Universitext

Universitext

Series Editors:

Sheldon Axler
San Francisco State University, San Francisco, CA, USA

Vincenzo Capasso
Università degli Studi di Milano, Milan, Italy

Carles Casacuberta
Universitat de Barcelona, Barcelona, Spain

Angus MacIntyre
Queen Mary, University of London, London, UK

Kenneth Ribet
University of California, Berkeley, Berkeley, CA, USA

Claude Sabbah
CNRS, École Polytechnique, Palaiseau, France

Endre Süli
University of Oxford, Oxford, UK

Wojbor A. Woyczynski
Case Western Reserve University, Cleveland, OH, USA

Universitext is a series of textbooks that presents material from a wide variety of mathematical disciplines at master's level and beyond. The books, often well class-tested by their author, may have an informal, personal, even experimental approach to their subject matter. Some of the most successful and established books in the series have evolved through several editions, always following the evolution of teaching curricula, into very polished texts.

Thus as research topics trickle down into graduate-level teaching, first textbooks written for new, cutting-edge courses may make their way into *Universitext*.

For further volumes:
www.springer.com/series/223

Alexandr A. Borovkov

Probability Theory

Edited by K.A. Borovkov
Translated by O.B. Borovkova and P.S. Ruzankin

Springer

Alexandr A. Borovkov
Sobolev Institute of Mathematics and
Novosibirsk State University
Novosibirsk, Russia

Translation from the 5th edn. of the Russian language edition:
'Teoriya Veroyatnostei' by Alexandr A. Borovkov
© Knizhnyi dom Librokom 2009
All Rights Reserved.

1st and 2nd edn. © Nauka 1976 and 1986
3rd edn. © Editorial URSS and Sobolev Institute of Mathematics 1999
4th edn. © Editorial URSS 2003

ISSN 0172-5939 ISSN 2191-6675 (electronic)
Universitext
ISBN 978-1-4471-5200-2 ISBN 978-1-4471-5201-9 (eBook)
DOI 10.1007/978-1-4471-5201-9
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2013941877

Mathematics Subject Classification: 60-XX, 60-01

© Springer-Verlag London 2013
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

The present edition of the book differs substantially from the previous one. Over the period of time since the publication of the previous edition the author has accumulated quite a lot of ideas concerning possible improvements to some chapters of the book. In addition, some new opportunities were found for an accessible exposition of new topics that had not appeared in textbooks before but which are of certain interest for applications and reflect current trends in the development of modern probability theory. All this led to the need for one more revision of the book. As a result, many methodological changes were made and a lot of new material was added, which makes the book more logically coherent and complete. We will list here only the main changes in the order of their appearance in the text.

• Section 4.4 "Expectations of Sums of a Random Number of Random Variables" was significantly revised. New sufficient conditions for Wald's identity were added. An example is given showing that, when summands are non-identically distributed, Wald's identity can fail to hold even in the case when its right-hand side is well-defined. Later on, Theorem 11.3.2 shows that, for identically distributed summands, Wald's identity is always valid whenever its right-hand side is well-defined.

• In Sect. 6.1 a criterion of uniform integrability of random variables is constructed, which simplifies the use of this notion. For example, the criterion directly implies uniform integrability of weighted sums of uniformly integrable random variables.

• Section 7.2, which is devoted to inversion formulas, was substantially expanded and now includes assertions useful for proving integro-local theorems in Sect. 8.7.

• In Chap. 8, integro-local limit theorems for sums of identically distributed random variables were added (Sects. 8.7 and 8.8). These theorems, being substantially more precise assertions than the integral limit theorems, do not require additional conditions and play an important role in investigating large deviation probabilities in Chap. 9.

• A new chapter was written on probabilities of large deviations of sums of random variables (Chap. 9). The chapter provides a systematic and rather complete exposition of the large deviation theory both in the case where the Cramér condition (rapid decay of distributions at infinity) is satisfied and where it is not. Both integral and integro-local theorems are obtained. The large deviation principle is established.

