. Aliev, F. A. (1998). Characterization of distributions through weak records, *Journal of Applied Statistical Science* (to appear).

4. Kirmani, S. N. U. A. and Beg, M. I. (1984). On characterization of distributions by expected records, *Sankhyā, Series A*, **49**, 463–465.

5. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*, New York: John Wiley & Sons.

6. Korwar, R. M. (1984). On characterizing distributions for which the second record value has a linear regression on the first, *Sankhyā, Series B*, **46**, 108–109.

7. Nagaraja, H. N. (1988). Record values and related statistics: A review, *Communications in Statistics—Theory and Methods*, **17**, 2223–2238.

8. Nevzorov, V. B. (1987). Records, *Theory of Probability and its Applications*, **32**, 201–228.

9. Nevzorov, V. B. and Balakrishnan, N. (1998). A record of records, In *Handbook of Statistics* – **16**: *Order Statistics: Theory and Methods* (Eds., N. Balakrishnan and C. R. Rao), pp. 515–570, Amsterdam, The Netherlands: North-Holland.

10. Stepanov, A. V. (1994). A characterization theorem for weak records, *Theory of Probability and its Applications*, **39**, 762–764.

11. Vervaat, W. (1973). Limit theorems for records from discrete distributions, *Stochastic Processes and its Applications*, 317–334.

22

Asymptotic Distributions of Statistics Based on Order Statistics and Record Values and Invariant Confidence Intervals

Ismihan G. Bairamov, Omer L. Gebizlioglu, and Mehmet F. Kaya

Ankara University, Ankara, Turkey
Ankara University, Ankara, Turkey
Selcuk University, Konya, Turkey

Abstract: In this chapter, we establish limit theorems for some statistics based on order statistics and record values. The finite-sample as well as asymptotic properties of statistics based on invariant confidence intervals are investigated and their use in statistical inference is outlined.

Keywords and phrases: Invariant confidence intervals, order statistics, record values, limit theorems

22.1 Introduction

Let $X_1, X_2, ..., X_n, ...$ be a sequence of independent and identically distributed (i.i.d.) random variables with continuous distribution function F. Let $X_{(1)}, X_{(2)}, ..., X_{(n)}$ be the order statistics obtained from $X_1, X_2, ..., X_n$. Define a sequence of random variables $U(n), n = 1, 2, ...$, as follows: $U(1) = 1$, $U(n) = \min\{j : j > U(n-1),\ X_j > X_{U(n-1)}, n > 1\}$. The random variables $U(1), U(2), ..., U(n), ...$ are called upper record times, and $X_{U(1)}, X_{U(2)}, ..., X_{U(n)}, ...$ are the record values of sequence $X_1,\ X_2, ...,\ X_n, ...$. Great interest in records exists because we often come across them in our everyday life so that singling out and fixing record values proves to be meaningful. In this chapter, limit distributions of some statistics based on order statistics and record values are obtained. These results are generalized for statistics based on invariant confidence intervals containing the main distributed mass of a general set. For more details on the theory of order statistics and records, one can refer to David (1981), Galambos (1987), Nevzorov (1987), Nagaraja (1988), Nevzorov and Balakrish-

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

nan (1998), Arnold, Balakrishnan and Nagaraja (1992, 1998), and Ahsanullah (1995), among others.

Let $X_1, X_2, ..., X_n$ be a sample from a continuous distribution with distribution function F, and $Y_1, Y_2, ..., Y_m$ be a sample from a continuous distribution with distribution function G. Let $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$ and $Y_{(1)} \leq Y_{(2)} \leq ... \leq Y_{(m)}$ be the respective order statistics. It is well known that under the hypothesis $H_0 : F = G$,

$$P\left\{Y_k \in (X_{(i)}, X_{(j)})\right\} = \frac{j-i}{n+1}, \; 1 \leq i < j \leq n; \; k = 1, 2, ..., m,$$

that is, the random interval $\delta_{ij} = (X_{(i)}, X_{(j)})$ is an invariant confidence interval containing the main distributed mass for a class of continuous distributions; see Bairamov and Petunin (1991).

Let us consider the following random variables:

$$\xi_k^{ij} = \begin{cases} 1 & \text{if } \; Y_k \in (X_{(i)}, X_{(j)}) \\ 0 & \text{if } \; Y_k \notin (X_{(i)}, X_{(j)}) \end{cases}, \quad 1 \leq i < j \leq n; \; k = 1, 2, ..., m.$$

Denote $S_m^{ij} = \sum_{k=1}^{m} \xi_k^{ij}$. It is clear that S_m^{ij} is the number of observations $Y_1, Y_2, ..., Y_m$ falling into interval $(X_{(i)}, X_{(j)})$. The following theorem is a special case of Theorem 22.2.2 that will be proved later in Section 22.2.

Theorem 22.1.1 *For any r and s satisfying $1 \leq r < s \leq n$,*

$$\lim_{m \to \infty} \sup_{0 \leq x \leq 1} \left| P\left\{\frac{S_m^{rs}}{m} \leq x\right\} - P\left\{G(X_{(s)}) - G(X_{(r)}) \leq x\right\}\right| = 0.$$

Corollary 22.1.1 *Under the hypothesis $H_0 : F = G$,*

$$\lim_{m \to \infty} \sup_{0 \leq x \leq 1} \left| P\left\{\frac{S_m^{rs}}{m} \leq x\right\} - P\left\{W_{rs} \leq x\right\}\right| = 0,$$

where $W_{rs} = F(X_{(s)}) - F(X_{(r)})$.

It is known that W_{rs} has the probability density function [see David (1981)]

$$f(w_{rs}) = \begin{cases} \frac{1}{B(s-r, n-s+r+1)} w_{rs}^{s-r-1} (1 - w_{rs})^{n-s+r} & 0 \leq w_{rs} \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

Let us consider the random intervals $\delta_0 = (-\infty, X_{(1)})$, $\delta_1 = (X_{(1)}, X_{(2)})$, ..., $\delta_n = (X_{(n-1)}, X_{(n)})$, $\delta_{n+1} = (X_{(n)}, \infty)$. For $W_{i-1,i} = F(X_{(i)}) - F(X_{(i-1)})$, one can clearly obtain from Corollary 22.1.1 that

$$P\{\, W_{i-1,i} \leq x\} = \begin{cases} 0 & \text{if } \; x \leq 0 \\ 1 - (1-x)^n & \text{if } \; x \in (0,1) \\ 1 & \text{if } \; x \geq 1 \end{cases},$$

where $X_{(0)} = -\infty, X_{(n+1)} = \infty$.

Theorem 22.1.2 *Under the hypothesis $H_0 : F = G$, for $1 \leq i \leq n+1$*

$$\lim_{n\to\infty}\lim_{m\to\infty}\left|P\left\{\frac{nS_m^{i-1,i}}{m} \leq x\right\} - F_0(x)\right| = 0,$$

where $F_0(x) = 1 - e^{-x}, x \geq 0$.

Let $X_1, X_2, ..., X_n, ...$ be a sequence of independent and identically distributed random variables with continuous distribution function F. Consider the r-th record value $X_{U(r)}$. Let $X_{U(r)+1}, X_{U(r)+2}, ..., X_{U(r)+m}$ be the next m observations that come after $X_{U(r)}$. It is not difficult to prove that $X_{U(r)}$, $X_{U(r)+1}, X_{U(r)+2}, ..., X_{U(r)+m}$ are mutually independent and $X_{U(r)+k}$ has the same distribution F for any $k = 1, 2, ..., m$. Let us define the following random variables for a given r:

$$\xi_i(r) = 1 \text{ if } X_{U(r)+i} < X_{U(r)} \text{ and } \xi_i(r) = 0 \text{ if } X_{U(r)+i} \geq X_{U(r)} \;\; i = 1, 2, ..., m,$$

and let $S_m(r) = \sum_{i=1}^{m}\xi_i(r)$. It is clear that $S_m(r)$ is the number of observations $X_{U(r)+1}, X_{U(r)+2}, ..., X_{U(r)+m}$ which are less than $X_{U(r)}$. Note that the random variables $\xi_1(r), \xi_2(r), ..., \xi_m(r)$ are generally dependent. In the work of Bairamov (1997) the finite-sample and asymptotic properties of the statistic $S_m(r)$ are given. We will mention here some of these results.

Theorem 22.1.3 *For any $m, r = 1, 2, ...,$*

$$P\{S_m(r) = k\} = \binom{m}{k}(r-1)!\int_0^\infty e^{-z(m-k+1)}(1-e^{-z})^k z^{r-1}dz$$

$k = 0, 1, 2, ..., m.$

Let us denote $S_m^*(r) = \frac{S_m(r) - ES_m(r)}{\sqrt{var(S_m(r))}}$. Then, $ES_m^*(r) = 0$, and $var(S_m^*(r)) = 1$. Denote $a = \frac{1}{2^r}$ and $b = \sqrt{\frac{1}{3^r} - \frac{1}{2^{2r}}}$.

Theorem 22.1.4 [Bairamov (1997)] *The statistic $S_m^*(r)$ has a continuous limiting distribution as $m \to \infty$, with probability density function $f^{\cdot}$ as*

$$f^*(x) = \begin{cases} \frac{b}{(r-1)!}\left[\ln\frac{1}{a-bx}\right]^{r-1} & if \;\; x \in \left[\frac{a-1}{b}, \frac{a}{b}\right] \\ 0 & if \;\; x \notin \left[\frac{a-1}{b}, \frac{a}{b}\right] \end{cases}.$$

Theorem 22.1.5 [Bairamov (1997)] *It is true that*

$$\lim_{m\to\infty}\sup_{0\leq x\leq 1}\left|P\left\{\frac{S_m(r)}{m} \leq x\right\} - \frac{1}{(r-1)!}\int_0^x\left[\ln\left(\frac{1}{1-u}\right)\right]^{r-1}du\right| = 0.$$

Analogous statistics based on invariant confidence intervals is considered in this chapter.

Let $X_1, X_2, ..., X_n$ be a sample from a distribution with distribution function $F \in \Im$, where $\Im$ is some class of distribution functions. Suppose $f_1(u_1, u_2, ..., u_n)$ and $f_2(u_1, u_2, ..., u_n)$ are two Borel functions with the property that

$$f_1(u_1, u_2, ..., u_n) \leq f_2(u_1, u_2, ..., u_n) \ \forall \ (u_1, u_2, ..., u_n) \in R^n. \tag{22.1}$$

Let X_{n+1} be a new sample point obtained from F which is independent of $X_1, X_2, ..., X_n$. If

$$P\{X_{n+1} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n)\} = a \quad \text{for all} \quad F \in \Im,$$

then $(f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))$ is called an invariant confidence interval containing the main distributed mass for class of distributions $\Im$ with confidence level a.

It is known that [see Bairamov and Petunin (1991)] if f_1 and f_2 are continuous, symmetric and different on every set with a non-zero Lebesgue measure functions of n arguments, only the order statistics form invariant confidence intervals for $\Im_c$, the class of all continuous distribution functions.

Properties of invariant confidence intervals for nonparametric class $\Im_c$ are used in many applications since a test statistic can be found and criteria can be established on the training samples for problems of classification of new observations [see Bairamov and Petunin (1991) and Bairamov (1992)]. Similar applications can also be extended to generalized Bernoulli schemes in variation statistics [see Matveichuk and Petunin (1990) and Matveichuk and Petunin (1991)].

The solution for the problem of the significance estimation of the indices used for diagnosis of the breast cancer on the basis of investigation of the DNA distribution in the interphase nuclei cells is obtained by using criteria and statistics introduced in by Petunin, Timoshenko and Petunina (1984) and using invariant confidence intervals and results obtained by Bairamov and Petunin (1991a).

In the next section, the finite-sample and asymptotic properties of statistics based on invariant confidence intervals are investigated and their use in statistical inference is also discussed.

22.2 The Main Results

Let $h(u_1, u_2, ..., u_n)$ be a real-valued integrable n-dimensional function. Consider a functional

$$H_F(h) = \int ... \int h(u_1, u_2, ..., u_n) dF(u_1) dF(u_2) .... dF(u_n) \ , \ F \in \mathbf{F} \ ,$$

where $\mathbf{F}$ is some class of distribution functions. The properties of the functional $H_F(h)$ are

(i) $H_F(1) = 1$

(ii) $H_F(c_1h_1(.) + c_2h_2(.)) = c_1H_F(h_1) + c_2H_F(h_2)$, where $h_j(.)$ are distinct functions and c_j's are real valued numbers.

Denote the random samples from the distributions $F(u)$ and $Q(u)$ as $(X_1, X_2, ..., X_n)$ and $(Y_1, Y_2, ..., Y_m)$, respectively. Let f_1 and f_2 be two functions as mentioned in (22.1). The probability of a random event

$$A_k = \{Y_k \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\}, \; k = 1, 2, ...m$$

is

$$p \equiv P(A_k) = \int ... \int [Q(f_2(u_1, u_2, ..., u_n)) - Q(f_1(u_1, u_2, ..., u_n))] \, dF(u_1)dF(u_2)....dF(u_n),$$

which is clearly independent of k. If we take the above definition of $H_F(h)$ into consideration, the required probability is calculated by

$$P(A_k) = p = H_F[Q(f_2(\bar{u})) - Q(f_1(\bar{u}))] \equiv H_F(Q_{f_1}^{f_2}(\bar{u})),$$

where $\bar{u} = (u_1, u_2, ..., u_n)$ and $Q(f_2(\bar{u})) - Q(f_1(\bar{u})) \equiv Q_{f_1}^{f_2}(\bar{u})$. Denoting

$$\xi_k = \begin{cases} 0, & \text{if random event } A_k \text{ is observed} \\ 1, & \text{if random event } A_k \text{ is not observed} \end{cases}$$

and defining a new random variable as $\nu_m = \xi_1 + \xi_2 + ... + \xi_m$, which can take values from the set $\{0, 1, 2, ..., m\}$, we can investigate the likelihood of having new sample values falling into a designated interval. Note that the random variables $\xi_1, \xi_2, ..., \xi_m$ are dependent.

Theorem 22.2.1 *For* $k = 0, 1, 2, ..., m$,

$$P\{\nu_m = k\} = C_m^k H_F\left(\left[Q_{f_1}^{f_2}(\bar{u})\right]^k \left[1 - Q_{f_1}^{f_2}(\bar{u})\right]^{m-k}\right),$$

where $C_m^k = \binom{m}{k} = \frac{m!}{k!(m-k)}$.

PROOF. Let $A_{ik} = \{Y_{ik} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\}$. The probability that k of $Y_1, Y_2..., Y_m$ fall in the interval $(f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))$ is then

$$P\{\nu_m = k\} = \sum_{i_1, i_2, .., i_m} P(A_{i1} \cap A_{i2} \cap ... \cap A_{ik} \cap \overline{A_{i_{k+1}}} \cap \overline{A_{i_{k+2}}} \cap ... \cap \overline{A_{i_m}}), \tag{22.2}$$

where $\overline{A}$ denotes the complement of A. Let us denote

$$P^{(k)}_{i_1,i_2,..,i_m} = P(A_{i1} \cap A_{i2} \cap ... \cap A_{ik} \cap \overline{A_{i_{k+1}}} \cap \overline{A_{i_{k+2}}} \cap ... \cap \overline{A_{i_m}}).$$

In this case, we have

$$P^{(k)}_{i_1,i_2,..,i_m} = \int \cdots \int_{\mathbf{A}} dF(y_{i_1}, y_{i_2}, ..., y_{i_m}, x_1, x_2, ..., x_n),$$

where

$$\mathbf{A} = \{(y_{i_1}, y_{i_2}, ..., y_{i_m}, x_1, x_2, ..., x_n) : -\infty < x_i < \infty \ , \ i = 1, 2, ..., n \ ;$$
$$f_1(x_1, x_2, ..., x_n) < y_{i_p} < f_2(x_1, x_2, ..., x_n), p = 1, 2, ..., k;$$
$$y_{i_j} \notin (f_1(x_1, x_2, ..., x_n), f_2(x_1, x_2, ..., x_n)), \ j = k+1, k+2, ..., m\Big\}.$$

$Y_1, Y_2..., Y_m$ and $X_1, X_2..., X_n$ have the following joint distribution function due to independence:

$$F(y_{i_1}, y_{i_2}, ..., y_{i_m}, x_1, x_2, ..., x_n) = Q(y_1)Q(y_2)...Q(y_m)F(x_1)F(x_2)...F(x_n).$$

So, we can write

$$\begin{aligned} P^{(k)}_{i_1,i_2,..,i_m} &= \int \cdots \int_{\mathrm{A}} [Q(f_2(u_1, u_2, ..., u_n)) - Q(f_1(u_1, u_2, ..., u_n))]^k \\ &\quad \times \ [1 - Q(f_2(u_1, u_2, ..., u_n)) + Q(f_1(u_1, u_2, ..., u_n)]^{m-k} \\ &\qquad \times \ dF(u_1)...dF(u_n) \\ &= H_F\left(\left[Q^{f_2}_{f_1}(u_1, u_2, ..., u_n)\right]^k \left[1 - Q^{f_2}_{f_1}(u_1, u_2, ..., u_n)\right]^{m-k}\right). \end{aligned}$$

This shows that the probabilities $P^{(k)}_{i_1,i_2,..,i_m}$ in (22.2) are independent of $i_1, i_2, ..., i_m$.

Hence, the theorem. ■

The low-order moments of interest for further uses are expressed as follows:

$$\begin{aligned} E(\nu_m) &= \sum_{k=0}^{m} kP\{\nu_m = k\} \\ &= \sum_{k=0}^{m} kC^k_m \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[Q^{f_2}_{f_1}(u_1, u_2, ..., u_n)\right]^k \\ &\qquad \times \left[1 - Q^{f_2}_{f_1}(u_1, u_2, ..., u_n)\right]^{m-k} dF(u_1)...dF(u_n) \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\sum_{k=0}^{m} kC^k_m \left[Q^{f_2}_{f_1}(u_1, u_2, ..., u_n)\right]^k \right. \\ &\qquad \left. \times \left[1 - Q^{f_2}_{f_1}(u_1, u_2, ..., u_n)\right]^{m-k}\right) dF(u_1)...dF(u_n). \end{aligned}$$

Now, we find that

$$\begin{aligned}
&\sum_{k=0}^{m} kC_m^k \left[Q_{f_1}^{f_2}(\bar{u})\right]^k \left[1-Q_{f_1}^{f_2}(\bar{u})\right]^{m-k} \\
&= m\sum_{k=1}^{m} C_{m-1}^{k-1} \left[Q_{f_1}^{f_2}(\bar{u})\right]^{k-1} \left[1-Q_{f_1}^{f_2}(\bar{u})\right]^{m-k} \left[Q_{f_1}^{f_2}(\bar{u})\right] \\
&= mQ_{f_1}^{f_2}(\bar{u}) \sum_{i=0}^{m-1} C_{m-1}^{i} \left[Q_{f_1}^{f_2}(\bar{u})\right]^{i} \left[1-Q_{f_1}^{f_2}(\bar{u})\right]^{(m-1)-i} \\
&= mQ_{f_1}^{f_2}(\bar{u})
\end{aligned}$$

using which, we obtain

$$\begin{aligned}
E(\nu_m) &= m\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} Q_{f_1}^{f_2}(u_1,u_2,...,u_n)dF(u_1)...dF(u_n) \\
&= mH_F\left[Q_{f_1}^{f_2}(u_1,u_2,...,u_n)\right].
\end{aligned}$$

Similarly, we also find

$$\begin{aligned}
E(\nu_m^2) &= m^2H_F\left[Q_{f_1}^{f_2}(u_1,u_2,...,u_n)\right]^2 - mH_F\left[Q_{f_1}^{f_2}(u_1,u_2,...,u_n)\right]^2 \\
&\quad + mH_F\left[Q_{f_1}^{f_2}(u_1,u_2,...,u_n)\right].
\end{aligned}$$

Therefore, the mean and the variance of ν_m are obtained as

$$E(\nu_m) = mH_F(Q_{f_1}^{f_2}(\bar{u}))$$

and

$$\begin{aligned}
var(\nu_m) &= m^2\left[\left(H_F(Q_{f_1}^{f_2}(\bar{u}))^2\right) - \left(H_F(Q_{f_1}^{f_2}(\bar{u}))\right)^2\right] \\
&\quad - m\left[\left(H_F(Q_{f_1}^{f_2}(\bar{u}))^2\right) - \left(H_F(Q_{f_1}^{f_2}(\bar{u}))\right)\right].
\end{aligned}$$

Lemma 22.2.1 *The characteristic function of ν_m is*

$$\varphi_{\nu_m}(t) = H_F\left(\left(1+\left(e^{it}-1\right)Q_{f_1}^{f_2}(u_1,u_2,...,u_n)\right)^m\right).$$

PROOF. By definition,

$$\begin{aligned}
\varphi_{\nu_m}(t) &= E\left(\exp(it\nu_m)\right) \\
&= \sum_{k=0}^{m} \exp(itk)P(\nu_m = k) \\
&= \sum_{k=0}^{m} \exp(itk)C_m^k H_F\left[Q_{f_1}^{f_2}(\bar{u})\right]^k \left[1-Q_{f_1}^{f_2}(\bar{u})\right]^{m-k} \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\sum_{k=0}^{m} \exp(itk)C_m^k \left[Q_{f_1}^{f_2}(u_1,u_2,...,u_n)\right]^k \\
&\quad \times\left[1-Q_{f_1}^{f_2}(u_1,u_2,...,u_n)\right]^{m-k} dF(u_1)...dF(u_n).
\end{aligned}$$

In order to carry out the necessary calculation, the summation term is first found as

$$\begin{aligned}
&\sum_{k=0}^{m} \exp(itk) C_m^k \left[Q_{f_1}^{f_2}(\bar{u})\right]^k \left[1 - Q_{f_1}^{f_2}(\bar{u})\right]^{m-k} \\
&= \sum_{k=0}^{m} C_m^k \left[\exp(it) Q_{f_1}^{f_2}(\bar{u})\right]^k \left[1 - Q_{f_1}^{f_2}(\bar{u})\right]^{m-k} \\
&= \left(\exp(it) Q_{f_1}^{f_2}(\bar{u}) + \left(1 - Q_{f_1}^{f_2}(\bar{u})\right)\right)^m \\
&= \left(1 - Q_{f_1}^{f_2}(\bar{u}) \left(1 - \exp(it)\right)\right)^m .
\end{aligned}$$

Then, we have

$$\begin{aligned}
\varphi_{\nu_m}(t) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(1 - Q_{f_1}^{f_2}(u_1, u_2, ..., u_n) \left(1 - \exp(it)\right)\right)^m \\
&\qquad \times dF(u_1)...dF(u_n) \\
&= H_F \left(1 + \left(\exp(it) - 1\right) Q_{f_1}^{f_2}(u_1, u_2, ..., u_n)\right)^m .
\end{aligned}$$

Hence, the lemma. ∎

Now, let us define the standardized form of ν_m as $\nu_m^* = \frac{\nu_m - E(\nu_m)}{\sqrt{var(\nu_m)}}$ with $E(\nu_m^*) = 0$ and $var(\nu_m^*) = 1$. Denote

$$\begin{aligned}
C(x) &= P\left\{Q_{f_1}^{f_2}(X_1, X_2, ..., X_n) \le x\right\} \\
&= P\left\{Q(f_2(X_1, X_2, ..., X_n)) - Q(f_1(X_1, X_2, ..., X_n)\ ) \le x\right\}.
\end{aligned}$$

Theorem 22.2.2 *Let f_1 and f_2 be continuous functions, and F and Q be continuous distribution functions. Then,*

$$\lim_{m\to\infty} \sup_{0\le x\le 1} \left| P\left\{\frac{\nu_m}{m} \le x\right\} - C(x)\right| = 0.$$

PROOF. By using Lemma 22.2.1, the characteristic function of $\frac{\nu_m}{m}$ can be written as

$$\varphi_{\frac{\nu_m}{m}}(t) = E(e^{i\frac{t}{m}\nu_m}) = H_F \left(1 + \left(\exp(i\frac{t}{m}) - 1\right) Q_{f_1}^{f_2}(\bar{u})\right)^m . \qquad (22.3)$$

Let us now denote $\Psi_m(t) = \left(1 + \left(\exp(i\frac{t}{m}) - 1\right) Q_{f_1}^{f_2}(\bar{u})\right)^m$. Using the Taylor expansions $e^x = 1 + x + o(x)$ and $\ln(1 + x) = x + o(x)$, we can write

$$\begin{aligned}
\ln \Psi_m(t) &= m \ln \left(1 + \left(\exp(i\frac{t}{m}) - 1\right) Q_{f_1}^{f_2}(\bar{u})\right) \\
&= m \ln \left(1 + \left(\frac{it}{m} + o(\frac{it}{m})\right) Q_{f_1}^{f_2}(\bar{u})\right)
\end{aligned}$$

$$\begin{aligned}
&= m \ln\left(1 + \left(\frac{it}{m} Q_{f_1}^{f_2}(\bar{u}) + o(\frac{t}{m})\right)\right) \\
&= m\left(\frac{it}{m} Q_{f_1}^{f_2}(\bar{u}) + o(\frac{t}{m})\right) \\
&= itQ_{f_1}^{f_2}(\bar{u}) + O(\frac{1}{m})
\end{aligned}$$

and so

$$\Psi_m(t) = \exp\left(itQ_{f_1}^{f_2}(\bar{u})\right) + O(\frac{1}{m}).$$

It then follows from (22.3) that

$$\varphi_{\frac{\nu_m}{m}}(t) = H_F(\Psi_m(t)) = H_F\left(\exp\left(itQ_{f_1}^{f_2}(\bar{u})\right)\right) + O(\frac{1}{m}). \tag{22.4}$$

Letting $m \to \infty$ in (22.4), we obtain

$$\lim_{m\to\infty} \varphi_{\frac{\nu_m}{m}}(t) = H_F\left(\exp\left(itQ_{f_1}^{f_2}(\bar{u})\right)\right) \equiv \Psi(t). \tag{22.5}$$

It easy to see that $\Psi(t)$ is continuous at $t = 0$. In fact, one has

$$\exp\left(itQ_{f_1}^{f_2}(\bar{u})\right) = 1 + itQ_{f_1}^{f_2}(\bar{u}) + \frac{i^2t^2\left(Q_{f_1}^{f_2}(\bar{u})\right)^2}{2!} + o(t^2)$$

and $\Psi(t) = H_F\left(\exp\left(itQ_{f_1}^{f_2}(\bar{u})\right)\right) \to 1 = \Psi(0)$ if $t \to 0$.

Let $F_m^*(x)$ be the distribution function of statistic ν_m, where $x = \frac{k}{m}$, $k = 0, 1, 2, ..., m$. By using Levy-Cramer theorem for characteristic functions [see Petrov (1975, Theorem 10, p. 15)], one can show that $F_m^*(x) \to F^*(x), x \in [0, 1]$, and F^* has a characteristic function

$$\Psi(t) = \int_0^1 e^{itx} dF^*(x). \tag{22.6}$$

On the other hand, from (22.5), we have

$$\begin{aligned}
\Psi(t) &= H_F\left(\exp\left(itQ_{f_1}^{f_2}(\bar{u})\right)\right) \\
&= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(itQ_{f_1}^{f_2}(u_1, u_2, ..., u_n)\right) dF(u_1)...dF(u_2) \\
&= E\left[\exp\left(itQ_{f_1}^{f_2}(X_1, X_2, ..., X_n)\right)\right] \\
&= \int_0^1 e^{itx} dP\left\{Q_{f_1}^{f_2}(X_1, X_2, ..., X_n) \le x\right\}.
\end{aligned} \tag{22.7}$$

Therefore, from (22.6) and (22.7), we have

$$\begin{aligned}
F^*(x) &= P\left\{Q_{f_1}^{f_2}(X_1, X_2, ..., X_n) \le x\right\} \\
&= P\left\{Q\left(f_2(X_1, ..., X_n)\right) - Q\left(f_1(X_1, ..., X_n)\right) \le x\right\}.
\end{aligned}$$

■

Corollary 22.2.1 *Let $(f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))$ be the invariant confidence interval for some class of distributions $\Im$ with confidence level α_1, that is,*

$$P_F\{X_{n+1} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\} = \alpha_1 \text{ for any } F \in \Im.$$

Let $\alpha_2 = P_F\{X_{n+1}, X_{n+2} \in (f_1(X_1, X_2, ..., X_n), f_2(X_1, X_2, ..., X_n))\}$, where $X_1, X_2, ..., X_n$, X_{n+1}, X_{n+2} is the random sample from distribution with distribution function $F \in \Im$. Let $F = Q$ and $F \in \Im$ and $\boldsymbol{X} = (X_1, X_2, ..., X_n)$. Then

$$\lim_{m\to\infty} \sup_x \left| P\left\{ \frac{\nu_m - m\alpha_1}{\sqrt{m^2(\alpha_2 - \alpha_1^2) - m(\alpha_2 - \alpha_1)}} \leq x \right\} - G_2(x) \right| = 0,$$

where

$$G_2(x) = \begin{cases} 0, & \text{if} \quad x \leq -\frac{\alpha_1}{\sqrt{\alpha_2 - \alpha_1^2}} \\ P\{F(f_2(\boldsymbol{X})) - F(f_1(\boldsymbol{X})) \leq x\}, & \text{if} \quad x \in (-\frac{\alpha_1}{\sqrt{\alpha_2-\alpha_1^2}}, \frac{1-\alpha_1}{\sqrt{\alpha_2-\alpha_1^2}}) \\ 1, & \text{if} \quad x \geq \frac{1-\alpha_1}{\sqrt{\alpha_2-\alpha_1^2}} \end{cases}.$$

Remark 22.2.1 Let $\mathbf{P} = \Im_c$, where $\Im_c$ is the family of all continuous distributions. Let $f_1(X_1, X_2, ..., X_n) = X_{(i)}$ and $f_2(X_1, X_2, ..., X_n) = X_{(j)}, 1 \leq i < j \leq n$. With these, we can then show that [see Bairamov and Petunin (1991)]

$$H_F\left(F_{u_{(i)}}^{u_{(j)}}(u_1, u_{2,...,}u_n)\right) = P\left\{X_{n+1} \in \left(X_{(i)}, X_{(j)}\right)\right\} = \frac{j-i}{n+1} \equiv \alpha_{i,j}$$

and

$$\begin{aligned} H_F\left[\left(F_{u_{(i)}}^{u_{(j)}}(u_1, u_{2,...,}u_n)\right)^m\right] &= P\left\{X_{n+1}, X_{n+2}, ..., X_{n+m} \in \left(X_{(i)}, X_{(j)}\right)\right\} \\ &= \frac{n!(m+j-i-1)!}{(j-i-1)!(m+n)!} \equiv \alpha_{ij}^{(m)}. \end{aligned}$$

If $i = 1$ and $j = n$, then $\alpha_{1,n} = \frac{n-1}{n+1}$, and $\alpha_{1,n}^{(2)} = \frac{(n-1)n}{(n+1)(n+2)}$.

Remark 22.2.2 Let $X_1, X_2, ..., X_n$ be a sample with distribution function $F \in \mathbf{P} = \Im_c$, where $\Im_c$ is the family of all continuous distributions. Let $f_1(X_1, X_2, ..., X_n) = X_{(i)}$ and $f_2(X_1, X_2, ..., X_n) = X_{(j)}$, $1 \leq i < j \leq n$. In this case, $C(x)$ in Theorem 22.2.2 takes the form $C(x) = P_F\left\{Q(X_{(j)}) - Q(X_{(i)}) \leq x\right\}$ and we have Theorem 22.1.1. If $F = Q$, then

$$C(x) = P\left\{F(X_{(j)}) - F(X_{(i)}) \leq x\right\} = P\{W_{ij} \leq x\}$$

and we have Corollary 22.1.1.

References

1. Ahsanullah, M. (1995). *Record Statistics*, Commack, NY: Nova Science Publishers.

2. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1992). *A First Course of Order Statistics*, New York: John Wiley & Sons.

3. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*, New York: John Wiley & Sons.

4. Bairamov, I. G. and Petunin, Yu. I. (1991a). Structure of invariant confidence intervals containing the main distributed mass, *Theory of Probability and Its Applications*, **35**, 15–26.

5. Bairamov, I. G. and Petunin, Yu. I. (1991b). Statistical criteria based on training samples, *Cybernetics*, **27**, 408–413.

6. Bairamov, I. G. (1992). Statistical criteria based on training samples in the problem of classification of observations, *Theory of Probability and Mathematical Statistics*, **46**, 13–17 (In Ukrainian).

7. Bairamov, I. G. (1997). Some distribution free properties of statistics based on record values and characterizations of the distributions through a record, *Journal of Applied Statistical Science*, **5**, 17–25.

8. David, H. A. (1981). *Order Statistics*, Second edition, New York: John Wiley & Sons.

9. Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics*, Second edition, Malabar, Florida: Krieger.

10. Matveychuk, S. A. and Petunin, Yu. I.(1990). The Bernoulli generalized schemes in the variation statistics. 1, *Ukrainskiy Mathematicheskiy Journal*, **42**, 518–528.

11. Matveychuk, S. A. and Petunin, Yu. I. (1991). Generalized Bernoulli circuits appearing in the variational statistics. 2, *Ukrainskiy Mathematicheskiy Journal*, **43**, 779–786.

12. Nagaraja, H. N. (1988). Record values and related statistics. A review, *Communication in Statistics—Theory and Methods*, **17**, 2223–2238.

13. Nevzorov, V. B. (1987). Records, *Theory of Probability and Its Applications*, **32**, 201–228.

14. Nevzorov, V. B. and Balakrishnan, N. (1998). A record of records, In *Handbook of Statistics* – **16**: *Order Statistics: Theory and Methods* (Eds., N. Balakrishnan and C. R. Rao), pp. 515–570, Amsterdam, The Netherlands: North-Holland.

15. Petunin, Yu. I., Timoshenko, Ja. G. and Petunina, M. Yu. (1984). Test for the identification of the general population for the finite class of the alternative hypotheses, *Dokladi Ukrainian Academy of Sciences, Series A*, **6**, 29–32.

16. Petrov, V. V. (1975). *Sums of Independent Random Variables*, New York: Springer-Verlag.

23

Record Values in Archimedean Copula Processes

N. Balakrishnan, L. N. Nevzorova, and V. B. Nevzorov

McMaster University, Hamilton, Ontario, Canada
St. Petersburg Humanitarian University of Trade Unions, St. Petersburg, Russia
St. Petersburg State University, St. Petersburg, Russia

Abstract: We investigate the asymptotic behavior of record values $X(n)$ for some types of Archimedean copula processes. It is shown that the set of all limit distribution functions for $X(n)$, normed and centered in a suitable way, under some restrictions on parameters of these processes, coincides with the corresponding set of asymptotic distributions of record values in the sequences of i.i.d. random variables.

Keywords and phrases: Archimedean copula process, extremes, record indicators, record times, record values

23.1 Introduction

Let $X_1, X_2, \ldots$ be a sequence of random variables and $M(n) = \max\{X_1, \ldots, X_n\}$ for $n = 1, 2, \ldots$. We define (upper) record times $L(n)$, record values $X(n)$ and record indicators ξ_n as follows:

$$\begin{aligned} L(1) &= 1, \quad L(n+1) = \min\{j > L(n)\, X_j > X_{L(n)}\}, \quad n = 1, 2, \ldots; \\ X(n) &= X_{L(n)} = M(L(n)), \quad n = 1, 2, \ldots; \\ \xi_1 &= 1, \quad \xi_n = 1_{\{M(n) > M(n-1)\}}, \quad n = 2, 3, \ldots. \end{aligned}$$

Consider also random variables

$$N(n) = \xi_1 + \cdots + \xi_n, \quad n = 1, 2, \ldots,$$

where $N(n)$ represents the number of records among $X_1, \ldots, X_n$.

Beginning from the pioneering paper by Chandler (1952), there are now about 350 papers on records. The most updated reviews of the literature on records can be found in Ahsanullah (1995), Arnold, Balakrishnan and Nagaraja (1998) and Nevzorov and Balakrishnan (1998).

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

Majority of the work on records consider record times and record values in sequences of independent identically distributed random variables $X_1, X_2, \ldots$. Moreover, the theory of records from dependent or/and nonidentically distributed X's have been connected with the independence property of record indicators ξ_n and maxima $M(n)$. This property, for example, was the basis for dependent record schemes of Ballerini and Resnick (1987), Deheuvels and Nevzorov (1994), Ballerini (1994), and Nevzorova, Nevzorov and Balakrishnan (1997). The last two of these are based on the so-called *Archimedean copula (AC) processes.*

A sequence $X_1, X_2, \ldots$ with marginal distribution functions $F_1, F_2, \ldots$ is said to be an AC process if, for any $n = 1, 2, \ldots$, the joint distribution function

$$H(t_1, \ldots, t_n) = P\{X_1 < t_1, \ldots, X_n < t_n\}$$

has the following form:

$$H(t_1, \ldots, t_n) = B(\sum_{i=1}^{n} A(F_i(t_i))), \tag{23.1}$$

where B is a completely monotone function such that $B(0) = 1$, and $A = B^{-1}$ is the inverse of the function B.

Ballerini (1994) has studied in detail the particular AC processes with

$$B(s) = \exp(-s^{1/\delta}), \quad \delta \geq 1,$$

and

$$F_i(x) = (F(x))^{\alpha(i)}, \quad i = 1, 2, \ldots, \tag{23.2}$$

where $\alpha(1), \alpha(2), \ldots$ are any positive constants and F is a continuous distribution function. When $\delta = 1$, Ballerini's scheme coincides with the so-called F^α–*scheme* which was initiated by Yang (1975) and developed by Nevzorov (1984, 1985, 1986, 1995), Pfeifer (1989, 1991), and Deheuvels and Nevzorov (1993, 1994). It is known that in the F^α–scheme, random variables $X_1, X_2, \ldots$ are independent and have distribution functions (23.2). Ballerini called his model as the *dependent* F^α*–scheme.* He proved that record indicators $\xi_1, \xi_2, \ldots, \xi_n$ and maximum $M(n)$, in the dependent F^α–scheme, are independent for any $n = 1, 2, \ldots$.

Nevzorova, Nevzorov and Balakrishnan (1997) investigated a much more general set of AC processes and found the necessary and sufficient conditions under which the random variables $\xi_1, \xi_2, \ldots, \xi_n$ and $M(n)$ are independent for any $n = 1, 2, \ldots$. It appears that this independence property takes place for AC processes with the joint distributions

$$H(t_1, \ldots, t_n) = B(\sum_{i=1}^{n} c_i A(F(t_i))), \tag{23.3}$$

where B and A are as defined in (23.1), F is any continuous distribution function, and c_i are any positive constants. This case includes Ballerini's situation, of course, when $B(s) = exp(-s^{1/\delta})$,

It is not difficult to prove that

$$p_k = P\{\xi_k = 1\} = 1 - P\{\xi_k = 0\} = 1 - \gamma(k-1)/\gamma(k), \; k = 1, 2, \ldots, \quad (23.4)$$

where $\gamma(n) = c_1 + \ldots + c_n$, for AC processes with dependence function (23.3).

The independence of record indicators has been used to examine some martingale and asymptotic properties of random variables $N(n)$ and $L(n)$ for the F^α–scheme; see, for example, Balakrishnan and Nevzorov (1997), Deheuvels and Nevzorov (1993), Nevzorov (1995), and Nevzorov and Stepanov (1988). Arguments which are in Section 23.3 show that almost all of these results can be reformulated for AC processes in (23.3).

For the F^α–scheme, using the independence of record indicators and maxima, Nevzorov (1995) has found the set of all possible limit distributions of record values $X(n)$, centered and normalized in a suitable way. Here we generalize his results for the AC processes. There are standard methods for proving such theorems; see, for example, Resnick (1973a,b) and Nevzorov (1995). The first step in this direction is to find the limit distribution of $X(n)$ for some convenient initial distribution (in Resnick's and Nevzorov's papers, the exponential distribution was used for this purpose). After that, arguments based on the Smirnov transformation and the well developed theory of extremes will enable one to construct the bridge from this special distribution to the general case. It appears that for AC processes in (23.3), unlike the classical record schemes, it is convenient to take the first step with the following analogue of the exponential distribution function:

$$F(x) = B(-\log(1 - \exp(-x))).$$

23.2 Main Results

Let a sequence $X_1, X_2, \ldots$ form an AC process with the joint distributions as given in (23.3). In this case,

$$P\{M(n) < x\} = B\{\gamma(n)A(F(x))\} = H\{G^{\gamma(n)}(x)\}, \quad (23.5)$$

where

$$H(x) = B(-\log x)$$

and $G(x) = \exp\{-A(F(x))\}$ is a distribution function. It follows from (23.5) that $P\{M(n) - a_n < b_n x\}$, as $n \to \infty$, converges to a nondegenerate limit distribution $R(x) = H(T(x))$, if

$$G^{\gamma(n)}(xb_n + a_n) \to T(x) \quad \text{as } n \to \infty. \tag{23.6}$$

It is well known that if $\gamma(n) = n$, then there are three classical types of limit distributions on in the RHS of (23.6), which have the following form:

$$T_i(x) = \exp(-\exp(-g_i(x))), \qquad i = 1, 2, 3,$$

where

$$g_1(x) = x, \tag{23.7}$$

$$g_2(x) = g_{2,\alpha}(x) = \alpha \log x \text{ if } x > 0, \text{ and } g_2(x) = -\infty \text{ if } x < 0, \tag{23.8}$$

$$g_3(x) = g_{3,\alpha}(x) = -\alpha \log(-x) \text{ if } x < 0, \text{ and } g_3(x) = \infty \text{ if } x > 0, \tag{23.9}$$

and $\alpha > 0$ in (23.8) and (23.9). In this situation (viz., $\gamma(n) = n$), all possible limit distributions for $M(n)$, centered and normalized in a suitable way, evidently have the form

$$H_i(x) = B(-\log(T(x))) = B(\exp(-g_i(x))), \quad i = 1, 2, 3,$$

where functions $g_i(x)$ are as defined in (23.7)–(23.9). Green (1976) has shown that for any fixed nondegenerate distribution function T, it is possible to find a continuous distribution function G and sequences a_n, b_n and $\gamma(n)$ such that (23.6) holds. His construction requires a very fast increase of coefficients $\gamma(n)$. In Green's example [see also Example 2.6.5 in Galambos (1978)]

$$\gamma(n) \sim \exp(\exp \lambda n), \quad \lambda > 0, \quad n \to \infty.$$

Note also that if $\gamma(n) \to \gamma, 0 < \gamma < \infty$, then for any fixed distribution function T, one can take $F(x) = B(A(T(x))/\gamma)$ and get the limit relation

$$\lim P\{M(n) < x\} = T(x) \quad \text{as } n \to \infty.$$

For record values $X(n) = M(L(n))$, one can easily prove that if

$$\lim P\{M(n) - a_n < b_n x\} = R(x) \quad \text{as } n \to \infty,$$

then

$$\lim P\{X(n) - a_{L(n)} < b_{L(n)} x\} = R(x) \quad \text{as } n \to \infty.$$

Let us consider the asymptotic behavior of $X(n)$ under nonrandom centering and normalizing.

Denote $A_n = \sum_{k=1}^{n} p_k$ and $B_n = \sum_{k=1}^{n} p_k^2$, $n = 1, 2, \ldots$, where p_n are as given in (23.4).

In the sequel, we will study AC processes in (23.4) with coefficients c_n satisfying one of two following sets of restraints:

$$\gamma(n) \to \infty, \quad p_n \to 0, \quad B_n/A_n^{1/2} \to 0 \quad \text{as } n \to \infty; \tag{23.10}$$

$$p_n \to p\ (0 < p < 1),\ \sum_{k=1}^{n}(p - p_k)/n^{1/2} \to 0,\ \sum_{k=1}^{n}(p - p_k)^2/n^{1/2} \to 0 \quad \text{as } n \to \infty. \tag{23.11}$$

Note here, in comparison to Green's construction where $\log \gamma(n)$ has the exponential rate of increasing, that $\gamma(n) \to \infty$ and $\log \gamma(n) = O(n)$ as $n \to \infty$, under conditions in (23.10) and (23.11).

First, we get some asymptotic results for the partial case when $G(x) = \exp\{-A(F(x))\}$ coincides with $1 - \exp(-x)$ for any positive x. It means that

$$F(x) = B(-\log(1 - \exp(-x))) \text{ if } x > 0 \text{ and } F(x) = 0 \text{ if } x < 0.$$

Let Φ denote the distribution function of the standard normal law. The following results then true.

Theorem 23.2.1 *If $F(x) = B(-\log(1 - \exp(-x)))$ and conditions in (23.10) hold, then for any fixed x and y,*

$$P\{X(n) - \log \gamma(L(n)) < x, \log \gamma(L(n)) - n < yn^{1/2}\} \to B(e^{-x})\Phi(y). \tag{23.12}$$

Theorem 23.2.2 *Let $F(x) = B(-\log(1 - \exp(-x)))$ and that conditions in (23.11) hold. Then for any fixed x and y,*

$$P\{X(n) - \log \gamma(L(n)) < x,\ p \log \gamma(L(n)) + n \log(1 - p) < y(-\log(1 - p))(n(1 - p))^{1/2}\} \to B(e^{-x})\Phi(y). \tag{23.13}$$

It is evident that for $B(s) = \exp(-s)$, which corresponds to the F^{α}–scheme, $F(x) = B(-\log(1 - \exp(-x))$ coincides with the standard exponential distribution function. Hence, Theorems 23.2.1– Theorem 23.2.2 generalize Theorems 7 and 8 of Nevzorov (1995).

Corollary 23.2.1 *Let $F(x) = B(-\log(1 - \exp(-x))$ and that conditions in (23.10) hold. Then for any fixed x,*

$$P\{X(n) - n < xn^{1/2}\} \to \Phi(x) \tag{23.14}$$

as $n \to \infty$.

In fact,

$$(X(n) - n)/n^{1/2} = \nu_n/n^{1/2} + \mu_n,$$

where $\nu_n = (X(n) - \log\gamma(L(n)))$ and $\mu_n = (\log\gamma(L(n)) - n)/n^{1/2}$. It is evident now that asymptotic normality of $(X(n) - n)/n^{1/2}$ easily follows from (23.13).

Corollary 23.2.2 *If $F(x) = B(-\log(1 - \exp(-x))$ and conditions in (23.11) hold, then for any fixed x,*

$$P\{pX(n) + n\log(1-p) < x(-log(1-p))((1-p)n)^{1/2}\} \to \Phi(x) \quad (23.15)$$

as $n \to \infty$.

In this situation,

$$(pX(n) + n\log(1-p))/(-\log(1-p))(n(1-p))^{1/2}$$
$$= p\eta_n/(-\log(1-p))(n(1-p))^{1/2} + \tau_n,$$

where

$$\eta_n = X(n) - \log\gamma(L(n))$$

and

$$\tau_n = (p\log\gamma(L(n)) + n\log(1-p))/(-\log(1-p))(n(1-p))^{1/2}.$$

All we need to do now is to apply (23.13).

The standard arguments based on the Smirnov transformation [see Resnick (1973a,b) and Nevzorov (1995)] allow us to obtain from (23.14) and (23.15) the limit distributions of record values $X(n)$ for any continuous distribution function F.

Theorem 23.2.3 *For AC processes in (23.3), with coefficients $c(n)$ satisfying (23.10) or (23.11), the record values $X(n)$, centered and normalized in a suitable fashion, can have only three types of nondegenerate asymptotic distribution functions*

$$R_i(x) = \Phi(g_i(x)), \quad i = 1, 2, 3,$$

where $g_i(x)$ are as given in (23.7)–(23.9) and Φ is the distribution function of the standard normal law.

23.3 Sketch of Proof

Let us compare indicators $\xi_1, \xi_2, \ldots$ and maxima $M(n)$ for AC processes in (23.3) with record indicators (denoted by $\xi_1^\star, \xi_2^\star, \ldots$) and maxima $M^\star(n)$ in the F^α–scheme with exponents $\alpha_n, n = 1, 2, \ldots$, which coincide with coefficients c_n of the corresponding AC process. Both the sets of indicators have the same distributions and these distributions do not depend on distribution function F, which enters in the definitions of the F^α–scheme and AC process in (23.3). Random variables $\xi_1, \xi_2, \ldots$ as well as indicators $\xi_1^\star, \xi_2^\star, \ldots$ are independent and

$$P\{\xi_n = 1\} = c_n/(c_1 + \cdots + c_n) = \alpha_n/(\alpha_1 + \cdots + \alpha_n) = P\{\xi_n^\star = 1\}.$$

Hence, for any $n = 1, 2, \ldots$, distributions of the vectors $(\xi_1, \ldots, \xi_n)$ and $(\xi_1^\star, \ldots, \xi_n^\star)$ coincide. It implies that the same situation is true for the random variables $N(n) = \xi_1 + \ldots + \xi_n$ and $N^\star(n) = \xi_1^\star + \ldots + \xi_n^\star$. Moreover, record times $L(n)$ and $L^\star(n)$ have the same distribution because we can apply the following inequalities:

$$\begin{aligned} P\{L^\star(n) > m\} &= P\{N^\star(m) < n\} = P\{\xi_1^\star + \ldots + \xi_m^\star < n\} \\ &= P\{\xi_1 + \ldots + \xi_m < n\} = P\{N(m) < n\} = P\{L(n) > m\}, \end{aligned} \tag{23.16}$$

which are valid for any $n = 1, 2, \ldots$ and $m = 1, 2, \ldots$. The random variables $L^\star(n)$ and $N^\star(n)$ in (23.16) as well as record values $X^\star(n)$ and maxima $M^\star(n)$ correspond to the F^α–scheme.

It therefore follows from (23.16) that all results which hold for record times $L^\star(n)$ and numbers of records $N^\star(n)$ in the F^α–scheme with exponents $\alpha_1, \alpha_2, \ldots$ can be reformulated for $L(n)$ and $N(n)$ in the case when these record statistics correspond to AC process in (23.3) with coefficients $c_n = \alpha_n$, $n = 1, 2, \ldots$.

Let us compare now record values $X^\star(n)$ and maxima $M^\star(n)$ for the F^α–scheme with record values $X(n)$ and maxima $M(n)$ for AC process in (23.3). From (23.5), one knows that the maximal value $M(n)$ for the AC process in (23.3) has the following form:

$$P\{M(n) < x\} = B\{\gamma(n)A(F(x))\} = H\{G^{\gamma(n)}(x)\},$$

where $H(x) = B(-\log x), G(x) = \exp\{-A(F(x))\}$ and $\gamma(n) = c_1 + \ldots + c_n$. For the F^α–scheme with exponents $\alpha_n = c_n$, $n = 1, 2, \ldots$, and distribution function $F^\star$,

$$M_n^\star(x) = P\{M^\star(n) < x\} = (F^\star(x))^{\gamma(n)}.$$

The formulae and arguments given above show that there exists a duality between AC process in (23.3) with coefficients c_n and distribution function F, and the F^α–scheme with exponents $\alpha_n = c_n$, $n = 1, 2, \ldots$, and distribution function $F^\alpha(x) = G(x) = \exp\{-A(F(x))\}$. As we mentioned above, distributions of record times coincide for these two cases. Besides, we have the following relation between distributions of $M(n)$ and $M^\star(n)$:

$$M_n(x) = P\{M(n) < x\} = H(M_n^\star(x)).$$

The analogous equalities are valid for asymptotic distributions. If $M_n^\star(xb_n + a_n)$ converges to a limiting distribution $T^\star(x)$ under some constants a_n and b_n, then $M_n(xb_n + a_n)$ converges to distribution function $T(x) = H(T^\star(x))$. Note also that in this situation, as $n \to \infty$,

$$P\{(X^\star(n) - a_{L^\star(n)})/b_{L^\star(n)} < x\} \to T^\star(x) \tag{23.17}$$

and

$$P\{(X(n) - a_{L(n)})/b_{L(n)} < x\} \to H(T^\star(x)). \tag{23.18}$$

One important fact that we must recall is that: for any $n = 1, 2, \ldots$, indicators $\xi_1^\star, \xi_2^\star, \ldots, \xi_n^\star$ and maximum $M^\star(n)$ are independent [see, for example, Ballerini and Resnick (1987)], as well as the indicators $\xi_1, \xi_2, \ldots, \xi_n$ and maximum $M(n)$ are independent in the case of AC processes in (23.3) [see Nevzorova, Nevzorov and Balakrishnan (1997)].

Therefore, we just need to trace the proof of Theorems 7 and 8 in Nevzorov (1995)) for the F^α–scheme which correspond to our Theorems 23.2.1 and 23.2.2 with some slight changes due to the differences in (23.17) and (23.18). In the theorems of Nevzorov (1995), relation (23.17) was used for the special case when $F^\star(x) = 1 - \exp(-x)$, $x > 0$, and it has the form

$$P\{X^\star(n) - \log\gamma(L^\star(n)) < x\} \to \exp(-\exp(-x)) \tag{23.19}$$

as $n \to \infty$. Then, $F(x) = B(-\log(1 - \exp(-x)))$ from Theorems 23.2.1 and 23.2.2 is a dual function for $F^\star(x)$, and in this situation, (23.18) can be rewritten as

$$P\{X(n) - \log\gamma(L(n)) < x\} \to H(\exp(-\exp(-x))) = B(e^{-x}) \tag{23.20}$$

as $n \to \infty$. All the other arguments are the same for both situations and the difference of the RHS in (23.19) and (23.20) is the only reason why the limit expressions in (23.12) and (23.13) have the form $B(e^{-x})\Phi(y)$, while the corresponding limit for the F^α–scheme is $\exp(-\exp(-x))\Phi(y)$.

There is a standard method to find the set of all possible nondegenerate limit distributions for record values $X(n)$, centered and normalized in a suitable way. In the classical situation, when the original random variables $X_1, X_2, \ldots$ are independent and identically distributed, this method was suggested by Resnick (1973 a,b). For the F^α–scheme, it was developed by Nevzorov (1995).

We will now consider the asymptotic distribution of $X(n)$ under the restrictions of Theorem 23.2.1. First, let $F(x) = F_0(x) = B(-\log(1 - \exp(-x)))$, $x > 0$. It follows from Corollary 23.2.1 that the corresponding record values (for the sake of convenience in this situation we denote them as $Z(n)$) satisfy the following relation:

$$P\{Z(n) - n < xn^{1/2}\} \to \Phi(x) \tag{23.21}$$

as $n \to \infty$. Then, due to Smirnov transformation, record values $X(n)$ for any continuous distribution function $F(x)$ can be expressed via random variables $Z(n)$ as follows: $X(n) = R(Z(n))$, where $R(x) = F^{-1}(B(-\log(1 - e^{-x})))$ and F^{-1} is the inverse function of F. If there exist some normalizing and centering constants a_n and b_n such that $P\{X(n) - a_n < xb_n\}$ converges to some distribution $T(x)$, then this fact can be rewritten as

$$P\{(Z(n) - n)/n^{1/2} < (U_F(a_n + xb_n) - n)/n^{1/2}\} \to T(x), \tag{23.22}$$

where $U_F(x) = -\log(1-\exp(-B^{-1}(F(x))))$ and B^{-1} is the inverse of B. Comparing (23.21) and (23.22), one can see that (23.22) holds if and only if there exists the following limit:

$$g(x) = \lim(U_F(a_n + xb_n) - n)/n^{1/2}, \qquad \text{as } n \to \infty. \tag{23.23}$$

Moreover, it means that all asymptotic distribution T for $X(n)$ must have the form $T(x) = \Phi(g(x))$. Now one can recall that the problem finding all possible limits in (23.23) has already been solved by Resnick. It appears that the set of limit functions g is restricted by three types of functions as in (23.7)–(23.9). This completes the proof of Theorem 23.2.3 under the restraints in (23.10).

The same arguments show that this theorem is also true if the conditions in (23.11) are posed. ■

Acknowledgement. This work was partially supported by the Russian Foundation of Fundamental Research (grants 96-01-00549 and 96-15-96199).

References

1. Ahsanullah, M. (1995) . *Record Statistics*, Commack, NY: Nova Science Publishers.

2. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*, New York: John Wiley & Sons.

3. Balakrishnan, N. and Nevzorov, V. (1997). Stirling numbers and records, In *Advances in Combinatorial Methods and Applications to Probability and Statistics* (Ed., N. Balakrishnan), pp. 189–200, Boston: Birkhäuser.

4. Ballerini, R. (1994). A dependent F^α-scheme, *Statistics & Probability Letters*, **21**, 21–25.

5. Ballerini, R. and Resnick, S. (1987). Embedding sequences of successive maxima in extremal processes, with applications, *Journal of Applied Probability*, **24**, 827–837.

6. Deheuvels, P. and Nevzorov, V. B. (1993). Records in the F^α–scheme. I. Martingale properties, *Notes of Sci. Semin. POMI*, **207**, 19–36 (in Russian). English version in: *Journal of Mathematical Science*, **81** (1996), 2368–2378.

7. Deheuvels , P. and Nevzorov, V. B. (1994). Records in the F^α–scheme. II. Limit laws, *Notes of Sci. Semin. POMI*, **216**, 42–51 (in Russian). English version in: *Journal of Mathematical Science*, **88** (1998), 2368–2378.

8. Galambos, J. (1978). *The Asymptotic Theory of Extreme Order Statistics,* New York: John Wiley & Sons.

9. Green, R. F. (1976). Partial attraction of maxima, *Journal of Applied Probability*, **13**, 159–163.

10. Nevzorov, V. B. (1984). Record times in the case of nonidentically distributed random variables, *Theory of Probability and Its Applications*, **29**, 808–809.

11. Nevzorov, V. B. (1985). Records and inter-record times for sequences of nonidentically distributed random variables, *Notes of Sci. Semin. LOMI*, **142**, 109–118 (in Russian). English version: in *Journal of Soviet Mathematics*, **36** (1987), 510–516.

12. Nevzorov, V. B. (1986). Two characterizations using records, In *Stability Problems for Stochastic Models* (Eds., V. V. Kalashnikov, B. Penkov and V. M. Zolotarev), Lecture Notes in Mathematics **1233**, pp. 79–85, Berlin: Springer-Verlag.

13. Nevzorov, V. B. (1995). Asymptotic distributions of records in nonstationary schemes, *Journal of Statistical Planning and Inference*, **44**, 261–273.

14. Nevzorov, V. B. and Stepanov, A. V. (1988). Records: Martingale approach to finding of moments, In *Rings and Modules. Limit Theorems of Probability Theory* (Eds., Z. I. Borevich and V. V. Petrov), **2**, 171–181, St.-Petersburg: St.-Petersburg State University.

15. Nevzorova, L. N., Nevzorov, V. B. and Balakrishnan, N. (1997). Characterizations of distributions by extremes and records in Archimedean

copula process, In *Advances in the Theory and Practice of Statistics: A Volume in Honor of Samuel Kotz* (Eds., N. L. Johnson and N. Balakrishnan), pp. 469–478, New York: John Wiley & Sons.

16. Pfeifer, D. (1989). Extremal processes, secretary problems and the 1/e law, *Journal of Applied Probability*, **26**, 722–733.

17. Pfeifer, D. (1991). Some remarks on Nevzorov's record model, *Advances in Applied Probability*, **23**, 823–834.

18. Resnick, S. I. (1973a). Record values and maxima, *Annals of Probability*, **1**, 650–662.

19. Resnick, S. I. (1973b). Limit laws for record values, *Stochastic Processes and their Applications*, **1**, 67–82.

20. Yang, M. C. K. (1975). On the distribution of the inter-record times in an increasing population, *Journal of Applied Probability*, **12**, 148–154.

24

Functional CLT and LIL for Induced Order Statistics

Yu. Davydov and V. Egorov

Université de Sciences et Technologies de Lille, Villeneuve d'Ascq, France
Electrotechnical University, St. Petersburg, Russia

Abstract: The asymptotic behavior of three processes defined by induced order statistics is studied. We prove for these processes a functional central limit theorem and a Strassen-type functional law of the iterated logarithm. The result about weak convergence is a large generalization of Bhattacharya's one (1974). The entropy technique and sharp estimates for weighted empirical processes are used. Here, we present an abridged version of our work.

Keywords and phrases: Functional limit theorem, Strassen-type FLIL, induced order statistics

24.1 Introduction

Let $Z_i = (X_i, Y_i)$, $i = 1, 2 \ldots$, be independent copies of a random vector $Z = (X, Y)$ such that $X \in \mathbf{R}^1$, $Y \in \mathbf{R}^d$. Denote $X_{n,1} \le X_{n,2} \le \cdots \le X_{n,n}$ for the order statistics of the sample $(X_i, i \le n)$ and $Y_{n,1}, Y_{n,2}, \ldots, Y_{n,n}$ for the corresponding values of the vectors Y. The random variables $(Y_{n,i}, i \le n)$ are called *induced order statistics* (IOS).

This generalization of order statistics was first introduced by David (1973) under the name of *concomitants* of order statistics and simultaneously by Bhattacharya (1974). While the asymptotic theory of induced order statistics have been discussed in great detail by David and Galambos (1974), Galambos (1987), Sen (1976), Yang (1977), Egorov and Nevzorov (1982, 1984), finite sample results have been presented by David *et al.* (1977) and Yang (1977). Barnett *et al.* (1976), David (1981), Gomes (1981), Kaminsky (1981), and Balakrishnan (1993) have provided further interesting discussions on these subjects. Interested readers may also refer to Bhattacharya (1984), David (1992) and David

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

and Nagaraja (1998) for extensive reviews on the developments of IOS. One should mention new unexpected connections of IOS with the problems of "convexification" of random walks consided by Davydov and Vershik (1998) (see below for more detailed discussion).

In this chapter, we study the asymptotic behavior of three processes constructed by the induced order statistics $Y_{n,i},\ i \leq n$:

$$\eta_n(t) = \sum_{j=1}^{n} Y_j \mathbf{1}_{[0,t)}(X_j), \tag{24.1}$$

$$\xi_n(t) = \frac{1}{n}\sum_{j=1}^{[nt]} Y_{n,j}, \tag{24.2}$$

$$\alpha_n(t) = \frac{1}{n}\sum_{j=1}^{[nt]} (Y_{n,j} - m(X_{n,j})), \tag{24.3}$$

where $m(t) = E(Y \mid X = t)$, $t \in [0,1]$.

It is not difficult to see (and it follows from the results of our work) that with probability one

$$\alpha_n \to 0,\ \xi_n \to f,\ \eta_n \to f$$

uniformly in $[0,1]$, where $f(t) = \int_0^t m(s)ds$.

Our main goal is to prove a functional central limit theorem (FCLT) and a Strassen-type functional law of the iterated logarithm (FLIL) for the processes

$$\tilde{\eta}_n = b_n(\eta_n - f),\quad \tilde{\xi}_n = b_n(\xi_n - f),\quad \tilde{\alpha}_n = b_n\alpha_n, \tag{24.4}$$

where $b_n = \sqrt{n}$ if the weak convergence is considered and $b_n = \sqrt{n/(2\log\log n)}$ in the case of convergence with probability one.

Our results about the weak convergence represent a large generalization of Bhattacharya's ones (1974, 1976): the moment condition is reduced to $E\|Y\|^2 < \infty$, the dimension d can be more than 1 and the regression function $m(x)$ is permitted to be unbounded. The result about the FLIF, as it seems, is the first result on this subject for IOS.

Finally, we would like to mention something about the methods applied in this chapter. We use entropy technique [modern results of Ossiander (1987) and Ledoux and Talagrand (1991)] and sharp estimates for weighted empirical processes [see Shorack and Wellner (1986)].

The Chapter is organized as follows. After notation (Section 24.2) in Section 24.3 we formulate and discuss Theorem 24.3.1. This theorem states the weak convergence (in the uniform topology) of the processes (24.4) to continuous Gaussian processes which admit integral representations in terms of the d-dimensional Wiener process and the Brownian bridge. Strassen balls for d-dimensional random Gaussian processes are described in Section 24.4. The

FLIL (Theorem 24.5.1) is investigated in Section 24.5. Section 24.6 is devoted to examples. Some simple lemmas are stated without proofs.

24.2 Notation

$X_{n1} \leq \cdots \leq X_{n,n}$ – order statistics for X_i, $i \leq n$
$Y_{n1}, \ldots, Y_{n,n}$ – induced order statistics
α_n, ξ_n, η_n – see (24.1), (24.2), (24.3)
$\tilde{\alpha}_n, \tilde{\xi}_n, \tilde{\eta}_n$ – see (24.4)
$F_n(t)$ – empirical distribution function
$U_n(t) = F_n(t) - t$
$V_n(t) = X_{n,[nt]} - t$
$V(t)$ – Brownian bridge
$W(t)$ – standard Wiener process
$m(s) = E(Y|X = s)$
$\sigma^2(s) = cov(Y|X = s)$ – conditional covariance matrix
$\sigma(s)$ – the positive square root of $\sigma^2(s) : \sigma(s)\sigma(s)^T = \sigma^2(s)$
$D\xi$ – variance of random variable ξ
$\|\ell\|$ – Euclidean norm in R^d
$\|X\|_a^b = \sup_{a<t<b} |X(t)|$
$B[0,1]$ – space of bounded functions with the norm $\|f\|_0^1$
$b_n = \sqrt{n/(2\log\log n)}, n > 3$
$\omega_f(a) = \sup_{|t-\tau|<a} |f(t) - f(\tau)|$
$C, C_1, \ldots$ – positive constants
IOS – induced order statistics
S_X – Strassen ball for the random process X
$f_+(x) = max(f(x), 0),\ f_-(x) = max(-f(x), 0)$

24.3 Functional Central Limit Theorem

We begin by supposing that X is uniformly distributed on the interval $[0, 1]$. The inverse probability transformation permits easily to pass from this case to general one. We recall that Y is supposed to have a finite second order moment. Let

$$\sigma^2(s) = E\{(Y - m(X))(Y - m(X))^T \mid X = s\}$$

be the conditional covariance matrix of Y and $\sigma(s)$ be the positive matrix such that $\sigma(s)\sigma(s)^T = \sigma^2(s)$.

All our limit processes being continuous, it is well known that in this case the convergence in uniform and Skorokhod topologies are equivalent [see Billingsley (1968)], hence we use the uniform metric. The symbol $\Rightarrow$ denotes the weak convergence in the uniform topology of the corresponding Skorokhod space or simply the convergence in distribution of random variables.

Theorem 24.3.1 *Suppose that* $E\|Y\|^2 < \infty$. *Then*

(1) $\tilde{\eta}_n \Rightarrow \tilde{\eta}$, $\tilde{\alpha}_n \Rightarrow \tilde{\alpha}$,
where

$$\tilde{\alpha}(t) = \int_0^t \sigma(s)d\bar{W}(s), \quad \tilde{\eta}(t) = \tilde{\alpha}(t) + \int_0^t m(s)dV(s), \tag{24.5}$$

$\bar{W}$ *is the d-dimensional standard Wiener process,* V *is the Brownian bridge independent of* $\bar{W}$.

(2) If, in addition, m *is continuous in the open interval* $(0,1)$ *and for some* $C > 0$, $a \in (0, 1/2)$

$$|m(t)| \leq Ct^{-a}(1-t)^{-a}, \; t \in (0,1), \tag{24.6}$$

then

$$\tilde{\xi}_n \Rightarrow \tilde{\xi}$$

where

$$\tilde{\xi}(t) = \tilde{\eta}(t) - m(t)V(t). \tag{24.7}$$

We set here $m(0)V(0) = m(1)V(1) = 0$.

Remark 24.3.1 Due to the condition $E\|Y\|^2 < \infty$, the function m and matrix-valued function σ are square integrable. It means that the integrals (24.5) are well defined.

Remark 24.3.2 If X has a continuous distribution function F of general type, we introduce the variables $U_j = F(X_j)$. Then the processes $\tilde{\alpha}_n, \tilde{\xi}_n, \tilde{\eta}_n$ coincide with ones of Theorem 24.3.1 constructed by the vectors (Y_j^T, U_j^T) with the functions m, σ replaced by $m \circ F^{-1}$, $\sigma \circ F^{-1}$ respectively. The same modifications are needed for the limit processes $\tilde{\alpha}, \tilde{\xi}, \tilde{\eta}$. The condition (24.6) turns into the following condition:

for some $C > 0$, $a \in (0, 1/2)$,

$$|m(s)| \leq CF^{-a}(s)(1-F(s))^{-a}, \; s \in \mathbf{R}^1. \tag{24.8}$$

Remark 24.3.3 The statements of Theorem 24.3.1 remain valid if the processes $\tilde{\alpha}_n, \tilde{\xi}_n$ and $\tilde{\eta}_n$ are constructed by means of the continuous polygonal interpolation.

PROOF OF THEOREM 24.3.1. Due to Cramer-Wold device, it is sufficient to prove this theorem only for $d = 1$.

Next we need two lemmas.

Lemma 24.3.1 *If (24.6) holds, then*

(a) $\sqrt{n}m(\cdot)U_n(\cdot) \Rightarrow m(\cdot)V(\cdot)$.

(b) For any random sequence γ_n *such that* $\gamma_n \downarrow 0$ *in probability*

$\sup_{t\in[0,\gamma_n]\cup[1-\gamma_n,1]} \sqrt{n}|m(t)U_n(t)| \xrightarrow{P} 0.$

(c) The statements (a) and (b) are still valid with U_n *replaced by* V_n.

PROOF OF LEMMA 24.3.1. First note that the Hölder condition for $V(t)$ is fulfilled a.s. for any λ, $0 < \lambda < 1/2$. Then, the random process $m(t)V(t)$ is a.s. continuous for $t \in [0,1]$.

The statement a) and the corresponding part of the statement c) follow from O'Reilly (1974) [see also Shorack and Wellner (1986, p. 462)], if we take in their result $q(t) = t^\beta(1-t)^\beta$ for some $\beta \in (a, 1/2)$ and notice that $m_1(t) = m(t)q(t)$ is continuous for $t \in [0,1]$.

The statement b) is a simple consequence of a). ∎

Lemma 24.3.2 *Let* β_n *be a sequence of random processes such that* $\beta_n \Rightarrow \beta$, *where* β *is a continuous random process. If the processes* $\{\tau_n(t),\ t \in [0,1]\}$ *are such that* $\tau_n(t) \in [0,1]$ *for any* $t \in [0,1]$ *and* $\|\tau_n - I\|_0^1 \xrightarrow{P} 0$ *where* $I(t) \equiv t,\ t \in [0,1]$, *then*

$$d_n = \|\beta_n(\tau_n(\cdot)) - \beta_n(\cdot)\|_0^1 \xrightarrow{P} 0.$$

1. Now consider the processes η_n, which are (generalized) empirical processes. Therefore, we can apply the general result of Ossiander (1987) for processes of this type to obtain the weak convergence.