• Assertions concerning the case of non-identically distributed random variables were added in Chap. 10 on "Renewal Processes". Among them are renewal theorems as well as the law of large numbers and the central limit theorem for renewal processes. A new section was written to present the theory of generalised renewal processes.

• An extension of the Kolmogorov strong law of large numbers to the case of non-identically distributed random variables having the first moment only was added to Chap. 11. A new subsection on the "Strong law of large numbers for generalised renewal processes" was written.

• Chapter 12 on "Random walks and factorisation identities" was substantially revised. A number of new sections were added: on finding factorisation components in explicit form, on the asymptotic properties of the distribution of the suprema of cumulated sums and generalised renewal processes, and on the distribution of the first passage time.

• In Chap. 13, devoted to Markov chains, a section on "The law of large numbers and central limit theorem for sums of random variables defined on a Markov chain" was added.

• Three new appendices (6, 7 and 8) were written. They present important auxiliary material on the following topics: "The basic properties of regularly varying functions and subexponential distributions", "Proofs of theorems on convergence to stable laws", and "Upper and lower bounds for the distributions of sums and maxima of sums of independent random variables".

As has already been noted, these are just the most significant changes; there are also many others. A lot of typos and other inaccuracies were fixed. The process of creating new typos and misprints in the course of one's work on a book is random and can be well described mathematically by the Poisson process (for the definition of Poisson processes, see Chaps 10 and 19). An important characteristic of the quality of a book is the intensity of this process. Unfortunately, I am afraid that in the two previous editions (1999 and 2003) this intensity perhaps exceeded a certain acceptable level. Not renouncing his own responsibility, the author still admits that this may be due, to some extent, to the fact that the publication of these editions took place at the time of a certain decline of the publishing industry in Russia related to the general state of the economy at that time (in the 1972, 1976 and 1986 editions there were much fewer such defects).

Before starting to work on the new edition, I asked my colleagues from our laboratory at the Sobolev Institute of Mathematics and from the Chair of Probability Theory and Mathematical Statistics at Novosibirsk State University to prepare lists of any typos and other inaccuracies they had spotted in the book, as well as suggested improvements of exposition. I am very grateful to everyone who provided me with such information. I would like to express special thanks to I.S. Borisov, V.I. Lotov, A.A. Mogul'sky and S.G. Foss, who also offered a number of methodological improvements.

I am also deeply grateful to T.V. Belyaeva for her invaluable assistance in typesetting the book with its numerous changes. Without that help, the work on the new edition would have been much more difficult.

A.A. Borovkov

Foreword to the Third and Fourth Editions

This book has been written on the basis of the Russian version (1986) published by "Nauka" Publishers in Moscow. A number of sections have been substantially revised and several new chapters have been introduced. The author has striven to provide a complete and logical exposition and simpler and more illustrative proofs. The 1986 text was preceded by two earlier editions (1972 and 1976). The first one appeared as an extended version of lecture notes of the course the author taught at the Department of Mechanics and Mathematics of Novosibirsk State University. Each new edition responded to comments by the readers and was completed with new sections which made the exposition more unified and complete.

The readers are assumed to be familiar with a traditional calculus course. They would also benefit from knowing elements of measure theory and, in particular, the notion of integral with respect to a measure on an arbitrary space and its basic properties. However, provided they are prepared to use a less general version of some of the assertions, this lack of additional knowledge will not hinder the reader from successfully mastering the material. It is also possible for the reader to avoid such complications completely by reading the respective Appendices (located at the end of the book) which contain all the necessary results.

The first ten chapters of the book are devoted to the basics of probability theory (including the main limit theorems for cumulative sums of random variables), and it is best to read them in succession. The remaining chapters deal with more specific parts of the theory of probability and could be divided into two blocks: random processes in discrete time (or random sequences, Chaps. 12 and 14–16) and random processes in continuous time (Chaps. 17–21).

There are also chapters which remain outside the mainstream of the text as indicated above. These include Chap. 11 "Factorisation Identities". The chapter not only contains a series of very useful probabilistic results, but also displays interesting relationships between problems on random walks in the presence of boundaries and boundary problems of complex analysis. Chapter 13 "Information and Entropy" and Chap. 19 "Functional Limit Theorems" also deviate from the mainstream. The former deals with problems closely related to probability theory but very rarely treated in texts on the discipline. The latter presents limit theorems for the convergence

of processes generated by cumulative sums of random variables to the Wiener and Poisson processes; as a consequence, the law of the iterated logarithm is established in that chapter.

The book has incorporated a number of methodological improvements. Some parts of it are devoted to subjects to be covered in a textbook for the first time (for example, Chap. 16 on stochastic recursive sequences playing an important role in applications).

The book can serve as a basis for third year courses for students with a reasonable mathematical background, and also for postgraduates. A one-semester (or two-trimester) course on probability theory might consist (there could be many variants) of the following parts: Chaps. 1–2, Sects. 3.1–3.4, 4.1–4.6 (partially), 5.2 and 5.4 (partially), 6.1–6.3 (partially), 7.1, 7.2, 7.4–7.6, 8.1–8.2 and 8.4 (partially), 10.1, 10.3, and the main results of Chap. 12.

For a more detailed exposition of some aspects of Probability Theory and the Theory of Random Processes, see for example [2, 10, 12–14, 26, 31].

While working on the different versions of the book, I received advice and help from many of my colleagues and friends. I am grateful to Yu.V. Prokhorov, V.V. Petrov and B.A. Rogozin for their numerous useful comments which helped to improve the first variant of the book. I am deeply indebted to A.N. Kolmogorov whose remarks and valuable recommendations, especially of methodological character, contributed to improvements in the second version of the book. In regard to the second and third versions, I am again thankful to V.V Petrov who gave me his comments, and to P. Franken, with whom I had a lot of useful discussions while the book was translated into German.

In conclusion I want to express my sincere gratitude to V.V. Yurinskii, A.I. Sakhanenko, K.A. Borovkov, and other colleagues of mine who also gave me their comments on the manuscript. I would also like to express my gratitude to all those who contributed, in one way or another, to the preparation and improvement of the book.

A.A. Borovkov

For the Reader's Attention

The numeration of formulas, lemmas, theorems and corollaries consists of three numbers, of which the first two are the numbers of the current chapter and section. For instance, Theorem 4.3.1 means Theorem 1 from Sect. 3 of Chap. 4. Section 6.2 means Sect. 2 of Chap. 6.

The sections marked with an asterisk may be omitted in the first reading.

The symbol $\square$ at the end of a paragraph denotes the end of a proof or an important argument, when it should be pointed out that the argument has ended.

The symbol $:=$, systematically used in the book, means that the left-hand side is *defined* to be given by the right-hand side. The relation $=:$ has the opposite meaning: the right-hand side is defined by the left-hand side.

The reader may find it useful to refer to the **Index of Basic Notation** and **Subject index**, which can be found at the end of this book.

Introduction

1. It is customary to set the origins of Probability Theory at the 17th century and relate them to combinatorial problems of games of chance. The latter can hardly be considered a serious occupation. However, it is games of chance that led to problems which could not be stated and solved within the framework of the then existing mathematical models, and thereby stimulated the introduction of new concepts, approaches and ideas. These new elements can already be encountered in writings by P. Fermat, D. Pascal, C. Huygens and, in a more developed form and somewhat later, in the works of J. Bernoulli, P.-S. Laplace, C.F. Gauss and others. The abovementioned names undoubtedly decorate the genealogy of Probability Theory which, as we saw, is also related to some extent to the vices of society. Incidentally, as it soon became clear, it is precisely this last circumstance that can make Probability Theory more attractive to the reader.