Let $H^B(u, S, \rho)$ denotes the metric entropy with bracketing and S, ρ, for the processes η_n are defined as follows:

$$\begin{aligned} S &= \{h_t(x,y),\ x \in [0,1], y \in R^1 : h_t(x,y) = y\mathbf{1}_{[0,t)}(x)\}, \\ \rho(h_t, h_s) &= (E(h_t(X,Y) - h_s(X,Y))^2)^{1/2}. \end{aligned}$$

According to this work, we must prove the inequality

$$\int_0^1 (\log H^B(u, S, \rho))^{1/2} du < \infty; \tag{24.9}$$

for more details, see Ossiander (1986). It is not difficult to show that

$$H^B(u, S, \rho) \le C \log \frac{1}{u}.$$

Hence, by Ossiander's theorem,

$$\tilde{\eta}_n \Rightarrow \tilde{\eta},$$

where $\tilde{\eta}$ is the Gaussian process, $E\tilde{\eta}(t) = 0$, $E(\tilde{\eta}(t), \tilde{\eta}(\tau)) = K_1(t, \tau)$, and $K_1(t, \tau)$ is the covariance function of every summand in the sum

$$n\eta_n(t) = \sum_1^n Y_i \mathbf{1}_{[0,t)}(X_i). \tag{24.10}$$

Elementary calculations show that

$$K_1(t, \tau) = \int_0^{t\wedge\tau} \sigma^2(s)ds + \Big(\int_0^{t\wedge\tau} m^2(s)ds - \int_0^t m(s)ds \int_0^\tau m(s)ds\Big)$$

and that $K_1(t, \tau)$ is the covariance function of the random process $\tilde{\eta}$ defined in (24.5).

2. The second step is to consider the processes ξ_n. Note that

$$\xi_n(t) = \frac{1}{n} \sum_{i=1}^n Y_i \mathbf{1}_{[0, X_{n,[nt]})}(X_i) = \eta_n(X_{n,[nt]}).$$

Hence, due to Lemma 24.3.2,

$$\|\tilde{\eta}_n(X_{n,[nt]}) - \tilde{\eta}_n(t)\|_0^1 \xrightarrow{P} 0$$

and

$$\xi_n(t) = \eta_n(t) + \int_t^{X_{n,[nt]}} m(s)ds + o\left(\frac{1}{\sqrt{n}}\right) \tag{24.11}$$

uniformly for $t \in [0, 1]$ in probability.

Using Lemma 24.5.4 , Lemma 24.5.5 and Part (c) of Lemma 24.3.1, we get

$$\sqrt{n}\Big\| \int_t^{X_{n,[nt]}} (m(s) - m(t))ds \Big\|_0^1 \xrightarrow{P} 0, \quad n \to \infty. \tag{24.12}$$

Due to (24.11) and (24.12),

$$\tilde{\xi}_n(t) = \tilde{\eta}_n(t) + \sqrt{n}m(t)V_n(t) + o(1). \tag{24.13}$$

We have proved already that the summands in (24.13) converge, hence they are tight. Then, $\tilde{\xi}_n$ is tight too. Thus, to prove the weak convergence $\tilde{\xi}_n$ to $\tilde{\xi}$ it is sufficient to show the convergence of their finite-dimensional distributions.

Moreover, since $\tilde{\xi}_n(t) \xrightarrow{P} 0 = \tilde{\xi}(0) = \tilde{\xi}(1)$ for $t = 0$ and $t = 1$, we can consider only $t \in (0,1)$. Then by virtue of the Kiefer-Bahadur theorem [see Shorack and Wellner (1986, p. 586)], we can replace in (24.13) V_n by $-U_n$. Hence,

$$\tilde{\xi}_n(t) = \eta_n(t) - \sqrt{n}m(t)U_n(t) + o(1). \tag{24.14}$$

We can also represent $\tilde{\xi}_n$ in the form

$$\tilde{\xi}_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (U(t; X_i, Y_i) - E(U(t; X_i, Y_i))) + o(1), \tag{24.15}$$

where

$$U(t; X, Y) = (Y - m(t))\mathbf{1}_{[0,t)}(X).$$

Now consider the k-dimensional distribution of $\tilde{\xi}_n$ for $0 < t_1 < \cdots < t_k < 1$. The multivariate central limit theorem for i.i.d. random variables gives the asymptotic normality with zero mean. The limiting covariance matrix is equal to the covariance matrix of the random vector $\{U(t_j; X, Y), j = 1, 2, \ldots, k\}$, that is, it is equal to the matrix $\{cov(U(t_i; X, Y), U(t_j; X, Y))\}_{i,j=1}^{k}$.

Hence, the covariance function of $\tilde{\xi}$ is equal to

$$\begin{aligned} K_2(t,\tau) &= K_1(t,\tau) - (m(t) + m(\tau)) \int_0^{t\wedge\tau} m(s)ds \\ &\quad + tm(t) \int_0^{\tau} m(s)ds + \tau m(\tau) \int_0^{t} m(s)ds + m(t)m(\tau)(t \wedge \tau - t\tau). \end{aligned}$$

It is easy to show that $K_2(t,\tau)$ is the covariance function of the process $\hat{\xi}$ defined by (24.7).

3. Consider the random processes $\tilde{\alpha}_n$. They are the same as the processes $\tilde{\xi}_n$ provided Y_n are replaced by $\tilde{Y}_n = Y_n - m(X_n)$, $n = 1, 2, \ldots$. Since $\tilde{m}(s) = E(\tilde{Y}_n | X_n = s) = 0$, the weak convergence for $\tilde{\alpha}_n$ follows from one for $\tilde{\xi}_n$. ■

24.4 Strassen Balls

In Section 24.5, we shall derive a Strassen-type functional law of the iterated logarithm (FLIL) for the processes defined by IOS. It is known that the limit sets in the FLIL are the Strassen balls for the corresponding limit Gaussian processes. That is why in this section we give a detailed description of the Strassen balls for multivariate Gaussian processes.

Consider a random Gaussian-centred element X of a separable Banach space B. Let K be its covariance operator, that is,

$$K : B^* \to B, \quad \langle Kx^*, y^* \rangle = E(\langle X, x^* \rangle \langle X, y^* \rangle),$$

for $x^*, y^* \in B^*$,where $\langle x, x^* \rangle = x^*(x)$ is the corresponding bilinear form. Let L be a Hilbert space, $\mathcal{I} : L \to B$ be a linear operator, $\mathcal{I}^* : \mathcal{B}^* \to \mathcal{L}$ be the operator which is adjoint for $\mathcal{I}$. Assume that $K = \mathcal{I}\mathcal{I}^*$.

Definition 24.4.1 The set

$$S_X = \{x \in B \mid x = \mathcal{I}(\ell),\ \ell \in \mathcal{L},\ \ \|\ell\|_{\mathcal{L}} \leq \infty\}$$

is called the *Strassen ball* of the Gaussian vector X.

The definition is correct because this set does not depend on the choice of the space L [see Lifshits (1995, Th. 9.4.)]. S_X is also called the *ellipsoid of concentration* of the distribution P_X of X.

There is an alternative approach to describing of Strassen balls based on the concept of reproducing kernel spaces [see Aronszajn (1950) and Vakhaniya, Tarieladze and Chobanyan (1985)] or in terms of abstract Wiener space [see Kuelbs and Le Page (1973)].

Consider now

$$X = \{X_t, t \in [0,1]\},\ \ X_t = (X_1(t), \ldots, X_d(t))^T,$$

a zero mean d-dimensional Gaussian process with continuous sample paths. Obviously, X is an element of the space $B = C^d[0,1]$. Then the covariance function of X is a matrix function

$$K(t,s) = (K_{i,j}(t,s))_{i,j=1}^d$$

where $K_{ij}(t,s) = EX_i(t)X_j(s)$.

Let L be some Hilbert space, L^d be the space of vectors $\ell = (\ell_1, \ldots, \ell_d)$, $\ell_i \in L$, $i \leq d$, $M^d(L)$ be the space of $d \times d$-matrix with the elements from L. Assume that for $G_1, G_2 \in M^d(L)$ and $\ell \in L^d$, the products G_1G_2 and $G_1\ell^T$ are defined as the usual matrix product provided the products of their elements $\ell_1, \ell_2 \in L$ equal $(\ell_1, \ell_2)_L$. In this case, we write simply $\ell_1\ell_2$. Obviously, L^d with the inner product $(\ell, k)_{L^d} = \ell k^T$ is a Hilbert space.

Definition 24.4.2 The set $\{G_t \mid G_t \in M^d(L), t \in [0,1]\}$ is called the *model* of the process X iff $K(t,s) = G_tG_s^T$ for $t, s \in [0,1]$.

If $d = 1$, then it is the usual model of the real Gaussian process X [see Lifshits (1995, §9)].

The notion of model is crucial for the constructive description of Strassen balls.

Proposition 24.4.1 *Let $\{G_t, t \in [0,1]\}$ be a model of X, S_X be the corresponding Strassen ball. Then,*

$$S_X = \{h \mid h \in C^d[0,1],\ \ h(t) = G_t\ell^T,\ \ \ell \in L^d,\ \ (\ell,\ell)_{L^d} \leq 1\},$$

where $L = L^2[0,1]$.

PROOF. It is clear that the dual space B^* can be interpreted as vector space with elements $\mu = (\mu_1, \mu_2, \ldots, \mu_d)$, where μ_i, $i \le d$, are finite charges. Thus if $x \in B = C^d[0,1]$, $\mu \in B^*$, then

$$\langle x, \mu \rangle = \int_0^1 \mu(dt) x(t) = \sum_{i=1}^{d} \int_0^1 x_i(t) \mu_i(dt). \tag{24.16}$$

Define the operators $\mathcal{I}^*$ and $\mathcal{I}$ by the formulas

$$\mathcal{I}^* :: B^* \to L^d, \; \mathcal{I}^* \mu = \int_0^1 \mu(dt) G_t = (\langle G_{t,1}, \mu \rangle, \langle G_{t,2}, \mu \rangle, \ldots, \langle G_{t,d}, \mu \rangle)$$

where $G_{t,i}$, $i \le d$, are the columns of G_t,

$$\mathcal{I} : L^d \to B, \quad \mathcal{I}\ell = G_t \ell^T.$$

It is not difficult to check that the operators $\mathcal{I}$ and $\mathcal{I}^*$ are adjoint. Furthermore,

$$\begin{aligned} \mathcal{I}(\mathcal{I}^* \mu)(t) &= G_t (\mathcal{I}^* \mu)^T = G_t \left(\int_0^1 \mu(ds) G_s \right)^T = \int_0^1 G_t G_s^T \mu^T(ds) \\ &= \int_0^1 K(t,s) \mu^T(ds) = E(\int_0^1 X(t) X^T(s) \mu^T(ds)). \end{aligned} \tag{24.17}$$

Then for $\mu, \; \nu \in B^*$,

$$\langle \mathcal{I}\mathcal{I}^* \mu, \nu \rangle = E(\int_0^1 \nu(dt) \int_0^1 X(t) X^T(s) \mu(ds)) = E(\langle X, \mu \rangle \langle X, \nu \rangle. \tag{24.18}$$

Hence $K = \mathcal{I}\mathcal{I}^*$. ■

Corollary 24.4.1 *Assume that*

$$X(t) = \int_0^1 g_t d\bar{W}, \tag{24.19}$$

where $\bar{W} = (W_1, \ldots, W_d)^T$ *is a standard d-dimensional Wiener process,* $g_t \in M^d(L)$, $L = L^2[0,1]$. *Then,* g_t *is a model of the process* X *and*

$$S_X = \{h \mid h \in B, \; h(t) = g_t \ell^T, \;\; \ell \in L^d, \;\; (\ell, \ell)_{L^d} \le 1\}.$$

Corollary 24.4.2 *Assume that*

$$X(t) = \int_0^1 r_t dW = \int_0^1 g_t d\bar{W}, \tag{24.20}$$

where W *is the one-dimensional Wiener process,* $r_t^T \in L^d$, $L = L^2[0,1]$, $r_t dw = (r_{t,1} dw, \ldots, r_{t,d} dw)^T$, $g_t = (g_{ij}(t,\cdot))_{i,j=1}^d$, $g_{ij}(t,\cdot) = r_i(t,\cdot)$ for $j = 1$, $g_{ij}(t,\cdot) = 0$ for $j = 2, \ldots, d$, $W(t) = W_1(t)$, *where* $W_1(t)$ *is the first component of the standard d-dimensional Wiener process* $\bar{W}(t) = (W_1(t), \ldots, W_d(t))^T$. *Then,* g_t *is a model of the process* X *and*

$$S_X = \{h \in B \mid h(t) = (h_1(t), \ldots, h_d(t))^T, h_i(t) = (r_{t,i}, \ell)_L, \; \ell \in L, \; (\ell, \ell)_L \le 1\}, \tag{24.21}$$

where $L = L^2[0,1]$, $r_t = (r_{t,1}, \ldots, r_{t,d})^T$.

Corollary 24.4.3 *Assume that*

$$X^T = \{(X_1^T(t), X_2^T(t)), t \in [0,1])\},$$

where X_1 and X_2 are independent random Gaussian processes, $X_1(t) \in R^{d_1}$, $X_2(t) \in R^{d_2}$. Let g_t be a model for X_1, h_t be a model for X_2. Then

$$\begin{pmatrix} g_t & 0 \\ 0 & h_t \end{pmatrix}$$

is a model for X.

Corollary 24.4.4 *Let $X^T = (X_1^T(t), X_2^T(t))$, where X_1, X_2 are independent d-dimensional random processes. Let $Y(t) = X_1(t) + X_2(t)$, g_t, h_t are models in $M^d(L)$ of X_1 and X_2. Suppose that $g_{i,j}(t,\cdot)$ and $h_{k,\ell}(s,\cdot)$ are orthogonal for any $t \neq s$, $i,j,k,\ell \leq d$. Then $g_t + h_t$ is a model of Y.*

Corollary 24.4.5 *Provided the hypotheses of Corollary 24.4.4 are valid, except of the orthogonality hypothesis, we have*

$$S_Y = \{h \mid h \in B, h = g_t\ell^T + h_t\tilde{\ell}^T,\ \ell, \tilde{\ell} \in L^d,\ (\ell,\ell)_{L^d} + (\tilde{\ell},\tilde{\ell})_{L^d} \leq 1\},$$

where $L = L^2[0,1]$.

Now we can give the representation of the Strassen balls for the processes $\tilde{\alpha}, \tilde{\eta}, \tilde{\xi}$ [see (24.5) and (24.7)].

Proposition 24.4.2 *There are the following Strassen balls representations:*

$$S_{\tilde{\alpha}} = \{h \in B \mid h(t) = \sigma_t\ell^T,\ \ell \in L^d, (\ell,\ell)_{L^d} \leq 1\} \tag{24.22}$$

where $\sigma_{t,i,j}(s) = \sigma_{i,j}(s)\mathbf{1}_{[0,t)}(s)$, $i,j = 1,\ldots,d$ and the product of the d-dimensional vector and the scalar is the componentwise inner product;

$$S_{\tilde{\eta}} = \{h \in B \mid h(t) = \sigma_t\ell^T + g(t)\tilde{\ell},\ \ell \in L^d, \tilde{\ell} \in L,\ (\ell,\ell)_{L^d} + (\tilde{\ell},\tilde{\ell})_L \leq 1\}, \tag{24.23}$$

where $g(t) = (g_1(t),\ldots,g_d(t))^T$, $g_i(t,s) = m_i(s)\mathbf{1}_{[0,t)}(s) - \int_0^t m_i(s)ds$, $i \leq d$;

$$S_{\tilde{\xi}} = \{h \in B \mid h(t) = \sigma(t)\ell^T + r(t)\tilde{\ell},\ \ell \in L^d, \tilde{\ell} \in L,\ (\ell,\ell)_{L^d} + (\tilde{\ell},\tilde{\ell})_L) \leq 1\}, \tag{24.24}$$

where $r(t) = (r_1(t),\ldots,r_d(t))^T$, $r_i(t,s) = (m_i(s) - m_i(t))\mathbf{1}_{[0,1)}(s) - \int_0^t m_i(s)ds + tm_i(t)$, $i \leq d$.

PROOF. (24.22) follows immediately from Corollary 24.4.1, (24.23) follows from Corollaries 24.4.2 and 24.4.5 and the representations

$$V(s) \stackrel{d}{=} W(s) - sW(1),$$

$$\int_0^t m_i(s)dV(s) \stackrel{d}{=} \int_0^1 (m_i(s)\mathbf{1}_{[0,t)}(s) - \int_0^t m_i(u)du)dW(s).$$

The proof of (24.24) is the same as that of (24.23). ■

Remark 24.4.1 Without loss of generality, we can suppose in (24.23) and (24.24) that $\int_0^1 \tilde{\ell}(s)ds = 0$. In fact, if $\tilde{\ell}^1(s) = \tilde{\ell}(s) - \int_0^1 \tilde{\ell}(s)ds$, then $(\tilde{\ell}^1, \tilde{\ell}^1) \leq 1$. If we replace $\tilde{\ell}(s)$ by $\tilde{\ell}^1(s)$, then h does not change and we can then put

$$g_i(t,s) = m_i(s)\mathbf{1}_{[0,t)}(s), \;\; r_i(t,s) = (m_i(s) - m_i(t))\mathbf{1}_{[0,t)}(s).$$

24.5 Law of the Iterated Logarithm

In this Section we give a Strassen type law of the iterated logarithm for the processes defined by IOS.

Definition 24.5.1 Let $\{X_n\}$, $X_n = (X_n(t), t \in [0,1])$, be a sequence of random processes with sample paths from a Banach functional space B. We say that this sequence *approximates* a function set $S \subset B$ and write $X_n \rightsquigarrow S$, iff with the probability one the set $\{X_n(\cdot)\}$ is relatively compact and the set of its limit points coincides with S.

Definition 24.5.2 Let $\{X_n\}$, $X_n = (X_n(t), t \in [0,1])$, be a sequence of centered random processes such that $X_n/\sqrt{n}$ converges weakly (in uniform topology) to a Gaussian process X. We say that the *functional law of the iterated logarithm* (FLIL) is valid for $\{X_n/a_n\}$ iff the sequence $\{X_n/a_n\}$, $a_n = \sqrt{2n\log\log n}$, approximates Strassen ball S_X of the process X.

Theorem 24.5.1 *Assume that the hypotheses of Theorem 24.3.1 are satisfied. Then the FLIL for the processes $\tilde{\alpha}_n, \tilde{\eta}_n, \tilde{\xi}_n$ are valid. The description of the corresponding Strassen balls $S_{\tilde{\alpha}}, S_{\tilde{\xi}}, S_{\tilde{\eta}}$ are given in Section 24.4 (Proposition 24.4.2).*

The proof of this theorem is based on Lemmas 24.5.1–24.5.6 below, on Theorem 4.2 of Ossiander (1988). It is long and tedious and therefore will be omitted here.

Lemma 24.5.1 *Let the conditions of Lemma 24.3.1 be valid. Set*

$$b_n = \sqrt{n/(2\log\log n)}.$$

Then

(a)
$$b_n m U_n \rightsquigarrow mK,$$

where

$$K = \{f\colon f(0) = f(1) = 0, f \;\; \textit{is absolute continuous}, \; \int_0^1 (f'(t))^2 dt \leq 1\},$$

$$mK = \{h\colon h(t) = m(t)f(t), f \in K\}.$$

(b) $\limsup_{n\to\infty} \sup_{t\in[0,\gamma]\cup[1-\gamma,1]} b_n|m(t)U_n(t)| \to 0$ *a.s., as* $\gamma \downarrow 0$.

PROOF OF LEMMA 24.5.1. First note that $m(t)f(t)$ is continuous for any $f \in K$, $t \in [0,1]$. It follows from the inequality $|f(t)| \le \sqrt{2t(1-t)}$, which is true for any $f \in K$.

(a) follows from James (1975) or Shorack and Wellner (1986 p. 517).

(b) We prove the statement (b) only for $t \in [0,\gamma]$ because for $t \in [1-\gamma,1]$ the proof is similar. Due to (a) and Shorack and Wellner (1986 p. 517), with probability one

$$\limsup_{n\to\infty} b_n\|mU_n\|_0^\gamma = \sup_{f\in K} \|mf\|_0^\gamma \le C \sup_{f\in K} \|t^{-a}f\|_0^\gamma \le C\gamma^{1/2-a} \to 0$$

as $\gamma \to 0$. ■

Lemma 24.5.2 *If*

$$\beta_n \rightsquigarrow K,$$

where K is a set of equicontinuous functions and the processes $\{\tau_n(t), t \in [0,1]\}$ are such that $\tau_n(t) \in [0,1]$ for any $t \in [0,1]$ and $\|\tau_n - I\|_0^1 \to 0$ with probability one, then

$$\|\beta_n(\tau_n(t)) - \beta_n(t)\|_0^1 \to 0 \ \ a.s.$$

Lemma 24.5.3 *Set $t_n = n^{-\beta}$, where $0 < \beta < 1$. Then*

$$X_{n,[nt_n]} = O(t_n) \ a.s. \ (n \to \infty).$$

PROOF OF LEMMA 24.5.3. Set $n_k = 2^k$, $m_k = [4n_k t_{n_k}]$. It is easy to check the inequality

$$D(X_{n_k,m_k}) \le \frac{m_k}{n_k^2} \le Cn_k^{-\beta-1}.$$

Chebyshev inequality therefore implies

$$\sum_{k=1}^{\infty} P(\max_{n_k\le n\le n_{k+1}} X_{n,[nt_n]} > 5t_{n_k}) \le \sum_{k=1}^{\infty} P(X_{n_k,m_k} > 5t_{n_k}) \le C\sum_{k=1}^{\infty} n_k^{\beta-1} \le \infty.$$

Lemma follows from these relations and the Borel-Cantelli lemma. ■

Lemma 24.5.4 *If $\gamma \in [0,1]$, $0 < y < x$, then*

$$|x^\gamma - y^\gamma| \le \frac{2(x-y)}{x^{1-\gamma} + y^{1-\gamma}}.$$

Lemma 24.5.5

(a) If $\beta_n(t) = \beta_n(t,\omega) \xrightarrow{P} 0$, $n \to \infty$ *for all positive* t, *then there exists a real sequence* $t_n \downarrow 0$ *such that* $\beta_n(t_n) \xrightarrow{P} 0$.

(b) If for all positive t $\beta_n(t) \to 0$ *with probability one, then there exists a sequence* $t_n = t_n(\omega)$ *such that* $\beta_n(t_n) \to 0$ *with probability one.*

(c) Assume that $\beta_n(t)$ *are deterministic functions such that* $\beta_n(t) \geq 0$, $\beta_n(t) \downarrow 0$ *for every* n *as* $t \downarrow 0$ *and* $\limsup_{n\to\infty} \beta_n(t) \leq \varphi(t)$ *where* $\varphi(t) \downarrow 0$, $t \downarrow 0$. *Then* $\beta_n(t_n) \to 0$ *as* $t_n \to 0$.

Lemma 24.5.6 *Let* B *be a Banach space. Let* $\{\rho_n\}$, $\{\kappa_n\}$ *and for any* $N \in \mathbf{N}$ $\{\rho_n^N\}$, $\{\kappa_n^N\}$ *be sequences of* B*-valued random vectors such that for some compact subsets* S, S^N *of* B *with the probability 1*

$$\kappa_n \rightsquigarrow S, \quad \rho_n^N \rightsquigarrow S^N, \quad \kappa_n^N \rightsquigarrow S^N, \tag{24.25}$$

$$\lim_{N\to\infty} \limsup_{n\to\infty} ||\rho_n - \rho_n^N|| = 0, \quad \lim_{N\to\infty} \limsup_{n\to\infty} ||\kappa_n - \kappa_n^N|| = 0. \tag{24.26}$$

Then,

$$\rho_n \rightsquigarrow S. \tag{24.27}$$

24.6 Applications

Statistical applications of our results involve the applications described by Bhattacharya (1974, 1976). He considered:

(1) tests for the regression function $m(x) = m_0(x)$,

(2) estimation of the function $f(t) = \int_0^t m(s)ds = E(Y; X \leq t)$,

and

(3) constructions of confidence sets for it.

Note that our results justify better these applications due to the less restrictive conditions. For example, for the processes $\tilde{\xi}_n$ we can not apply Bhattacharya theorems for a normal vector Z, with the exception of the case when X and Y are independent, since $m(x)$ is then not bounded.

Besides, by virtue of Theorem 24.5.1, the random processes ξ_n and η_n are consistent estimates with probability one of the function f, and the deviations from f have order $\sqrt{\log\log n/n}$. Due to Theorems 24.3.1 and 24.5.1, we can get similar results for multivariate vectors Y.

It is not difficult to extend the above mentioned applications to censored samples. Consider, for example, the test for regression function. Suppose that $d = 1$ and for a constant T, $T \leq 1$

$$Y_{n,1}, Y_{n,2}, \ldots, Y_{n,[nT]},$$

is a set of induced order statistics.

If we test the hypothesis $m(s) = m_0(s)$ against the general alternative, we can construct the test of the form

$$\left\{ \max_{0\le t\le T} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{[nt]} (Y_{n,i} - m_0(X_{n,i})) \right| > \lambda \right\} \stackrel{def}{=} A(\lambda).$$

Let $\Psi(t) = \int_0^t \sigma^2(s)ds$. Then, similar to Bhattacharya (1974), we have under the hypothesis

$$P(A(\lambda)) \approx P\{\max_{0\le t\le 1} |W(t)| > (\Psi(T))^{-1/2}\lambda\}.$$

Here, we can use instead of $\Psi(T)$ its estimate

$$\lambda_n(T) = \frac{1}{n} \sum_{i=1}^{[nT]} (Y_{n,i} - m_0(X_{n,i}))^2 \xrightarrow{P} \psi(T). \tag{24.28}$$

To prove the relation (24.28), set $\tilde{Y_i} = Y_i - m(X_i)$ and note that due to the law of large numbers

$$\beta_n(T) = \frac{1}{n} \sum_{i=1}^{n} \tilde{Y}_n^2 \mathbf{1}_{[0,T)}(X_i) \to E(\tilde{Y}_1^2 \mathbf{1}_{[0,T)}(X_i)) = \int_0^T \sigma^2(s)ds.$$

Then for any $\delta > 0$, with the probability one

$$\limsup_{n\to\infty} |\lambda_n(T) - \beta_n(T)| = \limsup_{n\to\infty} |\beta_n(X_{n,[nT]}) - \beta_n(T)|$$
$$\le \limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \tilde{Y}_i^2 \mathbf{1}_{[T-\delta,T+\delta]}(X_i) = \int_{T-\delta}^{T+\delta} \sigma^2(s)ds \to 0, \;\; \delta \to 0.$$

Hence, (24.28) is valid.

The convexification of random walks is another domain of applications for induced order statistics. Let $(Y_i, i \in \mathbf{N})$ be i.i.d. random variables in $\mathbf{R}^2$. Let (ρ_i, θ_i) be the polar coordinates of Y_i. We denote $\theta_{n1}, \ldots, \theta_{nn}$ for the values θ_i rearranged by growth and consider the polygonal continuous line L_n defined by the points $S_0^* = 0$, $S_k^* = \sum_{i=1}^k Y_{n,i}$, $k = 1, \ldots, n$. This line has the same origin and end as the initial random walk $S_0 = 0$, $S_k = \sum_{i=1}^k Y_i$, $k = 1, \ldots, n$, and represents exactly what one could call the *convexification* of the random walk (S_k). The asymptotic behavior of convexifications L_n and geometrical properties of their limits have been studied by Davydov and Vershik (1998).

It is clear that $Y_{n,i}$ are the order statistics induced by θ_s's and the processes L_n are analogues to processes ξ_n. Thus, this construction is a very particular case of one studied in this Chapter. For example, Theorem 5 of Davydov and Vershik (1998) is a direct consequence of our Theorem 24.3.1. On the other hand, the geometrical interpretation and the relations with Minkowski addition of convex sets used by Davydov and Vershik (1998) could be very useful for subsequent investigation of induced-order statistics when the second-order moment does not exist. In Davydov, Paulauskas and Rachkauskas (1998), this geometric construction is applied to the estimation of parameters of multivariate p-stable laws with $p < 1$.

Acknowledgments. This work was partially supported by Russian Foundation of Fundamental Research, grants N. 96-01-00549, N. 99-01-00724 and N. 96-01-0672. The authors express their gratitude to Professor Ya. Nikitin for the careful reading of the manuscript and valuable comments.

References

1. Aronszajn, N. (1950). Theory of reproducing kernels, *Transactions of the American Mathematical Society*, **68**, 337–404.

2. Balakrishnan, N. (1993). Multivariate normal distribution and multivariate order statistics induced by ordering linear combinations, *Statistics & Probability Letters*, **17**, 343–350.

3. Balakrishnan, N. and Cohen, A. C. (1991). *Order Statistics and Inference: Estimation Methods*, Boston: Academic Press.

4. Barnett, V. (1976). The ordering of multivariate data, *Journal of the Royal Statistical Society, Series A*, **139**, 318–354.

5. Barnett, V., Green, P. J. and Robinson, A. (1976). Concomitants and correlation estimates, *Biometrika*, **63**, 323–328.

6. Bhattacharya, P. K. (1974). Convergence of sample paths of normalized sums of induced order statistics, *The Annals of Statistics*, **2**, 1034–1039.

7. Bhattacharya, P. K. (1976). An invariance principle in regression analysis, *The Annals of Statistics*, **4,** 621–624.

8. Bhattacharya, P. K. (1984). Induced order statistics: Theory and applications, In *Handbook of Statistics* (Eds., P. R. Krishnaiah and P. K. Sen), **4**, pp. 383–403, Amsterdam: North-Holland.

9. Bickel, P. J. (1967). Some contributions to the theory of order statistics, In *Proceedings of the Fifth Berkeley Symposium on Mathematics, Statistics and Probability*, **1**, pp. 575–591, Berkeley, California: University of California Press.

10. Billingsley, P. (1968). *Convergence of Probability Measures*, New York: John Wiley & Sons.

11. David, H. A. (1973). Concomitants of order statistics, *Bulletin of the International Statististical Institute*, **45**, 295–300.

12. David, H. A. (1981). *Order Statistics*, Second edition, New York: John Wiley & Sons.

13. David, H. A. (1992). Concomitants of order statistics: Review and recent developments,

14. David, H. A. and Galambos, J. (1974). The asymptotic theory of concomitants of order statistics, *Journal of Applied Probability*, **11**, 762–770.

15. David, H. A. and Nagaraja, H. N. (1998).

16. David, H. A., O'Connell, M. J. and Yang, S. S. (1977). Distribution and expected value of the rank of a concomitant of an order statistic, *The Annals of Statistics*, **5**, 216–223.

17. Davydov, Yu. and Vershik, A. M. (1998). Réarrangements convexes des marches aléatoires, *Annales Inst. Henri Poincaré,* **34**, 73-95.

18. Davydov, Yu., Paulauskas, V. and Rachkauskas, A. (1998). More on p-stable random convex compact sets in Banach spaces, $0 < p < 1$, *Prepublications IRMA, Lille,* **45**.

19. Egorov, V. A. and Nevzorov, V. B. (1982). Some theorems on induced order statistics, *Theory of Probability and Its Applications*, **27**, 633–639.

20. Egorov, V. A. and Nevzorov, V. B. (1984). Convergence to the normal law of sums of induced order statistics, *Journal of Soviet Mathematics*, **25**, 1139–1146.

21. Galambos, J. (1987). *The Asymptotic Theory of Extreme Order Statistics,* Second edition, Malabar, Florida: Krieger.

22. Gomes, M. I. (1981). An i-dimensional limiting distribution function of largest values and its relevance to the statistical theory of extremes, In *Statistical Distribution in Scientific Work* (Eds., C. Taillie *et al.*), **6**, pp. 389–410, Dordrecht: Reidel.

23. James, B. (1975). A functional law of the iterated logarithm for weighted empirical distributions, *Annals of Probability*, **3**, 762–772.

24. Harter, H. L. (1961). Expected values of normal order statistics, *Biometrika*, **48**, 151–165. [Correction, *Biometrika*, **48**, 476.]

25. Harter, H. L. (1970). *Order Statistics and Their Use in Testing and Estimation,* Vols. **1** and **2**, Washington, DC: U.S. Government Printing Office.

26. Harter, H. L. and Balakrishnan, N. (1996). *CRC Handbook of Tables for Order Statistics and* , Boca Raton, Florida: CRC Press.

27. Kaminsky, K. S. (1981). A note on concomitants of order statistics in two bivariate normal samples, *43rd Session of International Statistical Institute at Buenos Aires,* **I**, 161–164.

28. Kuelbs, J. and Le Page, R. (1973). The law of the iterated logarithm for Brownian motion in a Banach space, *Transactions of American Mathematical Society*, **185**, 253–264.

29. Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces*, New York: Springer–Verlag.

30. Lifshits, M. A. (1995). *Gaussian Random Functions*, Amsterdam, The Netherlands: Kluwer.

31. Ossiander, M. (1987). A Central limit theorem under metric entropy with L_2 bracketing, *The Annals of Probability*, **15**, 897–919.

32. O'Reilly, N. (1974). On the weak convergence of empirical processes in sup-norm metrics, *The Annals of Probability*, **2**, 642–651.

33. Sen, P. K. (1976). A note on invariance principles for induced order statistics, *The Annals of Probability*, **4**, 474–479.

34. Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*, New York: John Wiley & Sons.

35. Vakhaniya, N. N., Tarieladze, V. I. and Chobanyan, S.A. (1985). *Probability Distributions in Banach Spaces*, Moscow: Nauka (in Russian).

36. Yang, S. S. (1977). General distribution theory of the concomitants of order statistics, *The Annals of Statistics*, **5**, 996–1002.

25

Notes on the KMT Brownian Bridge Approximation to the Uniform Empirical Process

David M. Mason

University of Delaware, Newark, Delaware

Abstract: Komlós, Major, and Tusnády [KMT] (1975) published a Brownian bridge approximation to the uniform empirical process. However, the majority of the often very technical details of the proof were left to the reader. This has sometimes discouraged the acceptance and informed use of this very powerful approximation tool. The aim of these notes is to gain a wider audience for this beautiful result by making its proof more accessible. This is done by providing the details of the proof and pointing the reader to published work where they can be found.

Keywords and phrases: Uniform empirical process, Brownian bridge, quantile function, exponential inequalities

25.1 Introduction

In two remarkable papers, Komlós, Major, and Tusnády [KMT] (1975, 1976) established powerful and nearly unimprovable results on the strong approximation of the partial sum process by a Wiener process via a diadic construction scheme. In the process, they also sketched how their methods yield strong approximations to the uniform empirical process by a Kiefer process and a Brownian bridge. Of most interest to statisticians has been their Brownian bridge approximation.

To formulate this result, let U, U_1, $U_2, \ldots$, be independent uniform $(0,1)$ random variables. For each integer $n \geq 1$, let

$$G_n(t) = n^{-1} \sum_{i=1}^{n} 1\{U_i < t\}, \text{ for } t \in \mathbb{R}, \tag{25.1}$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

denote the *empirical distribution function* based on $U_1, ..., U_n$, and

$$\alpha_n(t) = \sqrt{n}\{G_n(t) - t\},\ 0 \le t \le 1, \tag{25.2}$$

be the corresponding *uniform empirical process.* KMT (1975) stated the following Brownian bridge approximation to α_n, along with a brief description (less than three pages) of its proof.

Theorem 25.1.1 *There exists a probability space $(\Omega, \mathcal{A}, P)$ with independent uniform $(0,1)$ random variables $U_1, U_2, \ldots,$ and a sequence of Brownian bridges $B_1, B_2, ...,$ such that for all $n \ge 1$ and $x \in \mathbb{R}$*

$$P\left\{\sup_{0\le t\le 1} |\alpha_n(t) - B_n(t)| \ge n^{-1/2}(a \log n + x)\right\} \le b \exp(-cx), \tag{25.3}$$

where a, b and c are suitable positive constants.

Mason and van Zwet (1987) obtained the following refinement to Theorem 25.1.1 and in doing so published many of the missing details of the proof for the original approximation.

Theorem 25.1.2 *On the probability space of Theorem 25.1.1, for all $n \ge 1$, $1 \le d \le n$ and $x \in \mathbb{R}$,*

$$P\left\{\sup_{0\le t\le d/n} |\alpha_n(t) - B_n(t)| \ge n^{-1/2}(a \log d + x)\right\} \le b \exp(-cx), \tag{25.4}$$

with the same inequality holding when the supremum is taken over $[1 - d/n, 1]$ and where the constants a, b and c are as in Theorem 25.1.1.

[Bretagnolle and Massart (1989) and Rio (1994) have been able to provide specific values for the constants a, b and c.]

The main impetus behind Mason and van Zwet's work was to obtain the following weighted approximation result. For any $0 \le \nu < 1/2$ and $n \ge 2$, let

$$\Delta_{n,\nu} := \sup_{1/n\le t\le 1-1/n} \frac{n^\nu |\alpha_n(t) - B_n(t)|}{(t(1-t))^{1/2-\nu}}. \tag{25.5}$$

Using their inequality one can readily verifly that on the probability space of Theorem 25.1.1 one has

$$\Delta_{n,\nu} = O_p(1). \tag{25.6}$$

[Mason (1998) has recently shown that Theorem 25.1.2 leads to an exponential inequality for $\Delta_{n,\nu}$.]

Weighted approximations have proved to be very useful in establishing asymptotic distribution results in probability theory and statistics; see, for example, Part II of the proceedings volume edited by Hahn, Mason and Weiner (1991) and the monograph by M. Csörgő and Horváth (1993), along with the many references therein.

KMT (1975) described the diadic scheme which leads to the construction of the probability space of Theorem 25.1.1 and stated the basic quantile approximation result for binomial random variables upon which a proof of inequality (25.3) can be based. It was this approach that Mason and van Zwet (1987) followed to achieve their refined inequality (25.4).

Another approach was described in the Tusnády (1977) dissertation using an alternate quantile approximation especially tailored for binomial random variables. This is the so-called Tusnády lemma, cf. Lemma 4.4.2 of Csörgő and Révész (1981) and Lemma 4 of Bretagnolle and Massart (1989). It was through the Tusnády lemma that Bretagnolle and Massart (1989) and Rio (1994) obtained their results.

To state the two basic binomial quantile approximations, we need to introduce some notation. For any integer $m \geq 1$, let Y_m denote a random variable having the binomial distribution with parameters m and $1/2$, written $B(m, 1/2)$. Set for $x \in \mathbb{R}$

$$P_m(x) = P\{Y_m < x\}$$

and denote the *inverse distribution function* or *quantile function* of P_m by

$$H_m(s) = \sup\{x : P_m(x) \leq s\} \text{ for } 0 \leq s < 1,$$

and $H_m(1) = H_m(1-)$. Notice, in particular, that $H_m(s) = 0$ for $0 \leq s < P_m(1)$, and i for $P_m(i) \leq s < P_m(i+1)$, $i = 0, ..., m$.

Let Z denote a standard normal random variable, Φ be its distribution function and ϕ its density function. Since $\Phi(Z) =_d U$, we see that for each $m \geq 1$

$$H_m(\Phi(Z)) =_d Y_m.$$

For this reason, we will from now on write for convenience

$$H_m(\Phi(Z)) = Y_m \text{ and } S_m = Y_m - m/2.$$

KMT (1975) stated towards the bottom of page 130 of their paper the basic quantile result will lead to their Brownian bridge approximation to the uniform empirical process. We cite it here as Proposition 25.1.1.

Proposition 25.1.1 *There exist a $0 < C < \infty$ and an $0 < \varepsilon < \infty$ such that for all integers $m \geq 1$, whenever*

$$|S_m| \leq \varepsilon m, \tag{25.7}$$

we have

$$\left| S_m - \frac{\sqrt{m}Z}{2} \right| \leq \frac{CS_m^2}{m} + \frac{C}{4}. \tag{25.8}$$

Next, for comparison, we state the Tusnády (1977) lemma as Proposition 25.1.2.

Proposition 25.1.2 *For all integers $m \geq 1$*

$$|S_m| \leq 1 + \frac{\sqrt{m}|Z|}{2} \tag{25.9}$$

and

$$\left| S_m - \frac{\sqrt{m}Z}{2} \right| \leq 1 + \frac{Z^2}{8}. \tag{25.10}$$

Tusnády (1977) did not provide a fully detailed proof of his lemma. In fact, Csörgő and Révèsz (1981) remark in their monograph on strong approximations, "Although the proof of the inequality is elementary, it is not simple. It will not be given here however." When Bretagnolle and Massart (1989) published a proof of the Tusnády lemma, it indeed was not simple. [See Lemma 3.1.5 of M. Csörgő and Horváth (1993) for an extended version of their proof.]

KMT (1975) remark that the result stated in Proposition 25.1.1 can be deduced from their more general Lemma 1, which was designed for the purpose of obtaining their Wiener process strong approximation to the partial sum process. We will see in Section 25.2 that if one is only interested in the specific result stated in Proposition 25.1.1 for the $B(m, 1/2)$ quantile function, that it follows rather straightforwardly from an extended version of a large deviation theorem of Petrov (1972), along with some routine bounds on the tail of the standard normal distribution function.

In Section 25.2, we provide a detailed proof of Proposition 25.1.1, which in our opinion, is much easier than that of the Tusnády lemma. Then in Section 25.3, we describe the basic diadic scheme which leads to the probability space of the Theorem 25.1.1 and prove the main lemma that connects Proposition 25.1.1 to the diadic scheme. Next, in Section 25.4, we present proofs of some combinatorial results needed to fill in certain details in the Mason and van Zwet (1987) paper.

Since the KMT (1975) Brownian bridge approximation continues to have wide ranging applications in statistics and probability, it is essential that an accessible and detailed proof for this important result be available in the published literature. Bretagnolle and Massart (1989) follow the Tusnády (1977) path towards Theorem 25.1.1. The present notes, when combined with the Mason and van Zwet (1987) paper, provide a complete proof along the lines sketched in the original KMT (1975) paper, yielding the refinement of Theorem

25.1.2. In fact, these notes are best read in tandem with the Mason and van Zwet (1987) paper.

In writing these notes, the author benefited from the write-ups of the diadic scheme found in Delporte (1980) and Koltchinskii (1994). He also found Borisov (1978) useful as a guide as to what to do and the technical report of Einmahl (1986) helpful for some of the details. Much of Section 25.2 was adapted from the *Diplomarbeit* of Richter (1978). These notes were largely written while the author was visiting the University of Munich in 1984-1986 and to a lesser extent during his research stay at the University of Bielefeld in 1993–1994.

25.2 Proof of the KMT Quantile Inequality

In this Section, we prove Proposition 25.1.1. Let F_m denote the distribution function of S_m. Notice that since

$$\Phi^{-1}(F_m(S_m)) \leq Z \leq \Phi^{-1}(F_m(S_m+)),$$

to establish (25.8) subject to (25.7) it is enough to show that

$$\begin{aligned} -\frac{CS_m^2}{m} - \frac{C}{4} &\leq S_m - \frac{\sqrt{m}}{2}\Phi^{-1}(F_m(S_m+)) \\ &\leq S_m - \frac{\sqrt{m}}{2}\Phi^{-1}(F_m(S_m)) \leq \frac{CS_m^2}{m} + \frac{C}{4} . \end{aligned}$$

Writing $mx/2 = S_m$, this last inequality becomes

$$\begin{aligned} -\frac{Cmx^2}{4} - \frac{C}{4} &\leq \frac{mx}{2} - \frac{\sqrt{m}}{2}\Phi^{-1}(F_m(\frac{mx}{2}+)) \\ &\leq \frac{mx}{2} - \frac{\sqrt{m}}{2}\Phi^{-1}(F_m(\frac{mx}{2})) \leq \frac{Cmx^2}{4} + \frac{C}{4} , \end{aligned}$$

or

$$\Phi(\sqrt{m}x + u) \geq F_m(\frac{mx}{2}+) \geq F_m(\frac{mx}{2}) \geq \Phi(\sqrt{m}x - u), \tag{25.11}$$

where

$$u = \frac{C\sqrt{m}x^2}{2} + \frac{C}{2\sqrt{m}} =: D\sqrt{m}x^2 + \frac{D}{\sqrt{m}}.$$

To finish the proof of Proposition 25.1.1, we will need the following special case of Theorem A of KMT (1975) (stated there without proof), which is an extension of a large deviation theorem of Petrov (1972). This result can be obtained by a modification of the proof of Theorem 8.1.1 of Ibragimov and Linnik (1971) or that of Theorem 1 in Chapter VIII of Petrov (1972). For a detailed proof, see Einmahl (1986).

Theorem 25.2.1 *There exists an* $0 < \eta < \infty$ *such that for all* $0 \le x \le \eta$

$$P\{S_m > \frac{mx}{2}\} = (1 - \Phi(\sqrt{m}x)) \exp\left(mx^3\lambda(x) + O((x + m^{-1/2})\right), \quad (25.12)$$

and

$$P\{S_m < -\frac{mx}{2}\} = \Phi(-\sqrt{m}x) \exp\left(-mx^3\lambda(-x) + O((x + m^{-1/2})\right), \quad (25.13)$$

where $\lambda(x)$ *is a power series in* x *with coefficients depending on the cumulants of* Y_1.

We also require a number of lemmas.

Lemma 25.2.1 *For any fixed integer* $m_0 \ge 1$, *there exist* $0 < \varepsilon_1 < \infty$ *and* $0 < D_1 < \infty$ *such that for all* $x \in [-\varepsilon_1, \varepsilon_1]$ *and* $1 \le m \le m_0$ *we have*

$$\Phi(\sqrt{m}x + u) \ge F_m(\frac{mx}{2}+) \ge F_m(\frac{mx}{2}) \ge \Phi(\sqrt{m}x - u), \quad (25.14)$$

where $u = D_1(\sqrt{m}x^2 + m^{-1/2})$.

PROOF. Choose $\varepsilon_1 = 1/(2m_0)$. We see that for $|x| \le 1/(2m_0)$ and $1 \le m \le m_0$,

$$F_m(mx/2) \ge P\{Y_m < -1/4 + m/2\} \ge P\{Y_m = 0\} = 2^{-m} \ge 2^{-m_0}$$

and

$$\begin{aligned} F_m(mx/2+) &\le P\{Y_m \le 1/4 + m/2\} = P\{Y_m \le m/2\} \\ &= 1 - P\{Y_m > m/2\} \le 1 - P\{Y_m = 0\} = 1 - 2^{-m} \le 1 - 2^{-m_0}. \end{aligned}$$

Choose D_1 sufficiently large so that

$$\Phi(-\frac{1}{2\sqrt{m_0}} + \frac{D_1}{\sqrt{m_0}}) \ge 1 - 2^{-m_0} \quad \text{and} \quad \Phi(\frac{1}{2\sqrt{m_0}} - \frac{D_1}{\sqrt{m_0}}) \le 2^{-m_0}.$$

Since $|x| \le 1/(2m_0)$ and $1 \le m \le m_0$,

$$\Phi(\sqrt{m}x + D_1(\sqrt{m}x^2 + \frac{1}{\sqrt{m}})) \ge \Phi(-\frac{1}{2\sqrt{m_0}} + \frac{D_1}{\sqrt{m_0}})$$

and

$$\Phi(\sqrt{m}x - D_1(\sqrt{m}x^2 + \frac{1}{\sqrt{m}})) \le \Phi(\frac{1}{2\sqrt{m_0}} - \frac{D_1}{\sqrt{m_0}}),$$

we have (25.14). ■

Lemma 25.2.2 *For any* $x > 0$

$$\frac{1}{x}\left(1 - \frac{1}{x^2}\right) < \frac{1 - \Phi(x)}{\phi(x)} < \frac{1}{x}. \quad (25.15)$$

[This is the classic Mill's ratio. Refer, for instance, to Gänssler and Stute (1977).]

Lemma 25.2.3 *The function*

$$\Psi_1(x) := \frac{\phi(x)}{1-\Phi(x)} \text{ is increasing} \tag{25.16}$$

and the function

$$\Psi_2(x) := \frac{\phi(x)}{\Phi(x)} \text{ is decreasing.} \tag{25.17}$$

PROOF. First consider (25.16). We see that

$$\Psi_1'(x) = \Psi_1(x)(\Psi_1(x) - x),$$

which when $x \le 0$, is obviously > 0, and when $x > 0$ is positive by (25.15). Thus we have (25.16). Assertion (25.17) follows from the fact that $\Psi_1(x) := \phi(-x)/\Phi(-x) = \Psi_2(x)$. ■

Lemma 25.2.4 *For all $0 < A < \infty$, there exists an integer $m_0 \ge 1$, an $0 < \varepsilon_2 < \infty$ and a $0 < D_2 < \infty$ such that for all $0 \le x \le \varepsilon_2$ and $m \ge m_0$ we have*

$$\log\left(\frac{1-\Phi(\sqrt{m}x - u)}{1-\Phi(\sqrt{m}x)}\right) \ge A(mx^3 + x + m^{-1/2}), \tag{25.18}$$

where $u = D_2(\sqrt{m}x^2 + m^{-1/2})$.

PROOF. Notice that

$$\log\left(\frac{1-\Phi(\sqrt{m}x - u)}{1-\Phi(\sqrt{m}x)}\right) = \frac{u\phi(\xi)}{1-\Phi(\xi)}, \tag{25.19}$$

where $\xi \in [\sqrt{m}x - u, \sqrt{m}x]$.

First assume that $0 \le x \le m^{-1/2}$. Obviously for any $0 < D_2 < \infty$, we have $\xi \in [-2D_2/\sqrt{m}, 1]$. Thus by (25.16)

$$\frac{u\phi(\xi)}{1-\Phi(\xi)} \ge \frac{u\phi(-2D_2/\sqrt{m})}{1-\Phi(-2D_2/\sqrt{m})} = D_2(\sqrt{m}x^2 + m^{-1/2})\Psi_1(-2D_2/\sqrt{m}),$$

which, since $\sqrt{m}x \le 1$, is

$$\ge 2^{-1}D_2(\sqrt{m}x^2 + m^{-1/2} + x)\Psi_1(-2D_2/\sqrt{m}).$$

Now assume that $(4D_2)^{-1} \ge x \ge m^{-1/2}$. Since $x^2 m > 1$, we have

$$\sqrt{m}x - D_2(\sqrt{m}x^2 + m^{-1/2}) \ge \sqrt{m}x[1 - 2D_2x] \ge \sqrt{m}x/2. \tag{25.20}$$

Applying (25.15), we have

$$\frac{\phi(\sqrt{m}x-u)}{1-\Phi(\sqrt{m}x-u)} \geq \sqrt{m}x-u,$$

so using (25.16), (25.19) and (25.20) we get

$$\begin{aligned}\log\left(\frac{1-\Phi(\sqrt{m}x-u)}{1-\Phi(\sqrt{m}x)}\right) &\geq u(\sqrt{m}x-u)\\ &\geq u\sqrt{m}x/2 = 2^{-1}D_2(mx^3+x)\\ &\geq 4^{-1}D_2(mx^3+x+m^{-1/2}).\end{aligned}$$

Choose D_2 large enough so that $4^{-1}D_2 > A$ and $2^{-1}D_2\Psi_1(-1) > A$, and m large enough so that $-2D_2/\sqrt{m} > -1$. Hence with $\varepsilon_2 = 1/(4D_2)$, inequality (25.18) holds for all large enough m. ■

Lemma 25.2.5 *For all $0 < A < \infty$, there exists an integer $m_0 \geq 1$, an $0 < \varepsilon_2 < \infty$ and a $0 < D_2 < \infty$ such that for all $0 \leq x \leq \varepsilon_2$ and $m \geq m_0$ we have*

$$\log\left(\frac{1-\Phi(\sqrt{m}x+u)}{1-\Phi(\sqrt{m}x)}\right) \leq -A(mx^3+x+m^{-1/2}), \tag{25.21}$$

where $u = D_2(\sqrt{m}x^2+m^{-1/2})$.

PROOF. Let m_0, ε_2 and D_2 be chosen as in Lemma 25.2.4. Choose $0 \leq x \leq \varepsilon_2$ and $m \geq m_0$. We need only show that

$$\log\left(\frac{1-\Phi(\sqrt{m}x)}{1-\Phi(\sqrt{m}x+u)}\right) \geq \log\left(\frac{1-\Phi(\sqrt{m}x-u)}{1-\Phi(\sqrt{m}x)}\right). \tag{25.22}$$

Now

$$\log\left(\frac{1-\Phi(\sqrt{m}x-u)}{1-\Phi(\sqrt{m}x)}\right) = \frac{u\phi(\xi_1)}{1-\Phi(\xi_1)},$$

where $\xi_1 \in [\sqrt{m}x-u, \sqrt{m}x]$ and

$$\log\left(\frac{1-\Phi(\sqrt{m}x)}{1-\Phi(\sqrt{m}x+u)}\right) = \frac{u\phi(\xi_2)}{1-\Phi(\xi_2)},$$

where $\xi_2 \in [\sqrt{m}x, \sqrt{m}x+u]$.

Since $\xi_1 \leq \xi_2$, (25.22) follows from (25.16). Assertion (25.21) follows from (25.18). ■

Lemma 25.2.6 *For all $0 < A < \infty$, there exists an integer $m_0 \geq 1$, an $0 < \varepsilon_2 < \infty$ and a $0 < D_2 < \infty$ such that for all $0 \leq x \leq \varepsilon_2$ and $m \geq m_0$ we have*

$$\log\left(\frac{\Phi(-\sqrt{m}x+u)}{\Phi(-\sqrt{m}x)}\right) \geq A(mx^3+x+m^{-1/2}) \tag{25.23}$$

and

$$\log\left(\frac{\Phi(-\sqrt{m}x-u)}{\Phi(-\sqrt{m}x)}\right) \le -A(mx^3+x+m^{-1/2}), \tag{25.24}$$

where $u = D_2(\sqrt{m}x^2 + m^{-1/2})$.

PROOF. Assertions (25.23) and (25.24) follow from (25.18) and (25.21) and the facts that

$$\log\left(\frac{\Phi(-\sqrt{m}x+u)}{\Phi(-\sqrt{m}x)}\right) = \log\left(\frac{\Phi(\sqrt{m}x-u)}{1-\Phi(\sqrt{m}x)}\right)$$

and

$$\log\left(\frac{\Phi(-\sqrt{m}x-u)}{\Phi(-\sqrt{m}x)}\right) = \log\left(\frac{1-\Phi(\sqrt{m}x+u)}{1-\Phi(\sqrt{m}x)}\right).$$

■

We are now ready to complete the proof of Proposition 25.1.1. By Theorem 25.2.1 and Lemmas 25.2.4, 25.2.5 and 25.2.6, we can choose an integer $m_0 \ge 1$, an $0 < \varepsilon_2 < \infty$, an $0 < A < \infty$ and a $0 < D_2 < \infty$ such that for all $0 \le x \le \varepsilon_2$ and $m \ge m_0$ we have

$$F_n(-mx/2) \ge \Phi(-\sqrt{m}x)\exp(-A(mx^3+x+m^{-1/2})), \tag{25.25}$$

$$F_n(-mx/2+) \le \Phi(-\sqrt{m}x)\exp(A(mx^3+x+m^{-1/2})), \tag{25.26}$$

$$1-F_n(mx/2) \ge (1-\Phi(\sqrt{m}x))\exp(-A(mx^3+x+m^{-1/2})), \tag{25.27}$$

$$1-F_n(mx/2+) \le (1-\Phi(\sqrt{m}x))\exp(A(mx^3+x+m^{-1/2})), \tag{25.28}$$

and (25.18), (25.21), (25.23) and (25.24) hold, which imply that for all $0 \le x \le \varepsilon_2$ and $m \ge m_0$

$$\Phi(-\sqrt{m}x+u) \ge F_n(-mx/2+) \ge F_n(-mx/2) \ge \Phi(-\sqrt{m}x-u) \tag{25.29}$$

and

$$1-\Phi(\sqrt{m}x-u) \ge 1-F_m(mx/2) \ge 1-F_m(mx/2+) \ge 1-\Phi(\sqrt{m}x+u). \tag{25.30}$$

We easily see now from the preliminary discussion above that (25.8) holds with $C = 2D_2$ and $\varepsilon = \varepsilon_2/2$ as long as $m \ge m_0$. For $1 \le m \le m_0$, we apply Lemma 25.2.1 to obtain $0 < \varepsilon_1 < \infty$ and $0 < D_1 < \infty$ such that (25.14) holds, which in turn implies that (25.8) holds with $C = 2D_1$ and $\varepsilon = \varepsilon_1/2$. Letting $\varepsilon = \min\{\varepsilon_1/2, \varepsilon_2/2\}$ and $C = \max\{2D_1, 2D_2\}$, we see that (25.8) is valid for all $m \ge 1$ with this choice of C and ε. This completes the proof of Proposition 25.1.1. ■

Remark 25.2.1 Notice that the proof of Proposition 25.1.1 only uses the binomial assumption in Lemma 25.2.1. It is clear then that an extended version of Proposition 25.1.1 can be formulated for sums of i.i.d. nondegenerate random variables possessing a moment generating function finite in a neighborhood of zero. For versions of Proposition 25.1.1 under this more general assumption, where the inequality is asserted to hold for all large enough m, refer to Sakhanenko (1984) or Einmahl (1986).

Remark 25.2.2 Theorem 25.2.1 was only used to derive inequalities (25.25)–(25.28). Instead, we could have applied the somewhat less precise result given in Theorem 6.3.1 of Arak and Zaitsev (1988).

25.3 The Diadic Scheme

Let

$$\mathcal{Z} = \{Z\} \cup_{j=1}^{\infty} \{Z_i : i \in \{0,1\}^j\} \tag{25.31}$$

be an indexed set of independent standard normal random variables. As in the Introduction for each integer $m \geq 1$, let H_m denote the inverse distribution function of a $B(m, 1/2)$ random variable. Define $H_0 = 0$. Set

$$N_0 = H_n(\Phi(Z)) \text{ and } N_1 = n - H_n(\Phi(Z)).$$

Next for each $i \in \{0,1\}^j$, $j \geq 1$, let

$$N_{i,0} = H_{N_i}(\Phi(Z_i)) \text{ and } N_{i,1} = N_i - H_{N_i}(\Phi(Z_i)).$$

Notice that $n = N_0 + N_1$ and for each $i \in \{0,1\}^j$, $j \geq 1$, $N_i = N_{i,0} + N_{i,1}$; and N_0 is $B(1/2, n)$ and for each $i \in \{0,1\}^j, j \geq 1$, $N_{i,0}$ is $B(1/2, n_i)$ given $N_i = n_i$. This implies that for each $j \geq 1$

$$\{N_i : i \in \{0,1\}^j\} \text{ is multinomial } (2^{-j}, ..., 2^{-j}, n).$$

Given any set of standard normal random variables, we can construct for any $n \geq 1$ a uniform empirical distribution function G_n from this sequence of nested multinomials. Define for $j \geq 1$, $k = 1, ..., 2^j$,

$$G_n(\frac{k}{2^j}) = \frac{1}{n} \sum_{i \in A_{i,k}} N_i,$$

where

$$A_{i,k} = \{i \in \{0,1\}^j : \sum_{s=1}^{j} 2^{-s} i_s \leq k2^{-j} \text{ with } i = (i_1, ..., i_j)\}.$$

From this one readily constructs $G_n(t)$ for any $t \in [0,1]$ by taking limits. Next, one obtains the order statistics $U_{1,n} \le \cdots \le U_{n,n}$ of n independent uniform $(0,1)$ random variables by inverting G_n and then, as in KMT (1975), n independent uniform $(0,1)$ random variables $U_1, ..., U_n$ by taking a random permutation of the order statistics.

The trick is to do this in a way so that the corresponding empirical process is very close to a Brownian bridge. This is what KMT (1975) tell us how to do. This is accomplished by forming the set of standard normals $\mathcal{Z}$ through a fixed Brownian bridge B and verifying by means of a quantile inequality that the resulting processes are close with high probability.

To do this, we begin by setting $Z = 2B(1/2)$ and for $i \in \{0,1\}^j$, $j \ge 1$, we let

$$Z_i = 2^{j/2}\{2B(\frac{2k+1}{2^{j+1}}) - B(\frac{2k}{2^{j+1}}) - B(\frac{2(k+1)}{2^{j+1}})\}$$

where

$$k2^{-j} = \sum_{s=1}^{j} 2^{-s} i_s \text{ with } i = (i_1, ..., i_j). \tag{25.32}$$

It is readily checked that these form a set as in (25.31) of independent standard normal random variables. We define

$$nG_n(1/2) = H_n(\Phi(Z)) \text{ and } n\{G_n(1) - G_n(1/2)\} = n - H_n(\Phi(Z)),$$

and for $i \in \{0,1\}^j$, $j \ge 1$, with k as in (25.32),

$$N_{i,0} = n\{G_n((2k+1)/2^{j+1}) - G_n(2k/2^{j+1})\} = H_{N_i}(Z_i)$$

and

$$N_{i,1} = n\{G_n((2k+2)/2^{j+1}) - G_n((2k+1)/2^{j+1})\} = N_i - H_{N_i}(Z_i).$$

Now we switch to the notation of KMT (1975). Set for $k = 0, 1, ..., 2^j - 1$; $j = 1, 2, ...,$

$$V_{j,k} = B((k+1)/2^j) - B(k/2^j) \text{ and } \widetilde{V}_{j,k} = V_{j+1,2k} - V_{j+1,2k+1};$$

$$U_{1,0} = nG_n(1/2) \text{ and } U_{1,1} = n - nG_n(1/2),$$

$$U_{j,k} = n\{G_n((k+1)/2^j) - G_n(k/2^j)\},$$

and

$$\widetilde{U}_{j,k} = U_{j+1,2k} - U_{j+1,2k+1}.$$

The following lemma ties these two sets of random variables together. This is Lemma 2 of KMT (1975), where it is stated without proof. From this lemma Mason and van Zwet (1987) obtain Theorem 25.1.2 through a series of probability inequalities, which include an exponential inequality due to Yurinskiĭ (1976) and one for sums of squares of standardized multinomials.

Lemma 25.3.1 *There exist finite positive constants C_1, C_2 and η such that*

$$|\widetilde{U}_{j,k} - \sqrt{n}\widetilde{V}_{j,k}| \le C_1 2^j n^{-1}\{\widetilde{U}_{j,k}^2 + (U_{j,k} - n2^{-j})^2\} + C_2 \tag{25.33}$$

whenever

$$\left|\widetilde{U}_{j,k}\right| \le \eta n 2^{-j} \text{ and } |U_{j,k} - n2^{-j}| \le \eta n 2^{-j}. \tag{25.34}$$

PROOF. Write $n_{j,k} = U_{j,k}$. First notice that when $n_{j,k} \ge 1$

$$\begin{aligned}&|2^{-1}\widetilde{U}_{j,k} - 2^{-1}\sqrt{n}\widetilde{V}_{j,k}| \\ \le\ & \sqrt{\frac{n2^{-j}}{n_{j,k}}}|2^{-1}\widetilde{U}_{j,k} - 2^{-1}\sqrt{n_{j,k}}2^{j/2}\widetilde{V}_{j,k}| + \left|\sqrt{\frac{n2^{-j}}{n_{j,k}}} - 1\right| |2^{-1}\widetilde{U}_{j,k}|\end{aligned}$$

and

$$\begin{aligned}\left|\sqrt{\frac{n2^{-j}}{n_{j,k}}} - 1\right| |2^{-1}\widetilde{U}_{j,k}| &\le \left|\frac{n_{j,k} - n2^{-j}}{\sqrt{n2^{-j}}}\right| \left|\frac{2^{-1}\widetilde{U}_{j,k}}{\sqrt{n_{j,k}}}\right| \\ &\le \frac{(n_{j,k} - n2^{-j})^2}{2n2^{-j}} + \frac{(2^{-1}\widetilde{U}_{j,k})^2}{2n_{j,k}}.\end{aligned}$$

Observe that

$$2^{-1}\widetilde{U}_{j,k} = U_{j+1,2k} - n_{j,k}/2$$

and that given $n_{j,k}$, $U_{j+1,2k}$ is binomial with parameters $n_{j,k}$ and $1/2$. Moreover, $2^{j/2}\widetilde{V}_{j,k}$ is a standard normal random variable. Now choose ε and C as in Proposition 25.1.1. We see that as long as $|2^{-1}\widetilde{U}_{j,k}| \le \varepsilon n_{j,k}$ and $n_{j,k} \ge 1$, we have

$$\sqrt{\frac{n2^{-j}}{n_{j,k}}}|2^{-1}\widetilde{U}_{j,k} - 2^{-1}\sqrt{n_{j,k}}2^{j/2}\widetilde{V}_{j,k}| \le \frac{C}{4n_{j,k}}\sqrt{\frac{n2^{-j}}{n_{j,k}}}\widetilde{U}_{j,k}^2 + \frac{C}{4}.$$

Now let $\eta = \min\{\varepsilon, 1/2\}$. Notice that whenever

$$|n_{j,k} - n2^{-j}| = |U_{j,k} - n2^{-j}| \le \eta n 2^{-j}, \tag{25.35}$$

which insures that $n_{j,k} \ge 1$, and

$$|\widetilde{U}_{j,k}| \le \eta n 2^{-j}, \tag{25.36}$$

we have $1/2 \le n2^{-j}/n_{j,k} \le 2$, which implies that $|2^{-1}\widetilde{U}_{j,k}| \le \eta n_{j,k} \le \varepsilon n_{j,k}$. Thus whenever (25.35) and (25.36) hold, we get from the above inequalities that

$$\begin{aligned}&|\widetilde{U}_{j,k} - \sqrt{n}\widetilde{V}_{j,k}| \\ \le\ & C\sqrt{2}2^j n^{-1}(\widetilde{U}_{j,k})^2 + C/2 + 2^{-1}2^j n^{-1}(U_{j,k} - n2^{-j})^2 + 2^{-1}2^j n^{-1}(\widetilde{U}_{j,k})^2.\end{aligned}$$

Setting $C_1 = C\sqrt{2} + 1/2$ and $C_2 = C/2$, we see that the proof is complete. ■

25.4 Some Combinatorics

In this Section, we prove some combinatorial results needed to fill in some missing details in the Mason and van Zwet (1987) paper. We begin by verifying an identity that appears as equation (19) on page 880 of their paper. For any function f defined on $[0,1]$ and nonnegative integers i and l such that

$$(2l+1)/2^{i+1} < 1,$$

set

$$\Delta(i,l,f) = 2f((2l+1)/2^{i+1}) - f(l/2^i) - f((l+1)/2^i).$$

Then the following identity holds:

Proposition 25.4.1 *For nonnegative integers j, k and p such that*

$$2^{-(j+1)} < (2k+1)/2^p \leq 2^{-j},$$

$$\begin{aligned} f((2k+1)/2^p) &= \{2 - 2^{j+1-p}(2k+1)\}f(1/2^{j+1}) \\ &\quad + \{2^{j+1-p}(2k+1) - 1\}f(1/2^j) \\ &\quad + \sum_{i=j+1}^{p-1} c(i,p,k)\Delta(i,h(k,i,p),f), \end{aligned} \tag{25.37}$$

where $0 \leq c(i,p,k) \leq 1$ for $i = j+1,...,p-1$ are independent of f and

$$h(k,i,p) = [(2k+1)/2^{p-i}] \text{ for } i = j+1,...,p-1,$$

with $[x]$ denoting the integer part of x and the summation is defined to be 0 if $j = p$ and $k = 0$.

PROOF. We shall require a couple of lemmas.

Lemma 25.4.1 *Let k be a nonnegative integer. Whenever $2(k+1) = 2^l b$, where b is odd and $l \geq 1$ is an integer, then*

$$[b/2^s] = [(2k+1)/2^{l+s}], \text{ for } s = 1,2,... \tag{25.38}$$

PROOF. By the division algorithm and the fact that $2k+1$ is odd

$$2k+1 = [(2k+1)/2^{l+s}]2^{l+s} + r,$$

where $1 \leq r < 2^{l+s}$. Thus

$$2k+2 = [(2k+1)/2^{l+s}]2^{l+s} + r + 1,$$

so

$$b/2^s = (2k+2)/2^{s+l} = [(2k+1)/2^{l+s}] + (r+1)/2^{l+s}.$$

Since b is odd, we must have $0 < (r+1)/2^{l+s} < 1$, which gives (25.38). ■

Lemma 25.4.2 *Let k be a nonnegative integer. Whenever $2k = 2^m c$, where c is odd and $m \geq 1$ is an integer, then*

$$[c/2^s] = [(2k+1)/2^{m+s}], \text{ for } s = 1, 2, ... \tag{25.39}$$

PROOF. As above, we can write

$$2k+1 = [(2k+1)/2^{m+s}]2^{m+s} + r,$$

where $1 \leq r < 2^{m+s}$. Hence

$$2k = [(2k+1)/2^{m+s}]2^{m+s} + r - 1,$$

which gives

$$c/2^s = 2k/2^{s+m} = [(2k+1)/2^{m+s}] + (r-1)/2^{m+s}.$$

Observing that $(r-1)/2^{s+m} < 1$, we have (25.39). ■

We now turn to the proof of the identity. First it is easy to see that it holds when $k = 0$, that is, when $(2k+1)/2^p = 2^{-j}$. In this case, we have

$$f((2k+1)/2^p) = f(1/2^j), \text{ for } j = p$$

for any value of $p \geq 1$.

Next we assume that $k \geq 1$, i.e. $1/2^{j+1} < (2k+1)/2^p < 1/2^j$. In this case, necessarily $p - j \geq 2$, which implies

$$1/2^j - (2k+1)/2^p = (2^{p-j} - (2k+1))/2^p \geq 1/2^p,$$

and therefore $(2k+1)/2^p + 1/2^p \leq 1/2^j$, from which we get that

$$1/2^{j+1} < 2(k+1)/2^p \leq 1/2^j. \tag{25.40}$$

Also since $p - j \geq 2$

$$(2k+1)/2^p - 1/2^{j+1} = (2k+1-2^{p-j-1})/2^p \geq 1/2^p$$

which gives

$$1/2^{j+1} \leq (2k)/2^p \leq 1/2^j. \tag{25.41}$$

The proof of the identity (25.37) will be completed using induction. Keep $j \geq 0$ fixed and assume that for some $p \geq j+1$ the identity is true for all $k \geq 0$ and $p - 1 \geq p' \geq j$ such that

$$1/2^{j+1} < (2k+1)/2^{p'} \leq 1/2^j. \tag{25.42}$$

It is easy to check that (25.37) holds for the choice $p = j+1$.