The first text on Probability Theory was Huygens' treatise *De Ratiociniis in Ludo Alea* ("On Ratiocination in Dice Games", 1657). A bit later in 1663 the book *Liber de Ludo Aleae* ("Book on Games of Chance") by G. Cardano was published (in fact it was written earlier, in the mid 16th century). The subject of these treatises was the same as in the writings of Fermat and Pascal: dice and card games (problems within the framework of Sect. 1.2 of the present book). As if Huygens foresaw future events, he wrote that if the reader studied the subject closely, he would notice that one was not dealing just with a game here, but rather that the foundations of a very interesting and deep theory were being laid. Huygens' treatise, which is also known as the first text introducing the concept of mathematical expectation, was later included by J. Bernoulli in his famous book *Ars Conjectandi* ("The Art of Conjecturing"; published posthumously in 1713). To this book is related the notion of the so-called Bernoulli scheme (see Sect. 1.3), for which Bernoulli gave a cumbersome (cf. our Sect. 5.1) but mathematically faultless proof of the first limit theorem of Probability Theory, the Law of Large Numbers.

By the end of the 19th and the beginning of the 20th centuries, the natural sciences led to the formulation of more serious problems which resulted in the development of a large branch of mathematics that is nowadays called Probability Theory. This subject is still going through a stage of intensive development. To a large extent,

Probability Theory owes its elegance, modern form and a multitude of achievements to the remarkable Russian mathematicians P.L. Chebyshev, A.A. Markov, A.N. Kolmogorov and others.

The fact that increasing our knowledge about nature leads to further demand for Probability Theory appears, at first glance, paradoxical. Indeed, as the reader might already know, the main object of the theory is randomness, or uncertainty, which is due, as a rule, to a lack of knowledge. This is certainly so in the classical example of coin tossing, where one cannot take into account all the factors influencing the eventual position of the tossed coin when it lands.

However, this is only an apparent paradox. In fact, there are almost no exact deterministic quantitative laws in nature. Thus, for example, the classical law relating the pressure and temperature in a volume of gas is actually a result of a probabilistic nature that relates the number of collisions of particles with the vessel walls to their velocities. The fact is, at typical temperatures and pressures, the number of particles is so large and their individual contributions are so small that, using conventional instruments, one simply cannot register the random deviations from the relationship which actually take place. This is not the case when one studies more sparse flows of particles—say, cosmic rays—although there is no qualitative difference between these two examples.

We could move in a somewhat different direction and name here the uncertainty principle stating that one cannot simultaneously obtain exact measurements of any two conjugate observables (for example, the position and velocity of an object). Here randomness is not entailed by a lack of knowledge, but rather appears as a fundamental phenomenon reflecting the nature of things. For instance, the lifetime of a radioactive nucleus is essentially random, and this randomness cannot be eliminated by increasing our knowledge.

Thus, uncertainty was there at the very beginning of the cognition process, and it will always accompany us in our quest for knowledge. These are rather general comments, of course, but it appears that the answer to the question of when one should use the methods of Probability Theory and when one should not will always be determined by the relationship between the degree of precision we want to attain when studying a given phenomenon and what we know about the nature of the latter.

2. In almost all areas of human activity there are situations where some experiments or observations can be repeated a large number of times under the same conditions. Probability Theory deals with those experiments of which the result (expressed in one way or another) may vary from trial to trial. The events that refer to the experiment's result and which may or may not occur are usually called *random events*.

For example, suppose we are tossing a coin. The experiment has only two outcomes: either heads or tails show up, and before the experiment has been carried out, it is impossible to say which one will occur. As we have already noted, the reason for this is that we cannot take into account all the factors influencing the final position of the coin. A similar situation will prevail if you buy a ticket for each lottery draw and try to predict whether it will win or not, or, observing the operation of a complex machine, you try to determine in advance if it will have failed before or

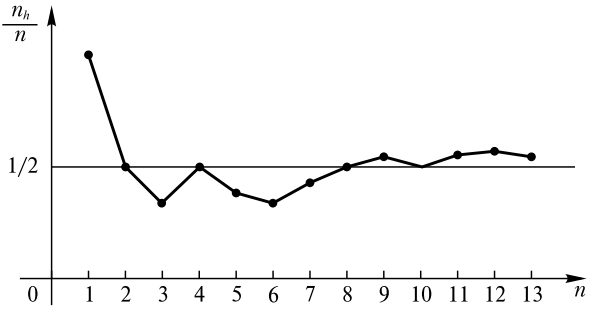

Fig. 1 The plot of the relative frequencies n_h/n corresponding to the outcome sequence *htthtthhhthht* in the coin tossing experiment

after a given time. In such situations, it is very hard to find any laws when considering the results of individual experiments. Therefore there is little justification for constructing any theory here.