We shall prove now that (25.37) holds whenever for some $k > 0$ inequality (25.42) holds with $p' = p$. Choose k such that

$$1/2^{j+1} < (2k+1)/2^p \le 1/2^j.$$

We can assume that $k \ge 1$. Notice that

$$\begin{aligned} f((2k+1)/2^p) &= f((2k+1)/2^p) - 2^{-1}(f(2k/2^p) + f(2(k+1)/2^p)) \\ &\quad + 2^{-1}(f(2k/2^p) + f(2(k+1))/2^p)), \end{aligned}$$

which equals

$$\begin{aligned} &2^{-1}\Delta(p-1,k,f) + 2^{-1}f(2k/2^p) + 2^{-1}f((2(k+1)/2^p) \\ &\quad = 2^{-1}\Delta(p-1,h(k,p-1,p),f) + 2^{-1}f(2k/2^p) + 2^{-1}f(2(k+1)/2^p). \end{aligned}$$

Write

$$2(k+1) = 2^l(2d+1) \ \text{ and } \ 2k = 2^m(2e+1),$$

and set $p_1 = p - l$ and $p_2 = p - m$. Since by (25.40) and (25.41)

$$1/2^{j+1} < (2d+1)/2^{p_1} \le 1/2^j \text{ and } 1/2^{j+1} \le (2e+1)/2^{p_2} \le 1/2^j,$$

we have by the inductive hypothesis that

$$\begin{aligned} f(2(k+1)/2^p) &= f((2d+1)/2^{p_1}) \\ &= \{2 - 2^{j+1-p_1}(2d+1)\}f(1/2^{j+1}) \\ &\quad + \{2^{j+1-p_1}(2d+1) - 1\}f(1/2^j) \\ &\quad + \sum_{i=j+1}^{p_1-1} c(i,p_1,d)\Delta(i,h(d,i,p_1),f) \end{aligned}$$

and

$$f(2k/2^p) = f((2e+1)/2^{p_2}) = f(1/2^{j+1}) \text{ , if } (2e+1)/2^{p_2} = 1/2^{j+1},$$

and otherwise by the inductive hypothesis we see that

$$\begin{aligned} f(2k/2^p) &= f((2e+1)/2^{p_2}) \\ &= \{2 - 2^{j+1-p_2}(2e+1)\}f(1/2^{j+1}) \\ &\quad + \{2^{j+1-p_2}(2e+1) - 1\}f(1/2^j) \\ &\quad + \sum_{i=j+1}^{p_2-1} c(i,p_2,e)\Delta(i,h(e,i,p_2),f). \end{aligned}$$

Notice that by Lemma 25.4.1, assuming $j+1 \le p_1 - 1$, we have for $i = j+1, ..., p_1 - 1$,

$$h(d,i,p_2) = [(2d+1)/2^{p_1-i}] = [(2k+1)/2^{p-i}] = h(k,i,p)$$

and similarly by Lemma 25.4.2, assuming $j+1 \le p_2 - 1$, we have for $i = j+1, ..., p_2 - 1$,

$$h(e, i, p_2) = h(k, i, p).$$

Observe that since either $k+1$ or k is odd, either $l = 1$ or $m = 1$, so that $\max\{p_1, p_2\} = p - 1$. Set

$$c(i, p_1, d) = 0 \text{ for } i \ge p_1 \text{ and } c(i, p_2, e) = 0 \text{ for } i \ge p_2,$$

and let

$$c(i, p, k) = 2^{-1}\{c(i, p_1, d) + c(i, p_2, e)\} \text{ for } i = j+1, ..., p-2$$

and $c(i, p, p-1) = 1/2$. By adding terms, we see that we have (25.37). ∎

Notice that when $f(0) = f(1) = 0$, the identity (25.37) becomes

$$f(1/2^p) = \sum_{j=0}^{p-1} 2^j \Delta(j, 0, f) 2^{-p}.$$

Next we turn to an essential distributional identity which is stated without proof on lines 9 and 10 from the bottom on page 882 of Mason and van Zwet (1987).

Proposition 25.4.2 *For any nonnegative integers j, k and p satisfying*

$$1/2^{j+1} < (2k+1)/2^p < 1/2^j, \tag{25.43}$$

$$\{(U_{i,k(i)}, (\widetilde{U}_{i,k(i)})^2) : i = j+1, ..., p-1\} =_d \{(U_{i,0}, (\widetilde{U}_{i,0})^2) : i = j+1, ..., p-1\}, \tag{25.44}$$

where $k(i) := h(k, i, p) = [(2k+1)/2^{p-i}]$ for $i = j+1, ..., p-1$.

PROOF. Notice that for any $x \ge 0$

$$[x] \le [2x]/2 < (2[x]+1)/2 \le [x] + 1,$$

with equality holding exactly on one end or the other. To see this, write $x = [x] + r$ and note that if $0 \le r < 1/2$, $[2x] = 2[x]$ and if $1/2 \le r < 1$, $[2x] = 2[x]+1$. Using this fact, one finds that for any $j+1 \le i \le p-2$ either

Case 1.

$$k(i) = k(i+1)/2,$$

in which case

$$\begin{aligned} k(i)/2^i &= k(i+1)/2^{i+1} < (2k(i+1)+1)/2^{i+2} < (2k(i)+1)/2^{i+1} \\ &= (k(i+1)+1)/2^{i+1} < (k(i)+1)/2^i; \end{aligned}$$

Case 2.

$$k(i) < k(i+1)/2,$$

in which case

$$\begin{aligned} k(i)/2^i &< k(i+1)/2^{i+1} = (2k(i)+1)/2^{i+1} < (2k(i+1)+1)/2^{i+2} \\ &< (k(i+1)+1)/2^{i+1} = (k(i)+1)/2^i. \end{aligned}$$

Write

$$I_i^{(1)} = (k(i)/2^i, (2k(i)+1)/2^{i+1}], \; I_i^{(2)} = ((2k(i)+1)/2^{i+1}, (k(i)+1)/2^i],$$

and $I_i^{(3)} = I_i^{(1)} \cup I_i^{(2)}$. In *Case 1*, both $I_{i+1}^{(1)}$ and $I_{i+1}^{(2)}$ are subsets of $I_i^{(1)}$ and $I_{i+1}^{(3)} = I_i^{(1)}$; and in *Case 2*, both $I_{i+1}^{(1)}$ and $I_{i+1}^{(2)}$ are subsets of $I_i^{(2)}$ and $I_{i+1}^{(3)} = I_i^{(2)}$. In either case,

$$I_{i+1}^{(3)} \subset I_i^{(3)} \text{ and } P\{U \in I_i^{(3)}\} = 2^{-i}. \tag{25.45}$$

For any subinterval $I \subset [0,1]$, write

$$nG_n(I) = \sum_{i=1}^{n} 1\{U_i \in I\}.$$

Notice that with this notation

$$U_{i,k(i)} = nG_n(I_i^{(3)}) = nG_n(I_i^{(1)}) + nG_n(I_i^{(2)})$$

and

$$\widetilde{U}_{i,k(i)} = 2nG_n(I_i^{(1)}) - nG_n(I_i^{(3)}).$$

Thus in *Case 1*

$$\widetilde{U}_{i,k(i)} = 2nG_n(I_{i+1}^{(3)}) - nG_n(I_i^{(3)})$$

and in *Case 2*

$$\widetilde{U}_{i,k(i)} = nG_n(I_i^{(3)}) - 2nG_n(I_{i+1}^{(3)}).$$

Therefore,

$$(U_{i,k(i)}, (\widetilde{U}_{i,k(i)})^2) = (nG_n(I_i^{(3)}), (2nG_n(I_{i+1}^{(3)}) - nG_n(I_i^{(3)}))^2). \tag{25.46}$$

Now clearly by (25.45)

$$(nG_n(I_{j+1}^{(3)}), ..., nG_n(I_p^{(3)})) =_d (nG_n(2^{-j-1}), ..., nG_n(2^{-p})).$$

Hence on account of (25.45) and (25.46), we have (25.44). ■

References

1. Arak, T. V. and Zaitsev, A. Yu. (1988). Uniform Limit Theorems for Sums of Independent Random Variables, translation of *Trudy. Mat. Inst. Steklov*, **174** 1986, *Proceedings of the Steklov Institute of Mathematics.*

2. Bretagnolle, J. and Massart, P. (1989). Hungarian constructions from the nonasymptotic viewpoint, *Annals of Probability*, **17**, 239–256.

3. Borisov, I. S. (1978). On the rate of convergence of in the "conditional" invariance principle, (Russian. English summary) *Teor. Verojatnost. i Primen.*, **23**, 67–79.

4. Csörgő, M. and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*, New York: John Wiley & Sons.

5. Csörgő, M., Csörgő, S., Horváth, L. and Mason, D. M. (1986). Weighted empirical and quantile processes, *Annals of Probability*, **14**, 31–85.

6. Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*, New York: Academic Press.

7. Delporte, J. (1980). Une construction de la fonction de repartition empirique à partir d'une suite de ponts browniens, *Pub. IRMA-Lille*, **2**, facs. 6-n^o 4.

8. Einmahl, U. (1986). A refinement of the KMT-inequality for partial sum strong approximation, *Technical Report* Laboratory for Research in Statistics and Probability, Carleton University-University of Ottawa, No. 88, September 1986.

9. Gänssler, P. and Stute, W. (1977). *Wahrscheinlichkeitstheorie*, Berlin: Springer-Verlag.

10. Hahn, M., Mason, D. M. and Weiner, D. (Eds.) (1991). *Sums, Trimmed Sums and Extremes*, Boston: Birkhäuser.

11. Ibragimov, I. A. and Linnik, Yu V. (1971). *Independent and Stationary Sequences of Random Variables*, Groningen: Wolter-Noordhoff Publishing.

12. Koltchinskii, V. I. (1994). Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions, *Journal of Theoretical Probability*, **7**, 73–118.

13. Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent rv's and the sample df I, *Z. Wahrsch. verw. Gebiete*, **32**, 111–131.

14. Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent rv's and the sample df II, *Z. Wahrsch. verw. Gebiete*, **34**, 33–58.

15. Mason, D. M. (2000). An exponential inequality for a weighted approximation to the uniform empirical process with applications, In *State of the Art in Probability and Statistics: Festschrift for Willem R. van Zwet* (Eds., M. C. M. de Gunst, C. A. J. Klaassen, A. W. van der Vaart), IMS Lecture Notes-Monograph Series, Hayward, California: Institute of Mathematical Statistics.

16. Mason, D. M. and Van Zwet, W. R. (1987). A refinement of the KMT inequality for the uniform empirical process, *Annals of Probability*, **15**, 871–884.

17. Petrov, V. V. (1972). *On Sums of Independent Random Variables*, Moscow: Nauka (in Russian).

18. Richter, H. (1978). *Über das dyadische Einbettungsschema von Komlós-Major-Tusnády*, Diplomarbeit, Ruhr-Universitaet Bochum, 1978; Referent: P. Gänssler

19. Rio, E. (1994). Local invariance principles and their application to density estimation, *Probability Theory and Related Fields*, **98**, 21–45.

20. Sakhanenko, A. I. (1984). Rate of convergence in the invariance principle for variables with exponential moments that are not identically distributed, In *Limit Theorems for Sums of Random Variables*, pp. 4–49, Trudy Inst. Mat., 3, "Nauka" Sibirsk. Otdel., Novosibirisk (in Russian).

21. Shorack, G. and Wellner, J. (1986). *Empirical Processes with Applications to Statistics*, New York: John Wiley & Sons.

22. Tusnády, G. (1977). *A Study of Statistical Hypotheses*, Dissertation, Budapest: The Hungarian Academy of Sciences (in Hungarian).

23. Yurinskiĭ, V. (1976). Exponential inequalites for sums of random vectors, *Journal of Multivariate Analysis*, **6**, 473–499.

26

Inter-Record Times in Poisson Paced F^α Models

H. N. Nagaraja and G. Hofmann

Ohio State University, Columbus, Ohio
Universidad de Concepcion, Concepcion, Chile

Abstract: Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables such that X_n has cdf $F^{\alpha(n)}$, where F is a continuous distribution function and $\alpha(n)$ is a positive constant. Suppose X_n is observed at the occurrence of the $(n-1)$th event of an independent Poisson process $\mathbf{P}$. We describe the exact as well as asymptotic distributions of the inter-arrival times of upper record values from such an F^α record model.

Keywords and phrases: Inter-record times, nonhomogeneous Poisson process, stochastic convergence, record values, exponential distribution

26.1 Introduction

Let $\{X_n, n \geq 1\}$ be a sequence of independent random variables such that X_n has cdf $F^{\alpha(n)}$, where F is a continuous distribution function and $\alpha(n)$ is a positive constant. An observation will be called an upper record if it exceeds all previous observations. We define the *record indices* T_n and the *record values* R_n as follows:

$$T_1 = 1,\ T_n = \min\{j > T_{n-1} : X_j > X_{T_{n-1}}\}, n \geq 2,$$

and $R_n = X_{T_n},\ n \geq 1$. This setup is known as the F^α *model*, formally introduced by Nevzorov (1985), and investigated quite extensively in recent years. Yang (1975) had earlier studied such a model in a very special case of geometrically increasing α's.

Now suppose X_1 is observed at time 0 and for $n \geq 2$, X_n is observed at the time of occurrence of the $(n-1)$th event in a Poisson process $\mathbf{P}$ with mean function $\Lambda(t)$ and positive intensity function $\lambda(t)$ $(t \geq 0)$, where $\mathbf{P}$ is independent of the X_n sequence. Let $\{V_n,\ n \geq 1\}$ be the sequence of inter-arrival times associated with $\mathbf{P}$. Record statistics arising from such a setup

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

was first studied by Pickands (1971) who took $\mathbf{P}$ to be a homogeneous Poisson process and assumed the X_n's to be i.i.d. continuous random variables. Since then, point process paced record models where $\mathbf{P}$ is a common simple point process have been studied by several researchers. However, almost all of this research has been directed to the i.i.d. case. For an overview of the F^α and the point process paced record models, see Arnold, Balakrishnan, and Nagaraja (1998; Chapters 6 and 7).

Recently, Hofmann (1997) and Hofmann and Nagaraja (1999) have considered the F^α setup in the context of point process models. Here, we consider the F^α model paced by an independent Poisson process observed over $[0, \infty)$. We investigate the properties of the sequence of record inter-arrival times $U_n = \sum_{i=T_n}^{T_{n+1}-1} V_i$, $n \geq 1$, and the sequence of record arrival times $W_n = \sum_{i=1}^n U_i$. In order that the U_n's are well defined, we assume that $\Lambda(t) \to \infty$ as $t \to \infty$. We also assume that $\Lambda(t)$ is finite for all real t. We present some exact and asymptotic results describing the distributional structure of the sequence of inter-arrival times of the records. For related work and proofs of some of the results reported here, we refer you to Hofmann and Nagaraja (1999).

Let $s(n) = \alpha(1) + \cdots + \alpha(n)$ and $S(t)$ be a function with positive derivative defined for all $t > 0$ such that for a positive integer n, $S(n) = s(n)$. (Later we provide an algorithm to produce such a function.) Let $p_n = \alpha(n)/s(n)$ represent the record accrual probability at the arrival of X_n, and $A(n) = \sum_{i=1}^n p_i$, $B(n) = \sum_{i=1}^n p_i^2$. Further, we let $\Psi(t) = \Lambda^{-1}(t) = \inf\{s : \Lambda(s) > t\}$, and $\tau(t) = \log S(\Lambda(t))$, $t > 0$. We denote by $\mathrm{Exp}(\theta)$ an exponential random variable with mean $1/\theta$.

26.2 Exact Distributions

Observe that $W_n^* = \Lambda(W_n), n \geq 1$, behave like the record arrival times associated with a homogeneous Poisson pacing process with unit intensity. Consequently, we have the following result.

Theorem 26.2.1 *For the Poisson process F^α model, the joint density of the inter-record times, $U_1, \ldots, U_n$, is given by*

$$\begin{aligned} & f_{\vec{U}}(u_1, \ldots, u_n) \\ = \; & \exp\left(-\Lambda(u_1 + \cdots + u_n)\right) \prod_{i=1}^n \lambda(u_1 + \cdots + u_i) \\ & \times \sum_{k_1, \cdots, k_n = 1}^{\infty} \frac{\alpha(1)}{s(1 + k_1 + \cdots + k_n)} \prod_{i=1}^n \frac{\alpha(1 + k_1 + \cdots + k_i)}{s(k_1 + \cdots + k_i)} \\ & \times \frac{1}{\Gamma(k_i)} \left(\Lambda(u_1 + \cdots + u_i) - \Lambda(u_1 + \cdots + u_{i-1})\right)^{k_i - 1}. \end{aligned} \tag{26.1}$$

PROOF. Basically the proof involves two steps. First we consider the homogeneous **P** with unit intensity and condition on the T_i and show that (26.1) holds when $\Lambda(t) = t$. Next we make a time transformation ($W_n = \Psi(W_n^*)$) to obtain the above equation for a general $\Lambda(t)$. Details may be found in Hofmann and Nagaraja (1999). ■

When $\alpha(1) = c\ (> 1)$, and $\alpha(n) = (c-1)c^{n-1}$, $n \geq 2$, $s(n) = c^n$ is increasing geometrically and p_n is a constant $(1 - c^{-1})$. Then (26.1) simplifies to

$$f_{\vec{U}}(u_1, \ldots, u_n) = \left(1 - c^{-1}\right)^n \exp\left(-\Lambda(u_1 + \cdots + u_n)\left(1 - c^{-1}\right)\right) \times \prod_{i=1}^{n} \lambda(u_1 + \cdots + u_i). \quad (26.2)$$

From (26.2) and the expression for the joint density of record values [see Arnold, Balakrishnan and Nagaraja (1998, p. 10)], it follows that the record arrival times (W's) behave like record values and the U's behave like spacings of record values generated from an i.i.d. sequence of random variables having the failure rate function given by $(1 - c^{-1})\lambda(t)$. This leads to characterizations of homogeneous Poisson process based on the properties of U_1 and U_2. For example, when $s(n) = c^n$, the following three statements are equivalent: (i) U_1 and U_2 are independent, (ii) U_1 and U_2 are identically distributed, (iii) $\lambda(t)$ is a constant, see, Hofmann and Nagaraja (1999, Theorem 4.1).

When **P** is a unit Poisson process and the record accrual probability p_n equals n^{-1}, there exists a distributional representation for the logarithms of the first n inter-record times:

$$(\log U_1, \ldots, \log U_n) \stackrel{d}{=} (C_1 + \log A_1, \ldots, C_n + \log A_n), \quad (26.3)$$

where A_j and B_j, $1 \leq j \leq n$, are all i. i. d. Exp(1), and $C_j = \sum_{i=1}^{j} B_i$ [see, for example, Arnold, Balakrishnan and Nagaraja (1998, Theorem 7.4.1)]. Similar representation holds even when p_n^{-1} is a linear function of n of the form $a(n-2) + b$ where $a > 0$ and $b > 1$. More precisely, the following result holds.

Theorem 26.2.2 *Let* **P** *be a homogeneous Poisson process with unit intensity and let the record rate* $p_n = c(d+n-2)^{-1}$, $0 < c < d$, *for* $n \geq 2$. *Suppose* A_j *are Exp(1),* $1 \leq j \leq n$, B_j *are i. i. d. Exp*(c), $2 \leq j \leq n$, *and* B_1 *has characteristic function*

$$E(e^{zB_1}) = \frac{\Gamma(d)\Gamma(c-z)}{\Gamma(c)\Gamma(d-z)}, \quad (26.4)$$

where z *is imaginary. Further, assume that all these random variables are mutually independent. Then*

$$(\log U_1, \ldots, \log U_n) \stackrel{d}{=} (C_1 + \log A_1, \ldots, C_n + \log A_n), \quad (26.5)$$

where $C_j = \sum_{i=1}^{j} B_i$.

PROOF. See Hofmann and Nagaraja (1999). ■

Remarks.

(a) The record rate of $p_n = c(d+n-2)^{-1}$, $n \geq 2$, corresponds to the α sequence given by

$$\alpha_2 = \frac{c\alpha_1}{d-c},$$

and

$$\alpha_n = \frac{c\alpha_1}{d-c+n-2} \prod_{i=0}^{n-3} \frac{d+i}{d-c+i}, \; n \geq 3.$$

(b) The characteristic function in (26.4) corresponds to a random variable B_1 that has the density function

$$f(x) = \frac{\Gamma(d)}{\Gamma(c)\Gamma(d-c)} e^{-cx}(1-e^{-x})^{d-c-1}, \; x > 0.$$

When d and c are positive integers, B_1 behaves like the $(d-c)$th order statistic from a random sample of size $(d-1)$ from a standard exponential distribution. When $d-c$ is a positive integer k, B_1 is the sum of k independent exponential random variables with parameters $c, \ldots, (d-1)$.

(c) Using the time axis transformation $\Lambda(t)$, we can obtain distributional representations for $(U_1, \ldots, U_n)$ for arbitrary Λ when $p_n = c(d+n-2)^{-1}$. It follows from (26.5) that

$$(U_1, U_2, \ldots, U_n) \stackrel{d}{=} \left(\Psi\left(A_1 e^{C_1}\right), \Psi\left(A_1 e^{C_1} + A_2 e^{C_2}\right) - \Psi\left(A_1 e^{C_1}\right), \ldots \right.$$
$$\left. \ldots, \Psi\left(\sum_{j=1}^{n} A_j e^{C_j}\right) - \Psi\left(\sum_{j=1}^{n-1} A_j e^{C_j}\right)\right),$$

where A_i and C_i are as described in Theorem 26.2.2.

26.3 Asymptotic Distributions

For the i.i.d. Poisson paced record model, assuming $\lim_{t\to\infty} \frac{\lambda(t)}{\Lambda(t)} = c \in (0, \infty)$, Bunge and Nagaraja (1992a) showed that U_n has an exponential limit. We will now look at some classes of combinations of α-structures and mean functions of the Poisson pacing process that lead to a limit for U_n.

We begin by presenting a smooth $S(t)$ that matches $s(n)$ for all $n \geq 1$. Let $s(0) = 0$, and for $0 \leq n \leq t \leq (n+1)$, define $S(t) = s(n) + \alpha(n+1)\int_n^t f_n(x)dx$, where f_n is a density function on $[n, n+1]$. This should produce a continuous function and differentiable almost everywhere. But we need a differentiable function with positive derivative everywhere (or at least for all large t) for our purpose. This would be accomplished by choosing $f_n(x)$ such that it is a density of the form $c_n e^{d_n x}$ and satisfies the condition $\alpha(n+1)f_n(n) = \alpha(n)f_{n-1}(n)$. The first constraint implies $c_n = \frac{d_n}{e^{nd_n}(e^{d_n}-1)}$ and the second ensures that the right and left derivatives of $S(t)$ at $t = n$ are equal. Such an $S(t)$ is presented in the following lemma.

Lemma 26.3.1 *Let*

$$\begin{aligned} S(t) &= \alpha(1)t, \; 0 \leq t \leq 1 \\ &= s(n) + \alpha(1)\exp\big(\sum_{i=0}^{n-1} d_i\big)\frac{e^{d_n(t-n)} - 1}{d_n}, \; n \leq t \leq n+1, \; n \geq 1, \end{aligned}$$

where $d_0 = 0$ and the d_n are determined sequentially using the relation

$$\begin{aligned} \frac{e^{d_n} - 1}{d_n} &= \frac{\alpha(n+1)}{\alpha(1)}\exp(-\sum_{i=0}^{n-1} d_i) \\ &= \theta_n, \text{ say.} \end{aligned} \tag{26.6}$$

The function $S(t)$ defined above is such that $S'(t) > 0$ for all $t > 0$ and $S(n) = s(n)$ for all positive integer n.

Since the function $g(y) = (e^y - 1)/y$ is strictly increasing for $y \in (-\infty, \infty)$ with range $(0, \infty)$, there is a unique d_n satisfying (26.6). When θ_n is greater (less) than 1, b_n is positive (negative), and when $\theta_n = 1$, b_n is to be interpreted as 0 and $S(t)$ takes on the form $s(n) + \alpha(n+1)(t-n)$. Also, in the interval $(n, n+1)$, S(t) is always either convex or concave (or linear). Further, $S'(t)$ is always bounded by limit superior and inferior of the sequence $\{\alpha(1)\exp(\sum_{i=1}^n d_i\}$, where d_i are defined by (26.6).

Recently, Hofmann and Nagaraja (1999) proved the following asymptotic result for the inter-record times that does not require any normalization of U_n.

Theorem 26.3.1 *Let* $\liminf_{t\to\infty}\frac{\lambda(t)}{\Lambda(t)} > 0$ *and* $\tau_0(t) = \log S([\Lambda(t)])$, *where* $[x]$ *is the greatest integer* $\leq x$. *Suppose*

$$L(a) = \lim_{t\to\infty}\frac{\tau_0(t)}{\tau_0(t+a)}$$

exists for all $a > 0$ (or equivalently, for all a in some interval $(0, \epsilon)$, $\epsilon > 0$). Then $U_n \xrightarrow{\mathcal{L}} U$, and the distribution function of U is given by $F_U(a) = 1 - L(a)$, $(a > 0)$. Further, one of the following holds:

$$U \equiv 0 \quad \textit{and} \quad \frac{\tau(t)}{t} \to \infty \, (t \to \infty) \tag{26.7}$$

$$U \equiv \infty \quad \textit{and} \quad \tau(t) = o(t) \tag{26.8}$$

$$U \sim \mathit{Exp}(c) \quad \textit{and} \quad \tau(t) \approx ct. \tag{26.9}$$

If (26.9) holds, $(cU_{n+1}, \dots, cU_{n+k}) \xrightarrow{\mathcal{L}} (E_1, \dots, E_k)$ for any finite k, where E_i are i. i. d. Exp(1) random variables.

The condition $\liminf_{t\to\infty} \frac{\lambda(t)}{\Lambda(t)} > 0$ in the above theorem is not necessary for the asymptotic exponential distribution of U_n. For example, when $\mathbf{P}$ is homogeneous Poisson and the record accrual probability p_n remains a constant (as it happens for geometrically increasing α's), U_j are i.i.d. exponential random variables for all j [Bunge and Nagaraja (1992b)].

When $U_n \to 0$ and $U_n \to \infty$, as in (26.7) and (26.8), respectively, we would like to know how fast this convergence occurs. In particular, we now ask whether it is possible to find norming constants a_n, $b_n (> 0)$ such that $(U_n - a_n)/b_n$ goes in law to a nondegenerate distribution. As shown in our next theorem, it is indeed possible with some additional conditions on the α-structure itself. However, first we need to establish the following lemma.

Lemma 26.3.2 *In a Poisson paced F^α model, suppose*
(a) $p_n = \frac{\alpha_n}{S(n)} \to 0$, $\frac{B(n)}{\sqrt{A(n)}} \to 0$,
(b) $\tau(t) = \log S(\Lambda(t))$ is differentiable and $\tau'(t) > 0$ for all large t,
(c) there exists a function h with $h(n) \to \infty$ as $n \to \infty$, and

$$\gamma_n = \sqrt{S^{-1}\left(e^{n+\sqrt{n}h(n)}\right) \log\left(n + \sqrt{n}h(n)\right)}\, \tau'(\tau^{-1}(n)) \sup_{\xi \in K_n} \Psi'(\xi) \to 0, \tag{26.10}$$

where

$$\begin{aligned} K_n \;=\; & \Bigg[S^{-1}\left(e^{n-\sqrt{n}h(n)}\right) - 2\sqrt{2S^{-1}\left(e^{n-\sqrt{n}h(n)}\right)\log\left(n - \sqrt{n}h(n)\right)}, \\ & S^{-1}\left(e^{n+\sqrt{n}h(n)}\right) + 2\sqrt{2S^{-1}\left(e^{n+\sqrt{n}h(n)}\right)\log\left(n + \sqrt{n}h(n)\right)} \Bigg]. \end{aligned}$$

Then

$$\left| \frac{U_n}{b_n} - \frac{\Psi(T_{n+1}) - \Psi(T_n)}{b_n} \right| \xrightarrow{P} 0,$$

where $b_n = \frac{1}{\tau'(\tau^{-1}(n))}$.

PROOF. Since $W_n^* = V_1^* + \cdots + V_{T_{n+1}-1}^*$ where the V_i^*'s are i. i. d. Exp(1) variables and $T_n \to \infty$ a. s., it follows from the law of the iterated logarithm [see, for example, Chung (1974, Theorem 7.5.1)] that

$$\limsup_{n\to\infty} \frac{|W_n^* - (T_{n+1}-1)|}{\sqrt{2(T_{n+1}-1)\log\log(T_{n+1}-1)}} \le 1 \text{ a. s. or}$$

$$\limsup_{n\to\infty} \frac{|W_n^* - T_{n+1}|}{\sqrt{2T_{n+1}\log\log T_{n+1}}} \le 1 \text{ a. s.}$$

Hence, there exists an n_0 such that

$$|W_n^* - T_{n+1}| \le 2\sqrt{2T_{n+1}\log\log T_{n+1}} \text{ a. s. for all } (T_n \ge)\, n > n_0,$$

$$|\Psi(W_n^*) - \Psi(T_{n+1})| \le 2\sqrt{2T_{n+1}\log\log T_{n+1}} \sup_{\xi\in J_n} \Psi'(\xi) \text{ a. s.,}$$

where

$$J_n = \left[T_{n+1} - 2\sqrt{2T_{n+1}\log\log T_{n+1}},\, T_{n+1} + 2\sqrt{2T_{n+1}\log\log T_{n+1}}\right],$$

$$\frac{|\Psi(W_n^*) - \Psi(T_{n+1})|}{b_n} \le 2\sqrt{2T_{n+1}\log\log T_{n+1}}\,\tau'(\tau^{-1}(n)) \sup_{\xi\in J_n} \Psi'(\xi) \text{ a. s.} \tag{26.11}$$

and

$$\frac{|\Psi(W_{n-1}^*) - \Psi(T_n)|}{b_n} \le 2\sqrt{2T_n\log\log T_n}\,\tau'(\tau^{-1}(n)) \sup_{\xi\in J_{n-1}} \Psi'(\xi) \text{ a. s.} \tag{26.12}$$

If the right-hand side goes to zero in probability, so does the left-hand side. However, from Nevzorov (1995),

$$\frac{\log S(T_n) - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1).$$

Therefore, for any $h(n) \to \infty$ $(n \to \infty)$,

$$\frac{\log S(T_n) - n}{\sqrt{n}h(n)} \xrightarrow{P} 0 \text{ and } \frac{\log S(T_{n+1}) - n}{\sqrt{n}h(n)} \xrightarrow{P} 0,$$

which imply that

$$P\left(S^{-1}\left(e^{n-\sqrt{n}h(n)}\right) \le T_{n+1} \le S^{-1}\left(e^{n+\sqrt{n}h(n)}\right)\right) \to 1 \text{ and}$$

$$P\left(S^{-1}\left(e^{n-\sqrt{n}h(n)}\right) \le T_n \le S^{-1}\left(e^{n+\sqrt{n}h(n)}\right)\right) \to 1.$$

Hence, the right hand sides of (26.11) and (26.12) go in probability to zero if $\gamma_n \to 0$. Recall that $U_n = \Psi(W_n^*) - \Psi(W_{n-1}^*)$. From (26.11) and (26.12), it now follows that

$$\begin{aligned}\left|\frac{U_n}{b_n} - \frac{\Psi(T_{n+1}) - \Psi(T_n)}{b_n}\right| &= \left|\frac{\Psi(W_n^*) - \Psi(T_{n+1})}{b_n} - \frac{\Psi(W_{n-1}^*) - \Psi(T_n)}{b_n}\right| \\ &\leq \left|\frac{\Psi(W_n^*) - \Psi(T_{n+1})}{b_n}\right| + \left|\frac{\Psi(W_{n-1}^*) - \Psi(T_n)}{b_n}\right| \xrightarrow{P} 0.\end{aligned}$$

∎

Note that in view of Lemma 26.3.1, we can always define $S(t)$ such that $\tau'(t) > 0$ as long as $\lambda(t) > 0$ for all t. Thus, condition (b) in Lemma 26.3.2 is always satisfied.

Theorem 26.3.2 *In a Poisson paced F^α model, let the conditions (a)-(c) of Lemma 26.3.2 hold. Further, suppose for all $x > 0$ there exists a function $g_x(n)$ with $g_x(n) \to \infty$ $(n \to \infty)$ such that*

$$b_n \inf_{t \in I_n} \tau'(t) \to 1 \text{ and } b_n \sup_{t \in I_n} \tau'(t) \to 1, \tag{26.13}$$

where

$$b_n = \frac{1}{\tau'(\tau^{-1}(n))}, \; I_n = \left[\tau^{-1}(n - \sqrt{n} g_x(n)), \tau^{-1}(n + \sqrt{n} g_x(n)) + b_n x\right].$$

Then

$$\frac{U_n}{b_n} \xrightarrow{\mathcal{L}} Exp(1).$$

PROOF. Let $x > 0$. Assuming $p_n \to 0$, and $\frac{B(n)}{\sqrt{A(n)}} \to 0$, Nevzorov (1986, 1995) has shown that $\frac{\log S(T_n) - n}{\sqrt{n}} \xrightarrow{\mathcal{L}} N(0,1)$. Therefore,

$$\frac{\log S(T_n) - n}{\sqrt{n} g_x(n)} = \frac{\tau(\Psi(T_n)) - n}{\sqrt{n} g_x(n)} \xrightarrow{P} 0.$$

Hence

$$P\left(-1 \leq \frac{\tau(\Psi(T_n)) - n}{\sqrt{n} g_x(n)} \leq 1\right) \to 1,$$

$$P\left(n - \sqrt{n} g_x(n) \leq \tau(\Psi(T_n)) \leq n + \sqrt{n} g_x(n)\right) \to 1,$$

$$P\left(\tau^{-1}(n - \sqrt{n} g_x(n)) \leq \Psi(T_n) \leq \tau^{-1}(n + \sqrt{n} g_x(n))\right) \to 1,$$

and

$$P\left(b_n \inf_{t \in I_n} \tau'(t) \leq b_n \tau'(\Psi(T_n) + \xi_n x b_n) \leq b_n \sup_{t \in I_n} \tau'(t)\right) \to 1 \;\; \forall \xi_n \in [0,1].$$

Since we assumed that both upper and lower bounds go to one,

$$b_n\tau'(\Psi(T_n)+\xi_n x b_n)\xrightarrow{P}1\,\forall\xi_n\in[0,1].$$

However,

$$\begin{aligned}Y_n &= \tau(\Psi(T_n)+b_n x)-\tau(\Psi(T_n))\\ &= x b_n\tau'(\Psi(T_n)+\xi_n x b_n)\text{ for some }\xi_n\in[0,1],\end{aligned}$$

which means $Y_n\xrightarrow{P}x$ and $e^{-Y_n}\xrightarrow{P}e^{-x}$. Since τ is strictly increasing, $Y_n\geq 0$. Hence $|e^{-Y_n}|\leq 1$ and e^{-Y_n} also converges in L^1 [Chung (1974, Theorem 4.1.4)]

$$\lim_{n\to\infty}Ee^{-Y_n}=e^{-x}\,\forall x>0. \tag{26.14}$$

Now let us look at (U_n/b_n). From Lemma 26.3.2, it follows that

$$\begin{aligned}\lim_{n\to\infty}P\left(\frac{U_n}{b_n}>x\right) &= \lim_{n\to\infty}P\left(\Psi(T_{n+1})-\Psi(T_n)>b_n x\right)\\ &= \lim_{n\to\infty}P\left(T_{n+1}>\Lambda(\Psi(T_n)+b_n x)\right)\\ &= \lim_{n\to\infty}\sum_{i=n}^{\infty}P\left(T_{n+1}>\Lambda(\Psi(T_n)+b_n x)|T_n=i\right)\\ &\quad\times P(T_n=i)\\ &= \lim_{n\to\infty}\sum_{i=n}^{\infty}\frac{S(i)}{S(\Lambda(\Psi(i)+b_n x))}P(T_n=i)\\ &= \lim_{n\to\infty}E\left[\frac{S(T_n)}{S(\Lambda(\Psi(T_n)+b_n x))}\right]\\ &= \lim_{n\to\infty}Ee^{\tau(\Psi(T_n))-\tau(\Psi(T_n)+b_n x)}\\ &= \lim_{n\to\infty}Ee^{-Y_n}=e^{-x}.\end{aligned}$$

■

Remark. The conditions (26.10) in Lemma 26.3.2 and (26.13) in Theorem 26.3.1 look cumbersome, but are usually not very hard to check. In all the situations we have considered, the choice of $g_x(n)=h(n)=n^\epsilon$, $\epsilon\in(0,1/2)$ and $g_x(n)=h(n)=\log n$ work. Below are a few examples.

Example 26.3.1 Let $S(x)=x$ (classical model), $\Lambda(t)=e^{(t^r)}-1$, $r>0$. It follows that $\tau(t)=\log\left(e^{(t^r)}-1\right)\approx t^r$, $\tau'(t)=rt^{r-1}\frac{e^{(t^r)}}{e^{(t^r)}-1}\approx rt^{r-1}$, and

$$b_n=\frac{1}{\tau'(\tau^{-1}(n))}=\frac{e^n}{e^n+1}\frac{1}{r}\left[\log\left(e^n+1\right)\right]^{\frac{1}{r}-1}\approx\frac{1}{r}n^{\frac{1}{r}-1}.$$

Note that $\tau'(t)$ is strictly increasing for large t ($r>1$), constant ($r=1$) or strictly decreasing ($r<1$). Therefore, to check (26.13), we only have to check whether $b_n\tau'(\tau^{-1}(n-\sqrt{n}g_x(n)))$ and $b_n\tau'(\tau^{-1}(n+\sqrt{n}g_x(n))+b_nx)$ go to one.

$$\begin{aligned} b_n\tau'(\tau^{-1}(n-\sqrt{n}g_x(n))) &\approx \frac{1}{r}n^{\frac{1}{r}-1}\,r\left\{\left[\log\left(e^{n-\sqrt{n}g_x(n)}+1\right)\right]^{\frac{1}{r}}\right\}^{r-1} \\ &\approx n^{\frac{1}{r}-1}\left[\log\left(e^{n-\sqrt{n}g_x(n)}+1\right)\right]^{1-\frac{1}{r}} \\ &\to 1 \text{ for all } g_x(n)=n^{\epsilon},\ \epsilon\in(0,1/2). \end{aligned}$$

Also,

$$b_n\tau'(\tau^{-1}(n+\sqrt{n}g_x(n))+b_nx) \approx \frac{1}{r}n^{\frac{1}{r}-1}\,r\left\{\left[\log(e^{n+\sqrt{n}g_x(n)}+1)\right]^{\frac{1}{r}}+\frac{x}{r}n^{\frac{1}{r}-1}\right\}^{r-1}$$

and hence converges to 1 as $n\to\infty$ for $g_x(n)=n^{\epsilon}$, $\epsilon\in(0,1/2)$. To check (26.9), note that

$$\Psi(t)=[\log(t+1)]^{\frac{1}{r}},\ S^{-1}(x)=x,\ \Psi'(t)=\frac{(\log(t+1))^{\frac{1}{r}-1}}{r(t+1)}.$$

Since $\Psi'(t)$ is strictly decreasing,

$$\sup_{\xi\in K_n}\Psi'(\xi)=\Psi'\left(e^{n-\sqrt{n}h(n)}-2\sqrt{2e^{n-\sqrt{n}h(n)}\log(n-\sqrt{n}h(n))}\right).$$

Hence,

$$\gamma_n\approx\sqrt{2e^{n-\sqrt{n}h(n)}\log(n-\sqrt{n}h(n))}\,rn^{1-\frac{1}{r}}\frac{[n-\sqrt{n}h(n)]^{\frac{1}{r}-1}}{re^{n-\sqrt{n}h(n)}}\to 0$$

$$\text{for } h(n)=n^{\epsilon},\epsilon\in(0,1/2).$$

It follows that $\frac{rU_n}{n^{\frac{1}{r}-1}}\xrightarrow{\mathcal{L}}\text{Exp}(1)$. Notice that b_n (and U_n) go to zero for $r>1$ and to ∞ for $r<1$.

Example 26.3.2 Let $S(x)=x^r$, $\Lambda(t)=e^{\frac{1}{r}e^t}-e^{\frac{1}{r}}$, $r>0$. It follows that $\tau(t)=r\log\left(e^{\frac{1}{r}e^t}-e^{\frac{1}{r}}\right)$, $\tau'(t)=\frac{e^{\frac{1}{r}e^t}}{e^{\frac{1}{r}e^t}-e^{\frac{1}{r}}}e^t\approx e^t$, $b_n\approx\frac{1}{n}$ $p_n=1-\frac{S(n-1)}{S(n)}=1-\frac{(n-1)^r}{n^r}\to 0$, $\frac{B(n)}{\sqrt{A(n)}}\to 0$ also holds. Since $\tau'(t)$ is strictly increasing for large t,

$$b_n\inf_{\xi\in I_n}\tau'(\xi)\approx\frac{1}{n}e^{\log(n-\sqrt{n}g_x(n))}\to 1 \text{ for } g_x(n)=n^{\epsilon},\ \epsilon\in(0,1/2).$$

Similarly, $b_n\sup_{\xi\in I_n}\tau'(\xi)\to 1$ for $g_x(n)$ as above. Now,

$$\Psi(t)=\log\left(r\log\left(t+e^{\frac{1}{r}}\right)\right),\ \Psi'(t)=\frac{1}{\left(t+e^{\frac{1}{r}}\right)\log\left(t+e^{\frac{1}{r}}\right)},$$

and $S^{-1}(x) = x^{\frac{1}{r}}$. Hence,

$$\sup_{\xi \in K_n} \Psi'(\xi) \approx e^{-\frac{n-\sqrt{n}h(n)}{r}} \frac{1}{\log\left(\frac{n-\sqrt{n}h(n)}{r}\right)},$$

$$\gamma_n \approx \sqrt{2e^{-\frac{n+\sqrt{n}h(n)}{r}} \log\left(\frac{n+\sqrt{n}h(n)}{r}\right) n e^{-\frac{n-\sqrt{n}h(n)}{r}} \frac{1}{\log\left(\frac{n-\sqrt{n}h(n)}{r}\right)}}$$

approaches 0 for $h(n) = n^\epsilon$, $\epsilon \in (0, 1/2)$. Therefore $nU_n \xrightarrow{\mathcal{L}} \text{Exp}(1)$.

Example 26.3.3 When $\tau(t) = \log t$, it follows that $\tau'(t) = \frac{1}{t}$ is strictly decreasing, $\tau^{-1}(t) = e^t$, $b_n = e^n$, and

$$b_n \sup_{\xi \in I_n} \tau'(\xi) = e^n e^{-(n-\sqrt{n}g_x(n))} = e^{\sqrt{n}g_x(n)} \to \infty \text{ for all } g_x \text{ with } g_x(n) \to \infty.$$

Hence, (26.13) fails to hold and Theorem 26.3.2 cannot be applied. This happens for the classical record model ($S(t) = t$) associated with the homogeneous Poisson pacing process ($\Lambda(t) = t$). In this case, recall from (26.3) that $U_n \stackrel{d}{=} A_n e^{C_n}$, where A_n ~Exp(1), C_n ~Gamma$(n, 1)$, and A_n and C_n are independent. Hence, $EU_n = \infty$ for all n, and indeed there cannot exist sequences a_n, b_n such that $\frac{U_n - a_n}{b_n}$ has an exponential limit.

Remark. When $p_n \to p \in (0, 1)$, techniques used in proving Lemma 26.3.2 and Theorem 26.3.1 do not work because $\frac{\Psi(W_n^*) - \Psi(T_{n+1})}{b_n} \stackrel{P}{\not\to} 0$. However, if we let $s(n) = c^n$ ($c > 1$), then $p_n = (c-1)c^{-1} \in (0, 1)$, and it follows from Bunge and Nagaraja (1992b) that $(1 - c^{-1})U_n \xrightarrow{\mathcal{L}} \text{Exp}(1)$. We conjecture therefore that there exist b_n's such that U_n/b_n has an exponential limit in a much larger class of combinations of α-structures and intensities of the Poisson process than given here.

Acknowledgement. The second author was supported in part by Fondo Nacional de Desarrollo Cientifico y Tecnologico (FONDECYT) grant no. 1990343 of Chile.

References

1. Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (1998). *Records*, New York: John Wiley.

2. Bunge, J. and Nagaraja, H. N. (1992a). Dependence structure of Poisson paced records, *Journal of Applied Probability*, **29**, 587–596.

3. Bunge, J. and Nagaraja, H. N. (1992b). Exact distribution theory for some point process record models, *Advances in Applied Probability*, **24**, 20–44.

4. Chung, K. L. (1974). *A Course in Probability Theory*, New York: Academic Press.

5. Hofmann, G. (1997). A Family of General Record Models, *Ph.D. Dissertation*, Ohio State University, Columbus, Ohio.

6. Hofmann, G. and Nagaraja, H. N. (1999). Random and Poisson paced record models in the F^{α} setup, *Journal of Applied Probablility* (to appear).

7. Nevzorov, V. B. (1985). Record and interrecord times for sequences of nonidentically distributed random variables, *Notes of Sci. Semin. LOMI*, **142**, 109–118. (English translation in *Journal of Soviet Mathematics*, **36**, 1987, 510-516.)

8. Nevzorov, V. B. (1986). The number of records in a sequence of nonidentically distributed random variables. In *Probability Distributions and Mathematical Statistics*, pp. 373–388, Fan, Tashkent. (English translation: *Journal of Soviet Mathematics*, 1987, 2375–2382.)

9. Nevzorov, V. B. (1995). Asymptotic distributions of records in nonstationary schemes, *Journal of Statistical Planning and Inference*, **45**, 261–273.

10. Pickands, J. (1971). The two dimensional Poisson process and extremal processes, *Journal of Applied Probability*, **8**, 745–756.

11. Yang, M. C. K. (1975). On the distribution of the inter-record times in an increasing population, *Journal of Applied Probability*, **12**, 148–154.

PART VII

Estimation of Parameters and Hypotheses Testing

27

Goodness-of-Fit Tests for the Generalized Additive Risk Models

Vilijandas B. Bagdonavičius and Mikhail S. Nikulin

University of Vilnius, Vilnius, Lithuania
Université Bordeaux-2, Bordeaux, France

Abstract: Goodness-of-fit tests for accelerated life models and generalizing or alternative to the additive risk model are proposed.

Keywords and phrases: Accelerated life testing, additive risk, goodness-of-fit, resource

27.1 Introduction

One of the models used to investigate the effects of stresses on failure occurrency is the additive risk (AR) model introduced by Aalen (1980). Under this model, the hazard rate function for the time to failure $T_{x(\cdot)}$ associated with a possibly time-varying stress $x(\cdot) = (x_1(\cdot), ..., x_m(\cdot))^T$ specifies the form

$$\alpha_{x(\cdot)}(t) = \alpha_0(t) + \beta^T x(t), \tag{27.1}$$

where $\alpha_0(\cdot)$ denotes the *unspecified* baseline hazard rate function and $\beta = (\beta_1, ..., \beta_m)^T$ is a vector of unknown regression parameters. This model is not natural in accelerated life testing for aging items. Really, if x_0 and x_1 are constant in time usual and accelerated stresses, respectively, that is, $S_{x_0}(t) \geq S_{x_1}(t)$ for all $t \geq 0$ and

$$x(t) = \begin{cases} x_1, & 0 \leq t < t_1, \\ x_0, & t \geq t_1, \end{cases}$$

is a step-stress, then under the AR model for all $t_1 \geq 0$ and $t \geq t_1$ we have $\alpha_{x(\cdot)}(t) = \alpha_{x_0}(t)$. So, for *any moment* t_1 of switch off from accelerated stress x_1 to the normal stress x_0, the hazard rate *after this moment* is the same as in

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

the case when the stress is normal until the moment t_1 and after it. So, it is not natural for aging items.

For the purpose of generalization, we formulate the AR model in other terms. Put

$$A_{x(\cdot)}(t) = \int_0^t \alpha_{x(\cdot)}(\tau)d\tau \quad \text{and} \quad A_0(t) = \int_0^t \alpha_0(v)dv.$$

Then the random variable $R = A_{x(\cdot)}(T_{x(\cdot)})$ has the standard exponential distribution which doesn't depend on $x(\cdot)$. So R is called the exponential resource and $A_{x(\cdot)}(t)$ is called the exponential resource used until the moment t under the stress $x(\cdot)$. In terms of resource usage, the AR model (27.1) can be written in the following manner :

$$A'_{x(\cdot)}(t) = A'_0(t) + \beta^T x(t). \tag{27.2}$$

This means that exponential resource usage rate under the stress $x(\cdot)$ at the moment t depends only on some baseline rate and the value of the stress in this moment.

Natural generalization of this model would be a model which states that the resource usage rate at the moment t could depend on the resource used up to this moment. We will consider two possible generalizations.

We formulate the *first generalized additive risk* (GAR) model

$$\alpha_{x(\cdot)}(t) = q\{A_{x(\cdot)}(t)\}(\alpha_0(t) + \beta^T x(t)), \tag{27.3}$$

where the nonnegative function q doesn't depend on $x(\cdot)$.

The function q can be specified or parametrized in some form as, for example,

$$q(u) = e^{\gamma u},\ \gamma \in \mathbf{R}; \quad q(u) = 1 + \gamma u,\ \gamma \geq 0; \quad q(u) = \frac{1}{1+\gamma u},\ \gamma \geq 0.$$

The function q can be specified by using relations between GAR and generalized additive (GA) [see Bagdonavičius and Nikulin (1995)] models. Really, suppose that G is some fixed survival function, $H = G^{-1}$ is the inverse function for G. If instead of $A_{x(\cdot)}(t) = -\ln S_{x(\cdot)}(t)$, we take $f^G_{x(\cdot)}(t) = (H \circ S_{x(\cdot)})(t)$, then the random variable $R^G = f^G_{x(\cdot)}(T_{x(\cdot)})$ has the survival function G, which doesn't depend on $x(\cdot)$ and R^G is called the G-resource used until the moment t. The AR model (27.2) can be generalized by considering the GA model:

$$\frac{\partial f^G_{x(\cdot)}(t)}{\partial t} = \frac{\partial f^G_0(t)}{\partial t} + \beta^T x(t), \tag{27.4}$$

where $\frac{\partial f^G_0(t)}{\partial t}$ denotes the unspecified baseline rate.

It is easy to show that the GA model is equivalent to the GAR model. Relations between the functions q and H are

$$q(u) = -e^u G' \circ H(e^{-u}), \quad H(u) = \int_0^{-\ln u} \frac{dv}{q(v)}. \tag{27.5}$$

So, through taking specified survival functions G, the function q can be specified.

Another possible generalization is obtained by supposing that the hazard rate at the moment t is influenced not only by the stress but also by the resource used until this moment:

$$\alpha_{x(\cdot)}(t) = \alpha_0(t) + \beta^T x(t) + \gamma A_{x(\cdot)}(t). \tag{27.6}$$

We'll call this model the *second generalized additive rate* model. Two-sample goodness-of-fit tests for the AR model was considered in Kim and Lee (1998). In this Chapter, we consider goodness-of-fit tests

(a) for the model (27.3) with specified q and the unidimensional stress;

(b) for the model (27.6) with possibly multidimensional stress.

27.2 Test for the First GAR Model Based on the Estimated Score Function

Consider construction of test for the GAR model when q is specified and $x(\cdot)$ is unidimensional.

Suppose that two groups of items are tested. The i-th group of n_i items is tested under a stress $x_i(\cdot)$; $x_1(\cdot)$ could be a constant in time under normal stress or a step-stress with alternating normal and accelerated constant stresses, $x_2(\cdot)$ could be a constant in time under accelerated or a step-stress with various alternating accelerated constant stresses.

Denote by T_{ij} and C_{ij} the failure and censoring times,

$$X_{ij} = T_{ij} \wedge C_{ij}, \quad \delta_{ij} = I\{T_{ij} \le C_{ij}\},$$

$$N_{ij}(t) = I\{T_{ij} \le t, \delta_{ij} = 1\}, \quad Y_{ij}(t) = I\{X_{ij} \ge t\},$$

$$N_i(t) = \sum_{j=1}^{n_i} N_{ij}(t), \quad Y_i(t) = \sum_{j=1}^{n_i} Y_{ij}(t), \quad J(t) = I(Y_1(t) + Y_2(t) > 0),$$

where I_A denotes the indicator of the event A. Then,

$$N(t) = N_1(t) + N_2(t) \quad \text{and} \quad Y(t) = Y_1(t) + Y_2(t)$$

are the numbers of observed failures in the interval $[0, t]$ and items "at risk" before the moment t, respectively, for the aggregated data. Denote by

$$\hat{A}_i(t) = \int_0^t \frac{dN_i(y)}{Y_i(y)}$$

the Nelson-Aalen estimator of the accumulated hazard function $A_i(t) = A_{x_i(\cdot)}(t)$,

$$\tilde{S}^{(0)}(u) = \sum_{i=1}^{2} Y_i(u) q(\hat{A}_i(u-)), \quad \tilde{S}^{(1)}(u) = \sum_{i=1}^{2} x_i(u) Y_i(u) q(\hat{A}_i(u-)),$$

$$\tilde{E}(u) = \frac{\tilde{S}^{(1)}(u)}{\tilde{S}^{(0)}(u)}.$$

Denote by $\mathbb{F} = \{\mathcal{F}_t, t \geq 0\}$ the filtration generated by the processes N_{ij} and Y_{ij} and suppose that censoring is independent, i.e. the $\mathbb{F}$-compensators Λ_{ij} of N_{ij} are absolutely continuous and

$$\Lambda_{ij}(t) = \int_0^u \alpha_{x_i}(u) Y_i(u) du.$$

Then [cf. Bagdonavičius and Nikulin (1995)] an estimator $\hat{\beta}$ of the parameter β can be obtained from estimating equations $U(\beta, \tau) = 0$, where

$$U(\beta, t) = \sum_{i=1}^{2} \int_0^t \{J(u)\{x_i(u) - \tilde{E}(u)\}\{dN_i(u) - \beta x_i(u) Y_i(u) q(\hat{A}_i(u-)) du\},$$

and it is in explicit form as

$$\hat{\beta} = \frac{\sum_{i=1}^{2} \int_0^t J(u)(x_i(u) - \tilde{E}(u)) dN_i(u)}{\sum_{i=1}^{2} \int_0^t J(u)(x_i(u) - \tilde{E}(u)) x_i(u) Y_i(u) q(\hat{A}_i(u-))}.$$

Denote by β_0 the true value of β.

Assumptions A:

(a) there exist nonnegative functions y_i, continuous and positive on $[0, \tau]$ such that

$$\sup_{0 \leq t \leq \tau} \left| \frac{Y_i(t)}{n} - y_i(t) \right| \xrightarrow{\mathbf{P}} 0, \quad \text{as} \quad n \to \infty, \qquad \frac{n_i}{n} \to l_i \in (0, 1);$$

(b) $A_0(\tau) < \infty$;

(c) the function $q(u)$ is positive and continuously differentiable on $[0, \tau]$;

(d) $S_{x_i(\cdot)}(\tau) > 0 \quad (i = 1, 2)$.

Under Assumptions A,

$$\sup_{u\in[0,\tau]} \mid \frac{1}{n}\tilde{S}^{(j)}(u) - s^{(j)}(u) \mid \xrightarrow{\mathbf{P}} 0, \quad n \to \infty,$$

$$\sup_{u\in[0,\tau]} \mid \tilde{E}(u) - e(u) \mid \xrightarrow{\mathbf{P}} 0, \quad n \to \infty,$$

where

$$s^{(0)}(u) = \sum_{i=1}^{2} y_i(u)q(A_i(u)); \quad s^{(1)}(u) = \sum_{i=1}^{2} x_i(u)y_i(u)q(A_i(u)),$$

$$e(u) = \frac{s^{(1)}(u)}{s^{(0)}(u)}.$$

Similar to Bagdonavičius and Nikulin (1997), it can be shown that

$$\frac{1}{\sqrt{n}}U(\beta_0, t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{2}\int_0^t h_i(v;\beta_0)dM_i(v) + o_p(1)$$

uniformly on $[0,\tau]$; here, $M_i(u) = N_i(u) - \alpha_{x_i}(u)Y_i(u)$ and

$$\begin{aligned} &h_i(u;\beta_0) \\ = \;& x_i(u) - e(u) + \frac{1}{y_i(u)}\int_u^t (x_i(v) - e(v))q'(A_i(v))y_i(v)(\alpha_0(v) + \beta x_i(v))dv. \end{aligned}$$

Put

$$s^{(2)}(u) = \sum_{i=1}^{2} x_i^2(u)y_i(u)q(A_i(u)), \quad v(u) = s^{(2)}(u) - s^{(1)}(u)e(u).$$

It can be verified that

$$\begin{aligned} &\frac{1}{\sqrt{n}}\hat{U}(t) \\ :=\;& \frac{1}{\sqrt{n}}U(\hat{\beta}, t) \\ =\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{2}\int_0^t J(u)\{x_i(u) \\ &- \tilde{E}(u)\}\{dN_i(u) - \hat{\beta}x_i(u)Y_i(u)q(\hat{A}_i(u-))du\} \\ =\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{2}\int_0^t h_i(u;\beta_0)dM_i(u) \\ &- \frac{1}{\sqrt{n}}\sum_{i=1}^{2}\int_0^\tau h_i(u;\beta_0)dM_i(u)(\int_0^t v(u)du)(\int_0^\tau v(u)du)^{-1} + o_p(1) \end{aligned}$$

uniformly on $[0,\tau]$. The predictable variation is

$$< \frac{1}{\sqrt{n}}\sum_{i=1}^{2}\int_0^t h_i(u;\beta_0)dM_i(u) >$$
$$= \frac{1}{n}\sum_{i=1}^{2}\int_0^t h_i^2(u;\beta_0)Y_i(u)q(A_i(u))(\alpha_0(u)+\beta x_i(u))du \xrightarrow{\mathbf{P}} g(t,\beta_0),$$

where

$$g(t,\beta_0) := \sum_{i=1}^{2}\int_0^t h_i^2(u;\beta_0)y_i(u)q(A_i(u))(\alpha_0(u)+\beta x_i(u))du.$$

Put

$$a(t) = \int_0^t v(u)du.$$

We have

$$\frac{1}{\sqrt{n}}\hat{U}(\cdot) \xrightarrow{\mathcal{D}} W(g(\cdot,\beta_0)) - \frac{a(\cdot)}{a(\tau)}W(g(\tau,\beta_0)) \quad \text{in} \quad D[0,\tau],$$

where W denotes the standard Wiener process. Then

$$\frac{\hat{U}(\cdot)}{\sqrt{ng(\tau)}} \xrightarrow{\mathcal{D}} W\left(\frac{g(\cdot)}{g(\tau)}\right) - \frac{a(\cdot)}{a(\tau)}W(1), \quad \text{in} \quad D[0,\tau].$$

Denote

$$\psi(u) = \frac{a(g^{-1}(g(\tau)u))}{a(\tau)}, \quad u \in [0,1].$$

Then

$$\sup_{t\in[0,\tau]} \mid \frac{\hat{U}(t)}{\sqrt{ng(\tau)}} \mid \xrightarrow{\mathcal{D}} \sup_{u\in[0,1]} \mid W(u) - \psi(u)W(1) \mid = V_\psi.$$

The function $\psi : [0,1] \to [0,1]$ is increasing, with $\psi(0) = 0$ and $\psi(1) = 1$. We have

$$T = \sup_{t\in[0,\tau]} \mid \frac{\hat{U}(t)}{\sqrt{n\hat{g}(t)}} \mid \xrightarrow{\mathcal{D}} V_\psi,$$

where

$$\hat{g}(t) = \frac{1}{n}\sum_{i=1}^{2}\int_0^t H_i^2(u;\hat{\beta})Y_i(u)q(\hat{A}_i(u))(d\hat{A}_0(u)+\hat{\beta}x_i(u)du),$$

$$H_i(u;\hat{\beta}) = x_i(u) - \tilde{E}(u) + \frac{1}{Y_i(u)}\int_u^t (x_i(v)-\tilde{E}(v))q'(\hat{A}_i(v))Y_i(v)$$
$$\left(d\hat{A}_0(v) + \hat{\beta}x_i(v)dv\right),$$

$$d\hat{A}_0(v) = \frac{dN(v) - \hat{\beta}\tilde{S}^{(1)}(v)dv}{\tilde{S}^{(0)}(v)}.$$

Denote by $V_{\psi,\alpha}$ the α-quantile of the random variable V_ψ,

$$\hat{\psi}(u) = a(\hat{g}^{-1}(\hat{g}(\tau)u))/a(\tau)),$$

where

$$\hat{g}^{-1}(s) = \sup\{u : \hat{g}(u) < s\}.$$

The quantiles $V_{\psi,\alpha}$ can be approximated by $V_{\hat{\psi},\alpha}$ which can be obtained by simulating the standard Wiener process in the jump points of $\hat{\psi}$. The approximate critical region with the significance level α is $T > V_{\hat{\psi},1-\alpha}$.

27.3 Tests for the Second GAR Model

Suppose that data are the same as in Section 27.2. Consider the model

$$\alpha_{x_i(\cdot)}(t) = \alpha_0(t) + \beta^T x_i(t) + \gamma A_{x_i(\cdot)}(t),$$

where $x_i(t) = (x_{i1}(t), ..., x_{im}(t))^T$, $\beta = (\beta_1, ..., \beta_m)^T$ and α_0 is unknown. Put $\theta = (\beta^T, \gamma)^T$, $z_i(t) = (x_i^T(t), A_{x_i(\cdot)}(t))^T$. So we treat the accumulated hazard rate function $A_{x_i(\cdot)}(t)$ as an additional covariate. Using the method of Lin and Ying (1994) generalized by Bagdonavičius and Nikulin (1995), we define the weighted estimator $\hat{\theta}_K$ of the parameter θ as the solution of estimation equations $U_K(\theta, \tau) = 0$, where

$$U_K(\theta, t) = \sum_{i=1}^{2} \int_0^t K(u)\{\tilde{z}_i(u) - \tilde{E}(u)\}\{dN_i(u) - Y_i(u)\theta^T \tilde{z}_i(u)du\},$$

where $\tilde{z}_i(u) = (x_i(t), \hat{A}_{x_i(\cdot)}(t))^T$, $\tilde{E}(u) = \sum_i \tilde{z}_i(u)Y_i(u)/Y(u)$ and the weight function $K(u)$ is a $\mathbb{F}$-predictable stochastic process that converges in probability to a nonnegative bounded function $K(u)$ uniformly in $u \in [0, \tau]$. For example, $K(u) = e^{-\hat{A}_0(u)}$, where $\hat{A}_0(u)$ is an estimator of

$$A_0(u) = \int_0^u \alpha_0(v)dv.$$

Denote by θ_0 the true value of θ. Under the model (27.6), both the unweighted estimator $\hat{\theta}_I$ (obtained when $K(u) = I(u) \equiv 1$) and the weighted estimator $\hat{\theta}_K$ ($K \neq I$) are asymptotically normal with the same mean θ_0. Under alternatives, both estimators $\hat{\theta}_I$ and $\hat{\theta}_K$ should be also asymptotically normal but with different means, and so a test statistic may be constructed in terms of the difference $\hat{\theta}_I - \hat{\theta}_K$.

Denote by

$$A_k = A_k(\tau) = -\sum_{i=1}^{2}\int_0^{\tau} k(u)(z_i(u) - e(u))y_i(u)z_i^T(u)du,$$

where $e(u) = \sum_i z_i(u)y_i(u)/y(u)$.

Proposition 27.3.1 *Suppose that the model (27.6) is true. Then under assumptions (a), (b), (d), and nonsingularity of A_k and A_I,*

$$n^{1/2}(\hat{\theta}_K - \hat{\theta}_I) \xrightarrow{\mathcal{D}} N(0, \Sigma_{kI}^{**}),$$

where

$$\Sigma_{kI}^{**} = \Sigma_{kk}^{*} - \Sigma_{kI}^{*} - \Sigma_{Ik}^{*} + \Sigma_{II}^{*}, \quad \Sigma_{kI}^{*} = A_k^{-1}\Sigma_{kI}(A_I^{-1})^T,$$

$$\Sigma_{kI} = \sum_{i=1}^{2}\int_0^t h_{ki}(v,\theta_0)h_{Ii}(v,\theta_0)(\alpha_0(v) + \theta^T z_i(v))dv,$$

$$h_{ki}(v;\theta_0) = k(v)(z_i(v) - e(v)) + \frac{\gamma}{y_i(v)}\int_v^t k(u)(z_i(u) - e(u))y_i(u)du.$$

SKETCH OF THE PROOF. Denote by

$$E(u) = \sum_i z_i(u)Y_i(u)/Y(u), \quad M_i(u) = N_i(u) - \int_0^u Y_i(v)\alpha_0(v)dv.$$

Using methods similar to those of Bagdonavičius and Nikulin (1997), we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}U_K(\theta,t) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{2}\int_0^t K(u)\{z_i(u) - E(u)\}\{Y_i(u)dA_0(u) \\
&\quad + \theta^T Y_i(u)(z_i(u) - \tilde{z}_i(u))du + dM_i(u)\} \\
&\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^{2}\int_0^t K(u)\{(\tilde{z}_i(u) - z_i(u)) \\
&\quad + (E(u) - \tilde{E}(u))\}\{Y_i(u)dA_0(u) \\
&\quad + \theta^T Y_i(u)(z_i(u) - \tilde{z}_i(u))du + dM_i(u)\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{2}\int_0^t h_{ki}(v)dM_i(v) + o_p(1)
\end{aligned}$$

uniformly on $[0,\tau]$, where

$$h_{ki}(v) = k(v)(z_i(v) - e(v)) + \frac{\gamma}{y_i(v)}\int_v^t k(u)(z_i(u) - e(u))y_i(u)du.$$

The predictable covariation is

$$< \frac{1}{\sqrt{n}}U_k, \frac{1}{\sqrt{n}}U_I > (t, \theta_0)$$
$$= \frac{1}{n}\sum_{i=1}^{2}\int_0^t h_{ki}(v,\theta_0)h_{Ii}(v,\theta_0)(\alpha_0(v)+\theta^T z_i(v))dv + o_p(1) \xrightarrow{\mathbf{P}} \Sigma_{kI}$$

and

$$\hat{\theta}_K \xrightarrow{\mathbf{P}} \theta_0, \quad \hat{\theta}_I \xrightarrow{\mathbf{P}} \theta_0.$$

So, taking into account that

$$\hat{A}_K = -\frac{1}{n}\frac{\partial U_k}{\partial \theta} = -\frac{1}{n}\int_0^\tau K(u)(\tilde{z}_i(u)-\tilde{E}(u))Y_i(u)\tilde{z}_i^T(u)d(u) \xrightarrow{\mathbf{P}} A_k,$$

we have

$$\mathbf{cov}(\sqrt{n}(\hat{\theta}_K-\theta_0), \sqrt{n}(\hat{\theta}_I-\theta_0)) \to A_k^{-1}\Sigma_{kI}(A_I^{-1})^T = \Sigma_{kI}^* \quad \text{as} \quad n \to \infty.$$

The limiting covariance matrix of $\sqrt{n}(\hat{\theta}_K - \hat{\theta}_I)$ is Σ_{kI}^{**}.

The test statistic can be defined as

$$T_n = (\hat{\theta}_K - \hat{\theta}_I)^T(\hat{\Sigma}_{KI}^{**})^{-1}(\hat{\theta}_K - \hat{\theta}_I),$$

where

$$\begin{aligned}
\hat{\Sigma}_{KI}^{**} &= \hat{\Sigma}_{KK}^* - \hat{\Sigma}_{KI}^* - \hat{\Sigma}_{IK}^* + \hat{\Sigma}_{II}^*, \quad \hat{\Sigma}_{KI}^* = \hat{A}_K^{-1}\hat{\Sigma}_{KI}(\hat{A}_I^{-1})^T, \\
\hat{\Sigma}_{KI} &= \frac{1}{n}\sum_{i=1}^{2}\int_0^\tau H_{Ki}(v,\hat{\theta})H_{Ii}(v,\hat{\theta})\{d\hat{A}_0(v)+\hat{\theta}^T\tilde{z}_i(v)dv\}, \\
H_{Ki}(v;\theta_0) &= K(v)(\tilde{z}_i(v)-\tilde{E}(v)) + \frac{\hat{\gamma}}{Y_i(v)}\int_v^t K(u)\{\tilde{z}_i(u)-\tilde{E}(u)\}Y_i(u)du, \\
d\hat{A}_0(v) &= \frac{dN(v)}{Y(v)} - \tilde{E}(v)dv.
\end{aligned}$$

The distribution of the statistic T_n is approximated by the chi-square distribution with $(m+1)$ degrees of freedom.

References

1. Aalen, O. (1980). A model for nonparametric regression analysis of counting processes, In *Mathematical Statistics and Probability Theory, Lecture Notes in Statistics* (Eds., W. Klonecki, A. Kozek and J. Rosinski), **2**, pp. 1–25, New York: Springer-Verlag.

2. Bagdonavičius, V. and Nikulin, M. (1995). *Semiparametric Models in Accelerated Life Testing*, Queen's Papers in Pure and Applied Mathematics **98**, Kingston, Ontario, Canada.

3. Bagdonavičius, V. and Nikulin, M. (1997). Semiparametric estimation in generalized additive-multiplicative model, *Proceedings of the Steklov Mathematical Institute*, St. Petersburg, **244**, 7–27.

4. Kim, J. and Lee, S.-Y. (1998). Two-sample goodness-of-fit tests for additive risk models with censored observations, *Biometrika*, **85**, 593–603.

5. Lin, D. Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model, *Biometrika*, **81**, 61–71.

28

The Combination of the Sign and Wilcoxon Tests for Symmetry and Their Pitman Efficiency

G. Burgio and Ya. Yu. Nikitin

University of Roma "La Sapienza," Rome, Italy
St. Petersburg State University, St. Petersburg, Russia

Abstract: We consider the efficiency properties of the test of symmetry based on the sequence of statistics $G_n = aS_n + bW_n$, where S_n and W_n are the classical sign and Wilcoxon statistics and a and b are some constants. Pitman efficiency of this test with respect to the student test in the case of location alternatives and its maximum with respect to a and b are calculated. This generalizes the classical results of Hodges and Lehmann (1956). We also find the form of the alternative distributions under which the new test is Pitman optimal.