However, if one turns to a *long* sequence of repetitions of such an experiment, an interesting phenomenon becomes apparent. While individual results of the experiments display a highly "irregular" behaviour, the average results demonstrate stability. Consider, say, a long series of repetitions of our coin tossing experiment and denote by n_h the number of heads in the first n trials. Plot the ratio n_h/n versus the number n of conducted experiments (see Fig. 1; the plot corresponds to the outcome sequence *htthtthhhthh*, where *h* stands for heads and *t* for tails, respectively).

We will then see that, as n increases, the polygon connecting the consecutive points $(n, n_h/n)$ very quickly approaches the straight line $n_h/n = 1/2$. To verify this observation, G.L. Leclerc, comte de Buffon,[1] tossed a coin 4040 times. The number of heads was 2048, so that the relative frequency n_h/n of heads was 0.5069. K. Pearson tossed a coin 24,000 times and got 12,012 heads, so that $n_h/n = 0.5005$.

It turns out that this phenomenon is universal: *the relative frequency of a certain outcome in a series of repetitions of an experiment under the same conditions tends towards a certain number* $p \in [0, 1]$ *as the number of repetitions grows*. It is an objective law of nature which forms the foundation of Probability Theory.

It would be natural to define the probability of an experiment outcome to be just the number p towards which the relative frequency of the outcome tends. However, such a definition of probability (usually related to the name of R. von Mises) has proven to be inconvenient. First of all, in reality, each time we will be dealing not with an infinite sequence of frequencies, but rather with finitely many elements thereof. Obtaining the entire sequence is unfeasible. Hence the frequency (let it again be n_h/n) of the occurrence of a certain outcome will, as a rule, be different for each new series of repetitions of the same experiment.

This fact led to intense discussions and a lot of disagreement regarding how one should define the concept of probability. Fortunately, there was a class of phenomena that possessed certain "symmetry" (in gambling, coin tossing etc.) for which one could compute *in advance*, prior to the experiment, the expected numerical values

[1] The data is borrowed from [15].

of the probabilities. Take, for instance, a cube made of a sufficiently homogeneous material. There are no reasons for the cube to fall on any of its faces more often than on some other face. It is therefore natural to expect that, when rolling a die a large number of times, the frequency of each of its faces will be close to 1/6. Based on these considerations, Laplace believed that the concept of *equiprobability* is the fundamental one for Probability Theory. The probability of an event would then be defined as the ratio of the number of "favourable" outcomes to the total number of possible outcomes. Thus, the probability of getting an odd number of points (e.g. 1, 3 or 5) when rolling a die once was declared to be 3/6 (i.e. the number of faces with an odd number of points was divided by the total number of all faces). If the die were rolled ten times, then one would have 6^{10} in the denominator, as this number gives the total number of equally likely outcomes and calculating probabilities reduces to counting the number of "favourable outcomes" (the ones resulting in the occurrence of a given event).

The development of the mathematical theory of probabilities began from the instance when one started defining probability as the ratio of the number of favourable outcomes to the total number of equally likely outcomes, and this approach is nowadays called "classical" (for more details, see Chap. 1).

Later on, at the beginning of the 20th century, this approach was severely criticised for being too restrictive. The initiator of the critique was R. von Mises. As we have already noted, his conception was based on *postulating* stability of the frequencies of events in a long series of experiments. That was a confusion of physical and mathematical concepts. No passage to the limit can serve as justification for introducing the notion of "probability". If, for instance, the values n_h/n were to converge to the limiting value $1/2$ in Fig. 1 too slowly, that would mean that nobody would be able to find the value of that limit in the general (non-classical) case. So the approach is clearly vulnerable: it would mean that Probability Theory would be applicable only to those situations where *frequencies have a limit*. But why frequencies would have a limit remained unexplained and was not even discussed.

In this relation, R. von Mises' conception has been in turn criticised by many mathematicians, including A.Ya. Khinchin, S.N. Bernstein, A.N. Kolmogorov and others. Somewhat later, another approach was suggested that proved to be fruitful for the development of the mathematical theory of probabilities. Its general features were outlined by S.N. Bernstein in 1908. In 1933 a rather short book "Foundations of Probability Theory" by A.N. Kolmogorov appeared that contained a complete and clear exposition of the axioms of Probability Theory. The general construction of the concept of probability based on Kolmogorov's axiomatics removed all the obstacles for the development of the theory and is nowadays universally accepted.

The creation of an axiomatic Probability Theory provided a solution to the sixth Hilbert problem (which concerned, in particular, Probability Theory) that had been formulated by D. Hilbert at the Second International Congress of Mathematicians in Paris in 1900. The problem was on the axiomatic construction of a number of physical sciences, Probability Theory being classified as such by Hilbert at that time.

An axiomatic foundation separates the mathematical aspect from the physical: one no longer needs to explain *how and where* the concept of probability comes

from. The concept simply becomes a primitive one, its properties being described by *axioms* (which are essentially the axioms of Measure Theory). However, the problem of how the probability thus introduced is related (and can be applied) to the real world remains open. But this problem is mostly removed by the remarkable fact that, under the axiomatic construction, the desired fundamental property that the frequencies of the occurrence of an event converge to the probability of the event does take place and is a precise mathematical result. (For more details, see Chaps. 2 and 5.)[2]

We will begin by defining probability in a somewhat simplified situation, in the so-called *discrete case*.

[2]Much later, in the 1960s A.N. Kolmogorov attempted to develop a fundamentally different approach to the notions of probability and randomness. In that approach, the measure of randomness, say, of a sequence 0, 1, 0, 0, 1, ... consisting of 0s and 1s (or some other symbols) is the complexity of the algorithm describing this sequence. The new approach stimulated the development of a number of directions in contemporary mathematics, but, mostly due to its complexity, has not yet become widely accepted.

Contents

1 Discrete Spaces of Elementary Events 1
1.1 Probability Space 1
1.2 The Classical Scheme 4
1.3 The Bernoulli Scheme 6
1.4 The Probability of the Union of Events. Examples 9

2 An Arbitrary Space of Elementary Events 13
2.1 The Axioms of Probability Theory. A Probability Space 13
2.2 Properties of Probability 20
2.3 Conditional Probability. Independence of Events and Trials 21
2.4 The Total Probability Formula. The Bayes Formula 25

3 Random Variables and Distribution Functions 31
3.1 Definitions and Examples 31
3.2 Properties of Distribution Functions. Examples 33
3.2.1 The Basic Properties of Distribution Functions 33
3.2.2 The Most Common Distributions 37
3.2.3 The Three Distribution Types 39
3.2.4 Distributions of Functions of Random Variables 42
3.3 Multivariate Random Variables 44
3.4 Independence of Random Variables and Classes of Events 48
3.4.1 Independence of Random Vectors 48
3.4.2 Independence of Classes of Events 50
3.4.3 Relations Between the Introduced Notions 52
3.5 On Infinite Sequences of Random Variables 56
3.6 Integrals 56
3.6.1 Integral with Respect to Measure 56
3.6.2 The Stieltjes Integral 57
3.6.3 Integrals of Multivariate Random Variables. The Distribution of the Sum of Independent Random Variables 59

4 Numerical Characteristics of Random Variables 65
4.1 Expectation 65
4.2 Conditional Distribution Functions and Conditional Expectations 70
4.3 Expectations of Functions of Independent Random Variables . . . 74
4.4 Expectations of Sums of a Random Number of Random Variables 75
4.5 Variance 83
4.6 The Correlation Coefficient and Other Numerical Characteristics 85
4.7 Inequalities 87
4.7.1 Moment Inequalities 87
4.7.2 Inequalities for Probabilities 89
4.8 Extension of the Notion of Conditional Expectation 91
4.8.1 Definition of Conditional Expectation 91
4.8.2 Properties of Conditional Expectations 95
4.9 Conditional Distributions 99