Keywords and phrases: Pitman efficiency, sign test, Wilcoxon test, U-statistic, rate of convergence, location alternative

28.1 Introduction

Consider the classical problem of testing of symmetry with respect to zero of the univariate sample $X_1, ..., X_n$. Most known and simple test statistics for this problem are the sign statistic

$$S_n = n^{-1} \sum_{i=1}^{n} 1_{\{X_i > 0\}}$$

and the Wilcoxon signed rank statistic which is asymptotically equivalent to the statistic

$$W_n = (C_n^2)^{-1} \sum_{1 \le i < j \le n} 1_{\{X_i + X_j > 0\}}.$$

Their properties are well explored and described in many sources; see, for example, Hájek and Šidák (1967), Lehmann (1975) and Hettmansperger (1984).

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

The aim of the present chapter is to consider the following combination G_n of these statistics:

$$G_n = aS_n + bW_n$$

for some constants a and b, $a^2 + b^2 > 0$. For definiteness, let us consider only $b \geq 0$.

The idea of taking the linear combination of S_n and W_n to obtain a more flexible test with better power and efficiency properties is certainly not new; see Hemelrijk (1950) and Hájek and Šidák (1967), and Section 3.5 of the latter for the discussion of early papers on the subject. A typical example of subsequent publications is the article of Doksum and Thompson (1971). They considered the statistic equivalent to

$$DT_n = W_n - (1/2)S_n \tag{28.1}$$

and established its asymptotic minimax properties for a special class of alternatives. Doksum and Thompson (1971) refer also to some other papers with similar considerations. Interesting examples of estimation via linearly combining two given statistics are presented in Baksalary and Kala (1983) and in Gross (1998).

However it seems that the general statistic G_n for testing of symmetry with arbitrary coefficients a and b has never been investigated from the point of view of Pitman efficiency and optimality.

Our approach is primarily based on the theory of U-statistics. This theory was initiated by Hoeffding (1948) and was strongly developed in last few decades. Considering G_n as a seqence of U-statistics with the kernel

$$\Phi_{a,b}(s,t) = a(1_{\{s>0\}} + 1_{\{t>0\}})/2 + b1_{\{s+t>0\}}, \tag{28.2}$$

we prove its asymptotic normality under arbitrary distribution of the initial sample. This enables us to find the general formula for the Pitman efficacy of the proposed test for contiguous parametric alternatives and to make some calculations for the location alternative generalizing the classical results known for the Pitman efficiency of the sign and Wilcoxon tests from the paper of Hodges and Lehmann (1956).

For any density f_0 of the observations under the null-hypothesis, we find the best values of the constants a and b for which the test based on G_n has the highest possible Pitman efficiency with respect to the t-test. This efficiency is essentially higher than the efficiency of S_n and W_n alone, and this is a serious argument in favor of the proposed combined test.

Moreover, using the results of Rao (1963, 1965), we explore the problem of asymptotic optimality of the proposed test. Under certain regularity conditions and some restrictions on parameters a and b, it is found that the test based on G_n is Pitman optimal iff the initial observations have the symmetric density

$$f_{a,b}(x) = \frac{\nu(b+a)(b+a/2)\exp(\nu|x|)}{((b+a)\exp(\nu|x|)+b)^2}, \nu > 0, x \in R^1, a+b > 0.$$

The particular cases of this three-parameter family of densities are the Laplace density when $a = 1, b = 0$ and the logistic density when $a = 0, b = 1$. It is well known that the sign and Wilcoxon test are locally most powerful and Pitman optimal signed-rank tests for these two densities in case of location alternatives [see, for example, Hájek and Šidák (1967) or Hettmansperger (1984)]. Thus, our characterization generalizes these classical results.

28.2 Asymptotic Distribution of the Statistic G_n

It is clear that the statistic G_n can be represented in the form

$$G_n = (C_n^2)^{-1} \sum_{1 \le i < j \le n} \Phi_{a,b}(X_i, X_j),$$

where the kernel $\Phi_{a,b}(s,t)$ is given by (28.2).

Denote by F the distribution function and by f the density of the initial observations $X_1, X_2, \ldots$. Let us find the expressions for some standard characteristics of this kernel. First of all, we need

$$\begin{aligned} \mu_{a,b} &= E_F \Phi_{a,b}(X_1, X_2) = aP_F(X_1 > 0) + bP_F(X_1 + X_2 > 0) \\ &= a(1 - F(0)) + b(1 - F * F(0)), \end{aligned}$$

where $*$ is the symbol of convolution.

We are also interested in the so-called canonical function of the kernel given by

$$\begin{aligned} \Psi_{a,b}(t) &= E_F[\Phi_{a,b}(X_1, X_2) | X_1 = t] \\ &= a(1_{\{t>0\}} + 1 - F(0))/2 + b(1 - F(-t)). \end{aligned}$$

Clearly,

$$E_F \Psi_{a,b}(X_1) = a(1 - F(0)) + b \int_{-\infty}^{+\infty} (1 - F(-x)) dF(x).$$

We can now calculate the variance

$$\begin{aligned} \sigma_{a,b}^2 = Var_F \Psi_{a,b}^2(X_1) = \quad & E_F\{a(1_{\{X_1>0\}} - (1 - F(0))/2 \\ & + b(1 - F(-X_1) - \textstyle\int_{-\infty}^{+\infty} (1 - F(-y)) dF(y))\}^2. \end{aligned}$$

Suppose that for given a, b and F we have $\sigma_{a,b}^2 > 0$ so that the kernel $\Phi_{a,b}$ is nondegenerate. Then the distribution of the U-statistic G_n may be approximated by the normal law. Using the central limit theorem for U-statistics from Hoeffding (1948) or Denker (1985), we see that uniformly in x

$$P_F(\sqrt{n}(G_n - \mu_{a,b})/2\sigma_{a,b} < x) \longrightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-u^2/2) du, \quad n \to \infty. \quad (28.3)$$

Relying on (28.3), we may construct the symmetry test based on G_n for sufficiently large samples.

28.3 Pitman Efficiency of the Proposed Statistic

In order to make the Pitman efficiency calculation, we should make the formulation of the statistical problem quite precise. Suppose that under the null hypothesis of symmetry H_0 the initial distribution function F is absolutely continuous and is symmetric with respect to zero, hence for every x

$$1 - F(x) - F(-x) = 0.$$

We suppose that under the alternative H_1 the observations have the common distribution function $F(x,\theta), \theta \geq 0$, such that $F(x,\theta) = F_0(x)$ for some symmetric distribution function F_0 with continuous density f_0 only for $\theta = 0$. For simplicity, we consider only the case of location alternative when

$$F(x,\theta) = F_0(x - \theta).$$

Other smooth parametric families may be treated in an analogous way as shown in Nikitin (1995, Ch. 6).

As usual, when calculating the Pitman efficiency we take the parameter θ in the form $\theta = \theta_n = \delta/\sqrt{n}$ for some $\delta \geq 0$. It is assumed that the condition

$$\int_{-\infty}^{+\infty} f_0^2(y)dy < \infty \tag{28.4}$$

is valid all along this work.

In the case of location alternative, the expressions for $\mu_{a,b}$ and $\sigma_{a,b}^2$ become simpler if we use the symmetry of F_0 as follows:

$$\begin{aligned}
\mu_{a,b}(\theta) &= aF_0(\theta) + bF_0 * F_0(2\theta), \\
\sigma_{a,b}^2(\theta) &= a^2 F_0(\theta)(1 - F_0(\theta))/4 \\
&\quad + ab \textstyle\int_0^{+\infty} F_0(y+\theta)dF_0(y-\theta) - abF_0(\theta) \int_{-\infty}^{+\infty} F_0(y+\theta)dF_0(y-\theta) \\
&\quad + b^2 (\textstyle\int_{-\infty}^{+\infty} F_0^2(y+\theta)dF_0(y-\theta) - (\int_{-\infty}^{+\infty} F_0^2(y+\theta)dF_0(y-\theta))^2).
\end{aligned}$$

Using Lemma 3.4 from Mehra and Sarangi (1967), we obtain due to condition (28.4) that

$$\frac{d}{d\theta}\mu_{a,b}(\theta) = af_0(\theta) + 2b\int_{-\infty}^{+\infty} f_0(y - 2\theta)f_0(y)dy. \tag{28.5}$$

Because of uniform convergence with respect to θ of the integral in (28.5), this expression is continuous in θ and hence

$$\mu'_{a,b}(0) = af_0(0) + 2b\int_{-\infty}^{+\infty} f_0^2(y)dy.$$

In the sequel, we always assume that a and b are such that

$$\mu'_{a,b}(0) = af_0(0) + 2b\int_{-\infty}^{+\infty} f_0^2(y)dy > 0 \tag{28.6}$$

(this condition ensuring the consistency of our test is always required when calculating the Pitman efficiency). Moreover, it is quite clear that the function $\theta \to \sigma^2_{a,b}(\theta)$ is continuous in θ and

$$\sigma^2_{a,b}(0) = a^2/16 + b^2/12 + 3ab/8 - ab/4 = (3a^2 + 6ab + 4b^2)/48 > 0. \tag{28.7}$$

We underline (it will be used in Section 28.4) that for sufficiently small θ one still has $\sigma^2_{a,b}(\theta) > 0$ and hence our statistic is asymptotically normal both under H_0 and H_1 for small θ. We see that all requirements necessary for the existence of Pitman efficiency are fulfilled [see, for example, Lehmann (1975) or Hettmansperger (1984)].

The measure of Pitman efficiency called *Pitman efficacy* equals therefore

$$\kappa^2(f_0) = (\mu'_{a,b}(0)/2\sigma_{a,b}(0))^2 = \frac{12(af_0(0) + 2b\int_{-\infty}^{+\infty} f_0^2(y)dy)^2}{3a^2 + 6ab + 4b^2}.$$

It is clear that the simple particular cases of this formula are the well-known efficacies $4f_0^2(0)$ for the sign test and $12(\int_{-\infty}^{+\infty} f_0^2(y)dy)^2$ for the Wilcoxon test if we take $a = 1, b = 0$ and $a = 0, b = 1$ correspondingly.

Now it is possible to compare our test with any other with known efficacy as Pitman efficiency is the ratio of efficacies.

The famous example first discussed by Hodges and Lehmann (1956) is the Pitman efficiency of the sign test and the Wilcoxon test with respect to the t-test. Arguing as in Hodges and Lehmann (1956), we get for the Pitman efficiency $e^P(G,t)$ the following formula:

$$e^P(G,t) = 12\sigma^2(af_0(0) + 2b\int_{-\infty}^{+\infty} f_0^2(y)dy)^2/(3a^2 + 6ab + 4b^2), \tag{28.8}$$

where σ^2 stands for the variance of the underlying density f_0. In case of standard normal density $f_0(x) = (2\pi)^{-1/2}\exp(-x^2/2)$, we have

$$e^P(G,t) = l(a,b) = 6(a + \sqrt{2}b)^2/[\pi(3a^2 + 6ab + 4b^2)].$$

Let us find the maximal value of this expression. As argued above, we may consider only the case when $a^2 + b^2 > 0$. Put $z = a/b$ and consider the auxiliary function

$$r(z) = (z + \sqrt{2})^2/(3z^2 + 6z + 4).$$

Clearly, it attains its maximal value when

$$z = z^* = (4 - 3\sqrt{2})/(3\sqrt{2} - 3) \approx -0.1953 < 0.$$

Substituting this value in $r(z)$, we obtain

$$\max_z r(z) = (10/3) - 2\sqrt{2}.$$

Consequently

$$\max e^P(G, t) = (20 - 12\sqrt{2})/\pi \approx 0.9643$$

and this maximum is attained for $a = bz^*$. It is well known from Hodges and Lehmann (1956) or Lehmann (1975) that in the case of the normal law, the Pitman efficiency of the sign test and the Wilcoxon test with respect to Student's test are correspondingly $2/\pi \approx 0.6366$ and $3/\pi \approx 0.9549$. Hence, we can improve by approximately 0.01 the already very high value of efficiency of Wilcoxon test due to the combination effect.

In the general case, similar arguments show that the maximum value in a and b of Pitman efficacy $\kappa^2(f_0)$ in terms of $u = f_0(0)$ and $v = \int_{-\infty}^{+\infty} f_0^2(y)dy$ is attained for

$$a/b = (6v - 4u)/(3u - 6v)$$

and is equal to

$$16(3v^2 - 3vu + u^2). \tag{28.9}$$

Hence, the relative Pitman efficiency of the "best" test statistic with respect to a and b (let us denote it by G_n^*) with respect to the t-test is

$$16\sigma^2(3v^2 - 3vu + u^2). \tag{28.10}$$

It is an interesting unsolved problem of finding the minimum in symmetric f_0 of this efficiency. This maxmin value would be the analogue of the famous lower bound 0.864 from Hodges and Lehmann (1956).

Now we want to calculate the Pitman efficiency of a best combined test given by (28.10) for the case of a very interesting family of densities, namely, for the normal distributions of order p with the density

$$f_p(x; \theta, \sigma^2) = (2p^{1/p}\Gamma(1 + 1/p)\sigma_p)^{-1} \exp(-|x - \theta|^p / p\sigma^p),$$

where

$$\sigma_p = \sigma p^{-1/p}[\Gamma((1/p)/\Gamma(3/p)]^{1/2},$$

$x \in R^1$, $\sigma > 0$ and θ are arbitrary real parameters and the "structural" parameter p is from $[1,\infty)$. Clearly, for $p = 2$ we have the classical normal distribution and for $p = 1$ the double-exponential or the Laplace distribution. See Burgio (1996) and Burgio and Nikitin (1998) for other efficiency calculations connected with this distribution.

Simple calculations give

$$f_p(0) = p\Gamma^{1/2}(3/p)/2\sigma\Gamma^{3/2}(1/p),$$
$$\int_{-\infty}^{+\infty} x^2 f_p(x)dx = \sigma^2,$$

$$\int_{-\infty}^{+\infty} f_p^2(x)dx = p\Gamma^{1/2}(3/p)/\sigma 2^{1+1/p}\Gamma^{3/2}(1/p).$$

Substituting these values in (28.9), we obtain for this family that

$$e^P(G,t) = 2^{2-2/p}p^2\Gamma(3/p)(2^{2/p} - 3\ 2^{1/p} + 3)/\Gamma^3(1/p).$$

Clearly, in the normal case of $p = 2$, we obtain the value of efficiency 0.9643 obtained above.

Now let us compare in efficiency the "best" test based on G_n^* with the efficacy (28.9) with the sign test and the Wilcoxon test.

We confine ourselves to weakly unimodal symmetric densities f_0 [see Hodges and Lehmann (1956, p. 327)]. Consider the density-quantile function

$$\phi(u) = f_0(F_0^{-1}(u)),\ 0 \le u \le 1. \tag{28.11}$$

The inequality

$$v = \int_{-\infty}^{+\infty} f_0^2(y)dy = \int_0^1 \phi(x)dx = 2\int_{1/2}^1 \phi(x)dx \le \phi(1/2) = f_0(0) = u$$

shows that either $u/v \ge 1$ or $v/u \le 1$. Hence, the Pitman efficiency of the "best" combined test with respect to the sign test satisfies the inequality

$$e(G^*, S) = 4(3v^2 - 3uv + u^2)/u^2 \ge 1,$$

the equality being attained for $v/u = 1/2$.

Quite analogously, in case of the Wilcoxon test we have

$$e(G^*, W) = 4(3v^2 - 3uv + u^2)/3v^2 \ge 1$$

with the equality for $u/v = 3/2$.

In practice, we do not know the true values of u and v. But we can estimate them from data. Denote the corresponding estimators u_n and v_n. It is an open question what the properties are of the test of combined type with

$$a/b = (6v_n - 4u_n)/(3u_n - 6v_n).$$

We may suppose that they are similar to that of G_n^*.

28.4 Basic Inequality for the Pitman Power

Consider again the sequence of i.i.d. observations $X_1, X_2, \ldots$ taking values in R^1 and having the common density $f(x,\theta), \theta \geq 0$. Denote by $\gamma_n(\theta)$ the power function, evaluated at the point $\theta = \theta_0 + \delta n^{-1/2}$, of a test for the hypothesis $H_0 : \theta = \theta_0$, based on $X_1, \ldots, X_n$, and with significance level $\alpha \in (0, 1)$. It is proved in Rao (1963), under some regularity conditions imposed on $f(x,\theta)$ and discussed below, that for every fixed $\delta \neq 0$

$$\limsup_{n\to\infty} \gamma_n(\theta_0 + \delta n^{-1/2}) \leq 1 - \Phi(z_\alpha - \delta\sqrt{i}), \tag{28.12}$$

where z_α is the quantile of order $1-\alpha$ of the standard normal distribution function Φ, and i is the Fisher information in the point θ_0, that is,

$$i = \int_{R^1} (\frac{\partial \ln f(x,\theta)}{\partial\theta})^2 f(x,\theta)dx \,|_{\theta=\theta_0}.$$

We will always assume that $0 < i < +\infty$.

The limit of the sequence of functions $\gamma_n(\theta_0 + \delta n^{-1/2})$, if it exists, is known as the Pitman power of the test and the inequality (28.12) gives the upper bound for the Pitman power of any test of level α for testing H_0.

We now turn to check the fulfilment of regularity assumptions given by Rao (1963), that are sufficient for the inequality (28.12). All these conditions are rather similar requirements imposed on the density $f(x,\theta)$. Therefore, we discuss only Assumption I from Rao (1963). The reader can find there the formulation of remaining conditions. See also Conti and Nikitin (1997) for similar research in the case of testing for independence.

For the sake of brevity, let us introduce the quantity

$$a(x,\theta) = \partial \ln f(x,\theta)/\partial\theta.$$

With these notation, the Fisher information is given by

$$i(\theta) = E_\theta a^2(X,\theta).$$

The first set of conditions in Rao (1963) looks as follows.

Assumption I. As $\theta \to 0$,

$$\begin{array}{ll} (i) & E_\theta a(X,0) = \theta i + o(\theta) \\ (ii) & Var_\theta a(X,0) = i + o(1) \\ (iii) & Cov_\theta[a(X,\theta), a(X,0)] = i + o(1). \end{array}$$

We see that this is a set of regularity conditions valid for many densities such as normal, Cauchy, and logistic. The same is true in case of other regularity assumptions.

Denote by $\mathcal{F}$ the class of densities $f(x, \theta)$ for which all these conditions are true. Hence, the inequality (28.12) is true for this class.

28.5 Pitman Power for G_n

In this Section, we find the Pitman power for G_n using general results from Section 7a.7 of Rao (1965). As a consequence of condition (28.6), we consider only one-sided tests of H_0, assuming significant large values of G_n. The generalization for the two-sided case is straightforward.

Denote by $\gamma_n^G(\theta)$ the power function of the test of H_0 against H_1 with significance level α in the point θ for the statistic G_n.

Our next aim is to prove that

$$\lim_{n\to\infty} \gamma_n^G(\delta/\sqrt{n}) = 1 - \Phi(z_\alpha - \delta\kappa(f_0)), \tag{28.13}$$

where

$$\kappa(f_0) = \sqrt{12}[af_0(0) + 2b\int_{-\infty}^{+\infty} f_0^2(y)dy]/\sqrt{3a^2 + 6ab + 4b^2}.$$

This relationship follows from the following theorem that is the combination of Conditions (II) and (III) in Section 7a.7 of Rao (1963).

Theorem 28.5.1 *Let T_n be a test statistic with the critical region of the form $\{T_n \geq \lambda_n\}$. Suppose that*

(a) $\lim_{n\to\infty} P_0(T_n \geq \lambda_n) = \alpha$ *for fixed* $\alpha \in (0, 1)$*;*

(b) There exist such functions $\mu(\theta)$ and $\sigma(\theta)$ that

$$P_\theta(\sqrt{n}(T_n - \mu(\theta)) < y\sigma(\theta)) = \Phi(y)$$

uniformly in $0 \leq \theta \leq \tau$, where $\tau > 0$ is arbitrary small;

(c) $\mu(\theta)$ has a positive derivative $\mu'(0)$ in the point 0 and $\sigma(\theta)$ is continuous in this point.

Then

$$\lim_{n\to\infty} \gamma_n(\delta/\sqrt{n}) = 1 - \Phi(z_\alpha - \delta\mu'(0)/\sigma(0)),$$

In order to verify the conditions of this theorem for the statistic G_n, observe that the asymptotic normality of the sequence G_n both under hypothesis and close alternative follows from (28.3) since the variance $\sigma^2_{a,b}(\theta)$ is positive. It follows that we can take as λ_n from (a) the value

$$\begin{aligned}\lambda_n &= \mu_{a,b}(0) + 2z_\alpha \sigma_{a,b}(0)\sqrt{n}(1+o(1)) \\ &= (a+b)/2 + z_\alpha \sqrt{n(3a^2+6ab+4b^2)/12}(1+o(1)), \quad n \to \infty.\end{aligned}$$

Moreover, the convergence in distribution to the normal law is uniform for sufficiently close alternatives because of the Berry-Esseen-type estimate for the rate of convergence in the Central Limit Theorem for nondegenerate U-statistics obtained by Callaert and Janssen (1978). Hence, we have (b). Finally, (c) follows from (28.6) and (28.7).

Comparing (28.12) and (28.13), we easily obtain the inequality

$$\kappa^2(f_0) = 12(af_0(0) + 2b\int_{-\infty}^{+\infty} f_0^2(y)dy)^2/(3a^2+6ab+4b^2) \le i. \quad (28.14)$$

In the left-hand side of (28.14), we recognize the Pitman efficacy of G_n as obtained above. The statistic G_n is Pitman-optimal when the right-hand side of (28.14) is equal to 1.

28.6 Conditions of Pitman Optimality

In this Section, we will explore the following rather natural question: under what density f_0 the sequence of statistic G_n is Pitman optimal? In other words, we should find for which f_0 the inequality (28.14) turns out to be equality. Clearly, we need consider only the densities f_0 from $\mathcal{F}$ with finite Fisher information i.

Analogous problems were discussed and solved in Nikitin (1995, Ch. 6) in another context connected with Bahadur efficiency.

To simplify the problem, let use once again the auxiliary function (28.10). It is known from Lemma 4.2.1 of Albers (1974), that if (28.4) holds and $i < \infty$ then

$$\phi(0) = \phi(1) = 0. \quad (28.15)$$

The inequality (28.14) in terms of ϕ may be rewritten as

$$12\left(\int_0^1 (a\phi'(u)1_{[0,1/2]}(u) + 2b\phi(y))dy\right)^2 \Big/ (3a^2+6ab+4b^2) \le \int_0^1 \phi'^{\,2}(u)du.$$

Let us minimize the functional $\int_0^1 \phi'^{\,2}(u)du$ subject to the norming condition

$$\int_0^1 \{a\phi'(u)1_{[0,1/2]}(u) + 2b\phi(y)\}dy = 1$$

and boundary conditions (28.15). This is the classical variational problem on conditional extremum. The existence of solution in the space $W_{2,1}[0,1]$ may be proved as in Nikitin (1995, Section 6.2). The application of Lagrange principle leads for any "test" function h to the following equation for the extremal function ϕ, where λ is the indeterminate Lagrange multiplier:

$$\int_0^1 \{2\phi'(t)h'(t) + a\lambda 1_{[0,1/2]}(t)h'(t) + 2b\lambda h(t))dt = 0.$$

It follows from the main lemma of variational calculus [see, for example, Young (1969, Lemma 7.1)] that

$$2\phi'(t) - 2b\lambda t + a\lambda 1_{[0,1/2]}(t) + C = 0,$$

where C is some constant. Integrating and using the boundary conditions, we get easily the solution of our extremal problem, namely,

$$\phi(s) = C((b - a/2)s - bs^2 + a\min(s, 1/2)). \tag{28.16}$$

Now let us return from function ϕ to the distribution function F_0 and the density f_0, namely consider the equations equivalent to (28.16), namely,

$$F_0' = C((b + a/2)F_0 - bF_0^2), F_0 \leq 1/2$$

$$F_0' = C((b - a/2)F_0 - bF_0^2 + a/2), F_0 \geq 1/2.$$

To avoid non-probabilistic solutions, let us now assume that $a + b > 0$. Hence, $a/2 + b > 0$ also.

To solve the first of our differential equations, let us rewrite it in the form

$$dx = dF_0/F_0(b + a/2) + bdF_0/(b + a/2)(b + a/2 - bF_0).$$

Integrating, we obtain easily

$$F_0(x) = (b + a/2)\exp(\nu x)/(b + a + b\exp(\nu x)), \quad x \leq 0.$$

The expression for the solution F_0 in the domain $x \geq 0$ may be obtained analogously. Finally, we get the following expression for the extremal density f_0^* giving Pitman optimality:

$$f_0^*(x) = \nu(b + a)(b + a/2)\exp(\nu|x|)/((b + a)\exp(\nu|x|) + b)^2, \ a + b > 0.$$

Clearly, for $b = 0$ we obtain the Laplace density known to give optimality for the sign test and for $a = 0$ we obtain the logistic density known to give optimality for the Wilcoxon test [see Hájek and Šidák (1967) or Hettmansperger (1984)]. Thus, we get another generalization of classical results.

In the case of Doksum-Thompson statistic (28.2), we obtain a curious bimodal density given by

$$\bar{f}_0(x) = 3\exp(|x|)/2(\exp(|x|)+2)^2, \qquad x \in R^1.$$

An open question is to find the density f_0 (if it exists) giving the Pitman optimality of the test based on the "best" choice of coefficients a and b.

Acknowledgement. This work was partially supported by Russian Foundation of Fundamental Research, grants No. 96-15-96199 and 98-01-04070.

References

1. Albers, W. (1974). Asymptotic expansions and the efficiency concept in statistics, *Mathematical Centre Tracts No. 58*, Amsterdam.

2. Baksalary, G. K. and Kala, R. (1983). Estimation via linearly combining two given statistics, *Annals of Statistics*, **11**, 691–696.

3. Burgio, G. (1996). L'effizienza asintotica di alcuni test non parametrici per caratteri distribuiti secondo la normale di ordine p, *Studi in Onore di Giampiero Landenna*, pp. 31–51. Giuffre' Ed.: Milano.

4. Burgio, G. and Nikitin, Ya. Yu. (1998). Goodness-of-fit tests for normal distribution of order p and their asymptotic efficiency, *Statistica*, **48** (to appear).

5. Callaert, H. and Janssen, P. (1978). The Berry-Esseen theorem for U-statistics, *Annals of Statistics*, **6**, 417–421.

6. Conti, P. L. and Nikitin, Ya. Yu. (1997). Pitman efficiency of independence tests based on Gini rank association coefficient and Spearman's footrule, Dipart. di Statistica Univ. di Roma "La Sapienza", Ricerche No. 15, 24 pp.

7. Doksum, K. and Thompson, R. (1971). Power bounds and asymptotic minimax results for one-sample rank tests, *Annals of Statistics*, **42**, 12–34.

8. Gross, J. (1998). Statistical estimation by a linear combination of two given statistics, *Statistics & Probability Letters*, **39**, 379–384.

9. Hájek, J. and Šidák, Z. (1967). *Theory of Rank Tests*, New York: Academic Press.

10. Hemelrijk, J. (1950). A family of parameter-free tests for symmetry with respect to a given point. I, II, *Indagationes Math.*, **12**, 340–350; 419–431.

11. Hettmansperger, T. P. (1984). *Statistical Inference Based on Ranks*, New York: John Wiley & Sons.

12. Hodges, J. and Lehmann, E. L. (1956). The efficiency of some nonparametric competitors of the t-test, *Annals of Mathematical Statistics*, **27**, 324–335.

13. Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution, *Annals of Mathematical Statistics*, **19**, 293–325.

14. Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*, San Francisco: Holden Day.

15. Mehra, K. L. and Sarangi, J. (1967). Asymptotic efficiency of certain rank tests for comparative experiments, *Annals of Mathematical Statistics*, **38**, 90–107.

16. Nikitin, Ya. Yu. (1995). *Asymptotic Efficiency of Nonparametric Tests*, New York: Cambridge University Press.

17. Rao, C. R. (1963). Criteria of estimation in large samples, *Sankhyā, Series A*, **25**, 189–206.

18. Rao, C. R. (1965). *Linear Statistical Inference and Its Applications*, New York: John Wiley & Sons.

19. Young, L. C. (1969). *Lectures on the Calculus of Variations and Optimal Control Theory*, Philadelphia: Saunders.

29

Exponential Approximation of Statistical Experiments

A. A. Gushchin and E. Valkeila

Steklov Mathematical Institute, Moscow, Russia
University of Helsinki, Helsinki, Finland

Abstract: We study conditions under which a sequence of statistical experiments can be approximated in a certain sense by experiments generated by exponential families with a *convex* canonical parameter space or weakly converges to such an experiment.

Keywords and phrases: Exponential family, Hellinger integral, Kullback information

29.1 Introduction

Exponential families, especially the Gaussian shift experiments, often occur in asymptotic statistics as limit models. They are also used as approximating models for the original experiments. Thus, it is of interest to study general conditions when a sequence of experiments weakly converges to the experiment corresponding to an exponential family or can be approximated by such experiments.

Here, we deal with the situation when exponential families are parameterized by their canonical parameter and the corresponding parameter space is an open convex subset of $\mathbb{R}^k$. This allows us to characterize exponential families via their extremal property which goes back to Čencov (1982) (Proposition 29.2.1). Using this characterization, we give necessary and sufficient conditions for a sequence of experiments to be "asymptotically exponential" (Theorem 29.3.1). Finally, we prove that such a sequence may be approximated in a stronger sense by exponential families if $\Theta = \mathbb{R}^k$ (Theorem 29.3.2).

Our methods rely heavily on the convexity of the parameter space. Other assumptions on the set of parameters are of minor importance (except Theorem

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

29.3.2); see Remark 29.2.1 below.

Since the canonical parameter space is assumed to be convex, our methods fit poorly to analyze curved exponential families. Curved exponential families often appear in limiting experiments of different models of stochastic processes, for example, if the model satisfies the local asymptotic mixed normality condition or, more generally, if it is locally asymptotically quadratic. In this situation, the just mentioned disadvantage of our approach can be avoided to some extent by modifying it in the context of filtered statistical experiments. Since this modified approach is quite complicated and technical — it is based on the heavy use of stochastic calculus — we shall discuss it elsewhere.

To avoid misunderstanding, let us mention that the expression "asymptotically exponential experiments" is used in the statistical literature not only in connection to exponential families but also in a completely different sense in connection to exponential distributions.

Let us fix a notation and recall some basic facts from the theory of statistical experiments, for which we refer to Strasser (1985), LeCam (1986), and Torgersen (1991).

For nonnegative measures μ and ν on a measurable space $(\Omega, \mathcal{F})$, we denote by $d\nu/d\mu$ the density of the absolutely continuous part of ν with respect to μ. Unless otherwise specified, $\|\cdot\|$ is the total variation norm

$$\|\nu - \mu\| = \int \Big\| \frac{d\nu}{d\lambda} - \frac{d\mu}{d\lambda} \Big\| \, d\lambda,$$

where λ dominates both μ and ν.

$\mathcal{L}(\xi|P)$ is the distribution of a random variable (or vector) ξ under P. The weak convergence of distributions is denoted by $\Rightarrow$.

If ξ_n and η_n are random variables on probability spaces $(\Omega^n, \mathcal{F}^n, P^n)$, we write $\xi_n \xrightarrow{P^n} 0$ as $n \to \infty$ if $\lim_{n\to\infty} P^n(|\xi_n| > \varepsilon) = 0$ for every $\varepsilon > 0$ and $\xi_n = \eta_n + o_{P^n}(1)$, $n \to \infty$, if $\xi_n - \eta_n \xrightarrow{P^n} 0$. These definitions also make sense if ξ_n take values in the extended real line. We say that the sequence (ξ_n, P^n) is uniformly integrable if

$$\lim_{a\to\infty} \sup_n \int_{|\xi_n|>a} |\xi_n| \, dP^n = 0.$$

A statistical experiment $\mathbb{E}$ is a collection $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$, where $(\Omega, \mathcal{F})$ is a measurable space and P_θ is a probability measure on $(\Omega, \mathcal{F})$ for each $\theta \in \Theta$. Sometimes, we use the notation $(P_\theta)_{\theta\in\Theta}$ to designate the experiment.

We say that a collection $\alpha = (\alpha_\theta)_{\theta\in\Theta}$ of nonnegative numbers is a *multi-index* if $I_\alpha \doteq \{\theta : \alpha_\theta \neq 0\}$ is finite and $\sum_{\theta\in I_\alpha} \alpha_\theta = 1$. $\mathcal{A}_+ = \mathcal{A}_+(\Theta)$ is the set of all multi-indices over Θ. If Θ is a subset of a linear space, $\langle\alpha\rangle$ is the barycentre of α: $\langle\alpha\rangle = \sum_{\theta\in I_\alpha} \alpha_\theta \theta$.

Let $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ be a statistical experiment. The *Hellinger transform* of $\mathbb{E}$ is defined by

$$H(\alpha; \mathbb{E}) = \int \prod_{\theta\in I_\alpha} \left(\frac{dP_\theta}{d\nu}\right)^{\alpha_\theta} d\nu, \quad \alpha \in \mathcal{A}_+,$$

where ν is any nonnegative measure that dominates all the P_θ, $\theta \in I_\alpha$.

If $\Theta = \{1, 2, \ldots, m\}$, then $\mathcal{A}_+(\Theta)$ can be identified with the simplex $S_m \doteq \{\alpha = (\alpha_1, \ldots, \alpha_m): \alpha_i \geq 0,\ i = 1, \ldots, m,\ \sum_{i=1}^m \alpha_i = 1\}$. In this case we alternatively write $H(\alpha; P_1, \ldots, P_m)$, $\alpha \in S_m$, instead of $H(\alpha; \mathbb{E})$, $\alpha \in \mathcal{A}_+$. Moreover, for $m = 2$ we write $H(u; P_1, P_2)$, $u \in (0, 1)$, instead of $H((u, 1-u); P_1, P_2)$.

The quantity $H(u; P, P')$, $u \in (0, 1)$, is the *Hellinger integral* of order u between P and P'. Define also the *Kullback information* $K(P', P)$ by

$$K(P', P) = \int \frac{dP'}{dP} \log\left(\frac{dP'}{dP}\right) dP$$

if P' is absolutely continuous with respect to P; otherwise, put $K(P', P) = +\infty$. Here and below we use the conventions $\log 0 = -\infty$, $0 \cdot \infty = 0$. It is well known [see, for example, Liese and Vajda (1987, Section 2)] that $\frac{1-H(u;P,P')}{u}$ is a decreasing finction in $u \in (0, 1)$ and

$$\lim_{u\downarrow 0} \frac{1 - H(u; P, P')}{u} = K(P', P). \tag{29.1}$$

Equivalence and weak convergence of experiments are denoted by $\sim$ and $\xrightarrow{w}$, respectively. Recall that the space of experiments (more exactly, of equivalence classes of experiments) over the same parameter space is compact with respect to the weak convergence, $\mathbb{E} \sim \mathbb{E}'$ if and only if $H(\alpha; \mathbb{E}) = H(\alpha; \mathbb{E}')$ for every $\alpha \in \mathcal{A}_+(\Theta)$, $E^n \xrightarrow{w} E$ if and only if $\lim_{n\to\infty} H(\alpha; \mathbb{E}^n) = H(\alpha; \mathbb{E})$ for every $\alpha \in \mathcal{A}_+(\Theta)$.

An experiment $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ is said to be *pairwise imperfect* if P_θ and P_η are not singular for all $\theta, \eta \in \Theta$. $\mathbb{E}$ is *regular* if for every finite set $\{\theta_1, \ldots, \theta_m\} \subseteq \Theta$ there is a nonnegative measure $\nu \neq 0$ such that ν is absolutely continuous with respect to all the P_{θ_i}, $i = 1, \ldots, m$. We note that $\mathbb{E}$ is regular if and only if $H(\alpha; \mathbb{E})$ does not vanish for all $\alpha \in \mathcal{A}_+$. Finally, $\mathbb{E}$ is *homogeneous* if P_θ and P_η are equivalent for all $\theta, \eta \in \Theta$.

Assume that $\mathbb{E} = (\Omega, \mathcal{F}, (P_1, \ldots, P_m))$ is a regular experiment. Then for any $\alpha = (\alpha_1, \ldots, \alpha_m) \in S_m$, there is a unique probability measure P such that

$$\frac{dP}{d\nu} = \frac{1}{H(\alpha; P_1, \ldots, P_m)} \prod_{i=1}^m \left(\frac{dP_i}{d\nu}\right)^{\alpha_i},$$

where ν is an arbitrary nonnegative measure dominating $P_1, \ldots, P_m$. According to Čencov (1982, Definition 18.2, p. 270), P is the *weighted geodesic* *mean* of $P_1, \dots, P_m$ with weights $\alpha_1, \dots, \alpha_m$. We denote this measure P by $G(\alpha; P_1, \dots, P_m)$. The weighted geodesic mean of P and P' with weights u and $1-u$, where $u \in (0,1)$, is denoted by $G(u; P, P')$. If $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ is a regular experiment with an arbitrary parameter set Θ, we denote by $G(\alpha; \mathbb{E})$, $\alpha \in \mathcal{A}_+(\Theta)$, the weighted geodesic mean of $P_{\theta_1}, \dots, P_{\theta_m}$ with weights $\alpha_{\theta_1}, \dots, \alpha_{\theta_m}$, where $\theta_1, \dots, \theta_m$ are the points that constitute I_α.

The following extremal property of weighted geodesic means is essentially due to Čencov (1982, Lemma 20.5, p. 296).

Lemma 29.1.1 *Let* $\mathbb{E} = (\Omega, \mathcal{F}, (P_1, \dots, P_m))$ *be a regular experiment and* R *a probability measure on* $(\Omega, \mathcal{F})$. *Then, for every* $\alpha = (\alpha_1, \dots, \alpha_m) \in S_m$,

$$K(R, G(\alpha; P_1, \dots, P_m)) = \sum_{i=1}^m \alpha_i K(R, P_i) + \log H(\alpha; P_1, \dots, P_m).$$

In particular,

$$\sum_{i=1}^m \alpha_i K(R, P_i) + \log H(\alpha; P_1, \dots, P_m) \geq 0$$

and the equality holds if and only if $R = G(\alpha; P_1, \dots, P_m)$.

The following lemma is a simple exercise. It is also valid for nets of experiments.

Lemma 29.1.2 *Assume that* $\mathbb{E}^n = (\Omega^n, \mathcal{F}^n, (P_\theta^n)_{\theta\in\Theta}) \xrightarrow{w} \mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$, *where* $\mathbb{E}$ *is a regular experiment. Define* $\mathbb{G} = (\Omega, \mathcal{F}, (G(\alpha; \mathbb{E}))_{\alpha\in\mathcal{A}_+(\Theta)})$, $\mathbb{G}^n = (\Omega^n, \mathcal{F}^n, (G(\alpha; \mathbb{E}^n))_{\alpha\in\mathcal{A}_+^n(\Theta)})$, $\mathcal{A}_+^n(\Theta) = \{\alpha \in \mathcal{A}_+(\Theta): H(\alpha; \mathbb{E}^n) > 0\}$. *Then,* $\mathbb{G}^n \xrightarrow{w} \mathbb{G}$.

Finally, let $\mathcal{N}(m, \sigma^2)$ be the normal distribution with mean m and variance σ^2, 1_A the indicator function of A, $a \vee b$ and $a \wedge b$ the maximum and the minimum of numbers a and b, and $\Delta^\top$ the transpose of a vector $\Delta \in \mathbb{R}^k$.

29.2 Characterization of Exponential Experiments and Their Convergence

Let Θ be an open convex set in $\mathbb{R}^k$. An experiment $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ is an *exponential experiment* if there are a nonnegative measure ν on $(\Omega, \mathcal{F})$ and measurable functions $h: \Omega \to \mathbb{R}_+$ and $S: \Omega \to \mathbb{R}^k$ such that $P_\theta \ll \nu$ and

$$\frac{dP_\theta}{d\nu} = C(\theta) h \exp(\theta^\top S) \tag{29.2}$$

for all $\theta \in \Theta$ for suitable constants $C(\theta)$; see Strasser (1985, Definition 26.1, p. 115).

Every exponential experiment is homogeneous.

The following lemma is a simple consequence of the above definition.

Lemma 29.2.1 *Let* $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ *and* $\mathbb{E}' = (\Omega', \mathcal{F}', (P'_\theta)_{\theta\in\Theta})$ *be exponential experiments. Assume that* I *is a subset of* Θ *such that its affine hull* aff I *contains* Θ. *Then,*

$$\mathbb{E} \sim \mathbb{E}' \qquad \textit{iff} \qquad (\Omega, \mathcal{F}, (P_\theta)_{\theta\in I}) \sim (\Omega', \mathcal{F}', (P'_\theta)_{\theta\in I}).$$

The idea to characterize exponential experiments using weighted geodesic means and their extremal property from Lemma 29.1.1 is due to Čencov (1982).

Proposition 29.2.1 *Let* $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ *be a pairwise imperfect experiment, where* Θ *is an open convex set in* $\mathbb{R}^k$. *The following properties are equivalent:*

(a) $\mathbb{E}$ *is an exponential experiment;*

(b) *for all* $\theta, \eta \in \Theta$ *and* $u \in (0,1)$,

$$P_{(1-u)\theta+u\eta} = G(u; P_\theta, P_\eta);$$

(c) *for all* $\theta, \eta \in \Theta$ *and* $u \in (0,1)$,

$$(1-u)K(P_{(1-u)\theta+u\eta}, P_\theta) + uK(P_{(1-u)\theta+u\eta}, P_\eta) + \log H(1-u; P_\theta, P_\eta) \le 0. \tag{29.3}$$

If these conditions are satisfied then, for every $\alpha \in \mathcal{A}_+$,

$$P_{\langle\alpha\rangle} = G(\alpha; \mathbb{E}) \tag{29.4}$$

and

$$\sum_{\theta\in I_\alpha} \alpha_\theta K(P_{\langle\alpha\rangle}, P_\theta) + \log H(\alpha; \mathbb{E}) = 0. \tag{29.5}$$

In particular, we have equality in (29.3).

Remark 29.2.1 There are variations on that proposition. First, whether or not Θ is open, the representation (29.2) implies other statements of the proposition. Next, assume that Θ is a convex subset of an arbitrary linear space V and $\mathbb{E}$ is pairwise imperfect. Then (b) and (c) are still equivalent and imply the regularity of $\mathbb{E}$, (29.4) and (29.5). But, in general, they do not imply the homogeneity of $\mathbb{E}$. This extra property is automatically satisfied under (b) or (c) if, for any two points in Θ, there is an open interval containing these points (if V is Euclidean, this means that Θ is relatively open in the affine hull generated by Θ). Finally, if Θ is a convex subset of a Euclidean space and the homogeneity assumption is satisfied, then (b) or (c) imply the representation (29.2).

PROOF. Properties (b) and (c) are equivalent due to Lemma 29.1.1. By the same reason, (29.4) and (29.5) are equivalent if $\mathbb{E}$ is regular. It is easy to see that (a) implies (29.4) and, in particular, (b).

It remains to show the implication (b)⇒(a). Since Θ is open and the measures $G(u; P_\theta, P_\eta)$, $u \in (0,1)$, are mutually absolutely continuous, $\mathbb{E}$ is homogeneous. Now an easy induction over $|I_\alpha|$ gives that (b) implies (29.4). To prove that (29.4) implies (a) is easy: see the details in Gushchin and Valkeila (1998). ■

Proposition 29.2.1 has important consequences.

Corollary 29.2.1 *Let Θ be an open convex set in $\mathbb{R}^k$ and $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ an exponential experiment. If $\mathbb{E}' = (\Omega', \mathcal{F}', (P'_\theta)_{\theta\in\Theta})$ is another experiment, then*

$$\mathbb{E} \sim \mathbb{E}' \qquad \textit{iff} \qquad (P_\theta, P_\eta) \sim (P'_\theta, P'_\eta) \quad \textit{for all } \theta, \eta \in \Theta.$$

PROOF. Only sufficiency has to be checked. By Proposition 29.2.1, (29.3) is satisfied for $\mathbb{E}$ and hence for $\mathbb{E}'$, which implies (29.5) both for $\mathbb{E}$ and $\mathbb{E}'$, and we obtain $H(\alpha; \mathbb{E}') = H(\alpha; \mathbb{E})$ for all $\alpha \in \mathcal{A}_+$. ■

Corollary 29.2.2 *Let Θ be an open convex set in $\mathbb{R}^k$ and $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ an exponential experiment. If $\mathbb{E}^n = (\Omega^n, \mathcal{F}^n, (P^n_\theta)_{\theta\in\Theta})$ is a sequence of experiments, then*

$$\mathbb{E}^n \xrightarrow{w} \mathbb{E} \qquad \textit{iff} \qquad (P^n_\theta, P^n_\eta) \xrightarrow{w} (P_\theta, P_\eta) \quad \textit{for all } \theta, \eta \in \Theta.$$

PROOF. The statement follows from Corollary 29.2.1 and weak compactness arguments. ■

For the notion of a Gaussian shift experiment on a Hilbert space, we refer to Strasser (1985, p. 343). It can be easily checked that Gaussian shift experiments satisfy property (b) in Proposition 29.2.1. Taking into account Remark 29.2.1 and the arguments in the proof of Corollaries 29.2.1 and 29.2.2, we obtain the following.

Corollary 29.2.3 *Let H be a Hilbert space with the norm $\|\cdot\|$.*

(1) *An experiment $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in H})$ is a Gaussian shift experiment if and only if*

$$\mathcal{L}\left(\log \frac{dP_\eta}{dP_\theta} \,\middle|\, P_\theta\right) = \mathcal{N}\left(-\frac{\|\eta-\theta\|^2}{2}, \|\eta-\theta\|^2\right) \qquad \textit{for all } \theta, \eta \in \Theta.$$

(2) *A sequence $\mathbb{E}^n = (\Omega^n, \mathcal{F}^n, (P^n_\theta)_{\theta\in H})$ weakly converges to a Gaussian shift experiment if and only if*

$$\mathcal{L}\left(\log \frac{dP^n_\eta}{dP^n_\theta} \,\middle|\, P^n_\theta\right) \Rightarrow \mathcal{N}\left(-\frac{\|\eta-\theta\|^2}{2}, \|\eta-\theta\|^2\right) \qquad \textit{for all } \theta, \eta \in \Theta.$$

29.3 Approximation by Exponential Experiments

In this Section, we describe the situation where a sequence of experiments (over an open convex parameter set) can be approximated in a certain sense by a sequence of exponential experiments.

Theorem 29.3.1 *Let $\mathbb{E}^n = (\Omega^n, \mathcal{F}^n, (P_\theta^n)_{\theta\in\Theta})$ be a sequence of experiments, where Θ is an open convex set in $\mathbb{R}^k$. Assume that*

$$\limsup_{n\to\infty} \|P_\theta^n - P_\eta^n\| < 2 \qquad \text{for all } \theta, \eta \in \Theta. \tag{29.6}$$

The following statements are equivalent:

(A1) *all weak accumulation points of the sequence $\mathbb{E}^n$ are exponential experiments;*

(A2) *there are probability measures P^n on $(\Omega^n, \mathcal{F}^n)$, random vectors Δ_n with values in $\mathbb{R}^k$, random variables γ_n and real-valued functions $\phi_n(\theta)$, $\theta \in \Theta$, such that, for all $\theta \in \Theta$, the sequences $(P_\theta^n)_{n\geq 1}$ and $(P^n)_{n\geq 1}$ are mutually contiguous and*

$$\log \frac{dP_\theta^n}{dP^n} = \theta^\top \Delta_n - \gamma_n - \phi_n(\theta) + o_{P^n}(1), \quad n \to \infty;$$

(B) *for all $\theta, \eta \in \Theta$ and $u \in (0,1)$,*

$$\lim_{n\to\infty} \|P^n_{(1-u)\theta+u\eta} - G(u; P_\theta^n, P_\eta^n)\| = 0;$$

(C) *for all $\theta, \eta \in \Theta$, $u \in (0,1)$, $\beta \in (0,1)$,*

$$\limsup_{n\to\infty} \left(\frac{1 - (1-u)H(\beta; P_\theta^n, P^n_{(1-u)\theta+u\eta}) - uH(\beta; P_\eta^n, P^n_{(1-u)\theta+u\eta})}{\beta} + \log H(1-u; P_\theta^n, P_\eta^n) \right) \leq 0. \tag{29.7}$$

Moreover, if these conditions are satisfied, then:

(1) *the sequence $\mathbb{E}^n$ is weakly sequentially compact;*

(2) (A2) *is valid with $P^n = P_\zeta^n$, where ζ is an arbitrary point in Θ; the vectors Δ_n and variables γ_n can be chosen in such a way that the laws $\mathcal{L}(\Delta_n | P_\zeta^n)$ and $\mathcal{L}(\gamma_n | P_\zeta^n)$ are tight in $\mathbb{R}^k$ and in $\mathbb{R}$, respectively;*

(3) *for every $\alpha \in \mathcal{A}_+$,*

$$\liminf_{n\to\infty} H(\alpha; \mathbb{E}^n) > 0, \tag{29.8}$$

$$\lim_{n\to\infty} \|P^n_{\langle\alpha\rangle} - G(\alpha; \mathbb{E}^n)\| = 0, \tag{29.9}$$

and

$$\limsup_{n\to\infty}\left(\frac{1-\sum_{\theta\in I_\alpha}\alpha_\theta H(\beta;P_\theta^n,P_{\langle\alpha\rangle}^n)}{\beta}+\log H(\alpha;\mathbb{E}^n)\right)\le 0 \tag{29.10}$$

for any $\beta\in(0,1)$.

Remark 29.3.1 The condition (29.6) guarantees that the measures $G(u;P_\theta^n, P_\eta^n)$ in (B) are well defined for n large enough. Similarly, (29.8) guarantees that the weighted geodesic means $G(\alpha;\mathbb{E}^n)$ in (29.9) are well defined for n large enough.

PROOF. Let $\{e_1,\dots,e_k\}$ be the canonical basis in $\mathbb{R}^k$.

(1) Assume (A1).

Take an arbitrary $\zeta\in\Theta$ and choose a $\delta>0$ such that $\zeta+\delta e_i\in\Theta$ for $i=1,\dots,k$. The sequence $\mathbb{E}^n$ (and every its subsequence) contains a subsequence $\mathbb{E}^{n_j}$ such that the experiments $(P_\theta^n)_{\theta\in\{\zeta,\zeta+\delta e_1,\dots,\zeta+\delta e_k\}}$ weakly converge as $j\to\infty$. Then the sequence $\mathbb{E}^{n_j}$ is weakly convergent since it has a unique accumulation point in view of Lemma 29.2.1. Thus, the sequence $\mathbb{E}^n$ is weakly sequentially compact.

If $\mathbb{E}^{n_j}\xrightarrow{w}\mathbb{E}=(\Omega,\mathcal{F},(P_\theta)_{\theta\in\Theta})$ for a subsequence (n_j), then $\mathbb{E}$ is an exponential experiment by the assumption. In particular, given $\alpha\in\mathcal{A}_+$, we have

$$\lim_{j\to\infty}H(\alpha;\mathbb{E}^{n_j})=H(\alpha;\mathbb{E})>0,$$

$$\lim_{j\to\infty}\|P_{\langle\alpha\rangle}^{n_j}-G(\alpha;\mathbb{E}^{n_j})\|=\|P_{\langle\alpha\rangle}-G(\alpha;\mathbb{E})\|=0$$

(passing to the limit is possible because of Lemma 29.1.2) and

$$\begin{aligned}\lim_{j\to\infty}&\left(\frac{1-\sum_{\theta\in I_\alpha}\alpha_\theta H(\beta;P_\theta^{n_j},P_{\langle\alpha\rangle}^{n_j})}{\beta}+\log H(\alpha;\mathbb{E}^{n_j})\right)\\&=\frac{1-\sum_{\theta\in I_\alpha}\alpha_\theta H(\beta;P_\theta,P_{\langle\alpha\rangle})}{\beta}+\log H(\alpha;\mathbb{E})\\&\le\sum_{\theta\in I_\alpha}\alpha_\theta K(P_{\langle\alpha\rangle},P_\theta)+\log H(\alpha;\mathbb{E})=0,\qquad\beta\in(0,1),\end{aligned}$$

by Proposition 29.2.1. Standard contradiction arguments together with the sequential compactness of $\mathbb{E}^n$ yield (29.8)–(29.10).

(2) Let us show that both (B) and (C) imply (A1). Let $\mathbb{E}=(\Omega,\mathcal{F},(P_\theta)_{\theta\in\Theta})$ be an accumulation point of the sequence $\mathbb{E}^n$. Then $\mathbb{E}$ is pairwise imperfect in view of (29.6). If (29.7) holds, then passing to the limit in this inequality we get

$$\frac{1-(1-u)H(\beta;P_\theta,P_{(1-u)\theta+u\eta})-uH(\beta;P_\eta,P_{(1-u)\theta+u\eta})}{\beta}+\log H(1-u;P_\theta,P_\eta)\le 0,$$

which implies (29.3) in view of (29.1). Similarly, (B) together with Lemma 29.1.2 implies $P_{(1-u)\theta+u\eta} = G(u; P_\theta, P_\eta)$. In the both cases, $\mathbb{E}$ is exponential by Proposition 29.2.1.

(3) The next step is to show that (A2) implies (A1). Take an arbitrary $\zeta \in \Theta$ and choose a $\delta > 0$ such that $\zeta + \delta e_i \in \Theta$ for $i = 1, \dots, k$. It follows from (A2) and standard contiguity arguments that, for all $\theta \in \Theta$, the sequences $(P_\theta^n)_{n\geq 1}$ and $(P_\zeta^n)_{n\geq 1}$ are mutually contiguous and

$$\log \frac{dP_\theta^n}{dP_\zeta^n} = (\theta - \zeta)^\top \Delta_n - \psi_n(\theta) + o_{P_\zeta^n}(1), \quad n \to \infty, \tag{29.11}$$

where $\psi_n(\theta) = \phi_n(\theta) - \phi_n(\zeta)$. This relation will not change if we replace Δ_n by $\Delta_n - \delta^{-1}(\psi_n(\theta + \delta e_1), \dots, \psi_n(\theta + \delta e_k))^\top$ and $\psi_n(\theta)$ by $\psi_n(\theta) - \delta^{-1}(\psi_n(\theta + \delta e_1), \dots, \psi_n(\theta + \delta e_k))(\theta - \zeta)$. After this replacement, $\psi_n(\theta + \delta e_i) = 0$ for $i = 1, \dots, k$. Now (29.11) implies the tightness of the laws $\mathcal{L}(\Delta_n | P_\zeta^n)$ since limiting distributions of the sequence $\mathcal{L}\left(\frac{dP_\theta^n}{dP_\zeta^n} \middle| P_\zeta^n\right)$ are carried by $(0, \infty)$ in view of the contiguity.

Let $\mathbb{E} = (\Omega, \mathcal{F}, (P_\theta)_{\theta\in\Theta})$ be an accumulation point of the sequence $\mathbb{E}^n$, that is, there is a net $(\mathbb{E}^{n_\lambda})_{\lambda\in\Lambda}$, Λ is a directed set, such that $\lim_\lambda n_\lambda = \infty$ and $\mathbb{E}_{n_\lambda} \xrightarrow{w} \mathbb{E}$. In view of the contiguity, the measures P_θ and P_ζ are mutually absolutely continuous. Using the weak convergence and (29.11) for $\theta \in \{\zeta + \delta e_1, \dots, \zeta + \delta e_k\}$, we get

$$\mathcal{L}(\Delta_{n_\lambda} | P_\zeta^{n_\lambda}) \Rightarrow \mathcal{L}(\Delta | P_\zeta), \qquad \Delta \doteq \delta^{-1}\left(\log \frac{dP_{\zeta+\delta e_1}}{dP_\zeta}, \dots, \log \frac{dP_{\zeta+\delta e_k}}{dP_\zeta}\right).$$

Applying (29.11) for an arbitrary $\theta \in \Theta$ and the contiguity, we obtain that

$$\lim_\lambda \psi_{n_\lambda}(\theta) = \psi(\theta) \doteq \log \int \exp((\theta - \zeta)^\top \Delta)\, dP_\zeta$$

and

$$\frac{dP_\theta}{dP_\zeta} = \exp((\theta - \zeta)^\top \Delta - \psi(\theta)).$$

The claim follows.

(4) It remains to prove the implication (A1)$\Rightarrow$(A2) with $P^n = P_\zeta^n$, where ζ is an arbitrary point in Θ. Fix a $\delta > 0$ such that $\zeta + \delta e_i \in \Theta$, $i = 1, \dots, k$. Put

$$z_\theta^n = \frac{dP_\theta^n}{dP_\zeta^n}, \quad \tilde{z}_\theta^n = \begin{cases} z_\theta^n, & \text{if } z_\theta^n > 0, \\ 1, & \text{otherwise,} \end{cases}$$

and define $\Delta_n = \delta^{-1}(\log \tilde{z}_{\zeta+\delta e_1}^n, \dots, \log \tilde{z}_{\zeta+\delta e_k}^n)^\top$.

Since all accumulation points of $(\mathbb{E}^n)$ are homogeneous experiments, the sequences (P_θ^n) and (P_ζ^n) are mutually contiguous for every $\theta \in \Theta$. In particular,

$P^n_\zeta(z^n_\theta \neq \tilde{z}^n_\theta) \to 0$ as $n \to \infty$. Therefore,

$$\log \frac{dP^n_\theta}{dP^n_\zeta} = (\theta-\zeta)^\top \Delta_n + o_{P^n_\zeta}(1), \quad n \to \infty, \quad \theta \in \{\zeta+\delta e_1, \dots, \zeta+\delta e_k\}. \tag{29.12}$$

Let θ belong to the convex hull Ξ generated by ζ and $\zeta+\delta e_i$, $i = 1, \dots, k$, i.e. there is an $\alpha = \alpha(\theta) \in \mathcal{A}_+$ such that $\theta = \langle\alpha\rangle$ and $I_\alpha \subseteq \{\zeta, \zeta+\delta e_1, \dots, \zeta+\delta e_k\}$. By the definition of weighted geodesic means, if $H(\alpha; \mathbb{E}^n) > 0$,

$$\log \frac{dG(\alpha; \mathbb{E}^n)}{dP^n_\zeta} = \sum_{\theta \in I_\alpha \setminus \{\zeta\}} \alpha_\theta \log \frac{dP^n_\theta}{dP^n_\zeta} - \log H(\alpha; \mathbb{E}^n).$$

Substituting (29.12), we get

$$\log \frac{dG(\alpha; \mathbb{E}^n)}{dP^n_\zeta} = (\theta - \zeta)^\top \Delta_n - \psi_n(\theta) + o_{P^n_\zeta}(1), \quad n \to \infty,$$

where $\psi_n(\theta) = \log H(\alpha; \mathbb{E}^n)$. In view of (29.9), which has been already proved under (A1), we obtain that

$$\log \frac{dP^n_\theta}{dP^n_\zeta} = (\theta - \zeta)^\top \Delta_n - \psi_n(\theta) + o_{P^n_\zeta}(1), \quad n \to \infty, \quad \theta \in \Xi. \tag{29.13}$$

Now assume that $\theta \in \Theta \setminus \Xi$. It is easy to find two vectors $\theta_1, \theta_2 \in \Xi$ and a number $u \in (0,1)$ such that

$$\theta_1 = u\theta + (1-u)\theta_2.$$

If $H(u; P^n_\theta, P^n_{\theta_2}) > 0$, then

$$\log \frac{dG(u; P^n_\theta, P^n_{\theta_2})}{dP^n_\zeta} = u \log \frac{dP^n_\theta}{dP^n_\zeta} + (1-u) \log \frac{dP^n_{\theta_2}}{dP^n_\zeta} - \log H(u; P^n_\theta, P^n_{\theta_2}),$$

thus

$$\log \frac{dP^n_{\theta_1}}{dP^n_\zeta} = u \log \frac{dP^n_\theta}{dP^n_\zeta} + (1-u) \log \frac{dP^n_{\theta_2}}{dP^n_\zeta} - \log H(u; P^n_\theta, P^n_{\theta_2}) + o_{P^n_\zeta}(1), \quad n \to \infty,$$

in view of (B) (which has also been proved under (A1)). Since (29.13) holds for θ_1 and θ_2, we get

$$\log \frac{dP^n_\theta}{dP^n_\zeta} = (\theta - \zeta)^\top \Delta_n - \psi_n(\theta) + o_{P^n_\zeta}(1), \quad n \to \infty,$$

with $\psi_n(\theta) = u^{-1}\psi_n(\theta_1) - u^{-1}(1-u)\psi_n(\theta_2) - u^{-1} \log H(u; P^n_\theta, P^n_{\theta_2})$. ∎

Let a sequence $(\mathbb{E}^n)$ satisfy the statements of the previous theorem. It follows from (29.9) that, for every set Ξ that is a convex hull of a finite set, one can construct a sequence $\widetilde{\mathbb{E}}^n = (\Omega^n, \mathcal{F}^n, (\widetilde{P}^n_\theta)_{\theta\in\Xi})$ of exponential experiments such that

$$\lim_n \|\widetilde{P}^n_\theta - P^n_\theta\| = 0, \quad \theta \in \Xi.$$

A natural question arises whether one can approximate $\mathbb{E}^n$ by exponential experiments in the same sense on the whole parameter set Θ. Such approximations are well known and play an important role in the case of local asymptotic normality. In the next theorem, we give a positive answer on the question posed if $\Theta = \mathbb{R}^k$. In the case where $(\mathbb{E}^n)$ is a weakly convergent sequence, another approximating construction was suggested in Gushchin and Valkeila (1998) for an arbitrary convex Θ.

Theorem 29.3.2 *Assume that* $\Theta = \mathbb{R}^k$ *and that a sequence of experiments* $\mathbb{E}^n = (\Omega^n, \mathcal{F}^n, (P^n_\theta)_{\theta\in\Theta})$ *satisfies conditions* (A1)–(C) *of Theorem* 1. *Then there is a sequence of exponential experiments* $\widetilde{\mathbb{E}}^n = (\Omega^n, \mathcal{F}^n, (\widetilde{P}^n_\theta)_{\theta\in\Theta})$ *satisfying*

$$\lim_{n\to\infty} \|\widetilde{P}^n_\theta - P^n_\theta\| = 0, \quad \theta \in \Theta. \tag{29.14}$$

The proof of Theorem 29.3.2 is based on the following lemma.

Lemma 29.3.1 *Let* ξ_n *and* η_n, $n \geq 1$, *be nonnegative random variables on a probability space* $(\Omega^n, \mathcal{F}^n, P^n)$. *Assume that*

$$\xi_n - \eta_n \xrightarrow{P^n} 0, \quad n \to \infty, \tag{29.15}$$

and the sequence (η_n, P^n) *is uniformly integrable. Then there are positive numbers* a_n *such that* $\lim_{n\to\infty} a_n = \infty$ *and the sequence* $(\xi_n \wedge a_n, P^n)$ *is uniformly integrable.*

PROOF. Let E^n be the expectation with respect to P^n. If $E^n\xi_n > E^n\eta_n$, define b_n as any nonnegative number such that $E^n(\xi_n \wedge b_n) = E^n\eta_n$; otherwise, put $b_n = +\infty$. First, we establish that

$$\xi_n - (\xi_n \wedge b_n) \xrightarrow{P^n} 0. \tag{29.16}$$

Assume the converse. Since the laws $\mathcal{L}(\eta_n|P^n)$ are tight, the laws $\mathcal{L}(\xi_n|P^n)$ are also tight in view of (29.15). Thus, there is a subsequence (n_j) such that the laws $\mathcal{L}(\xi_{n_j} \wedge b_{n_j}, \xi_{n_j} - (\xi_{n_j} \wedge b_{n_j})|P^{n_j})$ weakly converge, say, to $\mathcal{L}(\xi, \xi'|P)$, where $P(\xi' > 0) > 0$ by the assumption. In particular, (b_{n_j}) is bounded for sufficiently large j owing to the tightness of $\mathcal{L}(\xi_n|P^n)$. This implies

$$E\xi = \lim_{j\to\infty} E^{n_j}(\xi_{n_j} \wedge b_{n_j}) = \lim_{j\to\infty} E^{n_j}\eta_{n_j},$$

where the last equality follows from the choice of (b_n). On the other hand, $\mathcal{L}(\eta_{n_j}|P^{n_j}) \Rightarrow \mathcal{L}(\xi + \xi'|P)$, hence

$$\lim_{j\to\infty} E^{n_j}\eta_{n_j} = E(\xi + \xi'),$$

and we arrive at a contradiction.

Now we can conclude that $\lim_{n\to\infty} E^n|(\xi_n \wedge b_n) - \eta_n| = 0$ and therefore the sequence $(\xi_n \wedge b_n, P^n)$ is uniformly integrable. Indeed,

$$\begin{aligned} E^n|(\xi_n \wedge b_n) - \eta_n| &= E^n((\xi_n \wedge b_n) - \eta_n) + 2E^n(\eta_n - (\xi_n \wedge b_n))1_{\{\eta_n \geq \xi_n \wedge b_n\}} \\ &\leq 2E^n(\eta_n - (\xi_n \wedge b_n))1_{\{\eta_n \geq \xi_n \wedge b_n\}}. \end{aligned}$$

The right-hand side of the previous inequality tends to zero since it is non-negative and bounded by η_n and $\eta_n - (\xi_n \wedge b_n) \xrightarrow{P^n} 0$ in view of (29.15) and (29.16).

It follows from the above considerations that the sequence $((b_n+1)1_{\{\xi_n > b_n+1\}}, P^n)$ is uniformly integrable and $\lim_{n\to\infty} P^n(\xi_n > b_n + 1) = 0$. Therefore,

$$\lim_{n\to\infty} (b_n + 1)P^n(\xi_n > b_n + 1) = 0.$$

Now choose numbers $c_n \geq 1$ such that $\lim_{n\to\infty} c_n = \infty$ and

$$\lim_{n\to\infty} (b_n + 1)c_n P^n(\xi_n > b_n + 1) = 0. \tag{29.17}$$

Put $a_n \doteq (b_n+1)c_n$. Clearly, $\lim_{n\to\infty} a_n = \infty$. We have $\xi_n \wedge a_n \leq (\xi_n \wedge b_n) + 1 + a_n 1_{\{\xi_n > b_n+1\}}$. The sequence $(a_n 1_{\{\xi_n > b_n+1\}}, P^n)$ is uniformly integrable since its $L^1(P^n)$-norms tend to zero as $n \to \infty$ in view of (29.17). Therefore, $(\xi_n \wedge a_n, P^n)$ is also uniformly integrable. ■

PROOF OF THEOREM 29.3.2. Let $e_1, \ldots, e_k$ be the standard basis vectors in $\mathbb{R}^k$. Put $z_\theta^n = \frac{dP_\theta^n}{dP_0^n}$, $\theta \in \mathbb{R}^k$. According to Theorem 29.3.1, the sequences $(P_\theta^n)_{n\geq 1}$ and $(P_0^n)_{n\geq 1}$ are mutually contiguous. In other words, this means that, for every $\theta \in \mathbb{R}^k$,

$$\text{the sequence } (z_\theta^n, P_0^n)_{n\geq 1} \text{ is uniformly integrable,} \tag{29.18}$$

$$\lim_{n\to\infty} E_0^n z_\theta^n = 1 \tag{29.19}$$

(E_0^n is the expectation with respect to P_0^n) and

$$\text{limiting distributions of } \mathcal{L}(z_\theta^n \mid P_0^n) \text{ are concentrated on } (0, \infty). \tag{29.20}$$

The idea of our construction is standard. Let $0 < a_{n,i} \leq A_{n,i} < \infty$, $i = 1, \ldots, k$, $n = 1, 2, \ldots$, be numbers satisfying

$$\lim_{n\to\infty} a_{n,i} = 0, \quad \lim_{n\to\infty} A_{n,i} = \infty, \quad i = 1, \ldots, k. \tag{29.21}$$

Put

$$Y_{n,i} = \begin{cases} z_{e_i}^n, & \text{if} \quad a_{n,i} \leq z_{e_i}^n \leq A_{n,i}, \\ 1, & \text{otherwise.} \end{cases} \tag{29.22}$$

For $\theta = (\theta_1, \dots, \theta_k) \in \mathbb{R}^k$, we put

$$Y_\theta^n = \prod_{i=1}^k (Y_{n,i})^{\theta_i}, \quad Z_\theta^n = \frac{Y_\theta^n}{E_0^n Y_\theta^n}, \quad d\widetilde{P}_\theta^n = Z_\theta^n \, dP_0^n.$$

Due to (29.20)–(29.22),

$$\lim_{n\to\infty} P_0^n(Y_{n,i} \neq z_{e_i}^n) = 0, \quad i = 1, \dots, k. \tag{29.23}$$

It follows from (A2) (with $P^n = P_0^n$; see also (29.11)) and (29.20) that there are positive constants $\Psi_{n,\theta}$ such that

$$z_\theta^n = \frac{\prod_{i=1}^k (z_{e_i}^n)^{\theta_i}}{\Psi_{n,\theta}} + o_{P_0^n}(1), \quad n \to \infty, \tag{29.24}$$

and

$$0 < \liminf_{n\to\infty} \Psi_{n,\theta} \le \limsup_{n\to\infty} \Psi_{n,\theta} < \infty \tag{29.25}$$

for a given $\theta \in \mathbb{R}^k$. In view of (29.23),

$$z_\theta^n = \frac{Y_n^\theta}{\Psi_{n,\theta}} + o_{P_0^n}(1), \quad n \to \infty. \tag{29.26}$$

The trick of the proof is to choose numbers $a_{n,i}$ and $A_{n,i}$ in such a way that they satisfy (29.21) and

the sequence $(Y_\theta^n, P_0^n)_{n\ge1}$ is uniformly integrable for every $\theta \in \mathbb{R}^k$. (29.27)

Then the sequence $\widetilde{\mathbb{E}}^n = (\Omega^n, \mathcal{F}^n, (\widetilde{P}_\theta^n)_{\theta\in\mathbb{R}^k})$ satisfies the conditions of the theorem. Indeed, it is evident that $\widetilde{\mathbb{E}}^n$ is an exponential experiment. In view of (29.26), (29.18), (29.27), (29.25) and (29.19),

$$\lim_{n\to\infty} \frac{E_0^n Y_\theta^n}{\Psi_{n,\theta}} = 1,$$

hence

$$z_\theta^n - Z_\theta^n \xrightarrow{P_0^n} 0 \quad \text{and} \quad \lim_{n\to\infty} \|P_\theta^n - \widetilde{P}_\theta^n\| = 0, \quad \theta \in \mathbb{R}^k. \tag{29.28}$$

Fix a number $i \in \{1, \dots, k\}$. Since $Ne_i \in \Theta$ for every $N = 1, 2, \dots$,

$$z_{Ne_i}^n = \frac{(z_{e_i}^n)^N}{\Psi_{n,Ne_i}} + o_{P_0^n}(1), \quad n \to \infty,$$

in view of (29.24). Taking into account (29.18) and (29.25), we can apply Lemma 29.3.1 with $\xi_n = (z_{e_i}^n)^N$ and $\eta_n = z_{Ne_i}^n \Psi_{n,Ne_i}$ and to find numbers

$A_{n,i}^{(N)} \geq 1$ such that $\lim_{n\to\infty} A_{n,i}^{(N)} = \infty$ and the sequence $\left(\left(z_{e_i}^n \wedge A_{n,i}^{(N)}\right)^N, P_0^n\right)_{n\geq 1}$ is uniformly integrable.

The next step is to find numbers $A_{n,i} \geq 1$ such that

$$\lim_{n\to\infty} A_{n,i} = \infty \tag{29.29}$$

and

$$\text{the sequence } ((z_{e_i}^n \wedge A_{n,i})^N, P_0^n)_{n\geq 1} \text{ is uniformly integrable} \tag{29.30}$$

for every $N = 1, 2, \ldots$. This can be done as follows. Define

$$A_{n,i} = A_{n,i}^{(1)} \wedge \cdots \wedge A_{n,i}^{(N)} \quad \text{if} \quad M_N \leq n < M_{N+1},$$

where $(M_N)_{N\geq 1}$, $M_1 = 0$, is a strictly increasing sequence of integer numbers. Then we have (29.30) since $A_{n,i} \leq A_{n,i}^{(N)}$ for $n \geq M_N$. To ensure (29.29), define recursively $M_N \doteq 1 + \max\{n\colon A_{n,i}^{(1)} \wedge \cdots \wedge A_{n,i}^{(N)} \leq N\} \vee M_{N-1}$, $N \geq 2$, where $\max\emptyset \doteq 0$. By the construction, $M_{N-1} < M_N < \infty$ and $A_{n,i} \geq N$ if $n \geq M_N$.

Using the same construction for negative N, we construct similarly numbers $a_{n,i} \leq 1$ such that

$$\lim_{n\to\infty} a_{n,i} = 0$$

and the sequence $((z_{e_i}^n \vee a_{n,i})^N, P_0^n)_{n\geq 1}$ is uniformly integrable for every $N = -1, -2, \ldots$.

Now, if $Y_{n,i}$ are defined according to (29.22), we have $(Y_{n,i})^N \leq (z_{e_i}^n \wedge A_{n,i})^N + 1$ if $N = 1, 2, \ldots$, and $(Y_{n,i})^N \leq (z_{e_i}^n \vee a_{n,i})^N + 1$ if $N = -1, -2, \ldots$. This means that (29.27) holds if $\theta \in \{Ne_i\colon i = 1, \ldots, k,\ N = \pm 1, \pm 2, \ldots\}$ and, hence, if θ belongs to the coordinate axes. The same statement for all $\theta \in \mathbb{R}^k$ follows now from the inequality

$$Y_\theta^n \leq |\theta|^{-1} \sum_{i=1}^k |\theta_i| (Y_{n,i})^{|\theta| \operatorname{sign}\theta_i}, \quad \text{where} \quad |\theta| \doteq \sum_{i=1}^k |\theta_i|.$$

The proof is completed. ■

References

1. Čencov, N. N. (1982). *Statistical Decision Rules and Optimal Inference*, Translations of Mathematical Monographs, Vol. 53, Providence, Rhode Island: American Mathematical Society.

2. Gushchin, A. A. and Valkeila, E. (1998). On statistical experiments converging to exponential limits, *Preprint* 197, Department of Mathematics, University of Helsinki, Helsinki, Finland.

3. Janssen, A., Milbrodt, H. and Strasser, H. (1985). *Infinitely Divisible Statistical Experiments*, Lecture Notes in Statistics, Vol. 27, Berlin: Springer-Verlag.

4. LeCam, L. (1986). *Asymptotical Methods in Statistical Decision Theory*, Berlin: Springer-Verlag.

5. Liese, F. and Vajda, I. (1987). *Convex Statistical Distances*, Leipzig: Teubner.

6. Strasser, H. (1985). *Mathematical Theory of Statistics*, Berlin: de Gruyter.

7. Torgersen, E. (1991). *Comparison of Statistical Experiments*, Cambridge: Cambridge University Press.

30

The Asymptotic Distribution of a Sequential Estimator for the Parameter in an AR(1) Model With Stable Errors

Joop Mijnheer

Leiden University, Leiden, The Netherlands

Abstract: We consider the AR(1) model $X_k = \gamma X_{k-1} + \varepsilon_k$, $k = 1, 2, \ldots$, $X_0 = 0$ a.s. and ε_k are i.i.d. with a stable distribution. We obtain the limit distribution of a sequential estimator for γ .

Keywords and phrases: AR(1) model, stable errors, sequential estimator

30.1 Introduction

A random variable ε has a stable distribution function $F(\cdot\,;\alpha,\beta)$ if its characteristic function f is given by

$$\log f(t) = \begin{cases} -|t|^\alpha\{1 - i\beta\,\mathrm{sign(t)}\tan(\pi\,\alpha/2)\} & \text{if } \alpha \in (0,1)\cup(1,2) \\ \\ -|t| - i\beta(2/\pi)t\log(t) & \text{if } \alpha = 1 \end{cases} \tag{30.1}$$

and $|\beta| \leq 1$. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent and identically distributed (i.i.d.) with common distribution function $F(\cdot\,;\alpha,\beta)$. Then, for all n , we have, for $\alpha \neq 1$,

$$\varepsilon_1 + \cdots + \varepsilon_n \overset{\mathrm{d}}{=} n^{\frac{1}{\alpha}}\varepsilon_1 \tag{30.2}$$

and, for $\alpha = 1$,

$$\varepsilon_1 + \cdots + \varepsilon_n - (2/\pi)\beta n\log n \overset{\mathrm{d}}{=} n\,\varepsilon_1 \,. \tag{30.3}$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

Here, $\stackrel{d}{=}$ denotes "has the same distribution as." In the notation of Samorodnitsky and Taqqu (1994), we have $\varepsilon_1 \stackrel{d}{=} S_\alpha(1,\beta,0)$. In this chapter, we restrict ourselves to those values for (α,β) where the stability property (30.2) holds.

We consider the following AR(1) model.

Model I.

$$X_0 = 0 \qquad \text{a.s.}$$

$$X_k = \gamma X_{k-1} + \varepsilon_k \qquad k = 1,2,\ldots .$$

ε_k , $k = 1,2,\ldots$, are i.i.d. with distribution function $F(\,\cdot\,;\alpha,\beta)$ and satisfy the property (30.2).

In this case, we also have the following stability property.

For $s,t > 0$,

$$s^{\frac{1}{\alpha}}\varepsilon_1 + t^{\frac{1}{\alpha}}\varepsilon_2 \ \stackrel{d}{=}\ (s+t)^{\frac{1}{\alpha}}\varepsilon_1 \ . \tag{30.4}$$

In Section 30.2, we give the limit distribution of a non-sequential estimator of γ and in Section 30.3, for a sequential estimator.

30.2 Non-Sequential Estimation

The case $\alpha = 2$

In this case, we assume that ε_k, $k = 1,2,\ldots$, are i.i.d. with a standard normal distribution. Let $|\gamma| < 1$. Then, for $n = 1,2,\ldots$,

$$X_n \ = \ \gamma^{n-1}\varepsilon_1 + \ \cdots \ + \varepsilon_n \tag{30.5}$$

has a normal distribution with $E\,X_n = 0$ and $\sigma^2(X_n) = (1-\gamma^2)^{-1}(1-\gamma^{2n})$. For the maximum likelihood estimator or least-squares estimator $\hat{\gamma}_n$ for γ, we obtain

$$(\hat{\gamma}_n - \gamma) \ = \ \Big(\sum_{k=1}^{n} X_{k-1}^2\Big)^{-1} \sum_{k=1}^{n} X_{k-1}\varepsilon_k \ . \tag{30.6}$$

From the model, we obtain

$$X_n^2 + (1-\gamma^2)\sum_{k=1}^{n} X_{k-1}^2 \ = \ 2\gamma \sum_{k=1}^{n} X_{k-1}\varepsilon_k + \sum_{k=1}^{n} \varepsilon_k^2 \ . \tag{30.7}$$

First, we remark that $\sum_{k=1}^{n} X_{k-1}\varepsilon_k$ is a martingale so that the martingale central limit theorem implies, for $n \to \infty$,

$$n^{-\frac{1}{2}} \sum_{k=1}^{n} X_{k-1}\varepsilon_k \xrightarrow{\text{d}} (1-\gamma^2)^{-\frac{1}{2}} U \,. \tag{30.8}$$

Here, $\xrightarrow{\text{d}}$ denotes convergence in distribution, and U has a standard normal distribution.
As a consequence, we have

$$n^{-1} \sum_{k=1}^{n} X_{k-1}\varepsilon_k \xrightarrow{\text{P}} 0 \qquad \text{for } n \to \infty \,,$$

where $\xrightarrow{\text{P}}$ denotes convergence in probability.

We also have $n^{-1}X_n^2 \xrightarrow{\text{P}} 0$ for $n \to 0$. The strong law of large numbers gives us $n^{-1} \sum_{k=1}^{n} \varepsilon_k^2 \to 1$ a.s. for $n \to \infty$. As a consequence we have, for $n \to \infty$, by non-random normalization

$$n^{\frac{1}{2}}(\hat{\gamma}_n - \gamma) \xrightarrow{\text{d}} (1-\gamma^2)^{-\frac{1}{2}} U \,. \tag{30.9}$$

This result was proved – without martingale results – in Anderson (1959); see also Lai and Siegmund (1983) and Shiryaev and Spokoiny (1997). From (30.7) and the assertions above, we have for $n \to \infty$,

$$n^{-1} \sum_{k=1}^{n} X_{k-1}^2 \xrightarrow{\text{P}} (1-\gamma^2)^{-1} \,.$$

Therefore, by random normalization

$$\Big(\sum_{k=1}^{n} X_{k-1}^2\Big)^{\frac{1}{2}} (\hat{\gamma}_n - \gamma) \xrightarrow{\text{d}} U \,. \tag{30.10}$$

The case $\alpha \neq 2$

First, we make two remarks on differences with the case $\alpha = 2$.

Remark 30.2.1 If ε_1 and ε_2 have distribution function $F(\,\cdot\,;\alpha,\beta)$, then $\gamma\varepsilon_1 + \varepsilon_2$ has the same distribution as

$$(1+\gamma^{\alpha})^{\frac{1}{\alpha}}\,\varepsilon_1$$

for $\gamma > 0$ and

$$(1+|\gamma|^{\alpha})^{\frac{1}{\alpha}}\,\tilde{\varepsilon}$$

for $\gamma < 0$, where $\tilde{\varepsilon}$ has distribution function $F(\cdot\,;\alpha,\tilde{\beta})$ with $\tilde{\beta} = \beta\,(1-|\gamma|^{\alpha})\,(1+|\gamma|^{\alpha})^{-1}$. Thus, for negative values for γ the characteristic exponent α, measuring the "fatness" of the tail(s), does not change but the parameter β, measuring the "skewness", changes. This was already noticed in Mijnheer (1997b). In that paper, a related problem has been solved. To avoid this complication, we restrict ourselves from now on to the case $0 < \gamma < 1$.

Remark 30.2.2 For stable laws, we have $E|\varepsilon|^p = \infty$ for $p > \alpha$. Thus we can not use martingale techniques.

Using (30.5) and the stability property (30.4), we obtain that

$$X_n \stackrel{\mathrm{d}}{=} (1-\gamma^{n\alpha})^{\frac{1}{\alpha}}\,(1-\gamma^{\alpha})^{-\frac{1}{\alpha}}\,\varepsilon_1\,. \tag{30.11}$$

Thus, for $n \to \infty$,

$$n^{-\frac{2}{\alpha}} X_n^2 \xrightarrow{\mathrm{P}} 0\,. \tag{30.12}$$

Since ε_1^2 is in the domain of normal attraction of the stable law $F(\cdot\,;\frac{\alpha}{2},1)$, we obtain for $n \to \infty$,

$$c\,n^{-\frac{2}{\alpha}} \sum_{k=1}^{n} \varepsilon_k^2 \xrightarrow{\mathrm{d}} \text{stable law } F(\cdot\,;\tfrac{\alpha}{2},1) \tag{30.13}$$

for some constant c.

In order to obtain the limit distribution of $\hat{\gamma}_n$, we make use of the following proposition and theorem.

Proposition 30.2.1 *Let* ε_1 *and* ε_2 *be two independent random variables with distribution function* $F(\cdot\,;\alpha,\beta)$. *Let* $p^{\alpha}+q^{\alpha}=1$ *and* $\beta = p^{\alpha}-q^{\alpha}$. *Then,* $\varepsilon_1\varepsilon_2$ *is in the domain of (non-normal) attraction of the stable law* $F(\cdot\,;\alpha,\tilde{\beta})$, *where* $\tilde{\beta} = (p^{2\alpha}+q^{2\alpha})^{\frac{1}{\alpha}} - 2^{\frac{1}{\alpha}}pq$.

PROOF. Let Y_1 and Y_2 be two i.i.d. random variables with the distribution function given by

$$P(Y_1 \le y) = \begin{cases} 1-y^{-\alpha} & , y \ge 1 \\ 0 & , \text{otherwise}\,. \end{cases}$$

It is a simple exercise to show that, for $y > 1$,

$$P\,(Y_1Y_2 > y) = \alpha y^{-\alpha}\log y + y^{-\alpha}\,.$$

In the case $0 < \alpha < 1$ and $\beta = 1$, we have $P(\varepsilon_1 > 0) = 1$ and, for $x \to \infty$,

$$1 - F(x;\alpha,1) = c\,x^{-\alpha} + \mathcal{O}\left(x^{-2\alpha}\right); \tag{30.14}$$

see Chapter 2, Section 4 of Feller (1971). Then the result easily follows. In all other cases $0 < \alpha < 2$ and $|\beta| \neq 1$, we have the tail behavior as given in the right hand side (rhs) of (30.14) for both tails. In the case where $|\beta| = 1$, one of the tails has the behavior as given in (30.14) and this tail dominates the other tail. The assertion follows from the criterium for random variables to belong to the domain of attraction of a stable law. See, for example, Theorem 1 in Chapter 9, Section 8 in Feller (1971). ∎

Theorem 30.2.1 *Consider Model I as given in the Introduction. Let α, β and $\tilde{\beta}$ be as in Proposition 30.2.1. Then, for $n \to \infty$ and some constant c,*

$$c\,(n \log n)^{-\frac{1}{\alpha}} \sum_{k=1}^{n} X_{k-1}\varepsilon_k \xrightarrow{\ \mathrm{d}\ } S_1 ,$$

where S_1 has distribution function $F(\cdot\,;\alpha,\tilde{\beta})$.

PROOF. There are in the literature results on stable limits for partial sums of dependent random variables. Most of them are for stationary sequences of random variables; see, for example, Davis (1983) and Jakubowski (1993, 1997).

Consider Model II.

$$\begin{aligned} X_0 &= (1-\gamma^{\alpha})^{-\frac{1}{\alpha}}\,\varepsilon_0, \text{ and} \\ X_k &= \gamma X_{k-1} + \varepsilon_k , \qquad k = 1, 2, \ldots , \end{aligned}$$

$\varepsilon_0, \varepsilon_1, \ldots$ are i.i.d. with distribution function $F(\cdot\,;\alpha,\beta)$. In this model, we have stationarity. We have $X_k \stackrel{\mathrm{d}}{=} X_0$ for $k = 1, 2, \ldots$.

Let $\left(\sum_{k=1}^{n} X_{k-1}\varepsilon_k\right)_i$ stand for summation in model i, $i = \mathrm{I}$ or II. We write

$$\left(\sum_{k=1}^{n} X_{k-1}\varepsilon_k\right)_{\mathrm{II}} = \left(\sum_{k=1}^{n} X_{k-1}\varepsilon_k\right)_{\mathrm{I}} + R_n .$$

Then

$$R_n = X_0\left(\varepsilon_1 + \gamma\varepsilon_2 + \cdots + \gamma^{n-1}\varepsilon_n\right) \stackrel{\mathrm{d}}{=} (1-\gamma^{n\alpha})^{\frac{1}{\alpha}}\,\varepsilon_1\,\varepsilon_2\,(1-\gamma^{\alpha})^{-\frac{2}{\alpha}} .$$

Applying the result obtained in Proposition 30.2.1 we have, for $n \to \infty$, $(n \log n)^{-\frac{1}{\alpha}} R_n \xrightarrow{\ \mathrm{P}\ } 0$. The assertion follows from the results in Davis (1983). We may also use the results of Jakubowski (1993, 1997) to show that there exists an asymptotic independent representation and then apply classical results for sums of independent random variables. ∎

Remark 30.2.3 $2\left(\sum_{k=1}^{2n} X_{k-1}\varepsilon_k\right)_{\mathrm{I}}$ can be written as $2\sum_{k=1}^{2n}\sum_{l=1}^{k-1} \gamma^{l-1}\varepsilon_l\,\varepsilon_k$.

In order to make use of the matrix notation, we introduce $e'_n = (\varepsilon_1, \dots, \varepsilon_n)$, $\tilde{e}'_n = (\varepsilon_{n+1}, \dots, \varepsilon_{2n})$ and

$$\Gamma_n = \begin{pmatrix} 0 & 1 & \gamma & \dots & & \gamma^{n-2} \\ 1 & 0 & 1 & \dots & & \vdots \\ \vdots & \vdots & & \ddots & & \\ & & & & & 1 \\ \gamma^{n-2} & & & \dots & 1 & 0 \end{pmatrix} \quad \text{(symmetric)} .$$

With this notation, we have

$$e'_{2n}\, \Gamma_{2n}\, e_{2n} = e'_n \Gamma_n\, e_n + \tilde{e}'_n\, \Gamma_n\, \tilde{e}_n + Q$$

where

$$\begin{aligned} Q &= 2\left(\gamma^{n-1}\varepsilon_1 + \dots + \varepsilon_n\right)\left(\varepsilon_{n+1} + \dots + \gamma^{n-1}\varepsilon_{2n}\right) \\ &\stackrel{d}{=} 2\,(1-\gamma^{n\alpha})^{\frac{2}{\alpha}}\,(1-\gamma^{\alpha})^{-\frac{2}{\alpha}}\,\varepsilon_1\varepsilon_2 \; . \end{aligned}$$

Applying Proposition 30.2.1, we obtain $(n \log n)^{-\frac{1}{\alpha}}\, Q \xrightarrow{P} 0$ for $n \to \infty$.

From the independence of the random variables $\varepsilon_1, \dots, \varepsilon_{2n}$ and the structure of the matrix Γ_{2n}, we deduce that $e'_n\, \Gamma_n\, e_n$ and $\tilde{e}'_n\, \Gamma_n\, e_n$ are independent and identically distributed. Let Y_1 and Y_2 be two independent copies of $\lim_{n\to\infty} \{ (n \log n)^{-\frac{1}{\alpha}}\, e'_n\, \Gamma_n\, e_n \}$. From the foregoing assertions, we easily obtain $Y_1 + Y_2 \stackrel{d}{=} 2^{\frac{1}{\alpha}}\, Y_1$. This assertion may be extended to m independent copies $Y_1, \dots, Y_m$. Even if we take $m = m_n = \log n$, we can show

$$Y_1 + \dots + Y_{m_n} \stackrel{d}{=} m_n^{\frac{1}{\alpha}}\, Y_1 \; .$$

Although we consider Model I, in which $X_{k-1}\varepsilon_k$, $k = 1, 2, \dots$, is *not* a stationary sequence, we may use the foregoing results to check the two conditions in Theorem 9.1 in Jakubowski (1993) or Theorem 2.1 in Jakubowski (1997) applied to Model II in which $X_{k-1}\varepsilon_k$, $k = 1, 2, \dots$, is a stationary sequence.

Now we are ready to formulate (the first part) of the limit behavior of $\hat{\gamma}_n$. As a consequence of the assertion in Theorem 30.2.1 we have, for $n \to \infty$,

$$n^{-\frac{2}{\alpha}} \sum_{k=1}^{n} X_{k-1}\varepsilon_k \xrightarrow{P} 0 \; .$$

Using this result, together with (30.12) and (30.13), we obtain for some constant c and $n \to \infty$,

$$c\, n^{\frac{1}{\alpha}}\, (\log n)^{-\frac{1}{\alpha}}\, (\hat{\gamma}_n - \gamma) \xrightarrow{d} S_1/S_0$$

with S_0 having distribution function $F(\,\cdot\,; \frac{\alpha}{2}, 1)$ and S_1 as in Theorem 30.2.1.

Remark 30.2.4 This result was, in the case $\beta = 0$, obtained in Davis and Resnick (1986). They also proved that S_0 and S_1 are independent! This will be proved in the following theorem.

Theorem 30.2.2

a. $\sum_{k=1}^{n} \left(X_{k-1}\varepsilon_k , \varepsilon_k^2\right)_{\text{II}}$ *has an asymptotic independent representation.*

b. $\left((n \log n)^{-\frac{1}{\alpha}}\left(\sum_{k=1}^{n} X_{k-1}\varepsilon_k\right)_{\text{II}}, n^{-\frac{2}{\alpha}} \sum_{k=1}^{n} \varepsilon_k^2\right) \quad \xrightarrow{\text{d}} \quad (S_1, S_0)$ *where* $S_1 \stackrel{\text{d}}{=} F(\,\cdot\,;\alpha,\hat{\beta})$, $S_0 \stackrel{\text{d}}{=} F(\,\cdot\,;\frac{\alpha}{2},1)$ *and* S_0 *and* S_1 *are independent.*

PROOF. In the proof of Theorem 30.2.1 we have mentioned the papers by Jakubowski (1993, 1997). In Jakubowski, Nagaev and Zaigraev (1997) these results are generalized to multivariate stable distributions. In the same way one may generalize these results to bivariate stable distributions where the marginal distributions have different characteristic exponents (i.e., the parameter α in $F(\,\cdot\,;\alpha,\beta)$). I do not know if there exists a reference in the literature. Now we apply Theorem 4 of Resnick and Greenwood (1979). We still have to compute the limit in (4.14) of Part iii of that Theorem. This is done in Proposition 3.1 of Davis and Resnick (1986). ∎

30.3 Sequential Estimation

In the case of non-sequential estimation, we distinguish the cases $|\gamma| < 1$, $|\gamma| = 1$, $|\gamma| > 1$ and $\gamma_n = e^{\frac{a}{n}}$. In those cases, the least-squares estimator has different limit distributions. In the case of sequential estimation, we have in the first three cases the same limit distribution; see Shiryaev and Spokoiny (1997). They prove even stronger assertions. In Mijnheer (1997a), we summarize the results for the (non-sequential) least-squares estimator in the case of innovations with a stable distribution.

We define for each $h > 0$

$$\tau(h) \; = \; \inf\left\{n \geq 1 : \sum_{k=1}^{n} X_{k-1}^2 \geq h\right\} \tag{30.15}$$

For $\alpha = 2$ and $|\gamma| < 1$ we have

$$\lim_{n\to\infty} h^{-1}\,\tau(h) \; = \; 1 - \gamma^2 \,, \qquad P_\gamma \text{ a.s.} \tag{30.16}$$

This is Theorem 4.1, Part (i) of Lai and Siegmund (1983). In the same paper, they prove, for $|\gamma| \le 1$ (note the $\le$ sign)

$$\Big(\sum_{k=1}^{\tau(h)} X_{k-1}^2\Big)^{\frac{1}{2}} (\hat{\gamma}_{\tau(h)} - \gamma) \xrightarrow{\text{d}} U ,$$

for $h \to \infty$. In Theorem 2.1.a of Shiryaev and Spokoiny (1997), this assertion is extended to the case $|\gamma| > 1$. For $|\gamma| \ge 1$, the limit behavior of $\tau(h)$ is different from (30.16).

Now we consider the case $\alpha \neq 2$ and $0 < \gamma < 1$. The next theorem is the analogue of (30.16) in this case.

Theorem 30.3.1 *Let $\tau(h)$ be defined by (30.15). Then, for $x > 0$,*

$$\lim_{h\to\infty} P\,(h^{-\frac{\alpha}{2}}\tau(h) \le x) \;=\; P\left((1-\gamma^2)^{\frac{\alpha}{2}} S_0^{-\frac{\alpha}{2}} \le x\right) ,$$

where S_0 has distribution function $F(\,\cdot\,; \frac{\alpha}{2}, 1)$.

PROOF. Consider formula (30.7). One easily sees that $\sum_{k=1}^{n} X_{k-1}^2$ dominates X_n^2 and applying Theorem 30.2.2 we obtain that $\sum_{k=1}^{n} X_{k-1}\varepsilon_k$ is dominated by $\sum_{k=1}^{n} \varepsilon_k^2$. Take x fixed, then, for $h \to \infty$,

$$\begin{aligned} P\,(h^{-\frac{\alpha}{2}}\,\tau(h) \le x) \;&=\; P\Big(\sum_{k=1}^{\left[h^{\frac{\alpha}{2}}x\right]} X_{k-1}^2 \ge h\Big) \\ &\sim\; P\Big(\sum_{k=1}^{\left[h^{\frac{\alpha}{2}}x\right]} \varepsilon_k^2 \ge (1-\gamma^2)\,h\Big) . \end{aligned}$$

The assertion now easily follows if we notice that ε_k^2 is in the domain of normal attraction of $F(\,\cdot\,; \frac{\alpha}{2}, 1)$. ■

Remark 30.3.1 From the proof, it is clear that we also can define τ by

$$\tau(h) \;=\; \inf\,\{n \ge 1 : \varepsilon_1^2 + \cdots + \varepsilon_n^2 \ge h\} .$$

Since $E\varepsilon_1^2 = \infty$, the strong law of large numbers gives us $n^{-1}\sum_{k=1}^{n}\varepsilon_k^2 \to \infty$ a.s. Thus, $\tau(h) \to \infty$ a.s. for $h \to \infty$.

The main result of this Chapter is the following theorem.

Theorem 30.3.2 *Let $\tau(h)$ be defined by (30.15). Then,*

$$c(\tau(h))^{\frac{1}{\alpha}}(\log \tau(h))^{-\frac{1}{\alpha}}(\hat{\gamma}_{\tau(h)} - \gamma) \xrightarrow{\text{d}} S_1/S_0,$$

where S_0 and S_1 are as in Theorem 30.2.2.

PROOF. Applying Theorem 2.3.b and Theorem 2.5.7 of Embrechts, Klüppelberg and Mikosch (1997), we have

$$(\tau(h)\log \tau(h))^{-\frac{1}{\alpha}} \sum_{k=1}^{\tau(h)} X_{k-1}\varepsilon_k \xrightarrow{\text{d}} S_1 \ .$$

Now we use Theorem 30.3.1. Theorem 2.5.7 of Embrechts, Klüppelberg and Mikosch (1997) requires independence of $(X_{k-1}\varepsilon_k)_{k=1}^{\infty}$ and $\tau(h)$. Parts a and b Theorem 30.2.2 part a and b give us that this assumption causes no problems since there exists an asymptotic independent representation with independent components. ■

References

1. Anderson, T. W. (1959). On asymptotic distributions of estimates of parameters of stochastic difference equations, *Annals of Mathematical Statistics*, **30**, 676–687.

2. Davis, R. A. (1983). Stable limits for partial sums of dependent variables, *Annals of Probability*, **11**, 262–269.

3. Davis, R. and Resnick, S. (1986). Limit theory for sample covariance and correlation functions of moving averages, *Annals of Statistics*, **14**, 533–558.

4. Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events*, Berlin: Springer-Verlag.

5. Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Vol. II. Second edition, New York: John Wiley & Sons.

6. Jakubowski, A. (1993). Minimal conditions in p-stable limit theorems, *Stochastic Processes and their Applications*, **44**, 291–327.

7. Jakubowski, A. (1997). Minimal conditions in p-stable limit theorems - II, *Stochastic Processes and their Applications*, **68**, 1–20.

8. Jakubowski, A., Nagaev, A. V. and Zaigraev, A. (1997). Large deviation probabilities for sums of heavy-tailed dependent random vectors, *Communcations in Statistics—Stochastic Models*, **13**, 647–660.

9. Lai, T. L. and Siegmund, D.(1983). Fixed accuracy estimation of an autoregressive parameter, *Annals of Statistics*, **11**, 478–485.

10. Mijnheer, J. (1997a). Asymptotic inference for AR(1) processes with (non-normal) stable errors, *Journal of Mathematical Sciences*, **83**, 401–406.

11. Mijnheer, J. (1997b). Asymptotic inference for AR(1) processes with (non-normal) stable errors. IV. A note on the case of a negative unit root, *Submitted for publication.*

12. Resnick, S. and Greenwood, P. (1979). A bivariate stable characterization and domains of attraction, *Journal of Multivariate Analysis*, **9**, 206–221.

13. Shiryaev, A. N. and Spokoiny, V. G. (1997). On sequential estimation of an autoregressive parameter, *Stochastics and Stochastic Reports*, **60**, 219–240.

14. Samorodnitsky, G. and Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes*, New York: Chapman and Hall.

31

Estimation Based on the Empirical Characteristic Function

Bruno Rémillard and Radu Theodorescu

Université du Québec à Trois-Rivières, Trois-Rivières, Québec, Canada
Université Laval, Sainte-Foy, Québec, Canada

Abstract: There exist distributions for which standard estimation techniques based on the probability density function are not applicable. As an alternative, the characteristic function is used. Certain distributions whose characteristic functions can be expressed in terms of $|t|^\alpha$ are such examples. Tailweight properties are first examined; it is shown that these laws are Paretian, their tail index α being one of the parameters defining these laws. Estimators similar to those proposed by Press (1972) for stable laws are then used for the estimation of the parameters of such laws and asymptotic properties are proved. As an illustration, the Linnik distribution is examined.

Keywords and phrases: Characteristic function, consistent estimator, distribution theory, Linnik distribution, order statistic, Paretian distribution, Pólya distribution, stable distribution, Sibuya distribution, tailweight

31.1 Introduction

There exist distributions for which standard estimation techniques based on the probability density function are not applicable. As an alternative, the characteristic function is used.

The setting of this paper is more general than that in Jacques, Rémillard, and Theodorescu (1999). Its plan is as follows. In Section 31.2, we show that certain distributions whose characteristic functions can be expressed in terms of $|t|^\alpha$ are Paretian and, as a special case, we consider the symmetric two-parameter stable distribution, the three-parameter Linnik distribution, and the two-parameter Sibuya distribution. It turns out that α, one of the parameters of these laws, is their tail index. In Section 31.3, we examine the estimation of

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

the parameters of such distributions, by making use of some ideas concerning the adaptive estimators proposed by Press (1972) for stable laws [for an extensive survey see Csörgő (1984)]. This method essentially uses the characteristic function. We also prove asymptotic properties of our estimators. As an illustration, in Section 31.4 we consider the Linnik distribution. Finally, in Section 31.5, we mention numerical results and we comment on estimator efficiency. Our approach is valid for the univariate as well as for the multivariate case. Here we restrict ourselves to the univariate case; for the multivariate case, see Jacques, Rémillard, and Theodorescu (1999, Subsection 4.2).

31.2 Tailweight Behavior

Let α be a positive real number and let $\mathcal{P}_\alpha$ denote the family of all random variables X for which we have the following tailweight property:

$$\lim_{x\to\infty} x^\alpha P(X > x) = C,$$

where C is a positive constant, generally depending on α and eventually on other parameters. According to Mandelbrot (1962), we refer to $\mathcal{P}_\alpha$ as the *Paretian family of index* α.

Stable random variables are the most common representatives of the family $\mathcal{P}_\alpha$. For $\alpha \in (0,2]$, $|b| \le 1$, and $\gamma > 0$, let $S_{\alpha,b,\gamma}$ be a (centered) stable random variable with log characteristic function given by [Hall (1981)]

$$\log\phi(t,\alpha,b,\gamma) = -|\gamma t|^\alpha\left[1 - ib\,\mathrm{sgn}(t)\omega(t,\alpha)\right], \qquad t\in\mathbf{R}, \tag{31.1}$$

with

$$\omega(t,\alpha) = \begin{cases} \tan(\pi\alpha/2) & \text{for } \alpha \neq 1, \\ -2\pi^{-1}\log|t| & \text{for } \alpha = 1. \end{cases}$$

We have

$$\lim_{x\to\infty} x^\alpha P(S_{\alpha,b,\gamma} > x) = \frac{(1+b)\gamma^\alpha}{\pi}\Gamma(\alpha)\sin(\pi\alpha/2). \tag{31.2}$$

Consequently, $S_{\alpha,b,\gamma} \in \mathcal{P}_\alpha$ for every $\alpha \in (0,2)$.

Let now $\alpha \in (0,2]$, β, $\gamma > 0$ and let us consider the function $f : \mathbf{R} \to \mathbf{R}$ defined by

$$f(t,\alpha,\beta,\gamma) = (1 + |\gamma t|^\alpha)^{-\beta}, \quad t\in\mathbf{R}. \tag{31.3}$$

Linnik (1963, p. 67) showed that $f(\cdot,\alpha,1,1)$ is a characteristic function. Devroye (1990) proved in a short and elegant manner that if $L_{\alpha,\beta,\gamma}$ is a random variable distributed as the product $S_{\alpha,0,\gamma}T^{1/\alpha}$ of two independent random variables, where $T = T_\beta$ is gamma$(\beta,1)$ distributed, then it has the characteristic function (31.3). The symmetric distribution associated with (31.3) is called the *Linnik distribution* with parameters α, β, γ and is denoted by Linnik(α, β, γ). This distribution is continuous, symmetric, unimodal about zero, and its moment behavior is controlled by the tail index α. Also for $\beta \to \infty$ and an appropriate choice of γ, we are led to a symmetric stable random variable. Moreover, the distribution Linnik$(\alpha, 1, \gamma)$ is closed under geometric compounding [Anderson (1992)].

Devroye's (1990) representation plays an important part in simulating $L_{\alpha,\beta,\gamma}$. $S_{\alpha,0,\gamma}$ can be generated by means of the algorithm of Chambers, Mallows, and Stuck (1976) which is essentially based on a representation of $S_{\alpha,0,\gamma}$ due to Zolotarev (1966); for the gamma distributed T, there exist good generators.

We now examine the effect of multiplication on the asymptotic behavior of the upper tail probability. Let T be an arbitrary nonnegative random variable independent of $S_{\alpha,0,\gamma}$ and consider the product $S_{\alpha,0,\gamma}T^{1/\alpha}$. For such laws the asymptotic behavior of their upper tails is similar to that of the stable family. Indeed, if $E(T) < \infty$, we obtain [Jacques, Rémillard, and Theodorescu (1999, Proposition 2.5)] a result similar to (31.2):

$$\lim_{x\to\infty} x^{\alpha} P\left(S_{\alpha,0,\gamma}T^{1/\alpha} > x\right) = \frac{\gamma^{\alpha}}{\pi}\Gamma(\alpha)E(T)\sin(\pi\alpha/2).$$

In particular if $T = T_{\beta}$, then

$$\lim_{x\to\infty} x^{\alpha} P(L_{\alpha,\beta,\gamma} > x) = \frac{\beta\gamma^{\alpha}}{\pi}\Gamma(\alpha)\sin(\pi\alpha/2).$$

The characteristic function f of the product $S_{\alpha,0,1}T^{1/\alpha}$ has the form $L(|t|^{\alpha})$, where $L(u) = E[\exp(-uT)]$ is the Laplace transform of T [see also Pakes (1998, p. 214) in connection with mixture representations for Linnik laws]. Observe also that such characteristic functions arise in connection with limits of sums of a random number of random variables; see Gnedenko (1983).

Not all characteristic functions of the form $L(|t|^{\alpha})$ come from products of the form $S_{\alpha,0,1}T^{1/\alpha}$. For example, if $L(t)$ is one of the covariance functions of a stationary Gaussian reciprocal process [Carmichael, Massé, and Theodorescu (1982)], i.e., $L(t) = \max\{0, 1 - |t|\}$, then L is not a Laplace transform but

$$g(t, \alpha, \gamma) = \max\{0, 1 - |\gamma t|^{\alpha}\}, \qquad t \in \mathbf{R}, \tag{31.4}$$

is a characteristic function for any $\alpha \in (0, 1]$, $\gamma > 0$. Such expressions also concern Pólya-type distributions [Lukacs (1970, p. 87)] whose simulation was dealt with by Devroye (1986, pp. 186–190 and p. 718). By analogy with the discrete case [Devroye (1993, p. 350)], we shall call the distribution with the characteristic function (31.4) the *Sibuya distribution* with parameters α and γ and we shall denote it by Sibuya(α, γ).

Whenever the right derivative of L exists at 0 and is negative, then $n^{-1/\alpha}\sum_{i=1}^{n} X_i$ converges in law as $n \to \infty$ to $S_{\alpha,0,\gamma}$ with $\gamma^\alpha = -L'(0)$, where the X_i's are independent and identically distributed copies of a random variable X with characteristic function $L(|t|^\alpha)$. Under these hypotheses, it follows from Feller (1971, Theorem 1a, p. 313) that

$$\lim_{x\to\infty} x^\alpha P(X > x) = \frac{-L'(0)}{\pi}\Gamma(\alpha)\sin(\pi\alpha/2).$$

The converse is also true. If X is a symmetric random variable in the Paretian family $\mathcal{P}_\alpha$ and $n^{-1/\alpha}\sum_{i=1}^{n} X_i$ converges in law as $n \to \infty$ to $S_{\alpha,0,\gamma}$, then the right derivative of L at zero exists and $\gamma^\alpha = -L'(0)$.

31.3 Parameter Estimation

Standard parameter estimation techniques are not applicable to stable, Linnik, and Sibuya distributions since their probability density functions cannot be written in a simple form except for special cases. For these distributions, the parameter α is the tail index. So we may use any consistent tail index estimator, for instance de Haan's estimator, regardless whether the other parameters are known or not. This estimator is one possible choice among other estimators with nice properties.

Let $X_1, \ldots, X_n$ be a sample of size n from a random variable X and let $X_{n,1}, \ldots, X_{n,n}$ be the associated ordered sample. *De Haan's estimator* (1981) for α is given by

$$\breve{\alpha}_r = \frac{\log r}{\log X_{n,n} - \log X_{n-r,n}}.$$

If X is Paretian, then for a suitable sequence k_n the estimator $\breve{\alpha}_{k_n} \to \alpha \in (0,\infty]$ with probability one (Jacques, Rémillard, and Theodorescu (1999, Theorem 3.2)). Moreover $\log k_n(\alpha/\breve{\alpha}_{k_n} - 1)$ converges in law to an extreme value random variable with distribution function $\exp\{-e^{-x}\}$ [Jacques, Rémillard, and Theodorescu (1999, Theorem 3.4)].

A possible estimation procedure is based on the papers of Press (1972), Paulson, Holcomb, and Leitch (1975), and Leitch and Paulson (1975) who dealt with stable distributions. Essentially, this procedure reduces to a minimizing problem (in the parameters of the law) of a weighted norm of the difference $f_n(t) - f(t)$, where f_n is the empirical characteristic function. Anderson and Arnold (1993) examined the Linnik law with $\beta = \gamma = 1$.

Quite a different estimation procedure relies on some ideas concerning the adaptive estimators proposed by Press (1972) for stable distributions [for an extensive survey, see Csörgő (1984)]. From now on, we shall follow this path.

Since all estimations are based on the real part U_n of the empirical characteristic function f_n, we begin with a result on the asymptotic behavior of $U_n - f$.

From Csörgő (1981, Theorem 3, p. 133), we have the following.

Proposition 31.3.1 *Suppose that f is the characteristic function of a symmetric random variable X such that for some $\alpha > 0$, $x^\alpha P(X > x)$ is bounded. If $\tau > 0$ is given and $n \to \infty$, then $\sup_{t\in[0,\tau]} |U_n(t) - f(t)|$ converges a.s. to zero and $n^{1/2}(U_n(t) - f(t))$ converges in law in $C([0,\tau])$ to a centered Gaussian process $\mathcal{G}(t)$ with covariance function $v(t,s) = [f(t+s) + f(t-s) - 2f(t)f(s)]/2$ on $\mathbf{R}^2$.*

In view of (31.1), (31.3), and (31.4), we observe that the characteristic functions of symmetric stable, Linnik, and Sibuya distributions can be written in a unified form as functions of $|t|^\alpha$.

Estimating α and γ (L known)

Let $X_1, \ldots, X_n$ be a sample of size n from a distribution whose probability density function is of the form $f(t) = L(\phi(t))$, where $\phi(t) = |\gamma t|^\alpha$ and L is a known function. For the Linnik law, L is known when β is known. Suppose there exists an open interval $I = L^{-1}((0,1))$ such that L is continuously differentiable on I and is a one-to-one mapping of I onto $(0,1)$.

In order to formulate our results, we need the following notations:

$$Y(t) = \begin{cases} \log L^{-1}(f(t)) = \alpha(\log\gamma + \log|t|) & \text{for } f(t) \in (0,1), \\ 1 & \text{otherwise,} \end{cases} \tag{31.5}$$

$$Y_n(t) = \begin{cases} \log L^{-1}(U_n(t)) & \text{for } U_n(t) \in (0,1), \\ 1 & \text{otherwise,} \end{cases}$$

$$a(t) = -\frac{1}{L^{-1}(f(t))L'(L^{-1}(f(t))},$$

$$c = \frac{1}{q}\sum_{j=1}^{q} \log|t_j|,$$

$$b_i = \frac{\log|t_i| - c}{\sum_{j=1}^q lh \log|t_i| - c)^2}, \qquad \boldsymbol{b}^\top = (b_1, \ldots, b_q),$$

$$v(t,s) = [f(t+s) + f(t-s) - 2f(t)f(s)]/2,$$

$$w(t,s) = a(t)a(s)v(t,s)$$

for $f(t_i), U_n(t_i) \in (0,1)$, $1 \le i \le q$, $q \ge 2$, such that $\sum_{j=1}^q \left(\log|t_j| - c\right)^2 > 0$; $^\top$ stands for transposition.

For our three particular distributions, we have the following:

law	I	$L(t)$	$L^{-1}(t)$	$a(t)$
stable	$(0,\infty)$	e^{-t}	$-\log t$	$-\dfrac{1}{f(t)\log f(t)}$
Linnik	$(0,\infty)$	$(1+t)^{-\beta}$	$t^{-1/\beta}-1$	$\dfrac{f(t)^{-1/\beta-1}}{\beta(f(t)^{-1/\beta}-1)}$
Sibuya	$(0,1)$	$1-t$	$1-t$	$-\dfrac{1}{1-f(t)}$

Further, set

$$\alpha_n^{\mathrm{P}} = \sum_{j=1}^{q} Y_n(t_j) b_j$$

and

$$\gamma_n^{\mathrm{P}} = \exp\left\{ \frac{1}{q\alpha_n^{\mathrm{P}}} \sum_{j=1}^{q} Y_n(t_j) - c \right\}.$$

An argument similar to the one in Jacques, Rémillard, and Theodorescu (1999, Proposition 4.1) yields the following.

Proposition 31.3.2 *Let L be known. Then, α_n^{P} and γ_n^{P} are a.s. consistent estimators of α and γ, respectively.*

Further, by using a Slutski-type argument [Jacques, Rémillard, and Theodorescu (1999, Theorem 4.2)], we obtain the following normality property.

Theorem 31.3.1 *Let L be known. Then, $n^{1/2}(\alpha_n^{\mathrm{P}} - \alpha, \gamma_n^{\mathrm{P}} - \gamma)^\top$ converges in law as $n \to \infty$ to a bivariate centered Gaussian vector with covariance matrix $\boldsymbol{W}_1 = (w_1(i,j))$,*

$$\begin{aligned} w_1(1,1) &= \boldsymbol{b}^\top \boldsymbol{W} \boldsymbol{b}, \\ w_1(1,2) = w_1(2,1) &= \frac{\gamma}{\alpha}(\mathbf{1}/q - \boldsymbol{b}(c + \log\gamma))^\top \boldsymbol{W}\boldsymbol{b}, \\ w_1(2,2) &= \left(\frac{\gamma}{\alpha}\right)^2 (\mathbf{1}/q - \boldsymbol{b}(c+\log\gamma))^\top \boldsymbol{W} (\mathbf{1}/q - \boldsymbol{b}(c+\log\gamma)), \end{aligned}$$

where $\boldsymbol{W} = (w(t_i,t_j))$ and $\mathbf{1} = (1,\ldots,1)^\top$.

Remark 31.3.1 Simultaneous asymptotic confidence intervals for α and γ may be obtained by using Theorem 31.3.1. In practice, because of rounding errors, we recommend to estimate $\boldsymbol{W}_1$ in the following way: $n^{-1}\sum_{i=1}^{n}(\boldsymbol{Z}_n - \bar{\boldsymbol{Z}})(\boldsymbol{Z}_n - \bar{\boldsymbol{Z}})^\top$, where

$$\boldsymbol{Z}_n = \left(\sum_{j=1}^{q} a_n(t_j)b_j\cos(t_jX_i), \frac{\gamma_n^{\mathrm{P}}}{\alpha_n^{\mathrm{P}}}\sum_{j=1}^{q} a_n(t_j)(1/q - b_j(c+\log\gamma_n^{\mathrm{P}}))\cos(t_jX_i)\right)^\top,$$

and

$$a_n(t) = -\frac{1}{L^{-1}(U_n(t))L'(L^{-1}(U_n(t))}.$$

Remark 31.3.2 Suppose that for a certain family of distributions, the possible range of α is $(0, \alpha_0]$, where $0 < \alpha_0 \leq 2$. In this case, take $\alpha_n = \min\{\alpha_n^{\mathrm{P}}, \alpha_0\}$ and construct confidence intervals $[\alpha_n - 1.96\sigma_n n^{-1/2}, \alpha_n + 1.96\sigma_n n^{-1/2}]$ for α. Then the empirical significance level, i.e., the relative frequency, P_α of those intervals which do not contain α converges a.s. to 0.05 if $\alpha < \alpha_0$ and to 0.025 if $\alpha = \alpha_0$. Here, σ_n^2 is an estimation of $\boldsymbol{b}^\top \boldsymbol{W}\boldsymbol{b}$.

Estimating L (α or one of its consistent estimators known)

Suppose that $U(t) = \mathrm{Re}\, f(t)$ can be expressed as a unknown function L of $\phi(t) = t^\alpha$, where α or one of its consistent estimators is known; in either case, we set α_n to denote it. Here we use the same symbols L and ϕ with a slightly different meaning. For the symmetric stable law and the Sibuya law, this means that γ is unknown and for a Linnik law that β and γ are unknown.

Set $L_n(t) = U_n(t^{1/\alpha_n})$ and $\mathcal{L}(t) = \mathcal{G}(t^{1/\alpha})$.

Theorem 31.3.2 *Let α_n be a consistent estimator of α and suppose that*

$$s_n(\alpha_n - \alpha, U_n(t) - U(t)) \Rightarrow (\mathcal{A}, h\mathcal{G}(t)) \tag{31.6}$$

in $\mathbf{R} \times C([0,1])$ as $n \to \infty$, where $\Rightarrow$ stands for convergence in law, $\mathcal{A}$ is a random variable, and

$$s_n n^{-1/2} \to h \in [0, \infty).$$

If $P(\mathcal{A} \neq 0) > 0$, suppose in addition that L is continuously differentiable on $(0,\infty)$. Then, $s_n(\alpha_n - \alpha, L_n(t) - L(t))$ converges in $\mathbf{R} \times C([0,\infty))$ as $n \to \infty$ to $(\mathcal{A}, \mathcal{H}(t))$, where $\mathcal{H}(t) = \chi(t)\mathcal{A} + h\mathcal{L}(t)$ and $\chi(\cdot)$ is a continuous function defined by $\chi(t) = L'(t)t\alpha^{-1}\log t$, $t > 0$, and $\chi(0) = 0$.

If the differentiability condition is replaced by the assumption that L admits right and left derivatives $D^{\pm}L$ which are respectively right-continuous and left-continuous on $(0,\infty)$ and if $\lim_{t\downarrow 0} D^{\pm}L(t)t\log t = 0$, then for any $T > 0$,

$$\sup_{0\leq t\leq T}\Big| s_n(L_n(t) - L(t)) - \max\{0, s_n(\alpha_n - \alpha)\}\chi_+(t)$$
$$- s_n\left(U_n(t^{1/\alpha}) - U(t^{1/\alpha})\right) - \min\{0, s_n(\alpha_n - \alpha)\}\chi_-(t)\Big|$$

converges in probability to zero, where $\chi_\pm(\cdot)$ *is the function defined by* $\chi_\pm(t) = D^\pm L(t)t\alpha^{-1}\log t$, $t > 0$, *and* $\chi_\pm(0) = 0$.

PROOF. Since $s_n(L_n(t) - L(t))$ can be written as the sum of $s_n(L(t^{\alpha/\alpha_n}) - L(t))$ and $s_n(U_n(t^{1/\alpha_n}) - U(t^{1/\alpha_n}))$, and since

$$\sup_{t\in[0,T]} \left| s_n(U_n(t^{1/\alpha_n}) - U(t^{1/\alpha_n})) - s_n(U_n(t^{1/\alpha}) - U(t^{1/\alpha}))\right|$$

converges in probability to zero by Proposition 31.3.1, the result follows from the hypotheses on $D^\pm L$, from a Slutski-type argument [Jacques, Rémillard, and Theodorescu (1999, Theorem 4.2)], and from the following representation:

$$\begin{aligned} & s_n(L(t^{a_n}) - L(t)) \\ = \; & \max\{0, s_n(a_n - 1)\} \int_0^1 D^+L(t^{1+u(a_n-1)})t^{1+u(a_n-1)} \log t \, du \\ & + \min\{0, s_n(a_n - 1)\} \int_0^1 D^-L(t^{1+u(a_n-1)})t^{1+u(a_n-1)} \log t \, du, \end{aligned}$$

where $a_n = \alpha/\alpha_n$. ■

Remark 31.3.3 If $h = 0$, (31.6) is equivalent to the convergence in law of $s_n(\alpha_n - \alpha)$ to a random variable $\mathcal{A}$ as $n \to \infty$. If L is continuously differentiable on $(0, \infty)$, then the limit in Theorem 31.3.2 is $(\mathcal{A}, \chi(\cdot)\mathcal{A})$.

Remark 31.3.4 If α is known, we take $s_n = n^{-1/2}$ and (31.6) is satisfied with $\mathcal{A} = 0$ and $h = 1$. In this case $n^{-1/2}(L_n(t) - L(t))$ converges in $C([0, \infty))$ as $n \to \infty$ to $\mathcal{L}(t)$.

Remark 31.3.5 In view of applications, observe that if X is a random variable with characteristic function $L(|t|^\alpha)$ and if $E(|X|^p) < \infty$ for some $p > 1$, then L is differentiable, so $\chi_+(t) = \chi_-(t) = \chi(t)$ may be consistently estimated by

$$\chi_n(t) = U_n'(t^{1/\alpha_n})t^{1/\alpha_n}\alpha_n^{-2}\log t, \qquad t \geq 0,$$

since $\chi_n(t)$ converges in probability to $\chi(t)$, $t \geq 0$, as $n \to \infty$. An interesting application where we do not have to estimate χ is the test $H_0 : L = L_0$ versus $H_1 : L \neq L_0$, for a known function L_0.

Let us be more specific about the function L in a parametric context. Suppose that $U(t) = L(|t|^\alpha, \theta)$, where $\theta \in \Theta \subset \mathbf{R}^\mathbf{k}$ is an open subset. The goal is to estimate θ. To this end suppose that, for some $t_1, \ldots, t_k$, the mapping

$$\tilde{\theta} \mapsto \psi(\tilde{\theta}) = (L(t_1, \tilde{\theta}), \ldots, L(t_k, \tilde{\theta}))^\top$$

is of class C^1 on Θ and that $\psi'(\tilde{\theta})$ is invertible for all $\tilde{\theta} \in \Theta$. It follows that $\psi(\Theta)$ is open and that for fixed $\theta \in \Theta$ and $b = \psi(\tilde{\theta})$, there exist open sets U and V, depending on $\tilde{\theta}$ such that $\theta \in U$, $b \in V$, ψ is one-to-one on U, $\psi(U) = V$. Moreover if $\tau = \psi^{-1}$, it is uniquely defined on V by $\tau \circ \psi(\tilde{\theta}) = \tilde{\theta}$ for all $\tilde{\theta} \in U$, and τ is of class C^1 on V. Finally $\tau'(\tilde{b}) = (\psi')^{-1} \circ g(\tilde{b})$ for all $\tilde{b} \in V$.

In view of this discussion, set

$$b_n^\top = (L_n(t_1), \ldots, L_n(t_k)),$$

and define θ_n by $\tau(b_n) = \psi^{-1}(b_n)$ for $b_n \in \psi(\Theta)$ and $\theta_n = \theta_0$ otherwise for a fixed $\theta_0 \in \Theta$.

We obtain immediately the following.

Proposition 31.3.3 *Under the preceding hypotheses, θ_n is a consistent (a.s. consistent) estimator of θ if α_n is a consistent (a.s. consistent) estimator of α.*

The following asymptotic result also holds.

Theorem 31.3.3 *Under the hypotheses of Theorem 31.3.2, $s_n(\theta_n - \theta)$ converges in law to $\nabla\tau(b)(\mathcal{H}(t_1), \ldots, \mathcal{H}(t_k))^\top$ as $n \to \infty$, where $\nabla\tau$ is the matrix with entries $\frac{\partial \tau_i}{\partial s_j}$, $1 \leq i, j, \leq k$.*

31.4 An Illustration

To illustrate the last result, let us consider Linnik laws when β is not known. Here, $L(t, \beta, \dot{\gamma}) = (1 + \dot{\gamma} t)^{-\beta}$. Note that this notation is consistent with the previous one if we set $\gamma^\alpha = \dot{\gamma}$.

Suppose that α or one of its consistent estimators is known and in either case set it as α_n. Consider $\theta = (\beta, \gamma)^\top \in \Theta = (0, \infty)^2$ and for $0 < t_1 < t_2$, let us examine the mapping

$$(\tilde{\beta}, \tilde{\gamma})^\top \to \psi(\tilde{\beta}, \tilde{\gamma}) = \left((1 + \tilde{\gamma} t_1)^{-\beta}, (1 + \tilde{\gamma} t_2)^{-\beta}\right)^\top.$$

The range of ψ is

$$W = \left\{(s_1, s_2) : s_1 > s_2 > 0, \frac{\log s_1}{\log s_2} > \frac{t_1}{t_2}\right\}.$$

The mapping ψ is invertible if $(s_1, s_2) \in W$ or equivalently

$$\frac{\log L(t_1)}{\log L(t_2)} > \frac{t_1}{t_2},$$

where $s_i = L(t_i) = (1+\dot\gamma t_i)^{-\beta}$, $i = 1,2$.

Further, the mapping

$$\tilde\beta \to \frac{s_1^{-1/\tilde\beta}-1}{s_2^{-1/\tilde\beta}-1} \; in \left(0, \frac{\log s_1}{\log s_2}\right)$$

is increasing, so there is a unique solution $\tau_1(s_1,s_2)$ (in $\tilde\beta$) of the equation

$$\frac{s_1^{-1/\tilde\beta}-1}{s_2^{-1/\tilde\beta}-1} = \frac{t_1}{t_2}.$$

Set

$$\tau_2(s_1,s_2) = \frac{s_1^{-1/\tau_1(s_1,s_2)}-1}{t_1}$$

and

$$\beta_n^{\mathrm{P}} = \tau_1(L_n(t_1), L_n(t_2)), \qquad \dot\gamma_n^{\mathrm{P}} = \tau_2(L_n(t_1), L_n(t_2))$$

for

$$(L_n(t_1), L_n(t_2)) \in W \tag{31.7}$$

and $\beta_n^{\mathrm{P}} = \dot\gamma_n^{\mathrm{P}} = 1$ otherwise. Next, denote by D the expression

$$D = s_1^{-1/\beta}\log s_1 - \frac{t_1}{t_2} s_2^{-1/\beta}\log s_2.$$

Then the entries of the matrix $\nabla\tau$ are given by

$$\frac{\partial\tau_1}{\partial s_1} = \beta s_1^{-1-1/\beta}/D, \qquad \frac{\partial\tau_1}{\partial s_2} = -\beta\frac{t_1}{t_2}s_2^{-1-1/\beta}/D,$$

$$\frac{\partial\tau_2}{\partial s_1} = \frac{s_1^{-1/\beta}}{\beta t_1}\left\{\frac{\log s_1}{\beta}\frac{\partial\tau_1}{\partial s_1} - \frac{1}{s_1}\right\}, \qquad \frac{\partial\tau_2}{\partial s_2} = \frac{s_1^{-1/\beta}}{\beta^2 t_1}\frac{\partial\tau_1}{\partial s_2}\log s_1.$$

Proposition 31.4.1 *Let α or one of its consistent (a.s. consistent) estimators be known; in either case, denote it by α_n. If (31.7) holds, then β_n^{P}, $\dot\gamma_n^{\mathrm{P}}$, and $\gamma_n^{\mathrm{P}} = (\dot\gamma_n^{\mathrm{P}})^{1/\alpha_n}$ are consistent (a.s. consistent) estimators of β, $\dot\gamma$, and $\gamma = (\dot\gamma)^{1/\alpha}$ respectively.*

PROOF. We apply Proposition 31.3.3. ∎

Set $q(t) = (1+t)\log[(1+t)/t]$, $t > 0$. We also have the following.

Theorem 31.4.1 *Let α_n be a consistent estimator of α and suppose that*

$$s_n(\alpha_n - \alpha, U_n(t) - U(t)) \Rightarrow (\mathcal{A}, h\mathcal{G}(t))$$

in $\mathbf{R} \times C([0,1])$ *as* $n \to \infty$ *in such a way that*

$$s_n n^{-1/2} \to h \in [0, \infty).$$

Then, $s_n(\beta_n^{\mathrm{P}} - \beta, \dot{\gamma}_n^{\mathrm{P}} - \dot{\gamma})^\top$ *converges in law to*

$$(\mathcal{H}(t_1), \mathcal{H}(t_2)) \begin{pmatrix} \boldsymbol{\zeta} & \boldsymbol{\eta} \end{pmatrix} = h(\mathcal{L}(t_1), \mathcal{L}(t_2)) \begin{pmatrix} \boldsymbol{\zeta} & \boldsymbol{\eta} \end{pmatrix} + \boldsymbol{\xi}^\top \mathcal{A}$$

as $n \to \infty$, *where*

$$\boldsymbol{\zeta}^\top = \left(\frac{(1+\dot{\gamma}t_1)^{\beta+1}}{\dot{\gamma}t_1(q(\dot{\gamma}t_2) - q(\dot{\gamma}t_1))}, -\frac{(1+\dot{\gamma}t_2)^{\beta+1}}{\dot{\gamma}t_2(q(\dot{\gamma}t_2) - q(\dot{\gamma}t_1))} \right),$$

$$\boldsymbol{\eta}^\top = \left(-\frac{q(\dot{\gamma}t_2)(1+\dot{\gamma}t_1)^{\beta+1}}{\beta t_1(q(\dot{\gamma}t_2) - q(\dot{\gamma}t_1))}, \frac{q(\dot{\gamma}t_1)(1+\dot{\gamma}t_2)^{\beta+1}}{\beta t_2(q(\dot{\gamma}t_2) - q(\dot{\gamma}t_1))} \right),$$

and

$$\boldsymbol{\xi}^\top = \frac{\dot{\gamma}}{\alpha(q(\dot{\gamma}t_2) - q(\dot{\gamma}t_1))} \left(\frac{\beta}{\dot{\gamma}} \log(t_1/t_2), q(\dot{\gamma}t_1)\log t_2 - q(\dot{\gamma}t_2)\log t_1 \right); \quad (31.8)$$

here, $(\mathcal{L}(t_1), \mathcal{L}(t_2))^\top$ *is distributed as a centered bivariate Gaussian random vector with covariance matrix* $\boldsymbol{V} = (v(1 - t_i^{1/\alpha}, 1 - t_j^{1/\alpha}))$.

Moreover, $s_n(\beta_n^{\mathrm{P}} - \beta, \gamma_n^{\mathrm{P}} - \gamma)^\top$ *converges in law to*

$$h(\mathcal{L}(t_1), \mathcal{L}(t_2)) \begin{pmatrix} \boldsymbol{\zeta} & \tilde{\boldsymbol{\eta}} \end{pmatrix} + \tilde{\boldsymbol{\xi}}^\top \mathcal{A}$$

as $n \to \infty$, *where* $\tilde{\boldsymbol{\eta}} = \gamma^{1-\alpha}\boldsymbol{\eta}/\alpha$, $\tilde{\xi}_1 = \xi_1$, *and* $\tilde{\xi}_2 = \left(\gamma^{1-\alpha}\xi_2 - \gamma \log \gamma\right)/\alpha$.

PROOF. The result follows from the preceding calculations of $\nabla\tau$ and Theorem 31.3.3. ■

Remark 31.4.1 Since $L(t) = (1+\gamma t)^{-\beta}$ is the Laplace transform of a gamma random variable with parameters β and γ, observe that Proposition 31.4.1 and Theorem 31.4.1 concern the estimation of the parameters of a gamma distribution from estimates of its Laplace transform.

Remark 31.4.2 Take $\alpha_n = \min\{\check{\alpha}_{k_n}, 2\}$. In order to construct 95% confidence intervals for β and γ, we proceed as follows: if $\alpha_n = 2$, apply Theorem 31.4.1 with $s_n = n^{1/2}$ and $\mathcal{A} = 0$; if $\alpha_n < 2$, apply Theorem 31.4.1 with $h = 0$. Since $P(\alpha_n = 2) \to 1$ when the true value of α is 2 and $P(\alpha_n = 2) \to 0$ when the true value of α is less than 2, we obtain that the asymptotic significance of the confidence intervals for β and γ constructed by this method is 95%.

In the case $h = 0$, Theorem 31.4.1 can be restated in the following way.

Theorem 31.4.2 *Let α_n be a consistent estimator of α and suppose that $s_n(\alpha_n - \alpha)$ converges in law to a random variable $\mathcal{A}$ as $n \to \infty$ in such a way that $s_n n^{-1/2} \to 0$. Then, $s_n(\alpha_n - \alpha, \beta_n^{\mathrm{P}} - \beta, \dot{\gamma}_n^{\mathrm{P}} - \gamma)^{\top}$ converges in law to $\boldsymbol{\xi}'\mathcal{A}$, where $\boldsymbol{\xi}'^{\top} = (1, \boldsymbol{\xi}) \in \mathbf{R}^3$ with $\boldsymbol{\xi}$ given by* (31.8).

Remark 31.4.3 Theorem 31.4.2 yields simultaneous asymptotic confidence intervals for α, β, and γ.

31.5 Numerical Results and Estimator Efficiency

For Sibuya laws, we simulated samples of size $n = 10^2$, 5×10^2, 10^3, and we estimated α and γ using Theorem 31.3.1 with two points t_1 and t_2. The results are shown below. It appears that the estimations are quite good, even for small samples.

Empirical significance level P_α, P_γ corresponding to a theoretical significance level of 5%

(A) Simulated samples with $\alpha = 0.4$, $\gamma = 5.0$

n	t_1	t_2	P_α	P_γ
100	0.01	0.1	0.0561	0.0615
500	0.01	0.1	0.0539	0.0552
1000	0.01	0.1	0.0493	0.0499

(B) Simulated samples with $\alpha = 0.8$, $\gamma = 2.0$

n	t_1	t_2	P_α	P_γ
100	0.01	0.1	0.0407	0.0635
500	0.01	0.1	0.0416	0.0495
1000	0.01	0.1	0.0511	0.0498

(C) Simulated samples with $\alpha = 1.0$, $\gamma = 0.3$

n	t_1	t_2	P_α	P_γ
100	0.05	0.5	0.0322	0.2593
500	0.05	0.5	0.0265	0.0618
1000	0.05	0.5	0.0263	0.0358

For the estimation of α, we used the minimum between α_n^{P} and 1. The results are consistent with Remark 31.3.2.

In Jacques, Rémillard, and Theodorescu (1999, Subsection 4.3), we simulated several samples of size $n = 10^2, 5 \times 10^2, 10^3, 5 \times 10^3, 10^4, 10^5$ from univariate Linnik laws. We considered two points t_1 and t_2. Generally, we were satisfied with the estimations obtained for α, β, and γ for $n \geq 5 \times 10^3$, although the estimations showed a high degree of variability with respect to t_1 and t_2. When β was known, this variability diminished tremendously. For certain $\alpha < 1$, we obtained most satisfactory results even for $n = 10^2$. For 10^4 samples, we scored a satisfactory percentage of estimations within the 95% confidence interval. From the expression of Y_n and from (31.5), we deduce that t_1 and t_2 should be restricted to a region where $Y_n(t)$ is 'linear' with respect to $\log |t|$.

The asymptotic efficiency of the adaptive Press-type estimators depends on the chosen finite point set t_i, $1 \leq i \leq q$. A possible approach may follow the line indicated by Csörgő (1984) for stable laws, namely, take a 'good' estimator $\widehat{\theta}$ of the parameter θ we are concerned with, find the expression of its limiting variance and determine the t_i's which are minimizing the expression obtained by replacing in this variance f by U_n, say $t_i^*(n)$, $1 \leq i \leq q$, under appropriate conditions ensuring uniqueness. Let us go back to our estimator replacing the t_i's by the values $t_i^*(n)$'s and prove that consistency and other asymptotic properties still hold.

Acknowledgements. This work was supported by the Natural Sciences and Engineering Reseach Council of Canada and by the Fonds F.C.A.R. of the Province of Quebec. The authors wish to thank the referees for their helpful comments.

References

1. Anderson, D. N. (1992). A multivariate Linnik distribution, *Statistics & Probability Letters*, **14**, 333–336.

2. Anderson, D. N. and Arnold, B. C. (1993). Linnik distributions and processes, *Journal of Applied Probability*, **30**, 330–340.

3. Carmichael, J.-P., Massé J.-C., and Theodorescu, R. (1982). Processus gaussiens stationnaires réciproques sur un intervalle. *C. R. Acad. Sci. Paris, Sér. I*, **295**, 291–293.

4. Chambers, J. M., Mallows, C. L., and Stuck, B. W. (1976). A method for simulating stable random variables, *Journal of the American Statistical Association*, **71**, 340–344.

5. Csörgő, S. (1981). Limit behaviour of the empirical characteristic function, *Annals of Probability*, **9**, 130–144.

6. Csörgő, S. (1984). Adaptive estimation of the parameters of stable laws. In *Coll. Math. Soc. J. Bolyai 36. Limit Theorems in Probability and Statistics* (ed., P. Révész), pp. 305–368, Amsterdam: North-Holland.

7. Devroye, L. (1986). *Non-Uniform Random Variable Generation*, New York: Springer-Verlag.

8. Devroye, L. (1990). A note on Linnik's distribution, *Statistics & Probability Letters*, **9**, 305–306.

9. Devroye, L. (1993). A triptych of discrete distributions related to the stable law, *Statistics & Probability Letters*, **18**, 349–351.

10. Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol II, Second Edition, New York: John Wiley & Sons.

11. Gnedenko, B. V. (1983). On limit theorems for a random number of random variables. In *Probability Theory and Mathematical Statistics, Proc. 4th USSR - Jap. Symp. 1982*, Lecture Notes in Mathematics No. **1021**, pp. 167–176. New York: Springer-Verlag.

12. Haan, L. de (1981). Estimation of the minimum of a function using order statistics, *Journal of the American Statistical Association*, **76**, 467–469.

13. Hall, P. (1981). A comedy of errors: the canonical form for a stable characteristic function, *Bulletin of the London Mathematical Society*, **13**, 23–27.

14. Jacques, C., Rémillard, B., and Theodorescu, R. (1999). Estimation of Linnik law parameters. *Statistics and Decisions* (to appear).

15. Leitch, R. A. and Paulson, A. S. (1975). Estimation of stable law parameters: stock price behavior application, *Journal of the American Statistical Association*, **70**, 690–697.

16. Linnik, Yu. V. (1963). Linear forms and statistical criteria, I, II, *Selected Translations in Mathematical Statistics and Probability*, **3**, 1–90.

17. Lukacs, E. (1970). *Characteristic Functions*, Second Edition, New York: Hafner.

18. Mandelbrot, B. (1962). Paretian distributions and income maximization, *Quarterly Journal of Economics*, **76**, 57–85.

19. Pakes, A. G. (1998). Mixture representations for symmetric generalized Linnik laws, *Statistics & Probability Letters*, **37**, 213–224.

20. Paulson, A. S., Holcomb, E. W., and Leitch, R. A. (1975). The estimation of the parameters of the stable laws, *Biometrika*, **62**, 163–170.

21. Press, S. J. (1972). Estimation in univariate and multivariate stable distributions, *Journal of the American Statistical Association*, **67**, 842–846.

22. Zolotarev, V. M. (1966). On representation of stable laws by integrals, *Selected Translations in Mathematical Statistics and Probability*, **6**, 84–88.

32

Asymptotic Behavior of Approximate Entropy

Andrew L. Rukhin

University of Maryland at Baltimore County, Baltimore, Maryland

Abstract: In this chapter, a new concept of approximate entropy is modified and applied to the problem of testing the randomness of a string of binary bits. This concept has been introduced in a series of papers by Pincus and co-authors. The corresponding statistic is designed to measure the degree of randomness of observed sequences. It is evaluated through incremental contrasts of empirical entropies based on the frequencies of different patterns in the sequence. Sequences with large approximate entropy must have substantial fluctuation or irregularity. Alternatively, small values of this characteristic imply strong regularity, or lack of randomness. Tractable small sample distributions are hardly available, and testing randomness is based, as a rule, on fairly long strings. Therefore, to have rigorous statistical tests of randomess based on this approximate entropy statistic, one needs the limiting distribution of this characteristic under the randomness assumption. Until now, this distribution remained unknown and was thought to be difficult to obtain. The key step leading to the limiting distribution of approximate entropy is a modification of its definition based on the frequencies of different patterns in the augmented or circular version of the original sequence. It is shown that the approximate entropy as well as its modified version converges in distribution to a χ^2-random variable.

Keywords and phrases: Covariance, entropy, generalized inverse, χ^2-distribution

32.1 Introduction and Summary

In this chapter, we investigate the asymptotic behavior of approximate entropy which is used in the problem of testing for randomness a sequence of binary (or s-ary) bits. This problem is interpreted as testing for uniformity of the

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

distribution of all templates (words) of a given length. The suggested statistic is based on the comparison of the templates in the sequence that are similar with those corresponding to the length incremented by one unit.

This characteristic, the so-called approximate entropy, was suggested by Pincus and Singer (1996) as a measure of the degree of randomness of a sequence. Actually versions of it are studied in a series of papers by Pincus and co-authors [Pincus (1991), Pincus and Huang (1992), and Pincus and Kalman (1997)].

To elucidate the use of the approximate entropy in the problem of testing for randomness, denote by $\epsilon_k, k = 1, 2, \ldots, n$, the observed sequence of independent random variables taking values in the finite set $\{1, \ldots, s\}$. For a given word, $Y_i(m) = (\epsilon_i, \ldots, \epsilon_{i+m-1}), i \le n - m + 1$, let

$$D_i^m = \frac{1}{n+1-m} \# \{j : 1 \le j < n - m, Y_j(m) = Y_i(m)\}$$

and put

$$\Phi^{(m)} = \frac{1}{n+1-m} \sum_{i=1}^{n+1-m} \log D_i^m,$$

Thus, D_i^m is the relative frequency of occurrences of the template $Y_i(m)$ in the sequence, and $-\Phi^{(m)}$ is the entropy of the empirical distribution arising on the observed subset of the set of all s^m possible patterns of length m. For $Y_i(m) = (i_1, \ldots, i_m)$, let $\nu'_{i_1 \cdots i_m} = D_i^m$ denote the relative frequency of this pattern in our string. Then $\Phi^{(m)}$ can be written in an alternative form

$$\Phi^{(m)} = \sum_{i_1 \cdots i_m} \nu'_{i_1 \cdots i_m} \log \nu'_{i_1 \cdots i_m}.$$

The approximate entropy $ApEn$ of order $m, m \ge 1$, is defined as

$$ApEn(m) = \Phi^{(m)} - \Phi^{(m+1)}$$

with $ApEn(0) = -\Phi^{(1)}$. "$ApEn(m)$ measures the logarithmic frequency with which blocks of length m that are close together remain close together for blocks augmented by one position. Thus, small values of $ApEn(m)$ imply strong regularity, or persistence, in a sequence. Alternatively, large values of $ApEn(m)$ imply substantial fluctuation, or irregularity .." [Pincus and Singer (1996, p. 2083)].

Pincus and Kalman (1997) defined a sequence to be m-irregular (m-random) if its approximate entropy $ApEn(m)$ takes the largest possible value. They evaluated quantities $ApEn(m), m = 0, 1, 2$ for binary and decimal expansions of $e, \pi, \sqrt{2}$ and $\sqrt{3}$ with the surprising conclusion that the expansion of $\sqrt{3}$ demonstrated much more irregularity than that of π.

Since $-\Phi^{(m)}$ is the entropy of the empirical distribution which under randomness assumption must be approximately uniform, one should expect that for fixed m, $\Phi^{(m)} \sim -m \log s$ and $ApEn(m) = \Phi^{(m)} - \Phi^{(m+1)} \to \log s$. Indeed, this convergence follows from Theorem 2 in Pincus (1991).

Pincus and Huang (1992, p. 3072) indicate that "analytic proofs of asymptotic normality and especially explicit variance estimates for $ApEn$ appear to be extremely difficult." However, the limiting distribution of $ApEn(m) - \log s$ is needed to have rigorous statistical tests of randomness based on this statistic, especially since testing randomness is based, as a rule, on fairly long sequences.

32.2 Modified Definition of Approximate Entropy and Covariance Matrix for Frequencies

The important step leading to the form of the limiting distribution of approximate entropy is a modification of its definition as follows. The modified version of the empirical distribution entropy $-\Phi^{(m)}$ is

$$\tilde{\Phi}^{(m)} = \sum_{i_1 \cdots i_m} \nu_{i_1 \cdots i_m} \log \nu_{i_1 \cdots i_m}, \tag{32.1}$$

where $\nu_{i_1 \cdots i_m} = \omega_{i_1 \cdots i_m}/n$ denotes the relative frequency of the template $(i_1, \cdots, i_m)$ in the augmented (or circular) version of the original string, that is, in the string $(\epsilon_1, \ldots, \epsilon_n, \epsilon_1, \ldots, \epsilon_{m-1})$.

Observe that under this definition $\omega_{i_1 \cdots i_m} = \sum_k \omega_{i_1 \cdots i_m k}$, so that for any m, $\sum_{i_1 \cdots i_m} \omega_{i_1 \cdots i_m} = n$.

Define the modified approximate entropy as

$$\widetilde{ApEn}(m) = \tilde{\Phi}^{(m)} - \tilde{\Phi}^{(m+1)}. \tag{32.2}$$

By Jensen's inequality, $\log s \geq \widetilde{ApEn}(m)$ for any m, whereas it is possible that $\log s < ApEn(m)$. Therefore the largest possible value of $\widetilde{ApEn}(m)$ is merely $\log s$, which is attained when $n = s^m$ and the distribution of all m-patterns is uniform. This is a definite advantage of $\widetilde{ApEn}(m)$. Also when calculating the approximate entropy for several values of m, it is very convenient to have the sum of all frequencies of m-templates to be equal to n.

The maximally random sequences under the definition (32.2) have the empirical distribution of all patterns of a given length (in a circular version of the sequence) as close as possible to the uniform distribution. For example, from the point of view of $ApEn(1)$ the maximally random binary strings with $n = 5$, which have three zeros and two ones, are $(0, 0, 1, 1, 0)$ and $(0, 1, 1, 0, 0)$ [see Pincus and Singer (1996, p. 2084)]. According to $\widetilde{ApEn}(m)$, one should add two sequences $(1, 1, 0, 0, 0)$ and $(1, 0, 0, 0, 1)$.

Still when n is large, $ApEn(m)$ and $\widetilde{ApEn}(m)$ cannot differ much. Indeed, one has with $\omega'_{i_1\cdots i_m} = (n-m+1)\nu'_{i_1\cdots i_m}$

$$\sum_{i_1\cdots i_m} \omega'_{i_1\cdots i_m} = n - m + 1,$$

and $\omega_{i_1\cdots i_m} - \omega'_{i_1\cdots i_m} \leq m-1$. It follows that

$$\left|\nu_{i_1\cdots i_m} - \nu'_{i_1\cdots i_m}\right| \leq \frac{m-1}{n-m+1}, \tag{32.3}$$

which suggests that for a fixed m, $\Phi^{(m)}$ and $\tilde{\Phi}^{(m)}$ must be close for large n. Therefore, Pincus' approximate entropy and (32.2) also are close, and their asymptotic distributions must coincide.

The derivation of this distribution is based on the limiting covariance matrix of the joint distribution of $\omega_{i_1\cdots i_m}$. Clearly

$$\omega_{i_1\cdots i_m} = \sum_{j=1}^{n} \delta_{(i_1\cdots i_m),(\epsilon_j,\ldots,\epsilon_{j+m-1})}$$

with $\delta_{\mathbf{i},\ell}$ denoting the Kronecker symbol for two m-indices, $\mathbf{i}$ and ℓ. For any fixed m-pattern, $i_1\cdots i_m$, the random variables $\delta_{(i_1\cdots i_m),(\epsilon_j\ldots\epsilon_{j+m-1})}$ are m-dependent, so that for $|i-k| \geq m$

$$\mathbf{Cov}\left(\delta_{(i_1\cdots i_m),(\epsilon_i,\ldots,\epsilon_{i+m-1})}, \delta_{(j_1\cdots j_m),(\epsilon_k,\ldots,\epsilon_{k+m-1})}\right) = 0.$$

As $E\delta_{(i_1\cdots i_m),(\epsilon_i\ldots\epsilon_{i+m-1})} = s^{-m}$, one has for $r = |i-k| < m$ when $i \leq k$

$$\mathbf{Cov}\left(\delta_{(i_1\cdots i_m),(\epsilon_i,\ldots,\epsilon_{i+m-1})}, \delta_{(j_1\cdots j_m),(\epsilon_k,\ldots,\epsilon_{k+m-1})}\right)$$
$$= \frac{1}{s^{m+r}}\delta_{(i_{r+1}\cdots i_m),(j_1\ldots j_{m-r})} - \frac{1}{s^{2m}}.$$

Therefore,

$$\mathbf{Cov}\left(\omega_{i_1\cdots i_m}, \omega_{j_1\cdots j_m}\right) = \frac{n}{s^m}\delta_{(i_1\cdots i_m),(j_1\ldots j_m)} - \frac{n}{s^{2m}}$$
$$+ n\sum_{r=1}^{m-1}\left[\delta_{(i_{r+1}\cdots i_m),(j_1\ldots j_{m-r})}\frac{1}{s^{m+r}} + \delta_{(i_1\cdots i_{m-r}),(j_{r+1}\ldots j_m)}\frac{1}{s^{m+r}} - \frac{2}{s^{2m}}\right].$$

Now we introduce the matrix $\Sigma = \Sigma_m$ formed by $n^{-1}\mathbf{Cov}\left(\omega_{i_1\cdots i_m}, \omega_{j_1\cdots j_m}\right)$, i.e. by the elements

$$\sigma_{i_1\cdots i_m j_1\cdots j_m} = \frac{1}{s^m}\delta_{(i_1\cdots i_m),(j_1\ldots j_m)} - \frac{2m-1}{s^{2m}}$$
$$+ \sum_{r=1}^{m-1}\left[\delta_{(i_{r+1}\cdots i_m),(j_1\ldots j_{m-r})} + \delta_{(i_1\cdots i_{m-r}),(j_{r+1}\ldots j_m)}\right]\frac{1}{s^{m+r}}. \tag{32.4}$$

Observe that because of our definition of counting the frequencies, the elements of the matrix Σ_{m+1} are related to those of Σ_m by the formula

$$\sum_{ij} \sigma_{i_1\cdots i_m i j_1\cdots j_m j} = \sigma_{i_1\cdots i_m j_1\cdots j_m}.$$

The rank of the matrix Σ_{m+1} is $s^{m+1} - s^m$. Indeed, let the s^{m+1}-dimensional vector $\mathbf{e}_{m+1}$ have all coordinates equal to one. Then it belongs to the null space of Σ_{m+1} as for any $i_1, \ldots, i_{m+1}$

$$\sum_{j_1\cdots j_{m+1}} \sigma_{i_1\cdots i_{m+1} j_1\cdots j_{m+1}} = 0. \tag{32.5}$$

Consider the vectors whose $(i_1 \cdots i_{m+1})$-th coordinate has the form $\delta_{(i_1\cdots i_m),(k_1\cdots k_m)} - \delta_{(i_2\cdots i_{m+1}),(k_1\cdots k_m)}$ for some $k_1, \ldots, k_m$. Then

$$\begin{aligned}
&\sum_{j_1\cdots j_{m+1}} \sigma_{i_1\cdots i_{m+1} j_1\cdots j_{m+1}} \delta_{(j_1\cdots j_m),(k_1\cdots k_m)} \\
&= \sum_j \sigma_{i_1\cdots i_{m+1} k_1\cdots k_m j} \\
&= \frac{1}{s^{m+1}} \delta_{(i_1\cdots i_m),(k_1\cdots k_m)} - \frac{2m}{s^{2m+1}} \\
&\quad + \sum_{r=1}^{m} \delta_{(i_{r+1}\cdots i_{m+1}),(k_1\cdots k_{m+1-r})} \frac{1}{s^{m+r}} + \sum_{r=1}^{m-1} \delta_{(i_1\cdots i_{m-r}),(k_{r+1}\cdots k_m)} \frac{1}{s^{m+r+1}} \\
&= -\frac{2m}{s^{2m+1}} + \sum_{r=0}^{m-1} \left[\delta_{(i_{r+2}\cdots i_{m+1}),(k_1\cdots k_{m-r})} + \delta_{(i_1\cdots i_{m-r}),(k_{r+1}\cdots k_m)}\right] \frac{1}{s^{m+r+1}} \\
&= -\frac{2m}{s^{2m+1}} + \sum_{r=1}^{m} \left[\delta_{(i_{r+1}\cdots i_{m+1}),(k_1\cdots k_{m-r+1})} + \delta_{(i_1\cdots i_{m-r+1}),(k_r\cdots k_m)}\right] \frac{1}{s^{m+r}} \\
&= \sum_j \sigma_{i_1\cdots i_{m+1} j k_1\cdots k_m} = \sum_{j_1\cdots j_{m+1}} \sigma_{i_1\cdots i_{m+1} j_1\cdots j_{m+1}} \delta_{(j_2\cdots j_{m+1}),(k_1\cdots k_m)}.
\end{aligned} \tag{32.6}$$

Thus all these vectors also belong to the null space of Σ_{m+1}, and, as one can show, together with $\mathbf{e}_{m+1}$ they span this space. As there are $s^m - 1$ linearly independent vectors of the form above, the dimension of the null space is s^m.

Thus Σ_{m+1} is not invertible, but we show now that its generalized inverse Σ_{m+1}^- has the form

$$\Sigma_{m+1}^- = s^m \mathbf{Q} \tag{32.7}$$

with the $s^{m+1} \times s^{m+1}$ block-diagonal matrix $\mathbf{Q}$ formed by s^m blocks Q_0 of size $s \times s$,

$$Q_0 = s\mathbf{I}_1 - \mathbf{e}_1\mathbf{e}_1^T,$$

Here, $\mathbf{I}_1$ denotes the identity matrix of size $s \times s$. Thus

$$\mathbf{Q} = s\mathbf{I}_{m+1} - \mathbf{E}.$$

Here $\mathbf{E}$ is a block-diagonal matrix formed by s^m blocks formed by ones each of size $s \times s$.

Indeed we show now that

$$s^m \Sigma_{m+1} \mathbf{Q} \Sigma_{m+1} = \Sigma_{m+1}$$

or that

$$\Sigma_{m+1} \left[s^{m+1} \Sigma_{m+1} - \mathbf{I}_{m+1} \right] = s^m \Sigma_{m+1} \mathbf{E} \Sigma_{m+1}. \tag{32.8}$$

Because of (32.5), the elements of the matrix in the left-hand side have the form

$$\sum_{k_1 \cdots k_{m+1}} \sigma_{i_1 \cdots i_{m+1} k_1 \cdots k_{m+1}} \Big[-\frac{2m+1}{s^{m+1}} + \sum_{r=1}^{m} \left(\delta_{(k_{r+1} \cdots k_m),(j_1 \cdots j_{m-r})} + \delta_{(k_1 \cdots k_{m-r}),(j_{r+1} \cdots j_m)} \right) \frac{1}{s^r} \Big]$$
$$= \sum_{r=1}^{m} \sum_{k_1 \cdots k_{m+1}} \sigma_{i_1 \cdots i_{m+1} k_1 \cdots k_{m+1}} \left[\delta_{(k_{r+1} \cdots k_m),(j_1 \cdots j_{m-r})} + \delta_{(k_1 \cdots k_{m-r}),(j_{r+1} \cdots j_m)} \right] \frac{1}{s^r}$$

with $i_1 \cdots i_{m+1}$ and $j_1 \cdots j_{m+1}$ denoting the the row and the column respectively of this matrix.

On the other hand, the corresponding elements of $s^m \Sigma_{m+1} \mathbf{E} \Sigma_{m+1}$ are by (32.6)

$$s^m \sum_{k_1 \cdots k_{m+1} \ell} \sigma_{i_1 \cdots i_{m+1} k_1 \cdots k_{m+1}} \sigma_{k_1 \cdots k_m \ell j_1 \cdots j_{m+1}}$$
$$= \sum_{k_1 \cdots k_{m+1}} \sigma_{i_1 \cdots i_{m+1} k_1 \cdots k_{m+1}} \Big[-\frac{2m}{s^{m+1}} + \sum_{r=1}^{m} \left(\delta_{(k_{r+1} \cdots k_m),(j_1 \cdots j_{m-r})} + \delta_{(k_1 \cdots k_{m-r}),(j_{r+1} \cdots j_m)} \right) \frac{1}{s^r} \Big]$$
$$= \sum_{r=1}^{m} \sum_{k_1 \cdots k_{m+1}} \sigma_{i_1 \cdots i_{m+1} k_1 \cdots k_{m+1}} \times \left[\delta_{(k_{r+1} \cdots k_m),(j_1 \cdots j_{m-r})} + \delta_{(k_1 \cdots k_{m-r}),(j_{r+1} \cdots j_m)} \right] \frac{1}{s^r}.$$

Thus both matrices in (32.8) coincide, and a generalized inverse of Σ_{m+1} is given by (32.7).

We will use the formulas (32.4) and (32.7) in the next Section to obtain the limiting distribution of the approximate entropy. Note that (32.7) is a generalized inverse for a more general class of covariance matrices including Σ_{m+1}. For example, it is also a generalized inverse for the covariance matrix of frequencies corresponding to non-overlapping words, that is, to the covariance matrix of a multinomial distribution [Rukhin (1998)].

In this chapter, we show that the asymptotic distribution of $2n[\log s - \widetilde{ApEn}(m)]$ is the familiar χ^2-distribution. As $n[ApEn(m) - \widetilde{ApEn}(m)] = O_P\left(n^{-1}\right)$, the limiting distributions of Pincus' approximate entropy and of $\widetilde{ApEn}(m)$ coincide. This fact leads to an easily implementable statistical tests of randomness via the approximate entropy. See Knuth (1981) for a description of classical test procedures for randomness.

32.3 Limiting Distribution of Approximate Entropy

We prove here that the limiting distribution of $2n[\log s - \widetilde{ApEn}(m)]$ coincides with that of a χ^2-random variable, $\chi^2(s^{m+1} - s^m)$, with $s^{m+1} - s^m$ degrees of freedom.

Theorem 32.3.1 *For fixed m, as $n \to \infty$, one has the following convergence in distribution*

$$2n\left[\log s - \widetilde{ApEn}(m)\right] \to \chi^2(s^{m+1} - s^m).$$

Also

$$n[ApEn(m) - \widetilde{ApEn}(m)] = O_P\left(n^{-1}\right),$$

so that

$$2n[\log s - ApEn(m)] \to \chi^2(s^{m+1} - s^m).$$

PROOF. Let us start with the limit theorem for $\widetilde{ApEn}(m)$. One has

$$\tilde{\Phi}^{(m)} = \sum_{i_1 \cdots i_m} \nu_{i_1 \cdots i_m} \log \nu_{i_1 \cdots i_m}$$

with $\nu_{i_1 \cdots i_m}$ denoting the relative frequency of the pattern $(i_1, \cdots, i_m)$ in the string of bits $(\epsilon_1, \ldots, \epsilon_n, \epsilon_1, \ldots, \epsilon_{m-1})$.

Let

$$Z_{i_1 \cdots i_m} = \sqrt{n}\left[\nu_{i_1 \cdots i_m} - \frac{1}{s^m}\right].$$

Then by the Central Limit Theorem for m-dependent random vectors [Ibragimov and Linnik (1971)], the vector formed by $Z_{i_1 \cdots i_m}$ has the asymptotic multivariate normal distribution with zero mean and the covariance matrix Σ_m as in (32.4). Since with probability one, $\sum Z_{i_1 \cdots i_m} = 0$,

$$\begin{aligned} \tilde{\Phi}^{(m)} &= -\sum_{i_1 \cdots i_m}\left[\frac{1}{s^m} + \frac{Z_{i_1 \cdots i_m}}{\sqrt{n}}\right]\left[-m \log s + \frac{s^m Z_{i_1 \cdots i_m}}{\sqrt{n}}\right. \\ &\quad \left. - \frac{s^{2m} Z^2_{i_1 \cdots i_m}}{2n} + O_P\left(\frac{1}{n^{3/2}}\right)\right] \sim -m \log s + \frac{s^m}{2n}\sum_{i_1 \cdots i_m} Z^2_{i_1 \cdots i_m}. \end{aligned}$$

Using a similar notation for patterns of length $m+1$, let $\nu_{i_1 \cdots i_m i_{m+1}}$ be the relative frequencies, and let $Z = (Z_{i_1 \cdots i_m i_{m+1}})$ denote the vector formed by corresponding differences between empirical and theoretical probabilities. Then, because of our convention for counting the frequencies

$$Z_{i_1 \cdots i_m} = \sum_{k=1}^{s} Z_{i_1 \cdots i_m k}$$

and

$$\tilde{\Phi}^{(m+1)} \sim -(m+1)\log s + \frac{s^{m+1}}{2n} \sum_{i_1 \cdots i_m i_{m+1}} Z^2_{i_1 \cdots i_m i_{m+1}}.$$

Thus

$$\begin{aligned} & \tilde{\Phi}^{(m)} - \tilde{\Phi}^{(m+1)} \\ \sim \;& \log s - \frac{s^m}{2n}\left[\sum_{i_1 \cdots i_m}\left(\sum_k Z_{i_1 \cdots i_m k}\right)^2 - s \sum_{i_1 \cdots i_m i_{m+1}} Z^2_{i_1 \cdots i_m i_{m+1}}\right] \\ = \;& \log s - \frac{s^m}{2n} Z^T \mathbf{Q} Z \end{aligned}$$

with the $s^{m+1} \times s^{m+1}$ matrix $\mathbf{Q}$ defined by (32.7) so that $s^m\mathbf{Q}$ is a generalized inverse of Σ_{m+1}. It is well known [see, for example, Rao and Mitra (1971, Theorem 9.2.2)] that the asymptotic distribution of $Z^T \Sigma^-_{m+1} Z$ must be the χ^2-distribution with the degrees of freedom equal to the rank of Σ_{m+1}.

Therefore

$$2n\left[\log s - \widetilde{ApEn}(m)\right] \sim \chi^2(s^{m+1} - s^m).$$

The estimate (32.3) shows that if $Z'_{i_1 \cdots i_m} = \sqrt{n}\left[\nu'_{i_1 \cdots i_m} - s^{-m}\right]$, then

$$|Z'_{i_1 \cdots i_m} - Z_{i_1 \cdots i_m}| \le (m-1)\sqrt{n}/(n-m+1)$$

and

$$\left|\tilde{\Phi}^{(m)} - \Phi^{(m)}\right| \sim \frac{s^m}{2n}\left|\sum_{i_1 \cdots i_m} Z^2_{i_1 \cdots i_m} - \sum_{i_1 \cdots i_m} Z'^2_{i_1 \cdots i_m}\right| \le \frac{s^{2m}(m-1)^2}{2(n-m+1)^2},$$

which completes the proof of Theorem 32.3.1. ■

Theorem 32.3.1 provides the basis for statistical tests of randomness via the approximate entropy. Thus, say, if for the observed value $ApEn(m)$, $\chi^2(obs) = 2n\left[\log s - ApEn(m)\right]$, then the null hypothesis of randomness is rejected for large values of $\chi^2(obs)$.

The asymptotic distribution of the statistic $2n\left[\log s - \widetilde{ApEn}(m)(m)\right]$, evaluated under the alternative according to which the probability of the word

$i_1 \ldots i_{m+1}$ has the form $\pi_{i_1 \ldots i_{m+1}} = s^{-m-1} + n^{-1/2}\eta_{i_1 \ldots i_{m+1}}$ with $\eta_T \mathbf{e}_{m+1} = 0$, is the noncentral χ^2-distribution with $s^{m+1} - s^m$ degrees of freedom and the noncentrality parameter $\delta = s^{m+1}\left[\eta^T \eta\right]$. Indeed, the limiting distribution of $Z_{i_1 \cdots i_{m+1}}$ is normal with the mean μ formed by coordinates $\eta_{i_1 \ldots i_{m+1}}$, and the covariance matrix Σ_{m+1}. Thus, because of Theorem 9.2.3 of Rao and Mitra (1971), the distribution of the quadratic form $s^m Z^T \mathbf{Q} Z$ is that of a noncentral χ^2-random variable with $s^m \mathbf{tr}(\mathbf{Q}\Sigma_{m+1}) = s^{m+1} - s^m$ degrees of freedom and the noncentrality parameter equal to $s^m \mu^T \mathbf{Q}\mu$. An easy calculation shows that $\mathbf{Q}\mu = s\mu$, so that indeed $\delta = s^{m+1}\left[\eta^T \eta\right]$.

This fact allows for an approximate power function of the test of randomness based on approximate entropy. A similar test can be derived from the following considerations.

Let the block-diagonal matrix $\mathbf{R}$ of size $s^{m+1} \times s^{m+1}$ be formed by s^{m-1} blocks R_0 of size $s^2 \times s^2$,

$$R_0 = \mathbf{e}_2 \mathbf{e}_2^T.$$

Then the matrix

$$\mathbf{U} = s^m \left[s\mathbf{I}_{m+1} - 2\mathbf{E} + s^{-1}\mathbf{R}\right]$$

possesses the following property

$$\Sigma_{m+1}\mathbf{U}\Sigma_{m+1}\mathbf{U}\Sigma_{m+1} = \Sigma_{m+1}\mathbf{U}\Sigma_{m+1}. \tag{32.9}$$

Indeed, since $s^m\left[s\mathbf{I}_{m+1} - \mathbf{E}\right]$ is the genralized inverse of Σ_{m+1}, to prove (32.9) it suffices to show that

$$\Sigma_{m+1}\mathbf{V}\Sigma_{m+1}\mathbf{V}\Sigma_{m+1} = \Sigma_{m+1}\mathbf{V}\Sigma_{m+1} \tag{32.10}$$

with

$$\mathbf{V} = s^m\left[\mathbf{E} - s^{-1}\mathbf{R}\right].$$

This result can be derived from the fact that the matrix $\Sigma_{m+1}\mathbf{V}$ is idempotent.

Indeed, the elements of $\Sigma_{m+1}\mathbf{V}$ have the form

$$\begin{aligned}
s^{m-1} & \sum_{k_1 \cdots k_{m+1}} \sigma_{i_1 \cdots i_{m+1} k_1 \cdots k_{m+1}} \left(s\delta_{(k_1 \cdots k_m),(j_1 \ldots j_m)} - \delta_{(k_1 \cdots k_{m-1}),(j_1 \ldots j_{m-1})}\right) \\
&= s^{m-1} \sum_{k\ell} \left[\sigma_{i_1 \cdots i_{m+1} j_1 \cdots j_m k} - \sigma_{i_1 \cdots i_{m+1} j_1 \cdots j_{m-1} \ell k}\right] \\
&= \sum_{\ell}\Big[\delta_{(i_1 \cdots i_{m-1}),(j_1 \ldots j_{m-1})}\left[\delta_{i_m, j_m} - \delta_{i_m,\ell}\right]\frac{1}{s^2} \\
&\quad + \sum_{r=1}^{m} \delta_{(i_{r+1} \cdots i_{m+1}),(j_1 \ldots j_{m+1-r})} \frac{1}{s^{r+1}} \\
&\quad + \sum_{r=1}^{m-1} \delta_{(i_1 \cdots i_{m-r}),(j_{r+1} \ldots j_m)} \frac{1}{s^{r+2}}
\end{aligned}$$

$$
\begin{aligned}
&\quad - \sum_{r=2}^{m-1} \delta_{(i_{r+1}\cdots i_{m+1}),(j_1\cdots j_{m+1-r})}\frac{1}{s^{r+1}} \\
&\quad - \sum_{r=2}^{m-1} \delta_{(i_1\cdots i_{m-r}),(j_{r+1}\cdots j_{m-1}\ell)}\frac{1}{s^{r+2}} \\
&\quad - \delta_{(i_2\cdots i_{m+1}),(j_1\cdots j_{m-1}\ell)}\frac{1}{s^2} - \delta_{(i_1\cdots i_{m-1}),(j_2\cdots j_{m-1}\ell)}\frac{1}{s^3}\Big] \\
&= \sum_{\ell}\Big[\delta_{(i_1\cdots i_{m-1}),(j_1\cdots j_{m-1})}[\delta_{i_m,j_m} - \delta_{i_m,\ell}]\frac{1}{s^m} \\
&\quad + \sum_{r=1}^{m-1}[\delta_{(i_1\cdots i_{m-r}),(j_{r+1}\cdots j_m)} - \delta_{(i_1\cdots i_{m-r}),(j_{r+1}\cdots j_{m-1}\ell)}]\frac{1}{s^{r+2}} \\
&\quad + [\delta_{(i_2\cdots i_{m+1}),(j_1\cdots j_m)} - \delta_{(i_2\cdots i_{m+1}),(j_2\cdots j_{m-1}\ell)}]\frac{1}{s^3}\Big] \\
&= \delta_{(i_1\cdots i_{m-1}),(j_1\cdots j_{m-1})}[s\delta_{i_m,j_m} - 1]\frac{1}{s^2} \\
&\quad + \sum_{r=1}^{m-2} \delta_{(i_1\cdots i_{m-r-1}),(j_{r+1}\cdots j_{m-1})}[s\delta_{i_{m-r},j_m} - 1]\frac{1}{s^{r+2}} \\
&\quad + \delta_{(i_2\cdots i_m),(j_1\cdots j_{m-1})}[s\delta_{i_{m+1},j_m} - 1]\frac{1}{s^2} + [s\delta_{i_1,j_m} - 1]\frac{1}{s^{m+1}}.
\end{aligned}
$$

The form of these elements allows to check that $\Sigma_{m+1}\mathbf{V}\Sigma_{m+1}\mathbf{V} = \Sigma_{m+1}\mathbf{V}$, which implies that $(\Sigma_{m+1}\mathbf{V})^3 = (\Sigma_{m+1}\mathbf{V})^2$. Thus (32.10) is valid, and by Theorem 9.2.1 of Rao and Mitra (1971) the distribution of the quadratic form $Z^T\mathbf{U}Z$ is a χ^2-distribution with $\mathbf{tr}(\mathbf{U}\Sigma_{m+1}) = s^{m+1} - 2s^m + s^{m-1}$ degrees of freedom. This corresponds to the test statistic of the form $-\tilde{\Phi}^{(m-1)} + 2\tilde{\Phi}^{(m)} - \tilde{\Phi}^{(m+1)}$, having a χ^2-distribution with $s^{m+1} - 2s^m + s^{m-1}$ degrees of freedom. Thus, this statistic also can be readily used for testing randomness. However, this fact does not extend to higher order finite differences.

Observe that $-\tilde{\Phi}^{(m-1)} + \tilde{\Phi}^{(m)}$ also has a χ^2-distribution (with $\mathbf{tr}(\mathbf{V}\Sigma_{m+1}) = s^m - s^{m-1}$ degrees of freedom.)

References

1. Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables*, Wolkers Noordhoff: Groningen.

2. Knuth, D. E. (1981). *The Art of Computer Programming*, Volume 2, Second edition, Reading, MA: Addison-Wesley.

3. Pincus, S. and Huang, W.-M. (1992). Approximate entropy: statistical properties and applications, *Communication in Statistics—Theory and*

Methods, **21**, 3061–3077.

4. Pincus, S. and Kalman, R. E. (1997). Not all (possibly) "random" sequences are created equal, *Proceedings of the National Academy of Sciences*, **94**, 3513–3518.

5. Pincus, S. and Singer, B. H. (1996). Randomness and degrees of irregularity, *Proceedings of the National Academy of Sciences*, **93**, 2083–2088.

6. Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and its Applications*, New York: John Wiley & Sons.

7. Rukhin, A. L. (1998). Testing randomness via approximate entropy, In *Proceedings of Stochastics98 Prague Conference on Information Theory, Statistical Decision Functions and Random Processes* (Eds., M. Janzura and J. A. Visek), pp. 497–500, Prague: Academy of Sciences of Czech Republic.

PART VIII

Random Walks

33

Threshold Phenomena in Random Walks

A. V. Nagaev

Nicholas Copernicus University, Torun, Poland

Abstract: A parametric family of transient random walks is considered. We focus on the properties of the ladder functionals that appear when a transient random walk becomes closer and closer to a recurrent one. Within the adopted approach the crucial role is played by the so-called Spitzer series. The latter are studied with the help of a Fuk–Nagaev inequality for the large deviation probabilities.

Keywords and phrases: Spitzer series, transient and recurrent walk, ladder epoch and height, ruin probability, Cramér transformation

33.1 Introduction

Let $0, S_1, S_2, \ldots$ be a random walk in R^1 generated by the successive sums of i.i.d. variables $\xi_1, \xi_2, \ldots$, that is,

$$S_n = \xi_1 + \cdots + \xi_n, \; n \geq 1.$$

It is supposed that all the variables are defined on a measurable space $(\Omega, \mathcal{B})$. Consider a family of measures P_s, $s_- \leq s \leq s_+$, defined on $\mathcal{B}$. Denote by $a(s)$ the expectation of $S_1 = \xi_1$ with respect to P_s. We assume that $a(s)$ is continuous and strictly increases in $[s_-, s_+]$. Assume also that $0 \in [s_-, s_+]$ and $a(0) = 0$.

Obviously, s controls the behavior of the random walk. In particular, if $s = 0$, then the walk is *recurrent*, that is,

$$\inf_n S_n = -\infty, \; \sup_n S_n = \infty, \; P_0 - a.s..$$

It means that the walk visits both half-lines $(-\infty, 0)$ and $(0, \infty)$ infinitely many times. If $s \neq 0$, the walk is *transient*, that is, for all $s \in [s_-, 0)$

$$\inf_n S_n = -\infty, \; 0 \leq \sup_n S_n < \infty, \; P_s - a.s.$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

while for $s \in (0, [s_+]$

$$-\infty < \inf_n S_n \leq 0, \ \sup_n S_n = \infty, \ P_s - a.s..$$

The basic properties of the random walk can be found in Feller (1971, Ch. 12).

It is natural to regard the case $s = 0$ as *threshold* as well as the phenomena which arise as $s \to 0$. Threshold phenomena arise in queuing theory, branching processes, mathematical models of epidemics, etc. Within the context of the random walk, they were studied in Mogulski (1973) and Nagaev (1978, 1979).

Consider the boundary functionals that play a key role in the random walk theory

$$\nu = \min(n : \ S_n > 0), \ \chi = S_\nu.$$

They are called the *ladder epoch* and the *ladder height*, respectively. It is the threshold properties of the ladder pair (ν, χ) that are in the focus of our attention. Here, in contrast to Nagaev (1978, 1978), we study the case of the *lower* threshold behavior where $s \uparrow 0$. In this case, $P_s(\nu = \infty) > 0$, that is, the distribution of the ladder pair is *defective.*

Let $\psi_s(t, z)$ denote the moment generating function of (ν, χ). More precisely,

$$\psi_s(t, z) = E_s(e^{-t\nu - z\chi}, \nu < \infty).$$

Denote

$$M_\alpha = M_\alpha(s) = \sum_{n=1}^{\infty} n^\alpha P_s(S_n > 0) \tag{33.1}$$

and

$$M_{\alpha\beta} = M_{\alpha\beta}(s) = \sum_{n=1}^{\infty} n^\alpha \int_0^\infty u^\beta dP_s(S_n < u). \tag{33.2}$$

Following Lai (1976), we call the series (33.1) and (33.2) the Spitzer series. From the well–known representation [see, for example, Feller (1971, Lemma 18.1)]

$$\psi_s(t, z) = 1 - \exp\left(-\sum_{n=1}^{\infty} n^{-1} \int_{0+}^{\infty} e^{-zu} dP_s(S_n < u)\right),$$

it follows that for $s \in [s_-, 0)$

$$P_s(\nu = \infty) = e^{-M_{-1}}. \tag{33.3}$$

Furthermore

$$E_s(\nu, \nu < \infty) \ = \ M_0 e^{-M_{-1}}; \tag{33.4}$$

$$E_s(\nu^2, \nu < \infty) \ = \ (M_1 - M_0^2) e^{-M_{-1}}; \tag{33.5}$$

$$E_s(\nu^3, \nu < \infty) \ = \ (M_2 - 3M_0 M_1 + M_0^3) e^{-M_{-1}} \tag{33.6}$$

while

$$E_s(\chi, \nu < \infty) = M_{-1,1}e^{-M_{-1}}; \tag{33.7}$$

$$E_s(\chi^2, \nu < \infty) = (M_{-1,2} - M_{-1,1}^2)e^{-M_{-1}}; \tag{33.8}$$

$$E_s(\chi^3, \nu < \infty) = (M_{-1,3} - 3M_{-1,1}M_{-1,2} + M_{-1,1}^3)e^{-M_{-1}}. \tag{33.9}$$

Since $E_0\nu = \infty$, we expect that the moments (33.4)–(33.6) grow to infinity as $s \uparrow 0$. One of our goals is to study how fast they grow.

In Mogulski (1973) (see Theorem 1 and Corollary), it was shown that under certain conditions

$$E_s(\chi^k, \nu < \infty) \to E_0\chi^k. \tag{33.10}$$

Below, we try to establish (33.10) by means of direct asymptotic analysis of the Spitzer series (33.7)–(33.9).

If $s \in [s_-, 0)$, then the moment generating function of $\bar{S} = \max(0, S_1, S_2, \ldots)$ admits the following representation [see, for example, Feller (1971, Theorem 18.2)]

$$E_s e^{-t\bar{S}} = \exp\left(\sum_{n=1}^{\infty} n^{-1} \int_0^{\infty} (e^{-tu} - 1)dP_s(S_n < u)\right). \tag{33.11}$$

Let us also refer to

$$\Sigma(t) = \sum_{n=1}^{\infty} n^{-1} \int_0^{\infty} (e^{-tu} - 1)dP_s(S_n < u). \tag{33.12}$$

as a Spitzer series.

Below, by means of direct asymptotic analysis of (1.13), we establish the limiting, as $s \uparrow 0$, distribution for $\bar{S}$. In other words, we give a third proof of a fact proven earlier in Borovkov (1972) and Asmussen (1987).

It is worth noting that the asymptotic analysis of all the considered Spitzer series is implemented by the same scheme in which one of the well-known Fuk–Nagaev inequalities plays a dominant role. Such a scheme proved to be very useful in the case of the *upper* threshold behavior considered in Nagaev (1978, 1979).

The Chapter is organized as follows. In Section 33.2, we discuss a model of the risk process in which the threshold phenomena for the random walk appear. In Section 33.3, a number of auxiliary results is established. Sections 33.4 and 33.5 are devoted to the asymptotic analysis of the Spitzer series. In Section 33.6, on the basis of that analysis, we give some threshold properties of the basic boundary functionals.

33.2 Threshold Phenomena in the Risk Process

It should be noted that the case $s \downarrow 0$ was motivated in the already mentioned works Nagaev (1978, 1979). Here, we give an example that motivates our attention to the threshold phenomena arising as $s \uparrow 0$.

Let $\eta_j,\ j = 1, 2, \ldots$, be i.i.d. positive random variables interpreted as the successive claims on an insurance company. Denote by τ_j the occurence-time of the claim η_j. It is assumed that the inter-occurence times $\Delta_j = \tau_j - \tau_{j-1}$ are i.i.d. and $\tau_0 = 0$. It is assumed that the sequences $\{\eta_j\}$ and $\{\Delta_k\}$ are also independent.

Denote by $x(t)$ the capital of the company at time t, and let $x = x(0)$ be its initial capital. The inter-occurrence capital increment is determined as $c\Delta_j$, where c is called the *gross risk premium rate.* It is obvious that for $\tau_{j-1} \leq t < \tau_j$

$$x(t) = x(\tau_{j-1}) + c(t - \tau_{j-1})$$

and

$$x(\tau_j - 0) - x(\tau_j) = -\eta_j,$$

i.e., the sample paths of the *risk process* $x(t)$ are piecewise linear right-continuous and have jumps at the points τ_j. The company is ruined at $t = \tau_{j_0}$ if

$$\tau_{j_0} = \inf(t : x(t) < 0).$$

The ruin probability is defined as

$$R(x) = P(\inf_t x(t) < 0).$$

It is convenient to transfer to the process $y(t) = x - x(t)$ and the, related to it, embedded random walk $0, S_1, S_2, \ldots$, where $S_n = \xi_1 + \cdots + \xi_n$ and $\xi_j = \eta_j - c\Delta_j,\ j = 1, 2, \ldots$.

Let

$$\bar{S} = \sup(0, S_1, S_2, \ldots).$$

Then the ruin probability can be rewritten as

$$R(x) = P(\bar{S} > x).$$

It is evident that $R(x) < 1$ iff $E\eta_1 - cE\Delta_1 < 0$. The condition $c > c_0 = E\eta/E\Delta$ is called *safety loading.* For more details about the risk process, see Embrechts, Klüppelberg and Mikosch (1997), Grandell (1991) and Kalashnikov (1997).

Consider a problem that can be regarded as inverse to that of evaluating $R(x)$. Notice that the insurance company can to some extent control the risk process evolution by varying gross risk premium rate. If the company is eager to attract more clients, it should be possible reduce the gross risk premium rate under reasonable risk. However, choosing c close to c_0 makes the ruin probability greater. To compensate it, the company has to dispose of a greater initial capital. The question arises: *how great must be the initial capital given a gross risk premium rate?*

Let us state a formal set–up of the problem. Suppose that the above sequences $\{\eta_j\}$ and $\{\Delta_k\}$ are defined on a probability space $(\Omega, \mathcal{B}, \mathcal{P})$. Obviously, $\eta_j - c\Delta_j$, $j = 1, 2 \ldots$, and map $(\Omega, \mathcal{B})$ to a measurable space $(\Omega', \mathcal{B}')$. Denote by P_c, $c < c_0$, the probability measure generated by this mapping. Furthermore,

$$E_c\xi_1 = E\eta_1 - cE\Delta_1 = (c_0 - c)E\Delta_1,$$

where E_c and E correspond respectively to P_c and P. Then the ruin probability is

$$R(x, c) = P_c(\bar{S} > x).$$

Let

$$x_p(c) = \inf(u : R(u) \leq p), \qquad 0 < p < 1, \tag{33.13}$$

be the intial capital that guarantees a pregiven risk p. If $c \downarrow c_0$ then, obviously, $x_c(p) \to \infty$. The problem is to evaluate $x_p(c)$ or at least to establish an asymptotic expression for it.

Set $s = c_0 - c$. If $s \uparrow 0$, then $E_c\xi = a(s) = sE\Delta_1 \uparrow 0$. Therefore, we have a typical lower threshold case. Obviously, in order to judge how fast $x_p(c)$ grows, it suffices to establish the limiting distribution for $\bar{S}$. We will return to this problem in Section 33.7.

33.3 Auxiliary Statements

In what follows, the following condition is assumed to be fulfilled.

Condition A. We say that Condition A is fulfilled if

(A1) $a(s) \to 0$ as $s \uparrow 0$;

(A2) $0 < \inf_{s_- \leq s \leq 0} \sigma(s) \leq \sup_{s_- \leq s \leq 0} \sigma(s) < \infty$;

(A3) ξ^2 is uniformly integrable with respect to P_s, $s_- \leq s \leq 0$.

Sometimes, Condition A stands in combination with the following one

$$\sup_{s_- \leq s \leq 0} E_s \left(\max(0, \xi_1)\right)^\delta < \infty, \delta > 2. \tag{33.14}$$

We need an inequality that easily follows from the well-known Fuk–Nagaev inequalities.

Lemma 33.3.1 *If Condition A is fulfilled, then for any $\gamma \in (0,1)$*

$$P_s(S_n \geq u) \leq nP_s(\xi \geq \gamma(u+n|a|)) + c_\gamma n^{1/2\gamma}(u+n|a|)^{-1/\gamma}, \ 0 < c_\gamma < \infty.$$

PROOF. Setting in Eq. (47) of Fuk and Nagaev (1971)

$$t = 2, \ y_i = \gamma x, \ i = 1, \dots, n,$$

we obtain

$$P_s(S_n + n|a| \geq x) \leq nP_s(\xi \geq \gamma x)$$
$$+\frac{1}{2}\left(\frac{\gamma^2 x^2}{n\int_0^{\gamma x} u^2 dP_s(\xi < u)} + 1\right)^{-1/2\gamma} + \exp\left(-\frac{x^2}{8e^2 n\int_0^{\gamma x} u^2 P_s(\xi < u)}\right).$$

It remains to implement obvious estimates. The lemma is proved. ■

Lemma 33.3.2 *If Condition A is fulfilled, then there exist $c > 0$ and n_0 such that for $n \geq n_0$*

$$\sup_{s_- \leq s \leq 0} \sup_x P_s(x \leq S_n < x+1) \leq cn^{-1/2}.$$

PROOF. From Condition A, it follows that there exists n_0 such that

$$\sup_{s_- \leq s \leq 0} \sup_x P_s(x \leq S_{n_0} < x+1) \leq 1/2.$$

Set $n = kn_0 + r$, $0 \leq r \leq n_0 - 1$. For $n \geq 2n_0$, we have $k \geq \frac{n}{2n_0}$. From Kesten (1969) [see also Petrov (1975, Ch. III, 5.5)], it follows that there exists $c' > 0$ such that

$$\begin{aligned}\sup_x P_s(x \leq S_n < x+1) &\leq \sup_x P_s(x \leq S_{kn_0} < x+1) \leq \\ &< c'n^{-1/2}(1 - \sup_{s_- < s \leq 0} \sup_x P_s(x \leq S_{n_0} < x+1))^{-1/2}\end{aligned}$$

and the lemma follows. ■

From Lemma 33.3.2, it follows that the series

$$M = \sum_{n=1}^{\infty} n^{-1} P_s(S_n = 0)$$

converges uniformly in s, $s_- < s \leq 0$.

Lemma 33.3.3 *If Condition A is fulfilled, then*

$$\sum_{n=1}^{\infty} n^{-1}(P_s(S_n > na(s)) - 1/2)$$

converges uniformly in s, $s_- < s \leq 0$.

PROOF. In order to prove the lemma, one should verify that all the estimates established in Rośen (1962) hold uniformly in s. ■

33.4 Asymptotic Behavior of the Spitzer Series

From now on, c denotes any positive constant whose concrete value is of no importance. It means that, say, $c + c = c$, $c^2 = c$, etc. By $\omega(t)$, we denote any non-negative function such that $\lim_{t\to 0}\omega(t) = 0$ while θ denotes any variable varying within $[-1, 1]$.

We begin with the series (33.1).

Theorem 33.4.1 *Let Condition A hold. If $-1 < \alpha \le 0$, then as $s \uparrow 0$*

$$M_\alpha = 2^\alpha \pi^{-1/2}\Gamma(\alpha + 3/2)(\sigma^2(s)/a^2(s))^{\alpha+1}(1 + o(1)).$$

But if $\alpha > 0$, then the statement is valid under Condition A combined with (33.14) where $\delta = 2 + \alpha$.

Proof. Let us write for brevity a and σ instead of, respectively, $a(s)$ and $\sigma(s)$. Partition the sum on the right–hand side as

$$\sum_{n=1}^{\infty} = \sum_{\varepsilon \le na^2\sigma^{-2} \le 1/\varepsilon} + \sum_{na^2\sigma^{-2} > 1/\varepsilon} + \sum_{na^2\sigma^{-2} < \varepsilon} = \Sigma_1 + \Sigma_2 + \Sigma_3,$$

where $0 < \varepsilon < 1$. With the help of the central limit theorem, we have

$$\begin{aligned}
\Sigma_1 &= (\sigma/|a|)^{2(\alpha+1)} \sum_{\varepsilon \le na^2\sigma^{-2} \le 1/\varepsilon} (na^2\sigma^{-2})^\alpha (1 - \Phi(\sqrt{na^2\sigma^{-2}})) \\
&\quad \times (a^2\sigma^{-2})(1 + o(1)) \\
&= (\sigma/|a|)^{2(\alpha+1)} \int_\varepsilon^{1/\varepsilon} x^\alpha (1 - \Phi(\sqrt{x}))dx(1 + o(1)).
\end{aligned}$$

From Lemma 33.3.1, we obtain

$$\begin{aligned}
\Sigma_2 &\le c \sum_{na^2\sigma^{-2} > 1/\varepsilon} n^{\alpha+1} P_s(\xi \ge \gamma n|a|) + |a|^{-1/\gamma} \\
&\quad \times \sum_{na^2\sigma^{-2} > 1/\varepsilon} n^{\alpha - 1/2\gamma} = c(\Sigma_{21} + \Sigma_{22}).
\end{aligned}$$

Further,

$$\begin{aligned}
\Sigma_{21} &= |a|^{-\alpha-2} \sum_{n|a| \ge (1/\varepsilon)\sigma^2/|a|} (n|a|)^{\alpha+1} P_s(\xi \ge \gamma n|a|)|a| \\
&\le |a|^{-\alpha-2} \int_{c/|a|}^{\infty} u^{\alpha+1} P_s(\xi \ge \gamma u)du.
\end{aligned}$$

If $-1 < \alpha \le 0$, then

$$\int_{c/|a|}^{\infty} u^{\alpha+1} P_s(\xi \ge \gamma u) du \le c|a|^{-\alpha} \int_{c/|a|}^{\infty} u P_s(\xi \ge \gamma u) du.$$

From (A3), it follows that

$$\int_{c/|a|}^{\infty} u P_s(\xi \ge \gamma u) du \le \frac{1}{2} \int_{c/|a|}^{\infty} u^2 dP_s(\xi < u) = o(1)$$

as $s \uparrow 0$ and, therefore,

$$\Sigma_{21} = o(|a|^{-2(\alpha+1)}).$$

If $\alpha > 0$, then

$$\int_{v}^{\infty} u^{\alpha+1} P_s(\xi \ge u) du \le \frac{1}{\alpha+2} \int_{v}^{\infty} u^{\alpha+2} dP_s(\xi < u)$$

and in view of (33.14)

$$\Sigma_{21} = o(|a|^{-2(\alpha+1)}).$$

If we choose in Lemma 33.3.1 $0 < 2\gamma < 1/(\alpha+1)$, then

$$\Sigma_{22} \le \omega(\varepsilon)|a|^{-2(\alpha+1)}.$$

Thus for all sufficiently small $|s|$, we have

$$\Sigma_2 = \theta\omega(\varepsilon)|a|^{-2(\alpha+1)}.$$

It is obvious that

$$\Sigma_3 \le c\varepsilon^{\alpha+1}|a|^{-2(\alpha+1)} = \omega(\varepsilon)|a|^{-2(\alpha+1)}.$$

Since ε is arbitrary, we arrive at

$$M_\alpha = (\sigma/|a|)^{2(\alpha+1)} \int_0^\infty x^\alpha (1 - \Phi(\sqrt{x})) dx (1 + o(1)).$$

Straightforward calculations give

$$\int_0^\infty x^\alpha (1 - \Phi(\sqrt{x})) dx = 2^\alpha \pi^{-1/2} \Gamma(\alpha + 3/2).$$

The theorem is proved. ■

The series (33.2) need a little bit more complicated estimates.

Theorem 33.4.2 *Let $\alpha \geq -1$, $\beta > 1$. Assume that Condition A and (33.14), with $\delta = \alpha + \beta + 2$ are fulfilled. Then, as $s \uparrow 0$*

$$M_{\alpha\beta} = c_{\alpha\beta}\sigma(s)^{2\alpha+2\beta+2}|a|^{-2\alpha-\beta-2}(1+o(1)),$$

where

$$c_{\alpha\beta} = B(2\alpha+\beta+2, \beta+1)\int_{-\infty}^{\infty}|u|^{2(\alpha+\beta+1)}\varphi(u)du.$$

But if $-\alpha = \beta = 1$, then (33.14) is superfluous and the statement is valid under only Condition A.

PROOF. Integration by parts yields

$$\frac{1}{\beta}M_{\alpha\beta} = \sum_{n=1}^{\infty} n^{\alpha}\int_0^{\infty} u^{\beta-1}P_s(S_n \geq u)du.$$

Represent the series as

$$\frac{1}{\beta}M^{-}_{\alpha\beta} = \Sigma_1 + \Sigma_2,$$

where Σ_1 and Σ_2 correspond to the regions $((n,u) : n \geq 1,\ 0 < u < Kn^{1/2})$ and $((n,u) : n \geq 1,\ Kn^{1/2} \leq u < \infty)$. Partition Σ_1 as

$$\Sigma_1 = \sum_{\varepsilon \leq na^2\sigma^{-2} \leq 1/\varepsilon} + \sum_{na^2\sigma^{-2} > 1/\varepsilon} + \sum_{na^2\sigma^{-2} < \varepsilon} = \Sigma_{11} + \Sigma_{12} + \Sigma_{13}.$$

From the central limit theorem, it follows

$$\begin{aligned}
&\int_0^{Kn^{1/2}} u^{\beta-1}P_s(S_n \geq u)du \\
&= \int_0^{Kn^{1/2}} u^{\beta-1}\left(1 - \Phi\left(\frac{u+n|a|}{\sigma\sqrt{n}}\right)\right)du(1+o(1)) \\
&= \sigma^{\beta}n^{\beta/2}\int_{|a|\sigma^{-1}\sqrt{n}}^{K\sigma^{-1}+|a|\sigma^{-1}\sqrt{n}} (u - |a|\sigma^{-1}\sqrt{n})^{\beta-1}(1-\Phi(u))du(1+o(1)).
\end{aligned}$$

Hence,

$$\begin{aligned}
\Sigma_{11} &= \sigma^{2\alpha+2\beta+2}|a|^{-(2\alpha+\beta+2)}\sum_{\varepsilon \leq na^2\sigma^{-2} \leq 1/\varepsilon}(na^2\sigma^{-2})^{\alpha+\beta/2}a^2\sigma^{-2} \\
&\quad \times \int_{|a|\sigma^{-1}\sqrt{n}}^{K\sigma^{-1}+|a|\sigma^{-1}\sqrt{n}}(u-|a|\sigma^{-1}\sqrt{n})^{\beta-1}(1-\Phi(u))du(1+o(1)) \\
&= \sigma^{2\alpha+2\beta+2}|a|^{-(2\alpha+\beta+2)}\int_{\varepsilon}^{1/\varepsilon} x^{\alpha+\beta/2} \\
&\quad \times \left(\int_{\sqrt{x}}^{K\sigma^{-1}+\sqrt{x}}(u-\sqrt{x})^{\beta-1}(1-\Phi(u))du\right)dx(1+o(1)) \\
&= \sigma^{2\alpha+2\beta+2}|a|^{-(2\alpha+\beta+2)}(c'_{\alpha\beta} + \omega(1/K) + \omega(\varepsilon) + o(1)),
\end{aligned}$$

where

$$c'_{km} = \int\limits_0^\infty v^{\alpha+\beta/2}(\int_{\sqrt{v}}^\infty (u-\sqrt{v})^{\beta-1}(1-\Phi(u))du)dv.$$

By means of straightforward calculations, we easily obtain

$$c'_{\alpha\beta} = \frac{B(2\alpha+\beta+2,\beta)}{2(\alpha+\beta+1)} \int\limits_{-\infty}^\infty |u|^{2(\alpha+\beta+1)}\varphi(u)du.$$

Further,

$$\Sigma_{12} \le \frac{K^{\beta-1}}{\beta} \sum_{n\ge c\varepsilon^{-1}|a|^{-1}} n^{\alpha+\beta/2} P_s(S_n \ge 0).$$

From Lemma 33.3.1,

$$\begin{aligned}\Sigma_{12} &\le cK^{\beta-1}(\sum_{n\ge c\varepsilon^{-1}|a|^{-1}} n^{\alpha+\beta/2+1} P_s(\xi \ge \gamma|a|) \\ &\quad +|a|^{-1/\gamma} \sum_{n\ge c\varepsilon^{-1}|a|^{-1}} n^{\alpha+\beta/2-1/2\gamma}) \\ &= cK^m(\Sigma'_{12}+\Sigma''_{12}).\end{aligned}$$

Taking into account (33.14),

$$\begin{aligned}\Sigma'_{12} &\le |a|^{-\alpha-\beta/2-2} \int_{c|a|^{-1}\varepsilon^{-1}}^\infty u^{\alpha+\beta/2+1} P_s(\xi\ge u)du \\ &\le c|a|^{-\alpha-2} \int_{c|a|^{-1}\varepsilon^{-1}}^\infty u^{\alpha+\beta+1} P_s(\xi \ge u)du.\end{aligned}$$

Since $\alpha+\beta \ge 0$, we have

$$\Sigma'_{12} = o(|a|^{-2\alpha-\beta-2}).$$

If we choose in Lemma 33.3.1 $0<\gamma<1/(2\alpha+\beta+2)$, then

$$\Sigma'_{12} \le c|a|^{-1/\gamma}(\varepsilon a^2)^{1/2\gamma-\alpha-\beta/2-1} = \omega(\varepsilon)|a|^{-2\alpha-\beta-2}.$$

Thus, for all sufficiently small $|s|$

$$\Sigma_{12} \le \omega(\varepsilon)|a|^{-2\alpha-\beta-2}.$$

Further,

$$\Sigma_{13} \le \frac{K^\beta}{\beta} \sum_{n\le c\varepsilon a^{-2} n^{\alpha+\beta/2}} \le cK^\beta \varepsilon^{\alpha+\beta/2+1}|a|^{-2(\alpha+\beta/2+1)} = \omega(\varepsilon)|a|^{-2\alpha-\beta-2}.$$

It remains to estimate Σ_2. From Lemma 33.3.1, we get

$$\begin{aligned}\Sigma_2 &\le c(\sum_1^\infty n^{\alpha+1}\int_{Kn^{1/2}}^\infty u^{\beta-1}P_s(\xi \ge \gamma(u+n|a|))\,du \\ &\quad +\sum_1^\infty n^{\alpha+1/2\gamma}\int_{Kn^{1/2}}^\infty u^{\beta-1}(u+n|a|)^{-1/\gamma}du) \\ &= c(\Sigma_{21}+\Sigma_{22}).\end{aligned}$$

It is easily seen that

$$\Sigma_{21} \le c|a|^{-\alpha-2}\int_0^\infty v^{\alpha+1}(\int_{Kv^{1/2}|a|^{-1/2}}^\infty u^{\beta-1}P_s(\xi \ge \gamma(u+v))\,du)dv.$$

Since

$$\begin{aligned}&\int_0^\infty(\int_0^\infty v^{\alpha+1}u^{\beta-1}P_s(\xi\ge\gamma(u+v))du)dv \\ &= \int_0^{\pi/2}d\varphi\int_0^\infty r^{\alpha+\beta+1}P_s(\xi\ge\gamma r(\sin\varphi+\cos\varphi))dr \\ &\le (\pi/2)\int_0^\infty r^{\alpha+\beta+1}P_s(\xi\ge\gamma\sqrt{2}r)dr, \qquad (33.15)\end{aligned}$$

we have, taking into account (33.14),

$$\Sigma_{21} = o(|a|^{-\alpha-2}) = o(|a|^{-2\alpha-\beta-2}).$$

Further,

$$\begin{aligned}\Sigma_{22} &= \sum_{n=1}^\infty n^{\alpha+1/2}\int_{K+|a|\sqrt{n}}(u\sqrt{n}-n|a|)^{\beta-1}u^{-1/\gamma}du \\ &\le \sum_{n=1}^\infty n^{\alpha+\beta/2}\int_{K+|a|\sqrt{n}}u^{\beta-1-1/\gamma}du \\ &= c\sum_{n=1}^\infty n^{\alpha+\beta/2}(K+|a|\sqrt{n})^{\beta-1/\gamma} \\ &\le c|a|^{-2\alpha-\beta-2}\int_0^\infty u^{\alpha+\beta/2}(K+u)^{\beta-1/\gamma}du \\ &= \omega(1/K)|a|^{-2\alpha-\beta-2}\end{aligned}$$

provided, in Lemma 33.3.1, $0 < \gamma\min(1, 2/(2\alpha+3\beta))$. Since K and ε are arbitrary, the theorem follows. ■

It remains to study (33.12).

Theorem 33.4.3 *If Condition A is fulfilled, then as $s \uparrow 0$*

$$\Sigma(t|a|\sigma^{-2}) = 2tI(t) + o(1)$$

uniformly in t, $0 \le t \le T < \infty$, where

$$I(t) = \int_0^\infty \int_v^\infty e^{-tv(u-v)}(1-\Phi(u))dudv.$$

PROOF. Let us set for brevity $z = t|a|\sigma^{-2}$. Integration by parts yields

$$\Sigma(z) = z\sum_{n=1}^\infty n^{-1}\int_0^\infty e^{-zu}P_s(S_n \ge u)du.$$

Represent the series as

$$\Sigma(z) = \Sigma_1 + \Sigma_2,$$

where Σ_1 and Σ_2 correspond to the regions $((n,u) : n \ge 1,\ 0 < u < Kn^{1/2})$ and $((n,u) : n \ge 1,\ Kn^{1/2} \le u < \infty)$. Partition Σ_1 as

$$\Sigma_1 = \sum_{\varepsilon \le na^2\sigma^{-2} \le 1/\varepsilon} + \sum_{na^2\sigma^{-2} > 1/\varepsilon} + \sum_{na^2\sigma^{-2} < \varepsilon} = \Sigma_{11} + \Sigma_{12} + \Sigma_{13}.$$

From the central limit theorem, we have for $na^2 \asymp 1$

$$\begin{aligned}
&\int_0^{Kn^{1/2}} e^{-zu}P_s(S_n \ge u)du \\
&= \int_0^{Kn^{1/2}} e^{-zu}(1-\Phi(\frac{u+n|a|}{\sigma\sqrt{n}}))du(1+o(1)) \\
&= \sigma n^{1/2}\int_{|a|\sigma^{-1}\sqrt{n}}^{K\sigma^{-1}+|a|\sigma^{-1}\sqrt{n}} \exp(-t|a|\sigma^{-1}\sqrt{n}(u-|a|\sigma^{-1}\sqrt{n})) \\
&\quad \times(1-\Phi(u))du(1+o(1)).
\end{aligned}$$

Hence,

$$\begin{aligned}
\Sigma_{11} &= t\sum_{\varepsilon \le na^2\sigma^{-2} \le 1/\varepsilon} (na^2\sigma^{-2})^{-1/2}a^2\sigma^{-2} \\
&\quad \times \int_{|a|\sigma^{-1}\sqrt{n}}^{K\sigma^{-1}+|a|\sigma^{-1}\sqrt{n}} e^{-t|a|\sigma^{-1}\sqrt{n}(u-|a|\sigma^{-1}\sqrt{n})}(1-\Phi(u))du(1+o(1)) \\
&\quad + t\int_\varepsilon^{1/\varepsilon} v^{-1/2}\int_{\sqrt{v}}^{K/\sigma+\sqrt{v}} e^{-t\sqrt{v}(u-\sqrt{v})}(1-\Phi(u))du(1+o(1)) \\
&= t(2I(t) + \omega(1/K) + \omega(\varepsilon) + o(1)).
\end{aligned}$$

It is evident that

$$\Sigma_{12} \le cKz\sum_{na^2\sigma^{-2} \ge 1/\varepsilon} n^{-1/2}P_s(S_n \ge 0).$$

In view of Lemma 33.3.1, we obtain

$$\begin{aligned}\Sigma_{12} &\le cKz(\sum_{na^2\sigma^{-2}\ge 1/\varepsilon} n^{1/2}P_s(\xi \ge \gamma n|a|) \\ &\quad +|a|^{-1/\gamma}\sum_{na^2\sigma^{-2}>1/\varepsilon} n^{-1/2-1/2\gamma}) \\ &= cKz(\Sigma_{12}' + \Sigma_{12}'').\end{aligned}$$

Further,

$$\begin{aligned}\Sigma_{12}' &= |a|^{-3/2}\sum_{n|a|\ge c\varepsilon^{-1}|a|^{-1}} (n|a|)^{1/2}P_s(\xi \ge \gamma n|a|)|a| \\ &\le |a|^{-3/2}\int_{c\varepsilon^{-1}|a|^{-1}}^{\infty} u^{1/2}P_s(\xi\ge\gamma u)du \\ &\le c\varepsilon^{1/2}|a|^{-1}\int_{c\varepsilon^{-1}|a|^{-1}}^{\infty} uP_s(\xi\ge\gamma u)du.\end{aligned}$$

From (A3), we obtain

$$\Sigma_{12}' = o(|a|^{-1}).$$

Obviously

$$\Sigma_{12}'' \le c\varepsilon^{-1/2+1/2\gamma}|a|^{-1} = \omega(\varepsilon)|a|^{-1}.$$

Hence, for all sufficiently small $|s|$,

$$\Sigma_{12} \le K\omega(\varepsilon)t.$$

Further,

$$\Sigma_{13} \le Kz\sum_{1\le n\le \varepsilon ca^2} n^{-1/2} \le cKz\varepsilon^{1/2}|a|^{-1} = K\omega(\varepsilon)t.$$

Thus, for all sufficiently small $|s|$,

$$\Sigma_1 = t(2I(t) + \omega(1/K) + \theta K\omega(\varepsilon)).$$

It remains to estimate Σ_2. In accordance with Lemma 33.3.1,

$$\begin{aligned}\Sigma_2 &\le cz(\sum_1^\infty \int_{Kn^{1/2}}^{\infty} P_s(\xi\ge\gamma(u+n|a|))du \\ &\quad +\sum_1^\infty n^{-1+1/2\gamma}\int_{Kn^{1/2}}^{\infty}(u+n|a|)^{-1/\gamma}du) \\ &= cz(\Sigma_{21}+\Sigma_{22}).\end{aligned}$$

It is easily seen that

$$\Sigma_{21} \le c|a|^{-1}\int_0^\infty (\int_{Kv^{1/2}|a|^{-1/2}}^{\infty} P_s(\xi\ge\gamma(u+v))du)dv.$$

From (A3), it follows that [cf (33.15)]

$$\int_0^\infty \int_0^\infty P_s(\xi \geq \gamma(u+v))dudv \leq c\int_0^\infty rP_s(\xi \geq cr)dr < c$$

uniformly in s. Therefore,

$$\Sigma_{21} = o(|a|^{-1}).$$

Further,

$$\begin{aligned}
\Sigma_{22} &= c\sum_{n=1}^{\infty} n^{-1/2}(K + |a|\sqrt{n})^{-1/\gamma+1} \\
&\leq c|a|^{-1}\int_0^\infty u^{-1/2}(K+\sqrt{u})^{-1/\gamma+1}du \\
&= \omega(1/K)|a|^{-1}
\end{aligned}$$

provided $0 < \gamma < 2/3$. So

$$\Sigma_2 = t\theta\omega(1/K).$$

Since K and ε are arbitrary, the theorem follows. ■

33.5 The Asymptotic Behavior of M_{-1}

It turns out that the behavior of M_{-1} as $s \uparrow 0$ is very sensitive to local irregularities of $P_s(S_n < u)$.

Theorem 33.5.1 *If Condition A is fulfilled, then as $s \uparrow 0$*

$$M_- = \sum_{n=1}^{\infty} n^{-1}P_s(S_n > 0) = \ln(\sigma(s)/|a(s)|) + O(1).$$

But if additionally

$$P_s(S_n > na(s)) \to P_0(S_n > 0) \tag{33.16}$$

and

$$P_s(na(s) < S_n \leq 0) \to P_0(S_n = 0) \tag{33.17}$$

then

$$M_- = \ln(\sigma(s)/|a(s)|) + \ln 2^{-1/2} + \sum_{n=1}^{\infty} n^{-1}(1/2 - P_0(S_n < 0)).$$

PROOF. As in the proof of Theorem 33.4.1, we have

$$\sum_{na^2\sigma^{-2}>\varepsilon} n^{-1}P_s(S_n > 0) = \int_\varepsilon^\infty x^{-1}(1-\Phi(\sqrt{x}))dx + o(1).$$

So, it remains to estimate

$$\Sigma_3 = \sum_{na^2\sigma^{-2}<\varepsilon} n^{-1}P_s(S_n > 0).$$

We represent it as

$$\begin{aligned}\Sigma_3 &= \sum_{na^2\sigma^{-2}<\varepsilon} n^{-1}P_s(S_n - na \geq 0) - \sum_{na^2\sigma^{-2}<\varepsilon} n^{-1}P_s(na \leq S_n \leq 0)\\ &= \Sigma_{31} - \Sigma_{32}.\end{aligned}$$

Further,

$$\Sigma_{31} = \sum_{na^2\sigma^{-2}<\varepsilon} n^{-1}(P_s(S_n - na \geq 0) - 1/2) + \frac{1}{2}\sum_{na^2\sigma^{-2}<\epsilon} n^{-1}.$$

From Lemma 33.3.3, it follows that

$$\Sigma_{31} = \ln(\sigma/|a|) + \frac{1}{2}\ln\epsilon + R(s) + \frac{1}{2}E + o(1),$$

where

$$R(s) = \sum_{n=1}^\infty n^{-1}(P_s(S_n \geq na(s)) - 1/2)$$

and E is the Euler constant. From Lemma 33.3.2, we have

$$P_s(na < S_n \leq 0) < c\max(1, n|a|)n^{-1/2}.$$

Hence,

$$\sum_{1/|a|\leq n<\varepsilon a^{-2}\sigma^2} n^{-1}P_s(na < S_n \leq 0) < c|a| \sum_{n<\varepsilon a^{-2}\sigma^2} n^{-1/2} < c\varepsilon^{1/2}$$

while

$$\begin{aligned}&\sum_{1\leq n<1/|a|} n^{-1}P_s(na < S_n \leq 0)\\ &= \sum_{1\leq n\leq N} n^{-1}P_s(na < S_n \leq 0) + c\theta \sum_{N<n<1/|a|} n^{-3/2}.\end{aligned}$$

Thus,

$$\Sigma_{32} = \sum_{n=1}^N P_s(na < S_n \leq 0) + \omega(1/N) + \omega(\varepsilon).$$

Therefore,

$$\begin{aligned} M_{-1} &= \ln(\sigma/|a|) + \int_{\varepsilon}^{\infty} x^{-1}(1-\Phi(\sqrt{x}))dx + \frac{1}{2}\ln\varepsilon + R(s) \\ &\quad - \sum_{1\le n\le N} n^{-1}P_s(na < S_n \le 0) + \frac{1}{2}E + \omega(1/N) + \omega(\varepsilon) + o(1). \end{aligned}$$

Taking into account Lemmas 33.3.2 and 33.3.3, we arrive to the first assertion of the theorem.

Further, consider the function

$$\delta(x) = \begin{cases} 1/2 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\int_{\varepsilon}^{\infty} x^{-1}(1-\Phi(\sqrt{x}))dx + \frac{1}{2}\ln\varepsilon = \int_0^{\infty} x^{-1}(\Phi(-\sqrt{x}) - \delta(x))dx + \omega(\epsilon).$$

In Nagaev (1978), it was shown that

$$\int_0^{\infty} x^{-1}(\Phi(-\sqrt{x}) - \delta(x))dx + \frac{1}{2}E = \frac{1}{2}\ln 2.$$

It remains to take advantage of (33.16) and (33.17) and to recall that ε and N are arbitrary. The theorem is proved. ■

33.6 Threshold Properties of the Boundary Functionals

Let us apply the just proved theorems. From Theorem 33.5.1 and (33.3), it follows that under Condition A

$$P_s(\nu = \infty) \asymp |a(s)|.$$

If we assume also that (33.16) and (33.17) are fulfilled, then

$$P_s(\nu = \infty) = 2^{1/2}e^{-R(0)}(|a(s)|/\sigma(s))(1+o(1))$$

while in view of (33.4)–(33.6)

$$\begin{aligned} E_s(\nu, \nu < \infty) &= 2^{-1/2}e^{-R(0)}(\sigma(s)/|a(s)|)(1+o(1)); \\ E_s(\nu^2, \nu < \infty) &= 5\cdot 2^{-3/2}e^{-R(0)}(\sigma(s)/|a(s)|)^3(1+o(1)); \\ E_s(\nu^3, \nu < \infty) &= 43\cdot 2^{-5/2}e^{-R(0)}(\sigma(s)/|a(s)|)^5(1+o(1)). \end{aligned}$$

If solely Condition A takes place, then we may state only that

$$E_s(\nu^k, \nu < \infty) = O(|a(s)|^{-2k+1}).$$

According to (33.4) and (33.7), we have

$$\frac{E_s(\chi, \nu < \infty)}{E_s(\nu, \nu < \infty)} = \frac{M_{-1,1}}{M_0}.$$

Applying Theorems 33.4.1 and 33.4.2, we get

$$\frac{E_s(\chi, \nu < \infty)}{E_s(\nu, \nu < \infty)} = |a(s)|(1 + o(1)).$$

This relation may be regarded as an analogue of the well-known Wald identity $E\chi = E\xi E\nu$ that takes place when $E\xi > 0$.

Let Condition A and (33.16), (33.17) be fulfilled. Then from Theorems 33.4.2 and 33.5.1 we have, taking into account (33.7),

$$E_s(\chi, \nu < \infty) = 2^{-1/2} e^{-R(0)}(1 + o(1)),$$

i.e., [see Feller (1971, Th. 18.1)]

$$E_s(\chi, \nu < \infty) = E_0\chi + o(1).$$

It should be noted that we established the continuity of $E_s(\chi, \nu < \infty)$ with no assumption that $P_s(\xi_1 < u)$ weakly converges as $s \uparrow 0$ [see Mogulskii (1973, Theorem and Corollary)].

Unfortunately, Theorem 33.4.2 does not allow us to establish the continuity of the higher order moments. For example, from the theorem it follows only that in accordance to (33.7) and (33.8)

$$E_s(\chi^2, \nu < \infty) = o(|a(s)|^{-1}).$$

In order to establish the continuity of $E_s(\chi^k, \nu < \infty)$, $k \geq 2$, at $s = 0$, we need at least the next term of the asymptotic expansion for $M_{\alpha\beta}$.

33.7 The Limiting Distribution for $\bar{S}$

From (33.11), (33.12) and Theorem 33.4.3, it follows that the moment generating function of $|a|\sigma^{-2}\bar{S}$ converges to $g(t) = \exp(-2tI(t))$. Note that $I(t)$ does not depend on the underlying distribution. We are going to verify, as if there were no proofs of Borovkov and Asmussen, that $g(t)$ corresponds to the exponential distribution.

Consider a particular case. Let ξ_1 have the density of the form

$$p(u) = \begin{cases} \frac{e^{s-u}}{e^s+1} & \text{if } u \geq 0 \\ \frac{e^u}{e^s+1} & \text{if } u < 0. \end{cases} \tag{33.18}$$

It is easily seen that

$$a(s) = \frac{e^s - 1}{e^s + 1}$$

and

$$\sigma^2(s) = 2 + o(1) \qquad \text{as } s \uparrow 0.$$

From Theorem 33.5.1, it follows that under (33.18)

$$q_s = |a(s)|(1 + o(1)) \tag{33.19}$$

as $s \uparrow 0$.

If ξ_1 has the density (33.18), then

$$P_s(\nu = 1; \chi > x) = P_s(\nu = 1)e^{-x},$$

and for $k \geq 2$

$$\begin{aligned} P_s(\nu = k; \chi > x) &= \int_{-\infty}^{0} dP_s(S_1 \leq 0; \ldots; S_{k-2} \leq 0; S_{k-1} < u; \xi_k > x - u) \\ &= e^{-x} \int_{-\infty}^{0} dP_s(S_1 \leq 0; \ldots; S_{k-2} \leq 0; S_{k-1} < u; \xi_k > -u) \\ &= P_s(\nu = k)e^{-x}. \end{aligned}$$

Hence,

$$P_s(\chi > x | \nu < \infty) = e^{-x}$$

and

$$E_s(e^{-t\chi} | \nu < \infty) = \frac{1}{1+t}. \tag{33.20}$$

Let $(\nu_j,\ \chi_j) \stackrel{d}{=} (\nu,\ \chi)$ be i.i.d. variables. The random variable $\nu_1 + \cdots + \nu_k$ is called the kth ladder epoch of the random walk $0, S_1, S_2, \ldots$. If $s < 0$, then the total number of the ladder epochs is finite P_s- a.s. . Denote it by μ.

It is evident that

$$\bar{S} = \chi_1 + \cdots + \chi_\mu.$$

Therefore,

$$\begin{aligned} P_s(S > x) &= \sum_{k=1}^{\infty} P_s(\chi_1 + \cdots + \chi_k > x) \\ &= \sum_{k=1}^{\infty} q_s(1 - q_s)^k P_s(\chi_1 + \cdots + \chi_k > x | \mu = k) \\ &= \sum_{k=1}^{\infty} q_s(1 - q_s)^k P_s(\chi'_1 + \cdots + \chi'_k > x), \end{aligned}$$

where

$$P_s(\chi'_j < x) = P_s(\chi < x|\nu < \infty).$$

Then, in view of (33.20),

$$\varphi_s(t) = E_s e^{-t\bar{S}} = \sum_{k=1}^{\infty} q_s(1-q_s)^k(1+t)^{-k}.$$

It is easily seen that as $s \uparrow 0$

$$\varphi_s(tq_s) \to \frac{1}{1+t}$$

or, in view of (33.19),

$$\varphi_s(t|a|) \to \frac{1}{1+t}.$$

Thus, we have

$$\varphi_s(t|a|/2) \to g(t) = 1/(1+t/2).$$

On the other hand due to Theorem 33.4.3,

$$\varphi_s(t|a|/2) \to g(t) = \exp\left(-2tI(t)\right)$$

and we arrive at the curious identity

$$\exp(-2t\int_0^\infty \int_v^\infty e^{-tv(u-v)}\,(1-\Phi(u))\,dudv) = \frac{1}{1+t/2}. \tag{33.21}$$

Thus, we have proved that

$$P_s(|a(s)|\sigma^{-2}\bar{S} > z) \to e^{-2z}. \tag{33.22}$$

The identity (33.21) seems to be unknown.

It remains to make remarks on the problem discussed in Section 33.2. Assume that $Var(\eta_1)$ and $Var(\Delta_1)$ are finite. Then

$$Var({}_c\xi_1) = Var(\eta_1) + c_0^2 Var(\Delta_1) + o(1),\ c \to c_0.$$

Obviously, Condition A holds with $s = c_0 - c$. Then from (33.13) and (33.22), it follows that for any $p \in (0,1)$

$$x_p(c) \sim \frac{K\ln(1/p)}{c - c_0},$$

where

$$K = \frac{1}{2E\Delta_1}[Var(\eta_1) + c_0^2 Var(\Delta_1)].$$

References

1. Asmussen, S. (1987). *Applied Probability and Queues*, Chichester: John Wiley & Sons.

2. Borovkov, A. A. (1972). *Stochastic Processes in Queueing Theory*, Moscow: Nauka (in Russian).

3. Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling of Extremal Events*, Berlin: Springer-Verlag.

4. Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, Volume 2, New York: John Wiley & Sons.

5. Fuk, D. Kh. and Nagaev, S.V. (1971). Probabilistic inequalities for sums of independent random variables, *Probability Theory and Applications*, **16**, 660–675 (in Russian).

6. Grandell, J. (1991). *Aspects of Risk Theory*, New York: Springer-Verlag.

7. Kalashnikov, V. V. (1997), *Geometric Sums: Bounds for Rare Events with Applications*, Dordrecht: Kluwer Academic Publishers.

8. Kesten, H. (1969). A sharper form of the Doeblin–Lévy–Kolmogorov–Rogozin inequality for the concentration function, *Mathematica Scandinavica*, **25**, 133–144.

9. Lai, T. L. (1976). Asymptotic moments of random walks with applications to ladder variables and renewal theory, *Annals of Probability*, **4**, 51–66.

10. Mogulskii, A. A. (1973). Absolute estimates for moments of some boundary functionals, *Probability Theory and Applications*, **18**, 350–357 (in Russian).

11. Nagaev, A. V. (1978). Integral limit theorems for sums of ladder epochs, In *Stochastic Processes and Mathematical Statistics*, Fan, Tashkent, pp. 110–122 (in Russian).

12. Nagaev, A. V. (1979). Local limit theorems for sums of ladder epochs, In *Limit Theorems, Stochastic Processes and Their Applications*, Fan, Tashkent, 162–170 (in Russian).

13. Nagaev, A. V. (1985). On a method for the ladder heights moments calculating, *Probability Theory and Applications*, **30**, 535–538 (in Russian).

14. Petrov, V. V. (1975). *Sums of Independent Variables*, Berlin: Springer-Verlag.

15. Rosén, B. (1962). On the asymptotic distribution of sums of independent identically distributed random variables, *Arkiv for Matematik*, **4**, 323–332.

34

Identifying a Finite Graph by Its Random Walk

Heinrich v. Weizsäcker

University of Kaiserslautern, Kaiserslautern, Germany

Abstract: In this chapter, we illustrate by two examples and simplify the statement of the main result of a joint paper with Peter Scheffel [Scheffel and Weizsäcker (1997)].

Keywords and phrases: Finite graph, random walk, vertex set, Markov chain

Let $\Gamma_0 = (S, E_0)$ and $\Gamma_1 = (S, E_1)$ be two connected directed graphs with the same vertex set S and the two sets E_0, E_1 of edges. Suppose we observe $n+1$ points $X_0, \cdots, X_n$ of S which are produced by the random walk on one of the two graphs. We want to infer which graph was used and the mathematical goal is to compute the asymptotic behaviour of the error probabilities. As soon as the walk makes a step which is impossible for one of the two graphs, one knows it was the other graph. On the other hand, if all previous steps were possible for both models, then the number of competing possibilities becomes important.

This is a particular case of the following problem. Let π_0 and π_1 be two finite irreducible Markov transition matrices on the same state space. Fix an initial distribution μ and let $P_i^{(n)}$ denote the law on S^{n+1} of the Markov chain with initial measure μ and transition matrix π_i. Clearly, the two laws $P_0^{(n)}$ and $P_1^{(n)}$ become more and more singular to each other. The following result determines the exponential rate at which the overlap $2 - \|P_0^{(n)} - P_1^{(n)}\|$ (which can also be described as the sum of the error probabilities of the natural likelihood test) converges to zero. In the following, the symbol $\rho(A)$ denotes spectral radius of matrix A.

Theorem 34.1.1 *The laws $P_0^{(n)}$ and $P_1^{(n)}$ become singular at the rate*

$$r = \lim_{n\to\infty} \frac{1}{n} \log(2 - \|P_0^{(n)} - P_1^{(n)}\|) = \max_{I \in \mathcal{S}} \inf_{0<t<1} \log \rho(\pi_{t,I})$$

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

where $\pi_{t,I}$ *is the elementwise logarithmic convex combination*

$$(\pi_0(i,j)^{1-t}\pi_1(i,j)^t)_{i,j\in I}$$

and $\mathcal{S}$ *denotes the system of all subsets* I *of* S *which are maximal with respect to the property that* I *can be completely covered by a single path which has positive probability under both models.*

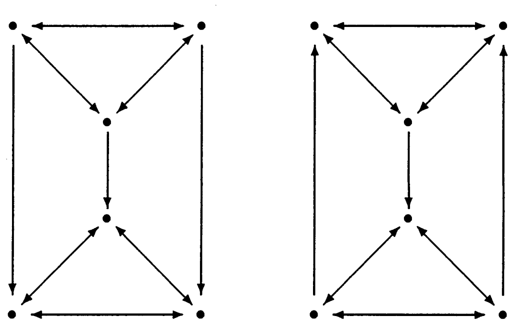
The proof uses an extension of the Large Deviation Theorem for the empirical pair distribution of ergodic Markov chains [see, for example, Dembo and Zeitouni (1993)]. This extension is needed for those cases in which some of the zero entries of one matrix are positive in the other matrix. This is always true in the graph problem. The above definition of the system $\mathcal{S}$ is much simpler than the one in Scheffel and Weizsäcker (1997).

If both π_0 and π_1 are strictly positive, or more generally if one can get from every point in S to every other point by transitions which are possible for both π_0 and π_1, then the only element of the system $\mathcal{S}$ is the set S itself and the result simplifies accordingly. As an example, for $|S| = 3$, consider the three matrices

$$\pi_0 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}, \quad \pi_1 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} & 0 \end{pmatrix}, \quad \pi_2 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

The first matrix differs from π_1 strongly in a single row and from π_2 in two rows but not so strongly. Intuitively, it is not clear which pair of Markov chains is better separated asymptotically. A numerical calculation of the corresponding spectral radii show that the rate of separation is given by $r \approx -.0115$ when comparing π_0 and π_1, and by $r \approx -.0096$ when comparing π_0 and π_2. This shows that the matrix π_0 is, empirically, more easily separated from π_1 than from π_2.

Now let us consider the following two directed graphs Γ_0 (left) and Γ_1 (right).



The only difference between the two graphs consists in the direction of the two long vertical edges. The random walk on these graphs leads to the transition matrices

$$\pi_0 = \begin{pmatrix} 0 & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}, \quad \pi_1 = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \end{pmatrix}.$$

and hence we have

$$\pi_t = \begin{pmatrix} 0 & \frac{3^{t-1}}{2^t} & \frac{3^{t-1}}{2^t} & 0 & 0 & 0 \\ \frac{3^{t-1}}{2^t} & 0 & \frac{3^{t-1}}{2^t} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & \frac{2^{t-1}}{3^t} & 0 & \frac{2^{t-1}}{3^t} \\ 0 & 0 & 0 & \frac{2^{t-1}}{3^t} & \frac{2^{t-1}}{3^t} & 0 \end{pmatrix}.$$

Consider a finite path which visits first all vertices in the upper part, then descends via the middle edge and finally visits all lower vertices. Such a path has positive probability under both models, provided the starting point has positive weight for the initial distribution μ. Therefore, in this case, the family $\mathcal{S}$ contains (only) the full set S. Thus, one has to compute the spectral radius of π_t and then pass to the infimum over t. Due to the block structure of π_t, its spectral radius is the maximum of the spectral radii of the upper left 3×3-submatrix and the similar lower right submatrix. It is easily seen that the upper left spectral radius is increasing in t and the lower right spectral radius decreases in t. Therefore, the infimum in t of the maximum of these two functions is attained at that value of t at which the two values coincide. If, however, the initial distribution is concentrated on the lower triangle, then the family $\mathcal{S}$ contains (only) the set which consists of these three lower points. In this case, one has to consider only the lower right submatrix of π_t. Due to the monotonicity mentioned above the rate r in this case is given by the logarithm of the spectral radius of the lower right submatrix of π_1.

In the general graph problem, the set $\mathcal{S}$ typically is of a more complex structure. It is easy to construct examples in which it contains two different sets which then will automatically have nontrivial relative complements with respect to each other.

References

1. Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis on the sum of observations, *Annals of Mathematical Statistics*, **23**, 493–507.

2. Dembo, A. and Zeitouni, O. (1993). *Large Deviations Techniques and Applications*, Boston: Jones and Bartlett Publishers.

3. Scheffel, P. and Weizsäcker, H. v. (1997). On risk rates and large deviations in finite parameter Markov chain experiments, *Mathematical Methods in Statistics*, **3**, 293–312.

PART IX

Miscellanea

35

The Comparison of the Edgeworth and Bergström Expansions

Vladimir I. Chebotarev and Anatolii Ya. Zolotukhin

Far-Eastern Branch of the Russian Academy of Sciences, Khabarovsk, Russia

Abstract: Uniform lower and upper bounds of remainders and terms of the Edgeworth and Bergström expansions in **R** are obtained. The case in which the Bergström expansion is allowed, but the Edgeworth one is absurd, is investigated.

Keywords and phrases: Edgeworth expansion, Bergström expansion, remainder terms, lower and upper bounds, uniform metric

35.1 Introduction and Results

It is well known that the Edgeworth expansion for the real random variables is formed with the help of the Chebyshev–Hermite polynomials, which depend on the cumulants of higher order when the number of the expansion terms is more [see, for instance, Petrov (1972, p. 173)]. On the other hand, the Bergström expansion terms are defined by the convolutions of a normal distribution with the difference between the initial distribution and the corresponding normal one [Bergström (1951)]. So in this expansion the order of the higher moment, on which the error depends, is defined not with the number of expansion terms, but with the number of the first moments of the initial distribution coinciding with the moments of the normal one. This circumstance may be a decisive cause influencing on the accuracy of the approximation. Furthermore, there are distributions for which, with the help of the Bergström expansion, one can obtain the approximation of any preassigned order (with respect to the number of the summands) while with the help of the Edgeworth expansion the approximation will be of at most of some fixed order.

We obtain lower and upper bounds of the remainders and the terms for these

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

expansions in $\mathbf{R}$. We also investigate the case when the Bergström expansion is allowed, but the Edgeworth one is absurd.

We shall use the following notations: Φ is the distribution function of the standard normal random variable α_1, $\Phi_{0,b}(x) = \Phi(x/b)$, $\varphi(x) = D_x\Phi(x)$, $H_m(x) = (-1)^m \exp\left\{\frac{x^2}{2}\right\} D_x^m \exp\left\{-\frac{x^2}{2}\right\}$ is the Hermite polynomial of the order m, $\sum\limits_{\{q\mu_q\}_{q=1}^{\nu}}$ is the sum over all sequences of non-negative integers $\mu_1, \mu_2, \ldots, \mu_\nu$, such that $\sum\limits_{q=1}^{\nu} q\mu_q = \nu$, $\kappa_q(\xi)$ is the cumulant of ξ of the order q,

$$Q_\nu(x;\xi) = \sum_{\{q\mu_q\}_{q=1}^{\nu}} h_{\nu+2s-1}(x) \prod_{q=1}^{\nu} \left[k_{q+2}(\xi)/(q+2)!\right]^{\mu_q} / \mu_q!, \tag{35.1}$$

where $h_j(x) = -\varphi(x)H_j(x)$, $s = \sum\limits_{q=1}^{\nu} \mu_q$.

Let X_1, $X_2, \ldots$ be independent copies of a random variable X, $\mathbf{E}X = 0$, $b^2 = \mathbf{D}(X)$, $\mathbf{E}|X|^{\mu+2} < \infty$. Recall that

$$\mathcal{E}_n(\mu; x; X) = \mathbf{P}\left(\frac{1}{\sqrt{n}}\sum\nolimits_{j=1}^{n} X_j < x\right) - \Phi_{0,b}(x) - \sum\nolimits_{\nu=1}^{\mu} n^{-\nu/2} Q_\nu(x/b; X/b)$$

is the remainder in the Edgeworth expansion, $\mu = 1, 2, \ldots$. Denote

$$Q_{\nu,n}(x/b; X/b) = \binom{n}{\nu} (F - \Phi_{0,b})^{*\nu} * \Phi_{0,b}^{*(n-\nu)} \left(x\sqrt{n}\right).$$

The quantity

$$\mathcal{B}_n(\mu; x; X) = \mathbf{P}\left(\frac{1}{\sqrt{n}}\sum\nolimits_{j=1}^{n} X_j < x\right) - \Phi_{0,b}(x) - \sum\nolimits_{\nu=1}^{\mu} Q_{\nu,n}(x/b; X/b) \tag{35.2}$$

is the remainder in the Bergström expansion, $\mu = 1, 2, \ldots$ [Bergström (1951)].

Define $v(t) = \mathbf{E}\exp\{itX\}$,

$$E(n;\mu;X) = \frac{e_0(\mu)}{b^{\mu+2}} \Bigg\{ 2\mathbf{E}\left[|X|^{\mu+2} I_{\{|X| \geq b\sqrt{n}\}}\right] + \frac{\mathbf{E}\left[|X|^{\mu+3} I_{\{|X| < b\sqrt{n}\}}\right]}{b\sqrt{n}}$$
$$+ b^{\mu+2} \left(\sup\left\{ |v(t)| : |t| \geq \frac{b^2}{12\,\mathbf{E}|X|^3} \right\} + \frac{1}{2n} \right)^n n^{\{\mu+(\mu+2)(\mu+3)\}/2} \Bigg\},$$

where $e_0(\mu)$ is the constant from Osipov's estimate [Osipov (1972), Petrov (1972, p. 173)]. In view of this estimate, the following formulas are valid:

$$\sup_x |\mathcal{E}_n(\mu; x; X)| \leq n^{-\mu/2} E(n;\mu;X) = o\left(n^{-\mu/2}\right) \text{ as } n \to \infty \tag{35.3}$$

if $\mathbf{E}|X|^{\mu+2} < \infty$, and

$$\sup\left\{|v(t)| : |t| \geq \frac{b^2}{12\,\mathbf{E}|X|^3}\right\} < 1. \tag{35.4}$$

Notice that the behavior of error in the Edgeworth expansion was investigated in many works; besides Osipov (1972) and Petrov (1972), see, for example, Rozovskii (1976, 1978) and Hall and Nakata (1986).

In Bergström (1951), the estimate $\mathcal{B}_n(\mu; x; X) = O\left(n^{-\frac{\mu+1}{2}}\right)$ is obtained under the condition that the distribution of X is non-singular and $\mathbf{E}|X|^3 < \infty$.

Before formulating the main statements, we introduce the following notations and definitions. Let $\{\widetilde{X}_n\}$ be a sequence of random variables (in what follows, we shall mean the triangular array). Define $\widetilde{F}_n(x) = \mathbf{P}(\widetilde{X}_n < x)$, $v_n(t) = \mathbf{E}\exp\{it\widetilde{X}_n\}$, $b_n = \mathbf{E}\widetilde{X}_n^2$,

$$\begin{aligned}
I_0(\widetilde{F}_n; n; \mu; N) &= b_n^{(\mu+1)N} \int_{-\infty}^{\infty} |t|^{(\mu+1)N-1} \left| v_n\left(\frac{t}{\sqrt{n}}\right) \right|^{\frac{n-\mu}{2}} dt, \\
B_1^\circ(n; \mu; 3) &= \frac{1}{\sqrt{n}} \left(\frac{\nu_{3,n}}{b_n^3}\right)^{\mu-1} \left[\frac{\nu_{4,n}}{b_n^4} + \frac{\nu_{3,n}}{b_n^3}\right], \\
B_2^\circ(n; \mu; 3) &= \frac{1}{\sqrt{n}} \left(\frac{\nu_{3,n}}{b_n^3}\right)^{\mu+1} I_0(\widetilde{F}_n; n; \mu; 3), \\
\nu_{r,n} \equiv \nu_r(\widetilde{X}_n, \alpha_1 b_n) &= \int_{-\infty}^{\infty} |x|^r \left|d\left(\widetilde{F}_n - \Phi_{0,b_n}\right)(x)\right|.
\end{aligned}$$

Theorem 35.1.1 *Let $\{\widetilde{X}_n\}$ be a sequence of random variables, such that* $\mathbf{E}\widetilde{X}_n = \mathbf{E}\widetilde{X}_n^3 = 0$, $\mathbf{D}\widetilde{X}_n = b_n^2$, $\mathbf{E}\widetilde{X}_n^6 < \infty$. *Let the following three conditions be fulfilled:*

$$|\kappa_6(\widetilde{X}_n)|/b_n^6 \underset{n\to\infty}{\to} \infty; \tag{35.5}$$

there exists a constant, such that for all n,

$$E(n; 4; \widetilde{X}_n) \leq c; \tag{35.6}$$

$$\lim_{n\to\infty}\left(B_1^\circ(n; \mu; 3) + B_2^\circ(n; \mu; 3)\right) = 0. \tag{35.7}$$

Then

$$n^2 \sup_x |\mathcal{E}_n(3; x; \widetilde{X}_n)| \underset{n\to\infty}{\to} \infty, \tag{35.8}$$

while for every $\mu \geq 1$,

$$\lim_{n\to\infty} n^{\mu/2} \sup_x |\mathcal{B}_n(\mu - 1; x; \widetilde{X}_n)| = 0. \tag{35.9}$$

The proof of Theorem 35.1.1 is based on the following two statements.
In the sequel $Q_\mu(x) = Q_\mu(x/b; X/b)$.

Lemma 35.1.1 *Let $\mathbf{E}|X|^{\mu+2} < \infty$ for some integer $\mu \geq 2$. Represent $Q_\mu(x)$ in the form*

$$Q_\mu(x) = \sum_{j=0}^{\mu-1} h_{2j+\mu+1}(x/b)\, b_{2j+\mu+1,\mu}, \tag{35.10}$$

where $b_{k,\mu}$ are coefficients, depending on the cumulants of X of order not more than $(k+1)\wedge(\mu+2)$. Then there exist positive numbers M_μ and $\overline{M}_\mu$ such that

$$M_\mu \bigvee_{j=0}^{\mu-1} |b_{\mu+2j+1,\mu}| - E(n;\mu) \leq n^{\mu/2} \sup_x |\mathcal{E}_n(\mu-1;x;X)|$$

$$\leq \overline{M}_\mu \bigvee_{j=0}^{\mu-1} |b_{\mu+2j+1,\mu}| + E(n;\mu), \tag{35.11}$$

where $E(n;\mu) = E(n;\mu;X)$. In particular, if $\mathbf{E}X^3 = 0$ and $\mathbf{E}X^6 < \infty$, then

$$M_4^\circ \left\{\frac{|\kappa_6|}{b^6} \bigvee \left(\frac{\kappa_4}{b^4}\right)^2\right\} - E(n;4) \leq n^2 \sup_x |\mathcal{E}_n(3;x;X)|$$

$$\leq \overline{M_4^\circ} \left\{\frac{|\kappa_6|}{b^6} \bigvee \left(\frac{\kappa_4}{b^4}\right)^2\right\} + E(n;4), \tag{35.12}$$

where $M_4^\circ > 4\cdot 10^{-5}$, $\overline{M_4^\circ} < 0.02$.

Define $h_j = \sup_x |h_j(x)|$,

$$I_1(F;n;\mu;N) = \frac{b^{(\mu+1)N}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |t|^{(\mu+1)N-1} \left| v\left(\frac{t}{\sqrt{n}}\right) \right|^{\frac{n-\mu}{2}} dt, \tag{35.13}$$

$$I(F;n;\mu;N) = I_1(F;n;\mu;N) \vee \left(\mathbf{E}|\alpha_1|^{(\mu+1)N-1} 2^{(\mu+1)N}\right), \tag{35.14}$$

$$B_1(n;\mu;3) = \sqrt{\frac{1}{\pi n}} \left(\frac{\nu_{3,n}}{b_n^3\, 3!}\right)^{\mu-1} \left[\frac{\nu_{4,n}}{b_n^4} + \frac{\nu_{3,n}}{b_n^3}\right] \frac{\mathbf{E}|\alpha_1|^{3\mu+1}}{(\mu-1)!\,3!} 2^{\frac{3\mu}{2}-1}(\mu+2), \tag{35.15}$$

$$B_2(n;\mu;3) = \sqrt{\frac{2}{\pi n}} \left(\frac{\nu_{3,n}}{b_n^3\, 3!}\right)^{\mu+1} \frac{1}{(\mu+1)!} I(F;n;\mu;3). \tag{35.16}$$

Lemma 35.1.2 *Let $\mathbf{E}X^4 < \infty$. Then for $2 \leq \mu+1 \leq n/2$,*

$$\left| n^{\frac{\mu}{2}} \sup_x |\mathcal{B}_n(\mu-1;x;X)| - \frac{h_{3\mu-1}}{\mu!} \left(\frac{|\mathbf{E}X^3|}{b^3\, 3!}\right)^\mu \right| \leq B_1(n;\mu;3) + B_2(n;\mu;3).$$

Now we shall give an example of the sequence $\{\widetilde{X}_n\}$ satisfying the conditions of Theorem 35.1.1.

Example. Consider $\widetilde{X}_n = \alpha_1 \sqrt{\lambda_n}$, where λ_n, $n = 1, 2, \ldots$, are random variables having density functions

$$f_n(x) = \begin{cases} p_n\, x^{-(4+\varepsilon_n)}, & x \in (1,\ 1/\varepsilon_n], \\ 0, & x \,\overline{\in}\, (1,\ 1/\varepsilon_n], \end{cases}$$

$n = 1, 2, \ldots,$ ε_n is a sequence, such that $\varepsilon_n \to 0$ monotonically and, moreover, $\varepsilon_n^{-1} = (n\,a_n)^\gamma \vee 3,\ 0 < \gamma < 1,\ a_n = \mathbf{E}\lambda_n = \frac{p_n}{2+\varepsilon_n}\left(1 - \varepsilon_n^{2+\varepsilon_n}\right)$, α_1 and λ_n are assumed to be independent.

Since $\mathbf{E}\lambda_n^3 \underset{n\to\infty}{\to} \infty$, $\lim\limits_{n\to\infty} a_n = \frac{3}{2}$, $\lim\limits_{n\to\infty} \mathbf{E}\lambda_n^2 = 3$, then $\kappa_3(\lambda_n) = \mathbf{E}(\lambda_n - a_n)^3 = \mathbf{E}\lambda_n^3 - a_n(3\mathbf{E}\lambda_n^2 - 2a_n^2) \underset{n\to\infty}{\to} \infty$. It follows from here that the formula (35.5) is fulfilled. One can show that $\mathbf{E}\left[\widetilde{X}_n^6 I_{\{|\widetilde{X}_n| \geq b_n\sqrt{n}\}}\right] \underset{n\to\infty}{\to} 0$, $\dfrac{\mathbf{E}\left[|\widetilde{X}_n|^7 I_{\{|\widetilde{X}_n| < b_n\sqrt{n}\}}\right]}{b_n\sqrt{n}} \leq c_1$ for all n, $\mathbf{E}\exp\{it\widetilde{X}_n\} = \mathbf{E}\exp\left\{-\frac{t^2\lambda_n}{2}\right\} \leq \exp\left\{-\frac{1}{2}\left(\frac{b_n^2}{12\,\mathbf{E}|\widetilde{X}_n|^3}\right)^2\right\}$ for $|t| \geq \dfrac{b_n^2}{12\,\mathbf{E}|\widetilde{X}_n|^3}$, $\mathbf{E}|\widetilde{X}_n|^3 = \mathbf{E}|\alpha_1|^3\,\mathbf{E}\lambda_n^{3/2} < c_2$. Therefore, (35.6) holds. The formula (35.7) follows from the boundedness of the sequences of the moments $\mathbf{E}|\widetilde{X}_n|^3$ and $\mathbf{E}\widetilde{X}_n^4$, and also from the inequality $\int_{-\infty}^{\infty} |t|^{3\mu+2} \left| v_n(t/\sqrt{n}) \right|^{\frac{n-\mu}{2}} dt \leq \int_{-\infty}^{\infty} |t|^{3\mu+2} \exp\left\{-\frac{t^2}{8}\right\} dt$, which is correct when $\mu \leq n/2$.

Similar results relating to the special case of mixtures of normal distributions are obtained in Chebotarev and Zolotukhin (1996) and Nagaev *et al.* (1997). The computer calculations in these papers and in the present one are provided by Zolotukhin.

35.2 Proof of Lemma 35.1.1

Lemma 35.2.1 *Let* $\mathbf{E}|X|^{\mu+3} < \infty$ *for some integer* $\mu \geq 1$. *If condition (35.4), is fulfilled then* $\lim\limits_{n\to\infty} n^{(\mu+1)/2}\,\mathcal{E}_n(\mu; x; X) = Q_{\mu+1}(x)$ *uniformly in* $x \in \mathbf{R}$. *Moreover,*

$$\left| n^{(\mu+1)/2} \sup_x |\mathcal{E}_n(\mu; x; X)| - \sup_x |Q_{\mu+1}(x)| \right| \leq E(n; \mu+1). \tag{35.17}$$

PROOF. It follows from (35.3) that

$$\sup_x \left| n^{(\mu+1)/2}\mathcal{E}_n(\mu; x; X) - Q_{\mu+1}(x) \right| \leq E(n; \mu+1). \tag{35.18}$$

Thus, we obtain the first statement of the lemma.

It is well known that if f, g are elements of a normed space (with the norm $\|\cdot\|$), then

$$\left| \|f\| - \|g\| \right| \leq \|f - g\|. \tag{35.19}$$

Denote $g(x) = Q_{\mu+1}(x)$, $f(x) = n^{(\mu+1)/2}\mathcal{E}_n(\mu;x;X)$, $\|\psi\| = \sup_x |\psi(x)|$. By (35.19) and (35.18), we have $\Big| \, \|f\| - \|g\| \, \Big| \le E(n;\mu+1)$, i.e. (35.17).

By virtue of Lemma 35.2.1 and (35.3), the problem of obtaining a lower estimate for $\sup_x |\mathcal{E}_n(\mu;x;X)|$ will be decided (under condition $\mathbf{E}|X|^{\mu+3} < \infty$), if we decide the same problem for $\sup_x |Q_{\mu+1}(x)|$, $\mu \le 1$.

Notice that taking into account (35.1), one can find with the help of induction that for $\nu = 1, 2, \ldots$ $Q_{2\nu-1}(x) = \sum_{j=\nu}^{3\nu-2} h_{2j}(x/b)\, b_{2j,2\nu-1}$, $Q_{2\nu}(x) = \sum_{j=\nu+1}^{3\nu} h_{2j-1}(x/b)\, b_{2j-1,2\nu}$, i.e. (35.10). We have, for instance, [Petrov (1972, p. 172)]

$$Q_1(x) = h_2(x/b)\frac{\kappa_3}{b^3\,3!}, \quad Q_2(x) = h_3(x/b)\frac{\kappa_4}{b^4\,4!} + h_5(x/b)\left(\frac{\kappa_3}{b^3\,3!}\right)^2\frac{1}{2},$$
$$Q_3(x) = h_4(x/b)\frac{\kappa_5}{b^5\,5!} + h_6(x/b)\frac{\kappa_3\,\kappa_4}{b^7\,3!\,4!} + h_8(x/b)\left(\frac{\kappa_3}{b^3\,3!}\right)^3\frac{1}{3!},$$
$$Q_4(x) = h_5(x/b)\frac{\kappa_6}{b^6\,6!} + h_7(x/b)\left[\frac{\kappa_3\,\kappa_5}{b^8\,3!\,5!} + \left(\frac{\kappa_4}{b^4\,4!}\right)^2\frac{1}{2}\right]$$
$$+h_9(x/b)\left(\frac{\kappa_3}{b^3\,3!}\right)^2\frac{1}{2}\,\frac{\kappa_4}{b^4\,4!} + h_{11}(x/b)\left(\frac{\kappa_3}{b^3\,3!}\right)^4\frac{1}{4!}, \tag{35.20}$$

where $\kappa_\nu \equiv \kappa_\nu(X)$.

In the following lemma, we suggest a method of lower estimation of $\sup_x |Q_\mu(x)|$ for arbitrary $\mu \ge 2$. Let $\psi(x)$ be a density function. The symbol $A = A(\mu;\psi)$ will denote a matrix $\mu \times \mu$ with elements

$$a_{jk} = \int_{-\infty}^{\infty} h_{2j+\mu+1}(x)\, h_{2k+\mu+1}(x)\, \psi(x)\, dx, \tag{35.21}$$

j, $k = 0, \ldots, \mu-1$. The integral $J(\mu;\psi) = \frac{1}{b}\int_{-\infty}^{\infty} Q_\mu^2(x)\, \psi(x/b)\, dx$ is the quadratic form with matrix A with respect to the coordinates of the vector $\vec{b}_\mu = (b_{\mu+1,\mu}, b_{\mu+3,\mu}, \ldots, b_{3\mu-1,\mu})$ [see (35.10)]. We assume that $\vec{b}_\mu \ne \vec{0}$. Since the function $Q_\mu(x)$ is the product of $\varphi(x/b)$ and the polynomial (of the power $3\mu-1$), then for every interval $(a,b) \subset \mathbf{R}$, there exists a point $x_0 \in (a,b)$ such that $Q_\mu(x_0) \ne 0$. Consequently, for any given density function ψ, A is a positive definite matrix [see also Cramer (1976, p. 149)]. Let $\rho(\mu;\psi)$ be the minimal eigenvalue of matrix A.

Lemma 35.2.2 *If* $\mathbf{E}|X|^{\mu+2} < \infty$, *then for every density function* $\psi(x)$,

$$\sup_x |Q_\mu(x)| \ge \sqrt{\rho(\mu;\psi)\sum_{j=0}^{\mu-1} b^2_{2j+\mu+1,\mu}} \ge \sqrt{\rho(\mu;\psi)}\,\bigvee_{j=0}^{\mu-1} |b_{2j+\mu+1,\mu}|.$$

PROOF. Evidently,

$$J(\mu;\psi) \le \sup_x Q_\mu^2(x). \tag{35.22}$$

We shall estimate $J(\mu;\psi)$ below. Notice that $J(\mu;\psi) = \sum_{j,k=0}^{\mu-1} a_{jk} d_j d_k$, where $d_j = b_{2j+\mu+1,\mu}$. Denote by $\vec{d}$ the column vector with the coordinates $d_0, \ldots, d_{\mu-1}$. Let U be an orthogonal matrix, reducing A to the diagonal form, i.e. $U\,A\,U' = V$, where V is a diagonal matrix, U' is the conjugate matrix. Denote by $v_0, v_1, \ldots, v_{\mu-1}$ the diagonal elements of V, and by $\widetilde{d}_0, \widetilde{d}_1, \ldots, \widetilde{d}_{\mu-1}$ the coordinates of the column vector $\vec{\widetilde{d}} \equiv U\,\vec{d}$. Then

$$J(\mu;\psi) = \sum_{j=0}^{\mu-1} v_j\, \widetilde{d}_j^2 \ge \left(\min_{0\le j\le \mu-1} v_j\right) \sum_{j=0}^{\mu-1} \widetilde{d}_j^2 = \rho(\mu;\psi) \sum_{j=0}^{\mu-1} d_j^2. \tag{35.23}$$

Lemma 35.2.2 follows from (35.22) and (35.23).

Lemma 35.2.3 *If* $\mathbf{E}X^6 < \infty$, *then there exists a constant* M_4, *viz.* $M_4 = 0.053\ldots$, *such that*

$$\sup_x |Q_4(x)| \ge M_4 \left\{ \frac{|\kappa_6|}{b^6\, 6!} \vee \left| \frac{\kappa_3\,\kappa_5}{b^8\, 3!\, 5!} + \left(\frac{\kappa_4}{b^4\, 4!}\right)^2 \frac{1}{2} \right| \vee \left(\left(\frac{\kappa_3}{b^3\, 3!}\right)^2 \frac{1}{2} \frac{|\kappa_4|}{b^4\, 4!} \right) \vee \left(\frac{\kappa_3}{b^3\, 3!}\right)^4 \right\}.$$

PROOF. Existence of $M_4 > 0$ follows from Lemma 35.2.2. Indeed, let us take in Lemma 35.2.2 $\psi(x) = \varphi(x)$. Calculations with the help of computer give us [see (35.21)] the matrix

$$A = \begin{pmatrix} 2.314\ldots & -13.340\ldots & 96.561\ldots & -838.564\ldots \\ -13.340\ldots & 83.221\ldots & -645.440\ldots & 5963.1\ldots \\ 96.561\ldots & -645.440\ldots & 5317.6\ldots & -51866.7\ldots \\ -838.564\ldots & 5963.1\ldots & -51866.7\ldots & 531152.9\ldots \end{pmatrix}$$

and its determinant $\det(A) = 258376.641\ldots$. In fact, the elements of the matrix are calculated with seventeen precise decimal points, and the determinant at least with four decimal points. The eigenvalues of A are calculated at least with four precise decimal points: $v_0 = 0.0029\ldots$, $v_1 = 0.6081\ldots$, $v_2 = 267.0463\ldots$, $v_3 = 536288.492\ldots$. By virtue of Lemma 35.2.2,

$$\sup_x |Q_4(x)| \ge \sqrt{\rho(4;\varphi)} \bigvee_{j=0}^{3} |b_{2j+5,4}| \ge 0.053 \bigvee_{j=0}^{3} |b_{2j+5,4}|.$$

Using (35.20) we obtain now the statement of Lemma 35.2.3.

Notice that there possibly exists a density function ψ such that $\rho(4;\psi)$ is considerably greater than 0.0029.

PROOF OF LEMMA 35.1.1. It follows from (35.17) that

$$\sup_x |Q_\mu(x)| - E(n;\mu) \le n^{\mu/2} \sup_x |\mathcal{E}_n(\mu-1;x;X)| \le \sup_x |Q_\mu(x)| + E(n;\mu).$$

By Lemma 35.2.2, $\sup_x |Q_\mu(x)| \ge \sqrt{\rho(\mu;\psi)} \bigvee_{j=0}^{\mu-1} |b_{2j+\mu+1,\mu}|$, where $\rho(\mu;\psi) \ne 0$. Moreover, it is easy to see from (35.10) that

$$\sup_x |Q_\mu(x)| \le \left\{ \bigvee_{j=0}^{\mu-1} |b_{2j+\mu+1,\mu}| \right\} \sup_x \left| \sum_{j=0}^{\mu-1} h_{2j+\mu+1}(x/b) \right|.$$

Thus we obtain (35.11).

Let $\mu = 4$, $\mathbf{E}X^3 = 0$. Then by (35.11), (35.20) and Lemma 35.2.3,

$$n^2 \sup_x |\mathcal{E}_n(3;x;X)| \ge M_4^\circ \left\{ \frac{|\kappa_6|}{b^6} \bigvee \left(\frac{\kappa_4}{b^4}\right)^2 \right\} - E(n;4),$$

where $M_4^\circ = \dfrac{0.053}{(4!)^2 2} > 4 \cdot 10^{-5}$. Moreover, by virtue of (35.20),

$$\sup_x |Q_4(x)| \le \frac{|\kappa_6|}{b^6\, 6!} h_5 + \left(\frac{\kappa_4}{b^4\, 4!}\right)^2 \frac{1}{2} h_7 \le 0.02 \left\{ \frac{|\kappa_6|}{b^6} \bigvee \left(\frac{\kappa_4}{b^4}\right)^2 \right\},$$

since $h_5 < 2.4$, $h_7 < 14.2$. Therefore, we have proved (35.12).

35.3 Proof of Lemma 35.1.2

Define $\mathbf{E}f(W) = \int_{-\infty}^{\infty} f(x)\, d(F - \Phi_{0,b})(x)$ for every measurable function f. We shall also use the abbreviation $s = it/\sqrt{n}$.

Lemma 35.3.1 *Let integer* $N \ge 3$, $\mathbf{E}W^p = 0$ *for* $p = 2, \dots, N-1$, $\mathbf{E}|X|^{N+1} < \infty$. *Then* $(\mathbf{E}\exp\{sW\})^\mu = \left(\frac{s^N}{N!}\mathbf{E}W^N\right)^\mu + R_1$ *for every integer* $\mu \ge 1$, *where* $|R_1| \le \mu\, |s|^{\mu N+1} \frac{\nu_{N+1}}{(N+1)!} \left(\frac{\nu_N}{N!}\right)^{\mu-1}$.

PROOF. By the Taylor formula,

$$\begin{aligned} R_0 &\equiv s^{-N}\, \mathbf{E}\exp\{sW\} = \frac{1}{(N-1)!} \int_0^1 (1-\theta)^{N-1} \mathbf{E}\left(W^N \exp\{\theta s W\}\right) d\theta \\ &= a_0 + a_1 s, \end{aligned}$$

where $a_0 = \frac{1}{N!}\mathbf{E}W^N$, and $a_1 = \frac{1}{N!}\int_0^1 (1-\theta)^N \mathbf{E}\left(W^{N+1}\exp\{\theta s\, W\}\right) d\theta$. Using the algebraic formula $(x+y)^\mu = x^\mu + y\sum_{j=0}^{\mu-1} x^j\,(x+y)^{\mu-1-j}$, we obtain

$$R_0^\mu = a_0^\mu + s\,a_2, \tag{35.24}$$

where $a_2 = a_1 \sum_{j=0}^{\mu-1} a_0^j\, R_0^{\mu-1-j}$. Since

$$|a_1| \le \frac{\nu_{N+1}}{(N+1)!}, \quad \left|\sum_{j=0}^{\mu-1} a_0^j\, R_0^{\mu-1-j}\right| \le \mu\left(|a_0| \vee |R_0|\right)^{\mu-1}, \quad |a_0| \vee |R_0| \le \frac{\nu_N}{N!},$$

we get

$$|a_2| \le \mu\,\frac{\nu_{N+1}}{(N+1)!}\left(\frac{\nu_N}{N!}\right)^{\mu-1}. \tag{35.25}$$

The statement of Lemma 35.3.1 follows from (35.24) and (35.25).

Define $B_{\mu,N}(x) = \frac{1}{\mu!}\left(\frac{\mathbf{E}W^N}{b^N\,N!}\right)^\mu h_{\mu N-1}(x)$. In the next lemmas, we shall use the notations (35.13)–(35.16).

Lemma 35.3.2 *Let the conditions of Lemma 35.3.1 be fulfilled. Then for every $x \in \mathbf{R}$ and integer μ such that $1 \le \mu \le n/2$,*

$$\left| n^{\frac{\mu}{2}(N-2)}\, Q_{\mu,n}(x/b; X/b) - B_{\mu,N}(x) \right| \le B_1(n;\mu;N), \tag{35.26}$$

where $B_1(n;\mu;N) = \frac{1}{\sqrt{\pi n}}\left(\frac{\nu_N}{b^N\,N!}\right)^{\mu-1}\left[\frac{\nu_{N+1}}{b^{N+1}} + \frac{\nu_N}{b^N}\right]\frac{\mathbf{E}|\alpha_1|^{\mu N+1}}{(\mu-1)!\,N!}2^{\frac{\mu N}{2}-1}(\mu+2).$

Proof. Denote $\Delta_{\mu,n}(x) = (F-\Phi_{0,b})^{*\mu} * \Phi_{0,b}^{*(n-\mu)}(x)$. By virtue of the inversion formula, for all x_1, $x_2 \in \mathbf{R}$

$$\begin{aligned}&\Delta_{\mu,n}\left(x_2\sqrt{n}\right) - \Delta_{\mu,n}\left(x_1\sqrt{n}\right)\\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\mathbf{E}\exp\{sW\}\right)^\mu \exp\left\{-\frac{(n-\mu)t^2b^2}{2n}\right\}\frac{\exp\{-itx_1\}-\exp\{-itx_2\}}{it}\,dt.\end{aligned} \tag{35.27}$$

We have

$$\exp\left\{-\frac{(n-\mu)t^2b^2}{2n}\right\} = \exp\left\{-\frac{t^2b^2}{2}\right\}(1+R_2), \tag{35.28}$$

where $R_2 = \frac{\mu}{2n}t^2b^2\exp\left\{\frac{\mu t^2b^2\theta_1}{2n}\right\}$, $0 < \theta_1 < 1$. Using the equality (35.28), Lemma 35.3.1 and the formula $\binom{n}{\mu} = \frac{n^\mu}{\mu!}(1+R_3)$, where

$$R_3 = -\frac{1}{n}\sum_{k=1}^{\mu-1} k \prod_{j\ne k}\left(1-\frac{j\theta_2}{n}\right) > -\frac{1}{n}\binom{\mu-1}{2}, \quad 0<\theta_2<1,$$

we obtain

$$\begin{aligned}
&(\mathbf{E}\exp\{sW\})^{\mu}\exp\Big\{-\frac{(n-\mu)t^2b^2}{2n}\Big\}\binom{n}{\mu}n^{\frac{\mu}{2}(N-2)}\\
&= \Big[\Big(\frac{s^N}{N!}\mathbf{E}W^N\Big)^{\mu}+R_1\Big]\exp\Big\{-\frac{(n-\mu)t^2b^2}{2n}\Big\}n^{\mu N/2}(1+R_3)\\
&= \frac{1}{\mu!}\Big(\frac{(it)^N}{N!}\mathbf{E}W^N\Big)^{\mu}\exp\Big\{-\frac{t^2b^2}{2}\Big\}+R_4, \qquad (35.29)
\end{aligned}$$

where

$$\begin{aligned}
R_4=\frac{1}{\mu!}\Big[&\Big(\frac{(it)^N}{N!}\mathbf{E}W^N\Big)^{\mu}\exp\Big\{-\frac{t^2b^2}{2}\Big\}R_2+\Big(\frac{(it)^N}{N!}\mathbf{E}W^N\Big)^{\mu}\exp\Big\{-\frac{(n-m)t^2b^2}{2n}\Big\}R_3\\
&+R_1\exp\Big\{-\frac{(n-m)t^2b^2}{2n}\Big\}n^{\mu N/2}(1+R_3)\Big].
\end{aligned}$$

Using Lemma 35.3.1, we obtain

$$\begin{aligned}
|R_4|\le \frac{1}{\mu!\sqrt{n}}\exp\Big\{-\frac{(n-\mu)t^2b^2}{2n}\Big\}&\Big\{\Big(\frac{|t|^N}{N!}\nu_N\Big)^{\mu}\frac{1}{\sqrt{n}}\Big[\frac{t^2b^2\mu}{2}+\binom{\mu-1}{2}\Big]\\
&+|t|^{\mu N+1}\mu\frac{\nu_{N+1}}{(N+1)!}\Big(\frac{\nu_N}{N!}\Big)^{\mu-1}\Big(1+\frac{1}{n}\binom{\mu-1}{2}\Big)\Big\}. \qquad (35.30)
\end{aligned}$$

By (35.27) and (35.29),

$$\begin{aligned}
&\Big(\Delta_{\mu,n}(x_2\sqrt{n})-\Delta_{\mu,n}(x_1\sqrt{n})\Big)\binom{n}{\mu}n^{\frac{\mu}{2}(N-2)}\\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}\Big[\frac{1}{\mu!}\Big(\frac{(it)^N}{N!}\mathbf{E}W^N\Big)^{\mu}\exp\Big\{-\frac{t^2b^2}{2}\Big\}+R_4\Big]\frac{\exp\{-itx_1\}-\exp\{-itx_2\}}{it}\,dt. \qquad (35.31)
\end{aligned}$$

Using (35.30) and the equality

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\Big\{-\frac{t^2c^2}{2}\Big\}|t|^l\,dt=\frac{\mathbf{E}|\alpha_1|^l}{c^{l+1}},\quad c>0, \qquad (35.32)$$

we find

$$\Big|\frac{1}{2\pi}\int_{-\infty}^{\infty}R_4\frac{\exp\{-itx_1\}-\exp\{-itx_2\}}{it}\,dt\Big|\le\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{|R_4|}{|t|}\,dt\le\widehat{B}_1(n;\mu;N), \qquad (35.33)$$

where

$$\begin{aligned}
\widehat{B}_1(n;\mu;N) &= \sqrt{\frac{2}{\pi n}}\Big(\frac{\nu_N}{b^N\,N!}\Big)^{\mu-1}\frac{\mathbf{E}|\alpha_1|^{\mu N+1}}{(\mu-1)!}\Big(\frac{n}{n-\mu}\Big)^{\frac{\mu N+1}{2}}\\
&\times\Big[\frac{\nu_{N+1}}{b^{N+1}(N+1)!}\Big(1+\frac{1}{n}\binom{\mu-1}{2}\Big)+\frac{\nu_N}{b^N\,N!\sqrt{n-\mu}}\Big],
\end{aligned}$$

Since $\frac{1}{2\pi}\int_{-\infty}^{\infty}(it)^k \exp\{-t^2/2 - itx\}\,dt = (-1)^k D_x^k\varphi(x) = -h_k(x)$, then

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}(it)^{\mu N-1}\exp\{-t^2b^2/2 - itx\}\,dt = \frac{1}{b^{\mu N}}\,h_{\mu N-1}(x/b).$$

Consequently,

$$\frac{1}{2\,\pi\,\mu!}\int_{-\infty}^{\infty}\left(\frac{(it)^N}{N!}\,\mathbf{E}W^N\right)^{\mu}\exp\left\{-\frac{t^2b^2}{2}\right\}\frac{\exp\{-itx_1\}-\exp\{-itx_2\}}{it}\,dt$$
$$= B_{\mu,N}(x_2/b) - B_{\mu,N}(x_1/b). \quad (35.34)$$

It follows from (35.31), (35.33) and (35.34) that

$$\left(\Delta_{\mu,n}\left(x_2\sqrt{n}\right) - \Delta_{\mu,n}\left(x_1\sqrt{n}\right)\right)\binom{n}{\mu}n^{\frac{\mu}{2}(N-2)}$$
$$= B_{\mu,N}(x_2/b) - B_{\mu,N}(x_1/b) + R_6(n;x_1;x_2), \quad (35.35)$$

where $|R_6(n;x_1;x_2)| \le \widehat{B}_1(n;\mu;N)$. Since $\lim_{u\to-\infty}\Delta_{\mu,n}(u) = 0$, $\lim_{u\to-\infty}B_{\mu,N}(u) = 0$, then putting $x_1 \to -\infty$, $x_2 = x$ in (35.35), we have for every $x \in \mathbf{R}$

$$\begin{aligned} n^{\frac{\mu}{2}(N-2)}\,Q_{\mu,n}(x/b;X/b) &= \Delta_{\mu,n}\left(x\sqrt{n}\right)\binom{n}{\mu}n^{\frac{\mu}{2}(N-2)} \\ &= B_{\mu,N}(x/b) + \lim_{x_1\to-\infty}R_6(n;x_1;x), \end{aligned}$$

where $\left|\lim_{x_1\to-\infty}R_6(n;x_1;x)\right| \le \widehat{B}_1(n;\mu;N)$. It is easy to see that $\widehat{B}_1(n;\mu;N) \le B_1(n;\mu;N)$ if $1 \le \mu \le n/2$. The lemma is proved.

Lemma 35.3.3. *Let the conditions of Lemma 35.3.1 be fulfilled. Then for every $x \in \mathbf{R}$ and integer μ such that $1 \le \mu \le n/2 - 1$,*

$$\left|\,n^{\frac{\mu}{2}(N-2)}\,\mathcal{B}_n(\mu-1;x;X) - B_{\mu,N}(x)\,\right| \le B_1(n;\mu;N) + B_2(n;\mu;N). \quad (35.36)$$

PROOF. It follows from the definition (35.2), that

$$\mathcal{B}_n(\mu-1;x;X) = Q_{\mu,n}(x/b;X/b) + \mathcal{B}_n(\mu;x;X). \quad (35.37)$$

Denote $\Delta_{\mu,\nu,n}(x) = (F-\Phi_{0,b})^{*(\mu+1)} * \Phi_{0,b}^{*\nu} * F^{*(n-\mu-\nu-1)}(x)$. The following identity holds

$$\mathcal{B}_n(\mu;x;X) = \sum_{\nu=0}^{n-\mu-1}\binom{\nu+\mu}{\mu}\Delta_{\mu,\nu,n}\left(x\sqrt{n}\right) \quad (35.38)$$

[see Bergström (1951)]. Similar to (35.27),

$$\begin{aligned}
&\Delta_{\mu,\nu,n}\left(x_2\sqrt{n}\right)-\Delta_{\mu,\nu,n}\left(x_1\sqrt{n}\right)\\
&=\frac{1}{2\pi}\int_{-\infty}^{\infty}(\mathbf{E}\exp\{sW\})^{\mu+1}\exp\left\{-\frac{\nu t^2b^2}{2n}\right\}\\
&\quad\times\left(v\left(\frac{t}{\sqrt{n}}\right)\right)^{n-\mu-\nu-1}\frac{\exp\{-itx_1\}-\exp\{-itx_2\}}{it}\,dt.
\end{aligned}$$

It follows from here that for all $x\in\mathbf{R}$

$$\begin{aligned}
&|\Delta_{\mu,\nu,n}\left(x\sqrt{n}\right)|\\
&\leq\frac{1}{\pi}\int_{-\infty}^{\infty}|\mathbf{E}\exp\{sW\}|^{\mu+1}\exp\left\{-\frac{\nu t^2b^2}{2n}\right\}\left|v\left(\frac{t}{\sqrt{n}}\right)\right|^{n-\mu-\nu-1}\frac{dt}{|t|}.
\end{aligned}\tag{35.39}$$

Since $\mathbf{E}\exp\{sW\}=\frac{s^N}{(N-1)!}\int_{-\infty}^{\infty}x^N\int_0^1(1-\theta)^{N-1}\exp\{\theta sx\}d\theta\,d(F-\Phi_{0,b})(x)$, then

$$|\mathbf{E}\exp\{sW\}|\leq\left(\frac{|t|}{\sqrt{n}}\right)^N\frac{\nu_N}{N!}.\tag{35.40}$$

By virtue of (35.39) and (35.40),

$$\begin{aligned}
&|\Delta_{\mu,\nu,n}\left(x\sqrt{n}\right)|\leq\frac{1}{\pi}\left(\frac{\nu_N}{N!}\right)^{\mu+1}\left(\frac{1}{\sqrt{n}}\right)^{(\mu+1)N}\\
&\times\int_{-\infty}^{\infty}|t|^{(\mu+1)N-1}\exp\left\{-\frac{\nu t^2b^2}{2n}\right\}\left|v\left(\frac{t}{\sqrt{n}}\right)\right|^{n-\mu-\nu-1}dt\leq\frac{1}{\pi}\left(\frac{\nu_N}{N!}\right)^{\mu+1}\left(\frac{1}{\sqrt{n}}\right)^{(\mu+1)N}\\
&\times\begin{cases}\displaystyle\int_{-\infty}^{\infty}|t|^{(\mu+1)N-1}\left|v\left(\frac{t}{\sqrt{n}}\right)\right|^{\frac{n-\mu}{2}}dt, & if\ \ 0\leq\nu\leq\frac{n-\mu-2}{2},\\[2ex]\displaystyle\int_{-\infty}^{\infty}|t|^{(\mu+1)N-1}\exp\left\{-\frac{(n-\mu-1)t^2b^2}{4n}\right\}dt, & if\ \ \frac{n-\mu-1}{2}\leq\nu\leq n-\mu-1.\end{cases}
\end{aligned}\tag{35.41}$$

Taking into account (35.32), we obtain from (35.41)

$$|\Delta_{\mu,\nu,n}\left(x\sqrt{n}\right)|\leq\sqrt{\frac{2}{\pi}}\left(\frac{\nu_N}{b^NN!}\right)^{\mu+1}\left(\frac{1}{\sqrt{n}}\right)^{(\mu+1)N}I(F;n;\mu;N).\tag{35.42}$$

Since $\sum_{\nu=0}^{n-\mu-1}\binom{\nu+\mu}{\mu}=\binom{n}{\mu+1}<\frac{n^{\mu+1}}{(\mu+1)!}$, then using (35.38) and (35.42) we find for all $x\in\mathbf{R}$

$$n^{\frac{\mu}{2}(N-2)}\mathcal{B}_n(\mu;x;X)\leq B_2(n;\mu;N).\tag{35.43}$$

The inequality (35.36) follows from (35.26), (35.37) and (35.43).

PROOF OF LEMMA 35.1.2. The statement of the lemma follows from (35.36) and (35.19).

35.4 Proof of Theorem 35.1.1

The formula (35.8) follows from (35.12) and the conditions (35.5) and (35.6). The equality (35.9) follows from Lemma 35.1.2 and the conditions $\mathbf{E}\widetilde{X}_n^3 = 0$ and (35.7). Theorem 35.1.1 is proved.

Acknowledgements. The research work of the authors was partially supported by the Russian Fund of Basic Researches grant N 99-01-00767.

References

1. Bergström, H. (1951). On asymptotic expansions of probability functions, *Skandinavisk Aktuarietidskrift*, **1–2**, 1–34.

2. Chebotarev, V. I. and Zolotukhin, A. Ya. (1996). The comparison of the Edgeworth and Bergström expansions. The case of normal mixtures, *Proceedings of the international conference and the Chebyshev readings, dedicated to 175 years from the birthday of P. L. Chebyshev*, **2**, pp. 359–363, Moscow: Moscow State University.

3. Cramer, G. (1976). *Mathematical Methods in Statistics*, Moscow: Mir. (in Russian).

4. Hall, P. and Nakata, T. (1986). On non-uniform and global descriptions of the rate of convergence of asymptotic expansions in the central limit theorem. *Journal of the Australian Mathematical Society, Series A*, **41**, 326–335.

5. Nagaev, S. V., Chebotarev, V. I., Zolotukhin, A. Ya. and Ali, M. M. (1997). On approximation of n-fold convolutions of mixtures of normal distributions, *Far Eastern Math. Reports. Vladivostok: Dal'nauka*, **3**, 23–48.

6. Osipov, L. V. (1972). On asymptotic expansions of the distribution function of the random variables sum with the non-uniform estimates of the remainder term, *Vestnik Leningrad. University*, **1**, 51–59 (in Russian).

7. Petrov, V. V. (1972). *Sums of Independent Random Variables*, Moscow: Nauka (in Russian).

8. Rozovskii, L. V. (1976). On the convergence of sums of independent random variables to the normal law in L_p, *Lietuvos Matem. Rink.*, **16**, **1**, 193–206 (in Russian).

9. Rozovskii, L. V. (1978). On lower estimate of the remainder term in central limit theorem, *Mathematical Notes*, **24**, **3**, 403–410 (in Russian).

36

Recent Progress in Probabilistic Number Theory

Jonas Kubilius

Vilnius University, Vilnius, Lithuania

Abstract: Methods of probabilistic number theory are discussed, and a review of the most important results is given.

Keywords and phrases: Asymptotic density, additive arithmetical function, probabilistic number theory

36.1 Results

I will discuss on the connection between the theory of numbers and the probability theory. At first sight, it seems impossible to say anything but trivial on this subject. Seemingly, there is nothing more determinate than the sequence of positive integers $1, 2, 3, \ldots$.

Long ago it was noticed that certain results of the number theory can be interpreted probabilistically. The reason for such an interpretation is the following. Let A be a finite or infinite sequence of positive integers. Denote by

$$\nu_n\{m \in A\} := \frac{\#\{m \le n, m \in A\}}{n},$$

the frequency of numbers of the sequence A, not exceeding n. The limit

$$\lim_{n\to\infty} \nu_n\{m \in A\} = D(A),$$

if it exists, is called *asymptotic density of the sequence* A.

Thus, the density of positive integers, divisible by a positive integer k, is equal to $1/k$.

Let $k_1, \ldots, k_r$ be positive integers. Suppose that every two of them are coprime. Then the density $D_{k_1 \ldots k_r}$ of positive integers, divisible by the product

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

$k_1 \dots k_r$, is equal to the product of the densities D_{k_j} $(j = 1, \dots, r)$ of positive integers, which are divisible by k_j:

$$D_{k_1 \cdots k_r} = D_{k_1} \cdots D_{k_r}.$$

So, we could speak on the probability that a positive integer, taken at random, is divisible by a given number k. The events that a positive integer, taken at random, is divisible by given numbers $k_1, \dots, k_r$, may be considered as independent in probability-theoretic sense, if every two of these numbers are coprime.

For a long time, the role of probabilistic arguments in the theory of numbers, based essentially on this remark, was almost exclusively that of a heuristic device. Sometimes, it led to erroneous conclusions.

I shall give an example. In 1905, Edmund Landau proved the following result. Denote by $w(m)$ the number of solutions of the equation $m = x^2 + y^2$ in integers. Landau proved that the frequency of positive integers $m \le n$ which are representable as the sum of two squares

$$\nu_n\{w(m) > 0\} \sim \frac{C}{\sqrt{\ln n}}$$

for some constant C. In 1939 Paul Lévy gave a simple heuristic derivation of this result. His argument also led him to the conjecture that if k is a positive integer, then

$$\frac{\nu_n\{w(m) = k\}}{\nu_n\{w(m) > 0\}} \sim \frac{e^{-\lambda}\lambda^k}{k!},$$

where $\lambda = c\sqrt{\ln n}$, c being a constant. In probabilistic terms, this means roughly that, out of the integers m for which $w(m) > 0$, those for which $w(m)$ has specified value, have a Poisson distribution. However, this is not true. The distribution of the specified values of $w(m)$ depends on the arithmetical structure of k. The frequency considered is equal to

$$\frac{C_k}{\sqrt{\ln n}} \cdot \begin{cases} (\ln \ln n)^{\alpha - 1}, & \text{if} \quad k = 2^\alpha k_1,\ 2 \not| k_1; \\ \dfrac{(\ln n)^\alpha}{\sqrt{n}}, & \text{if} \quad k = 3^\alpha k_1,\ 2 \not| k_1,\ 3 \not| k_1; \\ \dfrac{1}{\sqrt{n}}, & \text{if} \quad 2 \not| k,\ 3 \not| k; \end{cases}$$

here, C_k is a constant depending on k.

This example shows that heuristic probabilistic arguments are not always useful. Even such a distinguished mathematician as Paul Lévy was in the wrong way.

The main difficulty of applications of probability theory to the number theory consists in the fact that asymptotic density is not countably additive. For example, consider the finite sets $A_k = \{k\}$ $(k = 1, 2, \dots)$, where A_k consists of the single number k. Obviously, we have $D(A_k) = 0$. The union of the sets A_k is the set of all positive integers. The asymptotic density of it is 1. This phenomenon does not permit any non-trivial direct application of probability theory to number theory. Nevertheless, it is possible to obtain some interesting and deep results through some methods. We shall show this in the case of distribution theory of arithmetical functions.

As it is well known, an arithmetical function (in the simpliest case) is a sequence of real or complex numbers $h : \boldsymbol{N} \to \boldsymbol{C}$. There are two classes of arithmetical functions which are interesting for the theory of numbers, namely, additive and multiplicative functions. An arithmetical function $f(m)$ is called *additive* if for any pair of relatively prime integers m and n

$$f(mn) = f(m) + f(n).$$

If

$$m = p_1^{\alpha_1} \dots p_s^{\alpha_s}$$

is the canonical representation of m as the product of prime powers, then

$$f(m) = f(p_1^{\alpha_1}) + \dots + f(p_s^{\alpha_s}).$$

If the values of $f(p^\alpha) = f(p)$ for all p and all $\alpha = 1, 2, \dots$, then it is called *strongly additive*. In this case, we obviously have

$$f(m) = f(p_1) + \dots + f(p_s).$$

Similarly, an aritmetical function $g(m)$ is called *multiplicative* whenever

$$g(mn) = g(m)g(n)$$

provided m, n are coprime. Usually, one supposes that g(m) is not identically zero, or (what is the same) $g(1) = 1$. Analogously, we have the representation

$$g(m) = g(p_1^{\alpha_1}) \dots g(p_s^{\alpha_s}).$$

I shall give some examples.

1. For any fixed s, the function m^s is multiplicative.

2. Euler's function $\varphi(m)$ (the number of positive integers not greater than m and prime to m), the number $\tau(m)$ of all positive divisors of m, the Moebius' function $\mu(m)$ $(\mu(1) = 1,\ \mu(m) = 0$ if m is divisible by a square greater than 1; $\mu(m) = (-1)^r$ if m is the product of r different prime factors) are all multiplicative.

3. $w(m)/4$ is multipicative.

4. The logarithms of positive multiplicative functions (for example, $\log m$, $\log \varphi(m)$, $\log \tau(m)$) are additive.

5. The number $\omega(m)$ of distinct prime divisors of m is a strongly additive function.

6. Let us denote by $\Omega(m)$ the total number of prime divisors of m (multiple divisors being counted according to their multiplicity). It is an additive function.

From "canonical" representations, it follows that additive and multiplicative functions are completely determined by giving their values $f(p^\alpha)$, $g(p^\alpha)$ for prime powers p^α. It also follows that the values of such functions depend on the multiplicative structure of the argument and therefore the distribution of the values of these functions is very complicated. If we follow the change in the values of these functions as the argument runs through positive integers in order, we obtain a very chaotic picture which is usually observed when the additive and multiplicative properties of integers are examined jointly.

Nevertheless, it turns out that, in general, the distribution of values of many of these functions are subject to certain simple laws, which can be formulated and proved by using ideas and methods of probability theory.

In the classical research, in studying the distribution of values of number-theoretical functions, mathematicians usually limited themselves to two problems:

1. One looked for two simple (in some sense) functions $\psi_1(m)$ and $\psi_2(m)$ such that the inequalities

$$\psi_1(m) \leq h(m) \leq \psi_2(m)$$

hold for all or at least for all sufficiently large m. It is required that these inequalities were as exact as possible.

For example, it is easy to see that

$$1 \leq \omega(m) \leq m - 1 \quad (m = 2, 3, \ldots).$$

The equality $\omega(m) = 1$ holds for infinitely many m (for all prime powers). The upper estimate is not exact. One can prove that

$$\limsup_{m \to \infty} \frac{\omega(m) \ln \ln m}{\ln m} = 1.$$

It follows from the fact that in case m is product of k first primes, $m = p_1 \ldots p_k$,

$$\omega(m) = \frac{\ln m}{\ln \ln m}(1 + o(1)).$$

2. Beginning with Gauss and Lejeune–Dirichlet in studying the distribution of values of number-theoretical functions $h(m)$, mathematicians usually considered the behavior of the sum

$$\frac{1}{n}\sum_{m=1}^{n} h(m)$$

for $n \to \infty$ and looked for an asymptotic expression by means of simple functions of n.

For the function $\omega(m)$, we have

$$\frac{1}{n}\sum_{m=1}^{n} \omega(m) = \ln\ln n + O(1).$$

This sum is the mean of the function $\omega(m)$ on the segment $1, 2, \ldots, n$ of the sequence of positive integers. Thus, the mean value of the function $\omega(m)$ is approximately $\ln\ln n$ while the function can oscillate about the mean value within very wide bounds from 1 to approximately $\ln n/(\ln\ln n)$.

It is natural to consider the law of distribution

$$\nu_n\{h(m) \in B\}$$

where $h(m)$ is an arithmetical function, B is any Borel set on the real line and

$$\nu_n\{B\} = \frac{\#\{m \leq n, h(m) \in B\}}{n}$$

is the frequency of natural numbers m not exceeding n and satisfying the conditions written in the braces.

The most interesting cases are the following two: integral laws

$$\nu_n\{h(m) < x\}$$

or, more generally

$$\nu_n\{h(m) < C_n + D_n x\}$$

where C_n and D_n are normalizing constants, and local laws

$$\nu_n\{h(m) = a\}.$$

We are interested in investigating the asymptotic behavior of these laws as $n \to \infty$.

It is difficult now to give a full review of the results. My first book on this subject contained about 200 pages. And the book of Elliott, written two decades ago, has more than one thousand pages. At the present time, there are many new important results. In this regard, the names of many mathematicians, including first of all (in alphabetic order) Delange, Elliott, Erdös, Halász, Kac, Manstavičius, Ruzsa, Turán, Timofeev, must be mentioned.

I shall confine myself to integral laws of additive functions. I shall give principal ideas of some methods for the discussion of integral laws.

The simpliest method is the one of moments. Let us consider the distribution functions

$$F_n(x) = \nu_n\{f(m) < C_n + D_n x\}.$$

There exists the moments of all orders

$$\begin{aligned}\mu_k(n) &= \int_{-\infty}^{\infty} x^k d\nu_n\{f(m) < C_n + D_n x\} \\ &= \frac{1}{nD_n^k}\sum_{m=1}^{n}(f(m) - C_n)^k \qquad (k = 1, 2, \ldots).\end{aligned}$$

The calculation of the moments reduces to the sums

$$\begin{aligned}\sum_{m=1}^{n} f^k(m) &= \sum_{m=1}^{n}\Big(\sum_{p^\alpha \| m} f(p^\alpha)\Big)^k \\ &= \sum_{p_1^{\alpha_1}}\cdots\sum_{p_k^{\alpha_k}} f(p_1^{\alpha_1})\ldots f(p_k^{\alpha_k})\#\{m \le n, p_1^{\alpha_1}\|m, \ldots, p_k^{\alpha_k}\|m\}.\end{aligned}$$

If, for example, we could prove that $\mu_k(n)$, converges to 0 as n tends to infinity for all odd k's and to $(k-l)!!$ for all even k's, then it would follow that functions $F_n(x)$ converge to the standard normal law $\Phi(x)$.

However, there are many distribution functions which do not have moments.

Another method uses the ideas of sieve method. As mentioned earlier, the main difficulty in applying probability theory to number theory consists in the fact that asymptotic density of sets of natural numbers is not completely additive. Sometimes, it is possible to overcome this difficulty in the following way. I shall consider only strongly additive functions $f(m)$.

Let $\Omega = \boldsymbol{N} = \{1, 2, \ldots\}$ be the set of elementary events. Suppose that $r = r(n)$ is a positive function which tends to infinity with the tendency being not very quick:

$$\ln r(n) = o(\ln n).$$

Let

$$Q = \prod_{p \le r} p, \quad E(p) = \{m : p|m\}.$$

Let $\mathcal{A}$ be the set algebra generated by $E(p)$, $p \le r$. For every $k|Q$, let

$$E_k = \Big(\bigcap_{p|k} E(p)\Big) \cap \Big(\bigcap_{p|\frac{Q}{k}} E^c(p)\Big) = \Big\{m : k|m, \Big(m, \frac{Q}{k}\Big) = 1\Big\}.$$

The sets E_k are disjoint for different k. The algebra $\mathcal{A}$ consists of all finite unions of the sets E_k

$$A = \bigcup_{k_i} E_{k_i}.$$

We take two probability measures: $\nu_n\{A\}$ and

$$P(A) = \sum_{k_i} D(E_{k_i}).$$

By the sieve method, it is possible to prove that

$$\sup_{A\in\mathcal{A}} |\nu_n\{A\} - P(A)| \le C \exp\Big(- \frac{\ln n}{\ln r} \ln \frac{\ln n}{\ln r} + n^{-1/15}\Big).$$

Let

$$f(m)_r = \sum_{p|m, p\le r} f(p) = \sum_{p\le r} f^{(p)}(m)$$

where

$$f^{(p)}(m) = \begin{cases} f(p) & \text{if } \ p|m, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $f(m)_r$ and $f^{(p)}(m)$ are measurable with respect to $\mathcal{A}$. Thus, they are random variables with respect to $\{\boldsymbol{N}, \mathcal{A}, P\}$:

$$\xi_p = f^{(p)}(m) = \begin{cases} f(p) & \text{with probability } \ 1/p, \\ 0 & \text{with probability } \ 1 - 1/p. \end{cases}$$

So

$$\nu_n\{f(m)_r \in B\} = P\Big(\sum_{p\le r} \xi_p \in B\Big) + o(1)$$

uniformly for all Borel sets B on the real line. For a large class of functions, the quantity

$$\sum_{r(n)<p\le n} f^{(p)}(m)$$

may be neglected. By means of the inequality

$$\sum_{m=1}^{n} \Big(f(m) - \sum_{p\le n} \frac{f(p)}{p} \Big)^2 \le C_1 \sum_{p\le n} \frac{f^2(p)}{p} \ ,$$

it is possible to prove that in some cases the asymptotic laws for $f(m)$ and $f(m)_r$ are the same.

The third method uses characteristic functions and generating Dirichlet series. Let $F_n(x) = \nu_n\{f(m) < x\}$. Let us consider the Fourier transform

$$\varphi_n(t) = \int_{-\infty}^{\infty} e^{itx} dF_n(x) = \frac{1}{n} \sum_{m=1} e^{itf(m)}.$$

The summands are values of multiplicative functions. We may use the method of generating Dirichlet series

$$Z(s) = \sum_{m=1}^{\infty} \frac{e^{itf(m)}}{m^s} = \prod_p \left(1 + \frac{e^{itf(p)}}{p^s} + \frac{e^{itf(p^2)}}{p^{2s}} + \dots \right),$$

where $s = \sigma + i\tau$, $\sigma > 1$. Halász proved that there exists a function $\lambda = \lambda_n(t)$ such that

$$\varphi_n(t) = \frac{n^{i\lambda} Z(1 + 1/\ln n + \lambda)}{(1 + i\lambda)\zeta(1 + 1/\ln n)} + o(1).$$

In some cases, it is possible to prove that the estimate $o(1)$ is uniform for $t \in [-T, T]$.

Now we shall discuss integral laws for real additive arithmetical functions $f(m)$. Let C_n and $D_n > 0$ be two sequences of real numbers. We shall consider the distribution functions

$$\nu_n\{f(m) < C_n + D_n x\}. \tag{36.1}$$

It is proved that if they converge weakly to a distribution function, then D_n tends to a finite or infinite limit.

In the first case (when D_n tends to a finite limit), it is sufficient to consider the case $D_n \equiv 1$. In this case, necessary and sufficient conditions for the existence of the limit law are known.

Theorem 36.1.1 *There exist constants C_n such that*

$$\nu_n\{f(m) < C_n + x\} \tag{36.2}$$

converge weakly to a distribution function if and only if there exists a constant λ such that the series

$$\sum_p \frac{1}{p} \frac{\lambda_p^2}{1 + \lambda_p^2}$$

converges; here, $\lambda_p = f(p) - \lambda \ln p$. The constants C_n must have the form

$$C_n = \lambda \ln n + \sum_{p \le n,\ |\lambda_p| < 1} \frac{\lambda_p}{p} + C + o(1)$$

with C being a constant. The constants C_n are determined uniquely with exactness of the summand $C + o(1)$. The characteristic function of the limit law equals

$$\frac{e^{-iCt}}{1 + i\lambda t} \prod_p \left(\left(1 - \frac{1}{p}\right) e^{-it\lambda'_p} \sum_{\alpha=0}^{\infty} \frac{e^{itf(p^\alpha)}}{p^{\alpha(1+it\lambda)}} \right),$$

where

$$\lambda'_p = \begin{cases} \dfrac{\lambda_p}{p} & \text{if } \ |\lambda_p| < 1, \\ 1 & \text{otherwise.} \end{cases}$$

The limit law is degenerate if and only if $f(m) \equiv 0$. It is pure which means that it may be absolutely continuous or singular or discrete. All three cases really do exist.

The limit law is discrete if and only if the series

$$\sum_{f(p)\neq 0} \frac{1}{p} < \infty.$$

However there are no criteria to distinguish between the absolutely continuous and singular cases. It is interesting to note that the limit law is not always infinitely divisible.

Two special cases are of interest.

Theorem 36.1.2 *The distribution functions*

$$\nu_n\left\{f(m) < \sum_{p\leq n, |f(p)|<1} \frac{f(p)}{p} + x\right\}$$

converge weakly to a distribution function if and only if

$$\sum_p \frac{1}{p}\frac{f^2(p)}{1+f^2(p)} < \infty. \tag{36.3}$$

Theorem 36.1.3 *The distribution functions*

$$\nu_n\{f(m) < x\}$$

converge weakly to a distribution law if and only if the series (36.3) and the series

$$\sum_{|f(p)|<1} \frac{f(p)}{p}$$

converge.

Obviously, the convergence of the series (36.3) is equivalent to the convergence of the following two series

$$\sum_{|f(p)|\geq 1} \frac{1}{p}, \qquad \sum_{|f(p)|<1} \frac{f^2(p)}{p}.$$

So we have an analogue of the three series theorem. (It is called Erdös–Winter theorem.)

The case $D_n \to \infty$ is more complicated. We do not know necessary and sufficient conditions for the existence of the limit law for (36.1). It is known that D_n cannot increase quickly, namely

$$D_n = O(\log^a n).$$

This order is exact in the sense that for given D_n satisfying this condition, there exists an additive function $f(m)$ for which (36.1) with suitably chosen C_n have a limit law. It is known that the limit law is absolutely continuous or singular.

We know necessary and sufficient conditions for the convergence to the degenerate law

$$\varepsilon(x) = \begin{cases} 0, & \text{if} \quad x \le 0, \\ 1, & \text{if} \quad x > 0. \end{cases}$$

Theorem 36.1.4 *The frequencies*

$$\nu_n\Big\{\Big|\frac{f(m)}{D_n} - \alpha_n\Big| \ge \varepsilon\Big\} \to 0$$

as $n \to \infty$ if and only if there exists a sequence $\lambda_n = o(1)$ such that

$$\sum_{p \le n} \frac{\min\Big\{1, \Big(\frac{f(p)}{D_n} - \lambda_n \ln p\Big)^2\Big\}}{p} = o(1),$$

$$\alpha_n = \lambda_n + \sum_{\substack{p \le n \\ \left|\frac{f(p)}{D_n} - \lambda_n \ln p\right| < 1}} \frac{\frac{f(p)}{D_n} - \lambda \ln p}{p} + o(1)$$

as $n \to \infty$.

In general case, we have many results giving sufficient conditions. I shall mention some simple results. Let

$$A_n = \sum_{p \le n} \frac{f(p)}{p}, \qquad B_n^2 = \sum_{p \le n} \frac{f^2(p)}{p}.$$

Theorem 36.1.5 *If for every fixed $\varepsilon > 0$*

$$\frac{1}{B_n^2} \sum_{p \le n, |f(p)| > \varepsilon B_n} \frac{f^2(p)}{p} \to 0 \tag{36.4}$$

(an analogue of the Lindeberg condition), then

$$\nu_n\{fm) < A_n + B_n x\} \to \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du. \tag{36.5}$$

If (36.4) is true, then B_n is slowly increasing function of logarithm, $B_n = \chi(\ln n)$. Conversely, if B_n is such a function and (36.5) holds, then (36.4) is true.

There are examples with $B_n \sim \sqrt{\ln n}$ for which (36.5) holds, but (36.4) is not true.

It is possible to generalize this theorem.

Theorem 36.1.6 *Let $B_n = \chi(\ln n)$. The distribution functions*

$$\nu_n\{f(m) < A_n + B_n x\}$$

converge weakly to a distribution function with variance 1 if and only if there exists a non-decreasing function $K(u)$, $-\infty < u < \infty$, $K(-\infty) = 0$, $K(\infty) = 1$, such that for all $u \neq 0$ the sums

$$\frac{1}{B_n^2} \sum_{p \leq n, f(p) < uB_n} \frac{f^2(p)}{p}$$

converge to $K(u)$. The characteristic function of the limit law equals

$$\exp\Big(\int_{-\infty}^{\infty} (e^{itu} - 1 - itu) \frac{1}{u^2} dK(u) \Big).$$

The following example is of interest. The functions of distribution

$$\nu_n \Big\{ \sum_{p|m} (\ln p)^{\varrho} < x(\ln n)^{\varrho} \Big\}, \ \varrho > 1, \ \varrho \neq 1,$$

converge to a limit law with characteristic function

$$\frac{1}{2\pi i} \int_{1-i\infty}^{1+i\infty} \frac{e^z}{z} \exp\Big(\int_0^1 \frac{e^{itu^{\varrho}} - 1}{u} e^{-uz} du \Big) dz.$$

It is not infinitely divisible.

There are some results on the rate of convergence to the limit laws. I shall consider the case of the normal law.

Theorem 36.1.7 *Let $f(m)$ be a real-valued strongly additive arithmetical function and let*

$$\mu_n := \frac{\max_{p \leq n} |f(p)|}{B_n} = o(1) \tag{36.6}$$

as $n \to \infty$. Then

$$\nu_n\{f(m) < A_n + B_n x\} = \Phi(x) + O(\mu_n),$$

or more exactly

$$\nu_n\{f(m) < A_n + B_n x\} = \Phi(x) + O\Big(\frac{\mu_n}{1 + |x|^3}\Big)$$

uniformly for all real numbers x.

Drawing an analogy with the theory of summation of independent random variables, it was supposed that for any real-valued additive arithmetical function the remainder term in (36.6) may be changed into $O(\varrho_n)$, where

$$\varrho_n := \frac{1}{B_n^{3/2}} \sum_{p \le n} \frac{|f^3(p)|}{p}.$$

However, this is not always true. The following theorem holds.

Theorem 36.1.8 *Let $f(m)$ be a real-valued strongly additive arithmetical function and let*

$$E_n := \sum_{p \le n} \frac{f^2(p)}{p}\Big(1 - \frac{1}{p}\Big),$$

$$L_n := \frac{1}{D_n} \sum_{p,q \le n, pq > n} \frac{f_n(p) f_n(q)}{pq}$$

where the summation extend over all primes p, q satisfying the indicated conditions. Then

$$\nu_n\{f(m) < A_n + \sqrt{D_n}x\} = \Phi(x) + \frac{xL_n}{2\sqrt{2\pi}} e^{-x^2/2} + O(\varrho_n).$$

There are some theorems on the asymptotic expansion of the distribution laws. I shall give just one of them.

Theorem 36.1.9 *Let $f(m)$ be an integer-valued strongly additive arithmetical function. Suppose that there exists an $\varepsilon > 0$ such that*

$$\sum_{p \le n, f(p) \ne 1} \frac{\ln p}{p} = O(\ln n)^{1-\varepsilon}, \qquad \sum_{f(p) \ne 1} \frac{|fp|^s}{p} < \infty$$

for an $s \ge 1$. Then

$$\nu_n\Big\{f(m) < \ln\ln n + x\sqrt{\ln\ln n}\Big\} = \Phi(x) + \sum_{k=1}^{s} \frac{Q_k(x)}{(\ln\ln n)^{k/2}} + O\Big(\frac{\ln\ln\ln n}{(\ln\ln n)^{(s+1)/2}}\Big).$$

Here, $Q_k(x)$ are some functions not depending on n. One can write them explicitly.

I shall now give a theorem on large deviations.

Theorem 36.1.10 *Let $f(m)$ be an integer-valued strongly additive arithmetical function such that for a positive δ there exists an $\varepsilon > 0$ such that*

$$\sum_{p \le n, f(p) \ne 1} \frac{e^{\delta |f(p)|} \ln p}{p} = O(\ln n)^{1-\varepsilon}).$$

Then for $x = o(\sqrt{\ln\ln n})$,

$$\nu_n\Big\{f(m) < \ln\ln n + x\sqrt{\ln\ln n}\Big\}$$

in case $x \le 0$, *and*

$$\nu_n\Big\{f(m) > \ln\ln n + x\sqrt{\ln\ln n}\Big\}$$

in case $x \ge 0$, *are equal to*

$$e^{R_n(x)}\Phi(-|x|)\Big(1 + O\Big(\frac{|x|+1}{\sqrt{\ln\ln n}}\Big)\Big),$$

where

$$R_n(x) = \frac{x^2}{2} + (\zeta - (1+\zeta)\ln(1+\zeta))\ln\ln n$$

and

$$\zeta = \frac{x}{\sqrt{\ln\ln n}}.$$

PART X

Applications to Finance

37

On Mean Value of Profit for Option Holder: Cases of a Non-Classical and the Classical Market Models

O. V. Rusakov

St. Petersburg State University, St. Petersburg, Russia

Abstract: We consider an investor operating in a bond-stock market with derivatives who is going to buy/sell a stock. His decision problem is whether to buy or not to buy the corresponding option. We define the mean profit of such a potential option holder as a mean difference between his outlay for the stock in the cases of both owning the option or not owning. We consider market models: the classical Cox–Ross–Rubinstein, the classical Black–Scholes, and a non-classical continuous time model driven by the geometric integral Ornstein–Uhlenbeck process. For a defined class of contingent claims, we obtain conditions on non-negative/non-positive values of the mean profit.

Keywords and phrases: Mean profit, geometric integral Ornstein–Uhlenbeck process, the Black–Scholes formula, risk-neutral measures

37.1 Notation and Statements

In this chapter, we are interested in a problem of an investor's decision making. We consider an investor operating in a bond-stock (B,S) market. Suppose that the investor is going to buy a lot of shares of the stock. In the case he has a call type option with respect to this lot of the stock, will it be cheaper or not? We will consider this problem from the point of view of market expected value of buying stock at a horizon time T in cases when the investor owns the option or not.

Let X be a contingent claim given in a bond-stock market (B_t, S_t), $0 \leq t \leq T < \infty$, $B_t \in \mathbf{R}_+, S_t \in \mathbf{R}_+$.

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

At initial moment $t = 0$, a customer buys one share of stock option corresponding to X (thus he becomes an option holder) and at the terminal moment T, he is going to buy one share of stock S. The aim of this chapter is to consider the mean profit of such option holder, when the mathematical expectation ($\mathbf{E}$) is calculated under the market (not necessarily risk-neutral) probability measure.

Let us denote the random variable $\xi_T = S_T - X$ for the cost of one share of stock allowing for the contingent payment X. The mean profit for the option holder is defined as

$$\mathcal{R} \equiv \mathcal{R}(S, T, X) = \mathbf{E}\{B_T^{-1} S_T - B_T^{-1} \xi_T - \mathcal{C}_T\}, \tag{37.1}$$

where $\mathcal{C}_T$ is a rational value of the contingent claim X. In cases when $\mathcal{C}_T$ is a rational value calculated as mathematical expectation over risk-neutral probabilities ($\tilde{\mathbf{E}}$), elementary transformations of (37.1) give

$$\mathcal{R} = \mathbf{E}\{B_T^{-1} X\} - \tilde{\mathbf{E}}\{B_T^{-1} X\}. \tag{37.2}$$

Our interest lies in assumptions under which the mean profit $\mathcal{R}$ takes positive and negative values.

Assume that stochastically the contingent claim X depends only on the terminal value of stock $S_T \in \mathbf{R}$ and monotonically increasing function on S_T. Assume that in frames of the given (B, S)-model, there exists an unique rational value $\mathcal{C}_T$ of X at the initial moment $t = 0$. The class of such contingent claims is denoted by $\mathcal{X}$.

For the particular case of the standard European call option with a strike constant K, it is easy to calculate

$$\xi_T = S_T \mathbb{1}_{(-\infty,0)}(S_T - K) + K \mathbb{1}_{[0,\infty)}(S_T - K).$$

And it easy to check that the contingent claim for the standard European call option $(S_T - K)_+$ belongs to the class $\mathcal{X}$.

Remark that in the case of American options, a mean profit depends on a stopping time τ chosen by the investor. His mean profit for the standard American call option equals

$$\mathcal{R} = \mathbf{E}\{B_T^{-1} S_T - B_\tau^{-1} S_\tau + B_\tau^{-1}(S_\tau - K)_+ - \mathcal{A}_T\},$$

where $\mathcal{A}_T$ is the rational initial value of this option.

37.2 Models

In present chapter, we consider the following three models.

I. The classical Cox–Ross–Rubinstein (CRR) model [Cox, Ross and Rubinstein (1976)] :

$$\begin{cases} B_t &= (1+r)^t, \qquad t = 0,1,\ldots,T, \\ S_t &= S_0 \prod_{j=1}^{t} \zeta_t, \qquad t = 0,1,\ldots,T, \end{cases} \tag{37.3}$$

where the interest rate $r > 0$, and (ζ_t) is a sequence of i.i.d. binomial random variables with a common distribution: $\zeta_t = 1 + u$, with "market" probability p and $\zeta_t = 1 + d$, with "market" probability $(1 - p)$. The values "up" and "down" are connected by the inequality $-1 < d < r < u < \infty$. The risk-neutral probabilities for the CRR model are, respectively, given by

$$\tilde{p} = \frac{r-d}{u-d}, \qquad 1 - \tilde{p} = \frac{u-r}{u-d}. \tag{37.4}$$

II. The classical Black–Scholes model [Black and Scholes (1973)] :

$$\begin{cases} S_t &= S_0 \exp\left\{\mu t + \sigma W_t - \frac{\sigma^2}{2} t\right\}, \qquad 0 \le t \le T, \\ B_t &= \exp\{rt\}, \qquad 0 \le t \le T, \end{cases} \tag{37.5}$$

where W_t is the standard Brownian motion, $\mu \in \mathbf{R}$ is the expected return, $\sigma > 0$ is the volatility, and $r > 0$ is the interest rate.

For this model, we suppose that the (B, S)-market model defined on a complete stochastic basis $\{\Omega, (\mathcal{F}_t), \mathcal{F}, \mathbf{P}\}$, and $\tilde{\mathbf{P}}$ is a equivalent to $\mathbf{P}$ martingale measure.

III. The pricing model driven by the integral Ornstein–Uhlenbeck process:

$$\begin{cases} S_t &= S_0 \exp\left\{\mu t + \sigma \int_0^t U_s ds - \frac{\sigma^2}{2} t\right\}, \qquad 0 \le t \le T, \\ B_t &= \exp\{rt\}, \qquad 0 \le t \le T, \end{cases} \tag{37.6}$$

where U_s is the standard Ornstein–Uhlenbeck process with zero mean, the variance equaling to 1 and with a viscosity $m > 0$, i.e. $cov(U_s, U_t) = \exp\{-m|t-s|\}$. The coefficients $\mu \in \mathbf{R}$, $\sigma > 0$, and $r > 0$ play the same role as in the Black–Scholes model.

This model describes a behavior of a multiagents market and can be presented as a limit of mixtures of geometrical random walks. Each component of the mixture is naturally embedded into a binomial tree equipped by constant coefficients of "up" and "down" events. So we are able to determine risk-neutral probabilities for the limit process (37.6) of geometric integral Ornstein–Uhlenbeck process. Although this model is arbitrary, we can slightly perturb a behavior of this stock price by a process with independent increments (with small variations of the increments) and make the model arbitrarily free.

Vasićek (1977) introduced a conditional geometric integral of the Ornstein–Uhlenbeck process to describe behavior of prices of zero-coupon bonds. Fölmer and Schweizer (1993) gave from a view of stochastic differential equations a microeconomic motivation to consider the geometric Ornstein–Uhlenbeck process as a stochastic component in a class of stock price models. We justify the Ornstein–Uhlenbeck process from the view of limit theorems of probability [Rusakov (1998)]. More detailed description of a construction of the pricing model is introduced below.

Consider a market model with N agents, where N is a sufficiently large natural number. At any discrete time $t \in \{0, 1, \dots M\} \triangleq \mathbb{T}$, each agent buys or sells an asset which is called below a stock. Thus, each agent has a trading policy with respect to the stock, and the agents' trading policies form the stock prices. The moment M is a horizon time of our economy, possibly equal to infinity. In the market, there exists a factor which influences the trading policy of every agent. This factor has certain internal as well as external properties affecting the market agents. The role of the factor is that every agent may change his trading policy under its influence.

The main postulate of the model is that instantaneously, at any moment $t \in \mathbb{T}$, few agents are affected by the factor described above, and revise their trading policies accordingly. The remaining majority of agents keep their trading policies unchanged. So, every agent from the market does not react to events during a random period and continues his trading policy until a random moment when he is affected by the factor.

Further, we assume that all asset values are discounted, i.e., they are expressed in units of a risk-less security (bond).

We now describe explicitly the model of the market. At the moment $t \in \mathbb{T}$, each agent has his own trading policy on the market: to buy or to sell stock. By trading, each agent contributes to the evolution of the stock price:

> if the agent buys a stock, then the value of the stock is multiplied by $U > 1$;
>
> if the agent sells the stock, then the value of the stock is multiplied by $0 < D < 1$.

The values U, D depend on the time scale of the interval $\mathbb{T}$ in accordance with the standard asymptotic assumptions. This means that for the discrete time interval $\mathbb{T}$, the physical unit of scale is proportional to the square root of the value $U - D$. On the other hand, the probabilities of U and D are not dependant on n and are fixed: at any $t \in \mathbb{T}$, the probability of buying stock is $p > 0$, and the probability of selling stock is $q = 1 - p$. The values of U, D, p, q are the same for all agents.

Thus, we can assume that the trading policy of an i-th agent $(i = 1, \dots, N)$

at moment $n \in \mathbb{T}$ is a random variable with distribution

$$\xi_n^{(i)} = \begin{cases} U, & p \\ D, & q. \end{cases} \tag{37.7}$$

After receiving information about the factor, each agent decides for himself and only for himself whether to accept this information or remain indifferent whether not to accept this information. An agent who is not affected by the factor continues his previous trading strategy and an agent who is affected by the factor revises his trading policy and 'ruffle' (think about and decide anew) to buy or to sell a stock again. In the case when an agent is affected by the factor directly, i.e., for instance, he drives a bargain on the stock with another agent, we assume that such situations are equivalent to acceptance of the information about the factor.

At the initial moment $t = 0$, all trading policies over N agents are assumed to be independent in total. Agents affected by the factor retry to implement their trading policies independently. Thus, we suppose that trading policies $\xi_n^{(i)}$, $n = 0, 1, \dots, M$, $i = 1, \dots, N$, are conditionally independent given the 'information about the factor'. In view of these arguments, we can formulate two first key assumptions about the model.

(A) The random vector $(\xi_0^{(1)}, \xi_0^{(2)}, \dots, \xi_0^{(N)})$ consists of i.i.d. binomial random variables distributed according to (37.6).

(B) Let Υ_n denote the σ-algebra of events which are generated by influence of the factor to agents within moments $0, 1, \dots, n$. Given the condition Υ_n, the random vectors $\left(\xi_0^{(i)}, \xi_1^{(i)}, \dots, \xi_n^{(i)}\right)$ are conditionally independent over the family of indices $\{i\}$.

Next we define random variables $k_n^{(i)}$, $n = 0, 1, \dots, M-1, i = 1, \dots, N$, as indicators of the events: {At moment n the i-th agent is affected by the factor}. The equality $k_n^{(i)} = 1$ means that the i-th agent revises his trading policy from the moment n up to the moment $n+1$, and $k_n^{(i)} = 0$ means that the trading policy of the i-th agent remains the same.

In the model we suppose that $\forall n$ all influences of the factor until moment n inclusively are independent of influences of the factor at moment $n+1$. Therefore,

(C) the random vectors $\vec{k}_j \triangleq (k_j^{(1)}, k_j^{(2)}, \dots, k_j^{(N)})$; $j = 0, 1, \dots, M-1$, are independent in total over the family of indices $\{j\}$.

(D) Denote $\mathcal{K}_j \triangleq \sigma(k_j^{(1)}, k_j^{(2)}, \dots, k_j^{(N)})$, $j = 0, 1, \dots$. It is assumed that $\Upsilon_n = \mathcal{K}_0 \otimes \dots \otimes \mathcal{K}_n$.

In this chapter, we consider the case where the vectors $(k_0^{(1)}, k_0^{(2)}, \dots, k_0^{N})$ have the following identical distribution concentrated on $\{0, 1\}^N$:

(E) Every vector $\{\vec{k}_j\}$ consists of m units and $N-m$ nulls exactly. The rule of choice of units from the set $\{1,2,\ldots,N\}$ is random, uniform, and without replacements (without loss of generality).

The assumption **(E)** means that at any moment $n \in \mathbb{T}$, exactly m agents who have random numbers revise their trading policies. This assumption makes agents dependent through the factor. The condition "m is a constant" may be interpreted as follows: exactly m agents are dealing with the stock at any fixed moment $n \in \mathbb{T}$. The limit theorems below will also be obtained in the case where m is a random number with finite mathematical expectation as $N \to \infty$. Here, one must change m to $\mathbf{E}m$.

Equilibrium in the model is achieved in a certain sense. By the law of large numbers, the number of revised values 'buy' of trading policies is approximately equal to the number of revised values 'sell' during a sufficiently large period of time $\{0,1,\ldots\}$.

The sequence of trading policies can be described by the following two-dimensional table of random variables

$$\begin{array}{ccccc} \xi_0^{(1)} & \xi_0^{(2)} & \ldots & \xi_0^{(N-1)} & \xi_0^{(N)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \xi_n^{(1)} & \xi_n^{(2)} & \ldots & \xi_n^{(N-1)} & \xi_n^{(N)} \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{array} \tag{37.8}$$

such that:

(*a*) The initial row consists of independent random variables having the identical distribution as in (37.7).

(*b*) For $n = 1,2,\ldots$ and $\in \{1,\ldots,N\}$, the elements of the table are defined by the rule of replacement to independent copies

$$\xi_n^{(i)} = \begin{cases} \xi_{n-1}^i, & \text{if} \quad k_{n-1}^{(i)} = 0 \\ \\ \eta_{n;i}, & \text{if} \quad k_{n-1}^{(i)} = 1. \end{cases} \tag{37.9}$$

(*c*) Two-dimensional set of random variables (η) consists of mutually independent random variables.

(*d*) The initial row $(\xi_0^{(1)},\ldots,\xi_0^{(N)})$ is independent of totality of the random variables (η).

(*e*) The random variables (η) are i.i.d. with common distribution given by (37.7).

(*f*) The upper index of random variables ξ corresponds to the number of agent and it may be interpreted as a "local time", and the lower index of random variables ξ corresponds to the moment of real time.

(g) The number N is called the order of the table in (37.8), or order of the model.

It is easy to see that, for any N, the rows in the table in (37.8) form a strictly stationary and Markovian sequence of random vectors.

Next, we assume that every agent makes equivalent contribution to pricing of stock S by his trading policy.

(G) At any fixed moment $n \in \mathbb{T}$, each agent's contribution to pricing is defined as evolution of his trading policy during the time interval $\{0, 1, \ldots, n\}$

$$Z_n(i) = \prod_{j=0}^{n} \xi_j^{(i)}, \qquad i = 1, ..., N, \tag{37.10}$$

The following assumption is more important in the construction of model.

(F) At any moment $n \in \mathbb{T}$, the value of the stock is equal to *the geometric mean of contributions to pricing over N agents*

$$S_n = \overset{\circ}{S} \left(\prod_{i=1}^{N} Z_n(i) \right)^{1/N} = \overset{\circ}{S} \left(\prod_{i=1}^{N} \prod_{j=0}^{n} \xi_j^{(i)} \right)^{1/N}, \tag{37.11}$$

where $\overset{\circ}{S}$ is a starting point of the stock price.

The main objective of the chapter, is to investigate the limit behavior of the distribution of S_n as $N \to \infty$. We use the parameterization of time $n = [Nt]$ for any real $0 \leq t \leq T \leq \infty$ ($[\cdot]$ denotes an integer part). Thus, such parameterization implies that for any fixed $0 \leq t \leq T$ during time $0, 1, \ldots, n = [Nt]$ the number of revising points in the table in (37.8) is proportional to the number of agents N. In addition, we will consider the table in (37.8) in a sense of scheme of series (N being the index of the series) equipped with the constant probability measure (p, q) of trading policies.

(H) The parameter m defined by assumption **(E)** is assumed to be fixed as $N \to \infty$.

The assumption **(H)** can be interpreted as "the agents are strongly inert". At every fixed instant, a very small number of agents revise their trading policies. On the other hand, during the physical time interval $[0, t]$, a sufficiently significant proportion of agents revise their trading policies. Mean value of this proportion is equal in asymptotics to $1 - \exp\{-mt\}$ as $N \to \infty$. This is a straightforward implication of the theory of allocation of balls into cells e.g., Gnedenko (1983, pp. 167–176).

On the next step, we define smallness of the trading policies $\xi_n^{(i)}$ as $N \to \infty$. Order of the smallness is the same as in the classical limit transition from the CRR model to the Black–Scholes model.

So, let ξ be a random variable with distribution (37.7). Introduce parameters a and $v > 0$ by the following equalities

$$\left.\begin{array}{rcl} v^2 &=& (U-D)^2\,pq\,, \\ a\,v^2+1 &=& Up + Dq\,. \end{array}\right\} \tag{37.12}$$

It is easy to see that $v^2 = \mathbf{D}\xi$. By analogy to the classical limit transition, we assume that $v^2 \equiv v^2\,(N) = \sigma_0^2/N$, where σ_0 is a volatility constant independent of N. The parameter a, named as *relative risk*, characterizes difference between "market" measure (p,q) and "risk-neutral" measure $(\tilde p, \tilde q)$, i.e. the numbers $\tilde p$ and $\tilde q$ are defined by $U\tilde p + D\tilde q = 1$ $\tilde q = 1 - \tilde p$. In correspondence with the classical case, we assume that

$$a \equiv a(N) \to \frac{\mu - r}{\sigma_0^2}, \quad N \to \infty; \qquad (\mu - r) \in \mathbf{R}, \;\; r > 0\,, \tag{37.13}$$

where μ is an expected rate of stock return, and r is an interest rate of discount bond: $B_t = \exp\{rt\}$.

In accordance with the assumption of time parameterization and with assumption **(G)**, for the order of the model N we define the value of the stock at time moment $t \in [0,T]$ by the following geometric mean

$$S_{[Nt]} \;\triangleq\; \overset{\circ}{S}\left(\prod_{i=1}^{N} Z_{[Nt]}\,(i)\right)^{1/N} \;=\; \overset{\circ}{S}\left(\prod_{i=1}^{N}\prod_{j=0}^{[Nt]} \xi_j^{(i)}\right)^{1/N}\,. \tag{37.14}$$

So, by (37.14) we have defined a family of piecewise broken lines. Let $\mathbf{D}_{[0,T]}(\mathbb{C})$ denote the space of function from $\mathbf{D}_{[0,T]}$ equipped with the uniform metric. We consider the broken lines (37.14) as elements of $\mathbf{D}_{[0,T]}(\mathbb{C})$.

The market probabilities (p,q) induces a distribution $\mathcal{P}_N \equiv \mathcal{P}_N(S_{[Nt]}, 0 \le t \le T)$ concentrated on the broken lines $\{S_{[Nt]}\}$. Analogously, the risk-neutral measure $(\tilde p, \tilde q)$ induces a distribution $\tilde{\mathcal{P}}_N \equiv \tilde{\mathcal{P}}_N(S_{[Nt]},\, 0 \le t \le T)$ concentrated on the same broken lines $\{S_{[Nt]}\}$.

Recall, U_s denotes the standard Ornstein–Uhlenbeck process with the viscosity coefficient m.

Theorem 37.2.1 *Convergences of the broken lines to the geometric integral Ornstein–Uhlenbeck process and the rational option value.*

(i) *(The case of the market probabilities)*
Under the above assumptions about the model, the following weak convergence is valid as $N \to \infty$:

$$\mathcal{P}_N \Longrightarrow \overset{\circ}{S}\,\exp\Big\{\,(\mu - r)t \,+\, \sigma_0 \int_0^t U_s ds \,-\, \frac{\sigma_0^2}{2}\,t\Big\}, \qquad 0 \le t \le T. \tag{37.15}$$

(ii) *(The case of the risk-neutral probabilities for discounted values of "up" and "down")*
Under the assumptions about the model for the family of distributions $\{\tilde{\mathcal{P}}_N\}$, the following weak convergence is valid as $N \to \infty$

$$\tilde{\mathcal{P}}_N \Longrightarrow \overset{\circ}{S} \exp\Big\{\sigma_0 \int_0^t U_s ds \;-\; \frac{\sigma_0^2}{2} t\Big\}, \qquad 0 \le t \le T. \tag{37.16}$$

(iii) *(Analogue of the Black–Scholes formula for the model.)*
Let C_T denote the rational value at moment 0 of the Standard European Call Option with the contingent claim $(S_T - K)_+$. Denote the variance of the integral of the Ornstein–Uhlenbeck process by

$$v_m^2(t) \triangleq \mathbf{D}\Big\{\int_0^t U_s ds\Big\} = \frac{2\sigma_0^2}{m^2}(mt - 1 + e^{-mt}), \qquad 0 \le t \le T.$$

Then there exists an unique rational value C_T that is calculated as mathematical expectation of the limit (37.16) of the risk-neutral distributions $\tilde{\mathcal{P}}_N$, and it equals

$$C_T = \overset{\circ}{S} \exp\left\{\frac{v_m^2(T)}{2} - \frac{\sigma_0^2}{2} T\right\} \Phi(\rho + v_m(T)) \;-\; K \exp\{-rT\}\Phi(\rho), \tag{37.17}$$

where

$$\rho \triangleq \frac{\ln \overset{\circ}{S} - \ln K + rT - \sigma_0^2 T/2}{v_m(T)};$$

and $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal law.

For details and a proof of Theorem 37.2.1, we refer to Rusakov (1998).

37.3 Results

Proposition 37.3.1 *Let the contingent claim $X \in \mathcal{X}$.*

(i) *In the case of the Cox–Ross–Rubinstein model, the mean profit $\mathcal{R} \ge 0$ iff $p \ge \tilde{p}$.*

(ii) *In the case of the Black–Scholes model, the mean profit $\mathcal{R} \ge 0$ iff the market expected return μ of the stock is not less than the interest rate r.*

(iii) *Consider the case above of the Inert Market Agents model. Then the mean profit $\mathcal{R} \ge 0$ iff the coefficient μ is not less than the interest rate r.*

Analogous relations take place for the standard European put options under changing the inequality $\mathcal{R} \geq 0$ to $\mathcal{R} \leq 0$.

PROOF. Prove Proposition 37.3.1. Assume for simplicity that $S_0 = 1$ for all considered cases of the models I-III.

The case of the CRR model is trivial since the left-side tail of the binomial distribution $\sum_{j=0}^{k} C_T^j x^{T-j}(1-x)^j$ is a strictly monotonic function on $x \in (0,1)$ for any $0 < k \leq T$.

Consider the Black–Scholes model. Let Z_t be the process of density of the equivalent martingale measure $\tilde{\mathbf{P}}$ with respect to market probabilities $\mathbf{P}$,

$$Z_t = \exp\left\{\frac{\mu - r}{\sigma} W_t - \frac{t}{2}\left(\frac{\mu - r}{\sigma}\right)^2\right\},$$

and $\forall A \in \mathcal{F}_T$ risk-neutral probability is defined $\tilde{\mathbf{P}}(A) = \int_A Z_t d\mathbf{P}$. With respect to $\tilde{\mathbf{P}}$, the process $\tilde{W}_t = W_t + (\mu - r)t/\sigma$ is the standard Brownian motion. Under the measure $\tilde{\mathbf{P}}$, the stock price process represents

$$\tilde{S}_t = \exp\left\{rt + \sigma \tilde{W}_t - \frac{\sigma^2}{2} t\right\}.$$

Let us denote the function $\mathcal{E}(x)$ based on the Dolean exponent

$$\mathcal{E}(x) = \exp\{\sigma x - \frac{\sigma^2}{2} T\}, \qquad x \in \mathbf{R}.$$

So, for the Black–Scholes model, the terminal values of the stock $\tilde{S}_T$ and S_T can be represented in a form

$$\tilde{S}_T = \mathcal{E}(\tilde{W}_T)\exp\{rT\}, \qquad S_T = \mathcal{E}(W_T)\exp\{\mu T\} \tag{37.18}$$

and random variables $S_T \exp\{-\mu T\}$ and $\tilde{S}_T \exp\{-rT\}$ are identically distributed.

Now we consider the model of type III. From (37.15) and (37.16), it follows that the risk-neutral stock price process forms

$$\tilde{S}_t = \exp\left\{rt + \sigma \int_0^t \tilde{U}_s ds - \frac{\sigma^2}{2} t\right\},$$

where $\tilde{U}_s$ is also the standard Ornstein–Uhlenbeck process with the viscosity m.

So, Theorem 37.2.1 allows us to use the chang in the market measure by the risk-neutral probabilities and to conclude analogously (37.18), and that random variables $S_T \exp\{-\mu T\}$ and $\tilde{S}_T \exp\{-rT\}$ are identically distributed.

As $X \in \mathcal{X}$, then one can represent X in a form $X = g(S_T)$, where $g(s)$ is a monotonicly increasing function as $s \uparrow$. Change probabilities from risk-neutral to the market ones,

$$\begin{aligned}\tilde{\mathbf{E}}\{g(\tilde{S}_T)\} &= \tilde{\mathbf{E}}\{g(e^{rT} \cdot (\tilde{S}_T e^{-rT}))\} \\ &= \mathbf{E}\{g(e^{rT} \cdot (S_T e^{-\mu T}))\}.\end{aligned}$$

Substitute ξ_T in the definition of $\mathcal{R}(S,T,X)$ and change measures,

$$\begin{cases} B_T\mathcal{R}(S,T,X) &= \mathbf{E}\{g(S_T)\} - \tilde{\mathbf{E}}\{g(e^{rT}\cdot(\tilde{S}_T e^{-rT}))\} \\ &= \mathbf{E}\{g(S_T)\} - \mathbf{E}\{g(e^{rT}\cdot(S_T e^{-\mu T}))\} \\ &= \mathbf{E}\{g(e^{-rT}S_T) - g(e^{(r-\mu)T}S_T)\}. \end{cases}$$

It involves that $\mathcal{R}(S,T,X) \geq 0$ iff $\mu \geq r$. ∎

Acknowledgements. The author expresses his deep appreciation to A. A. Gushchin, A. V. Melnikov, A. A. Novikov, A. N. Shiryaev, E. Valkeila, and S. A. Vavilov for helpful discussions. The financial support provided by RFBR grant N 96-01-01809 is gratefully acknowledged.

References

1. Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637–659.

2. Cox, J. C., Ross, R. A. and Rubinstein, M. (1976). Option pricing: A simplified approach, *Journal of Financial Economics*, **3**, 229–263.

3. Fölmer, H. and Schweizer, M. (1993). A microeconomic approach to diffusion models for stock prices, *Mathematical Finance*, **3**, 1–23.

4. Gnedenko, B. V. (1983). On the limits theorems for a random number of random variables, *Lecture Notes in Mathematics*, No. 1021, Berlin: Springer-Verlag.

5. Rusakov, O. V. (1998). A model of market pricing with randomly distributed information and the geometric integral of the Ornstein–Uhlenbeck process, *Preprint 184*, University of Helsinki.

6. Vasićek, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics*, **5**, 177–188.

38

On the Probability Models to Control the Investor Portfolio

S. A. Vavilov

St. Petersburg State University, St. Petersburg, Russia

Abstract: We introduce the robust feed-back control schemes based on the current dynamics of the security prices to provide the stable increasing of the portfolio value. The portfolio value is defined as the value of securities plus the amount of cash, which can be negative as the borrowing of money is not prohibited. It is assumed that the price of securities follows the class of stochastic processes. The control is realized by varying respectively the number of securities and cash in the portfolio.

Keywords and phrases: Investor portfolio, stochastic control

38.1 Introduction

There are at least three basic strategies to operate with the investor portfolio. The first strategy prescribes to arrange portfolio by making use of the expectations of security prices and their variances on the fixed moment of time in the future. Such approach can be regarded as passive one because the scheme does not imply any operation with the portfolio over the corresponding time interval. It is possible to realize the pointed out strategy following the number of works which include Markowitz (1952) and Tobin (1958) as basic ones.

The second strategy is based on the relatively frequent variation of the types of securities and consequently their number in the portfolio. Such variation is realized due to the currently expected dynamics of prices to provide constantly the maximum value of the portfolio on the whole operating period. Usually, the resources to buy the new types of securities are mainly provided by selling the available and less profitable at the present moment assets from the portfolio so that the borrowing of money is not necessary and the number of cash in the

N. Balakrishnan et al. (eds.), *Asymptotic Methods in Probability and Statistics with Applications*

portfolio is equal to zero. The machinery of technical analysis as presented for instance in Murphy (1986) for the basis for such an approach.

The third strategy which is scrutinized in the present chapter deals as the first one with the fixed types of securities. However, the control in this scheme is to be realized perpetually over the whole operating period by varying respectively the number of securities and cash in the portfolio. It is based on the current dynamics of security prices and is to provide the stable increasing of the portfolio value when the borrowing of money and short sales are not prohibited. Here, the portfolio value is defined as the value of the securities plus the amount of cash which is negative when the money is borrowed. In contrast to the first and the second strategy, this approach is based not on the prediction of prices but on the feed-back control basic principle. In the present chapter, two such robust feed-back control schemes are constructed. The first one (Section 38.2) is derived to operate with the portfolio consisting of zero coupon bonds and cash. The second one (Section 38.3) is conceived to be realized for arbitrary securities under the reasonable assumptions on the price dynamics.

Both algorithms imply the current price of security to follow the class of random processes which is described by a stochastic differential equation of the form

$$dx_t = c(t, (F)(t))x_t dt + \hat{\sigma}(t) x_t dW_t \ , \tag{38.1}$$

where W_t is a standard Wiener process. Coefficient c is supposed to be an arbitrary function of t and rather general operator F which can describe the effects of delay, memory, stochastic integration, and so forth. The only reasonable restriction on the right-hand side of (38.1) is the existence of a solution to the corresponding Cauchy problem.

To consider the case of different types of securities constituting portfolio, we will supply the corresponding notions in (38.1) with index i

$$dx_{ti} = c_i(t, (F)(t))x_{ti} dt + \sum_{j=1}^{n} \hat{\sigma}_{ij}(t) x_{ti} dW_{tj} \ , \quad i = 1, .., n \ . \tag{38.2}$$

A number of models of market pricing including geometric Brownian motion [Black and Scholes (1973)] and those as presented in Vasićek (1977) and Rusakov (1998) are relevant to (38.1). It is worth noting that the pricing models based on the Ornstein–Uhlenbeck process and the multivariate Ornstein–Uhlenbeck process which have been employed lately for a description of security value evolution by a number of authors [for example, Lo and Wang (1995) and Fölmer and Schweizer (1993)] belong to the pointed out class of stochastic processes. In Section 38.4, it will be shown that the pricing model $x_t = exp(h_t)$, where h_t is the continuous analogue of the autoregression of each finite order, is also covered by Eq. (38.1).

It is remarkable that the stochastic control schemes studied in this chapter do not use, at least explicitly, the basic notion of the cost function [Davis (1977)]. On the other hand, the straightforward applications of the dynamic programming methods [Fleming and Rishel (1975)] are well-known in financial economics [Merton (1971)]. On these grounds, we can suppose that the questions arising in the investor portfolio management pose new problems in the stochastic control theory which have no analogues in the classical domains of it's applications, for instance, in physics and mechanics.

This chapter can be regarded as the extended and remade version of the results published in Vavilov (1998).

38.2 Portfolio Consisting of Zero Coupon Bonds: The First Scheme

Consider portfolio at an arbitrary moment t consisting of a number a_t of a zero coupon bond and an amount of cash. The bond is characterized by the maturity date corresponding to the time interval $[0, T^*]$ and the unit maturity value β [Hull (1993)]. Without loss of generality, the starting point of the portfolio control can be chosen at $t = 0$. Along with T^*, we introduce the quantity T which is taken in a close left-hand side vicinity of T^*. It is assumed that on the time interval $[0, T]$ the bond price follows the class of random processes which is described by the stochastic differential equation (38.1). Moreover, we suppose that the bond price belongs to the strip $[\alpha, \beta]$ with a relatively narrow width $\delta = \beta - \alpha$ (δ/α is small).

In this section, we dwell on the trading strategy which implies the portfolio value f_t to satisfy the relationship

$$df_t = a_t dx_t. \tag{38.3}$$

It means that the instant variance of the portfolio value is due only to the instant variance of the bond price.

The quantity

$$u_t = f_t - a_t x_t \tag{38.4}$$

may be interpreted as an amount of cash in the portfolio.

Here, we do not impose any restrictions on the sign of a_t and u_t so that short sale of bonds and borrowing money is not prohibited. Beside this, transaction costs and the interest rate for borrowing are not taken into account.

The problem under consideration is to construct the robust feed-back control scheme based on the current dynamics of the bond price which provides the stable increasing of the portfolio value. The control is to be realized by varying respectively the number of bonds and cash in the portfolio.

Along with (38.3), by making use of the Ito formula [Ikeda and Watanabe (1989)] for the function $f(t,x)$, where x is supposed to satisfy (38.1), one can write down

$$\begin{aligned} df &= (f'_t + cxf'_x + \frac{1}{2}\hat{\sigma}^2 x^2 f''_{xx})dt + \hat{\sigma} x f'_x dW_t \\ &= (f'_t + \frac{1}{2}\hat{\sigma}^2 x^2 f''_{xx})dt + f'_x(cxdt + \hat{\sigma} x dW_t) \\ &= (f'_t + \frac{1}{2}\hat{\sigma}^2 x^2 f''_{xx})dt + f'_x dx. \end{aligned} \tag{38.5}$$

By comparing (38.5) and (38.3), one can easily see that the trading strategy (38.3) implies the relationships

$$f'_t + \frac{1}{2}\hat{\sigma}^2(t) x^2 f''_{xx} = 0, \tag{38.6}$$

$$a_t = f'_x. \tag{38.7}$$

It is remarkable that (38.6) depends on $\hat{\sigma}$ but not on c in (38.1). This fact turns out to be the crucial point for operating at financial markets where the dynamics of price trends is hardly predictable. Moreover, because the width δ of the corresponding strip for the bond price is relatively small, Eq. (38.6) will be substituted by

$$f'_t + \frac{1}{2}\sigma^2(t) f''_{xx} = 0, \tag{38.8}$$

where $\sigma(t) = \hat{\sigma}(t)\alpha$.

To specify the appropriate function $f(t,x)$ satisfying (38.8), one can pose boundary conditions by making use of the following arguments.

When the price of the bond tends to the lower bound of the strip, the reasonable trading strategy prescribes to invest money in bonds. On the other hand, when the price of the bond tends to the unit maturity value β, it is lucrative to convert bonds in cash. Keeping in mind (38.4) and (38.7), the latter remarks yield the fulfilment of boundary conditions

$$f(t,\alpha) = 0 \tag{38.9}$$

and

$$\frac{\partial f}{\partial x}(t,\beta) = 0. \tag{38.10}$$

Eventually when $t = T$, we set the ultimate profile of the portfolio

$$f(T,x) = \varphi(x), \tag{38.11}$$

where the choice of function $\varphi(x)$ will be specified below.

By solving (38.8) subject to the boundary conditions (38.9), (38.10) and condition (38.11) [Tikhonov and Samarskii (1963)], the expressions for the portfolio value and the number of bonds can be written down, respectively, as

$$f(t,x) = \sum_{k=0}^{\infty} c_k e^{-\frac{1}{2\delta^2}(\frac{\pi}{2}+\pi k)^2 \int_0^{T-t} \sigma^2(T-s)ds} \sin\left[\frac{1}{\delta}(\frac{\pi}{2}+\pi k)(x-\alpha)\right], \qquad (38.12)$$

$$a_t = \sum_{k=0}^{\infty} c_k(\frac{\pi}{2}+\pi k)\frac{1}{\delta} e^{-\frac{1}{2\delta^2}(\frac{\pi}{2}+\pi k)^2 \int_0^{T-t} \sigma^2(T-s)ds} \cos\left[\frac{1}{\delta}(\frac{\pi}{2}+\pi k)(x-\alpha)\right], \qquad (38.13)$$

where

$$c_k = \frac{2}{\delta}\int_\alpha^\beta \varphi(x)\sin\left[\frac{1}{\delta}(\frac{\pi}{2}+\pi k)(x-\alpha)\right]dx. \qquad (38.14)$$

Formulas (38.13) and (38.14) constitute the desirable feed-back control trading strategy as one can easily see from (38.12) the stable exponential growth of the portfolio value in time. The choice of concrete function $\varphi(x)$ can be realized to provide the optimal (in some sense) growth of the portfolio value in time.

To pose the problem using condition (38.11) is reasonable when we define the initial number of bonds to be acquired in order to provide the desirable portfolio value $\varphi(\beta)$ at the date of maturity. On the other hand, under concrete circumstances it is possible to modify the statement of the problem. For instance, consider that the investor has already N_0 bonds at the moment $t = 0$. To realize the stochastic control of such initial portfolio in the future, one can follow Eq. (38.8) subject to the conditions

$$f(t,\alpha) = 0,\ \frac{\partial f}{\partial x}(t,\beta) = 0,\ \frac{\partial f}{\partial x}(0,x) = N(x) \qquad (38.15)$$

while $N(x_0) = N_0$. It is easy to write down the solution to this problem in a form analogous to (38.12). For instance, when

$$N(x) = N_0\frac{\cos[\frac{1}{\delta}\frac{\pi}{2}(x-\alpha)]}{\cos[\frac{1}{\delta}\frac{\pi}{2}(x_0-\alpha)]},$$

the solution to the problem (38.15) contains only one harmonic in the corresponding Fourier series and we arrive at the relationship

$$a_t = N_0 e^{\frac{\pi^2}{8\delta^2}\int_0^t \sigma^2(s)ds}\frac{\cos[\frac{1}{\delta}\frac{\pi}{2}(x_t-\alpha)]}{\cos[\frac{1}{\delta}\frac{\pi}{2}(x_0-\alpha)]}$$

while

$$f(t,x) = \frac{2\delta}{\pi}N_0 e^{\frac{\pi^2}{8\delta^2}\int_0^t \sigma^2(s)ds}\frac{\sin[\frac{1}{\delta}\frac{\pi}{2}(x-\alpha)]}{\cos[\frac{1}{\delta}\frac{\pi}{2}(x_0-\alpha)]}.$$

The choice of the concrete initial number of bonds profile $N(x)$ can be realized as earlier on the same grounds of the acceptable growth of the portfolio value by the date of maturity.

Now consider the situation when the portfolio includes n different types of zero coupon bonds. The analogue of trading strategy (38.3) can be written down as

$$df_t = \sum_{i=1}^{n} a_{ti} dx_{ti}. \tag{38.16}$$

Then by making use of the Ito formula for the function $f(t, x_1, ..., x_n)$, where x_i are supposed to satisfy (38.2), and reiterating the same arguments as in (38.5), it is possible to write down

$$df = (f'_t + \frac{1}{2} \sum_{i,j,k=1}^{n} \sigma_{ik}(t)\sigma_{jk}(t) \frac{\partial^2 f}{\partial x_i \partial x_j})dt + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i \,, \tag{38.17}$$

where $\sigma_{ik}(t) = \hat{\sigma}_{ik}(t)\alpha_i$.

By comparing (38.16) and (38.17), the following relationships are obtained:

$$f'_t + \frac{1}{2} \sum_{i,j,k=1}^{n} \sigma_{ik}(t)\sigma_{jk}(t) \frac{\partial^2 f}{\partial x_i \partial x_j} = 0 \,, \tag{38.18}$$

$$a_{ti} = \frac{\partial f}{\partial x_i} \,. \tag{38.19}$$

Introducing the new variable $s = T - t$ and following the previous one-dimensional case arguments, we arrive at the initial-value problem

$$f'_s = \frac{1}{2} \sum_{i,j,k=1}^{n} \sigma_{ik}(T-s)\sigma_{jk}(T-s) \frac{\partial^2 f}{\partial x_i \partial x_j} \,, \tag{38.20}$$

$$f(0, x_1, ..., x_n) = \varphi(x_1, ..., x_n) \tag{38.21}$$

subject to the boundary conditions analogous to (38.9) and (38.10) :

$$f(s, x_1, .., x_i, .., x_n) = 0 \tag{38.22}$$

when $x_i = \alpha_i$, $i = 1, ..., n$, and

$$\frac{\partial f}{\partial x_i}(s, x_1, .., x_i, .., x_n) = 0 \tag{38.23}$$

when $x_i = \beta_i$, $i = 1, ..., n$.

It is worth noting that the fulfilment of condition (38.22) implies to invest cash and securities of all other types in the i-th type security when the price of the latter one tends to the lower bound of the strip $[\alpha_i, \beta_i]$.

Unfortunately, because of the mixed partial derivatives in Eq. (38.20), the separation of variables is not possible and generally speaking the problem in (38.20)–(38.23) is to be solved numerically.

Anyhow, in the case when correlation matrix in (38.2) is diagonal the cross-terms in (38.20) disappear. If additionally, $\sigma_{ii}(t)$ are constants, the separation of variables turns out to be valid and the solution to the problem in (38.20)–(38.23) can be written down explicitly [Tikhonov and Samarskii (1963)] in the form analogous to the representation (38.12), (38.14).

38.3 Portfolio Consisting of Arbitrary Securities: The Second Scheme

In this Section, we consider portfolio consisting of a number of arbitrary securities whose price follows the class of random processes described by the stochastic differential equation (38.1). Additionally, it is supposed that the price of security during the period of control belongs to the interval $[\alpha, \beta]$ with a relatively narrow width $\delta = \beta - \alpha$ (δ/α is small).

It is worth noting that the interval $[\alpha, \beta]$, as one can see further, is not to be given in advance for the whole time of the control but can be varied as a step function in the process of operating with the investor portfolio.

In contrast to (38.3), another trading strategy which will be scrutinized in the present Section implies that

$$df_t = a_t dx_t + v_t dt, \tag{38.24}$$

where the presence of the additional term $v_t dt$ means that the instant variance of the portfolio value is also due to some given in advance schedule of "pumping" cash into the portfolio. The latter remark leads to the relationship

$$v_t = v(t, x_t) = l(t), \tag{38.25}$$

where the velocity of such "pumping" $l(t)$ is a known function.

Turning to the trading strategy defined by formula (38.24) and reiterating the arguments of the previous Section, we arrived at the differential equation

$$f_t' + \frac{1}{2}\sigma^2(t) f_{xx}'' = v(t, x) \tag{38.26}$$

subject to the boundary conditions

$$f(t, \alpha) = 0 \tag{38.27}$$

and

$$\frac{\partial f}{\partial x}(t, \beta) = 0. \tag{38.28}$$

Now in contrast to strategy (38.3), the initial portfolio is supposed to be empty

$$f(0,x) = 0. \tag{38.29}$$

The fulfilment of boundary conditions (38.27) and (38.28) yields specification of function $v(t,x)$

$$v(t,x) = \sum_{k=0}^{N} v_k(t) sin\left[\frac{1}{\delta}(\frac{\pi}{2} + \pi k)(x-\alpha)\right], \tag{38.30}$$

where N is an arbitrary positive integer.

By making use of (38.30), the solution to the problem in (38.26)–(38.29) can be written down as [Tikhonov and Samarskii (1963)]

$$f(t,x) = \sum_{k=0}^{N} \int_0^t e^{\frac{1}{2\delta^2}(\frac{\pi}{2}+\pi k)^2 \int\limits_0^{t-\tau} \sigma^2(s)ds} v_k(\tau)d\tau \sin\left[\frac{1}{\delta}(\frac{\pi}{2} + \pi k)(x-\alpha)\right] \tag{38.31}$$

and, consequently, the number of securities in the portfolio is defined by the formula

$$a_t = \sum_{k=0}^{N} \int_0^t e^{\frac{1}{2\delta^2}(\frac{\pi}{2}+\pi k)^2 \int\limits_0^{t-\tau} \sigma^2(s)ds} v_k(\tau)d\tau \frac{1}{\delta}(\frac{\pi}{2} + \pi k) \cos\left[\frac{1}{\delta}(\frac{\pi}{2} + \pi k)(x-\alpha)\right]. \tag{38.32}$$

Along with (38.13), we have also obtained the desired trading strategy (38.32) as formula (38.31) provides the stable exponential growth of the portfolio value in time.

To define functions $v_k(\tau)$, one should maximize the right-hand side of (38.31) for t, x predesigned under the restriction

$$\sum_{k=0}^{N} v_k(\tau) \sin\left[\frac{1}{b}(\frac{\pi}{2} + \pi k)x_\tau\right] = l(\tau), \tag{38.33}$$

where $l(\tau)$ is a given function and x_τ is a current security price on the interval $[0,t]$.

When $N = 0$, the latter problem is trivial and instead of (38.32), one arrives at the relationship

$$a_t = \frac{\pi}{2\delta} \int_0^t e^{\frac{\pi^2}{8\delta^2} \int_0^{t-\tau} \sigma^2(s)ds} \frac{l(\tau)}{\sin\left[\frac{1}{\delta}\frac{\pi}{2}(x(\tau)-\alpha)\right]} d\tau \cos\left[\frac{1}{\delta}\frac{\pi}{2}(x-\alpha)\right]. \tag{38.34}$$

It is remarkable that even in the case of the Markovian character of the stochastic process defined by (38.1), the trading strategy (38.34) turns out to memorize the dynamics of prices on the whole operating period.

By making use of the machinery elaborated in Section 38.2, one can easily write down the analogue of the problem in (38.26)–(38.30) in the situation when the portfolio contains different types of securities. For the sake of brevity, we confine ourselves to the case when the multidimensional version of the representation (38.30) contains the first harmonic only. Reiterating the arguments of Section 38.2, one arrives at the equation

$$f_t' + \frac{1}{2}\sum_{i,j,k=1}^{n} \sigma_{ik}(t)\sigma_{jk}(t)\frac{\partial^2 f}{\partial x_i \partial x_j}$$
$$= \frac{l(t)}{\prod_{i=1}^{n} sin[\frac{\pi}{2\delta_i}(x_i(t)-\alpha_i)]}\prod_{i=1}^{n} sin[\frac{\pi}{2\delta_i}(x_i-\alpha_i)] \qquad (38.35)$$

subject to the initial

$$f(0, x_1, ..., x_n) = 0 \qquad (38.36)$$

and boundary conditions

$$f(t, x_1, ..., x_i, ..., x_n) = 0 \qquad (38.37)$$

when $x_i = \alpha_i$, $i = 1, ..., n$, and

$$\frac{\partial f}{\partial x_i}(t, x_1, ..., x_i, ..., x_n) = 0 \qquad (38.38)$$

when $x_i = \beta_i$, $i = 1, ..., n$.

In the case when correlation matrix in (38.2) is diagonal and $\sigma_{ii}(t)$ are constants, the cross-terms in (38.35) disappear and the separation of variables towards the problem in (38.35)–(38.38) is applicable. Then the solution to the latter problem can be written down explicitly [Tikhonov and Samarskii (1963)] in the form analogous to the representation (38.31) when $N = 0$.

38.4 Continuous Analogue of the Finite-Order Autoregression

The goal of this Section is to prove that the pricing model $x_t = exp(h_t)$, where h_t is the continuous analogue of each finite order autoregression, is covered by Eq. (38.1).

Consider the n-order autoregression of the form

$$h_t = c_1 h_{t-\Delta} + c_2 h_{t-2\Delta} + ... + c_n h_{t-n\Delta} + \varepsilon_t, \qquad (38.39)$$

where ε_t is a white noise and Δ is a difference interval.

It is well-known [Box and Jenkins (1970)] that the relationship (38.39) can be rewritten in differences with one-to-one correspondence between coefficients c_i and a_i

$$a_1\nabla^n h_t + a_2\nabla^{n-1}h_t + ... + a_{n-1}\nabla h_t + a_n h_t = \varepsilon_t, \tag{38.40}$$

where $\nabla^{k+1}h_t = \nabla(\nabla^k h_t)$ and $\nabla h_t = h_t - h_{t-\Delta}$.

Write down the continuous analogue of the difference equation (38.40) as follows:

$$a_1\Delta^n \frac{d^n h_t}{dt^n} + a_2\Delta^{n-1}\frac{d^{n-1}h_t}{dt^{n-1}} + ... + a_{n-1}\Delta\frac{dh_t}{dt} + a_n h_t = \varepsilon_t \ . \tag{38.41}$$

The solution to Cauchy problem for Eq. (38.41) on the interval $[0, t]$ can be presented in the form

$$h_t = \varphi(t) + \int_0^t G(t-s)dW_s, \tag{38.42}$$

where $\varphi(t)$ is a deterministic function and $G(t-s)$ is the function of response as an integrand in Ito's integral.

Consider t to be fixed and introduce the new notation $F(s) = G(t-s)$ for the convenience. As $F(s)$ is a deterministic function, it is possible to apply integration by parts to Ito's integral in the expression (38.42). Thus, the following calculations are valid:

$$\begin{aligned}\int_0^t G(t-s)dW_s &= \int_0^t F(s)dW_s \\ &= F(t)W_t - \int_0^t W_s\dot{F}(s)ds \\ &= G(0)W_t + \int_0^t W_s\dot{G}(t-s)ds \\ &= G(0)\int_0^t dW_s + \int_0^t W_s\dot{G}(t-s)ds\end{aligned}$$

and we arrive at the relationship

$$d[\int_0^t G(t-s)dW_s] = G(0)dW_t + W_t\dot{G}(0)dt + \{\int_0^t W_s\ddot{G}(t-s)ds\}dt.$$

Ultimately, keeping in mind (38.42), the stochastic representation for h_t is derived as

$$dh_t = [\dot{\varphi}(t) + W_t\dot{G}(0) + \int_0^t W_s\ddot{G}(t-s)ds]dt + G(0)dW_t.$$

Evidently, $x_t = exp(h_t)$ satisfies stochastic differential equation of the form (38.1).

38.5 Conclusions

Considering variety of the pointed out strategies, we should recognize that the choice of one or the other is strongly motivated by the concrete situation at the financial market. For instance, when there are strongly pronounced trends in the dynamics of prices, it seems that one will prefer the first two strategies to the third one. On the contrary, when the trends of security prices are feebly marked it seems reasonable to choose the latter approach. Nevertheless, as it has been shown in the present study, the third strategy can provide the increasing of the portfolio value even in the case when the trend of a security price goes down. Moreover, the approach analyzed in this chapter prescribes the number of a security in the portfolio to depend not only on the price but on the time interval of the trading also. For example, it makes reasonable to launch an even number of portfolios operating with the same securities but separated in time to diminish the amount of borrowed money. Furthermore, the proposed approach may be helpful to arrange hedging strategies by striking futures and options. Ultimately, the rapid swap of cash flows inside one system operating with different strategies is probably not the worst solution while gambling.

Acknowledgements. The author would like to thank Dr. O. V. Rusakov for his stimulating interest in the research which led to this paper, fruitful discussions and valuable comments.

I am grateful to Professors Ya. I. Belopolskaya and A. A. Gushin for their helpful comments and suggestions.

Special thanks to my student I. V. Sosnushkin who constructed the software to operate with the real investor portfolio on the basis of the proposed schemes.

The financial support provided by RFBR grant N 96-01-01809 is gratefully acknowledged.

References

1. Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**, 637–659.

2. Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis. Forecasting and Control*, San Francisco, California: Holden-Day.

3. Davis, M. H. A. (1977). *Linear Estimation and Stochastic Control*, London, England: Chapman and Hall.

4. Fleming, W. H. and Rishel, R. W. (1975). *Deterministic and Stochastic Optimal Control*, New York: Springer-Verlag.

5. Fölmer, H. and Schweizer, M. (1993). A microeconomic approach to diffusion models for stock prices, *Mathematical Finance*, **3**, 1–23.

6. Hull, J. (1993). *Options, Futures and Other Derivative Securities*, Prentice-Hall.

7. Ikeda, N. and Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, Amsterdam: North-Holland.

8. Lo, A. W. and Wang, J. (1995). Implementing option pricing models when asset returns are predictable, *The Journal of of Finance*, **50**, 87–129.

9. Markowitz, H. M. (1952). Portfolio selection, *Journal of Finances*, **7**, 77–91.

10. Merton, R. C. (1971). Optimal consumption and portfolio rules in a continuous time model, *Journal of Economic Theory*, **3**, 373–413.

11. Murphy, J. J. (1986). *Technical Analysis of the Futures Markets: A Comprehensive Guide to Trading Methods and Applications*, New York Institute of Finance, A Prentice-Hall Company.

12. Rusakov, O. V. (1998). A model of market pricing with randomly distributed information and the geometric integral of the Ornstein–Uhlenbeck process, *Preprint 184*, University of Helsinki.

13. Tikhonov, A. N. and Samarskii, A. A. (1963). *Equations of Mathematical Physics*, International Series of Monographs on Pure and Applied Mathematics, **39**, Oxford: Pergamon Press.

14. Tobin, D. (1958). Liquidity preference as behaviour toward risk, *Rev. of Econ. Studies*, **25**, 65–86.

15. Vasićek, O. (1977). An equilibrium characterization of the term structure, *Journal of Financial Economics*, **5**, 177–188.

16. Vavilov, S. A. (1998). On the Stochastic Control of Investor Portfolio when the Price of Security Follows the Ornstein- Uhlenbeck Process, *International Conference on "Asymptotic Methods in Probability and Mathematical Statistics", St. Petersburg,* pp. 298–302.

Index