5 Sequences of Independent Trials with Two Outcomes 107
5.1 Laws of Large Numbers 107
5.2 The Local Limit Theorem and Its Refinements 109
5.2.1 The Local Limit Theorem 109
5.2.2 Refinements of the Local Theorem 111
5.2.3 The Local Limit Theorem for the Polynomial Distributions 114
5.3 The de Moivre–Laplace Theorem and Its Refinements 114
5.4 The Poisson Theorem and Its Refinements 117
5.4.1 Quantifying the Closeness of Poisson Distributions to Those of the Sums S_n 117
5.4.2 The Triangular Array Scheme. The Poisson Theorem . . . 120
5.5 Inequalities for Large Deviation Probabilities in the Bernoulli Scheme 125

6 On Convergence of Random Variables and Distributions 129
6.1 Convergence of Random Variables 129
6.1.1 Types of Convergence 129
6.1.2 The Continuity Theorem 134
6.1.3 Uniform Integrability and Its Consequences 134
6.2 Convergence of Distributions 140
6.3 Conditions for Weak Convergence 147

7 Characteristic Functions 153
7.1 Definition and Properties of Characteristic Functions 153
7.1.1 Properties of Characteristic Functions 154
7.1.2 The Properties of Ch.F.s Related to the Structure of the Distribution of ξ 159
7.2 Inversion Formulas 161

7.2.1 The Inversion Formula for Densities . . . 161
7.2.2 The Inversion Formula for Distributions . . . 163
7.2.3 The Inversion Formula in L_2. The Class of Functions that Are Both Densities and Ch.F.s . . . 164
7.3 The Continuity (Convergence) Theorem . . . 167
7.4 The Application of Characteristic Functions in the Proof of the Poisson Theorem . . . 169
7.5 Characteristic Functions of Multivariate Distributions. The Multivariate Normal Distribution . . . 171
7.6 Other Applications of Characteristic Functions. The Properties of the Gamma Distribution . . . 175
7.6.1 Stability of the Distributions $\mathbf{\Phi}_{\alpha,\sigma^2}$ and $\mathbf{K}_{\alpha,\sigma}$ . . . 175
7.6.2 The $\mathbf{\Gamma}$-distribution and its properties . . . 176
7.7 Generating Functions. Application to Branching Processes. A Problem on Extinction . . . 180
7.7.1 Generating Functions . . . 180
7.7.2 The Simplest Branching Processes . . . 180

8 Sequences of Independent Random Variables. Limit Theorems . . . 185
8.1 The Law of Large Numbers . . . 185
8.2 The Central Limit Theorem for Identically Distributed Random Variables . . . 187
8.3 The Law of Large Numbers for Arbitrary Independent Random Variables . . . 188
8.4 The Central Limit Theorem for Sums of Arbitrary Independent Random Variables . . . 199
8.5 Another Approach to Proving Limit Theorems. Estimating Approximation Rates . . . 209
8.6 The Law of Large Numbers and the Central Limit Theorem in the Multivariate Case . . . 214
8.7 Integro-Local and Local Limit Theorems for Sums of Identically Distributed Random Variables with Finite Variance . . . 216
8.7.1 Integro-Local Theorems . . . 216
8.7.2 Local Theorems . . . 219
8.7.3 The Proof of Theorem 8.7.1 in the General Case . . . 222
8.7.4 Uniform Versions of Theorems 8.7.1–8.7.3 for Random Variables Depending on a Parameter . . . 225
8.8 Convergence to Other Limiting Laws . . . 227
8.8.1 The Integral Theorem . . . 230
8.8.2 The Integro-Local and Local Theorems . . . 235
8.8.3 An Example . . . 236

9 Large Deviation Probabilities for Sums of Independent Random Variables . . . 239
9.1 Laplace's and Cramér's Transforms. The Rate Function . . . 240

9.1.1 The Cramér Condition. Laplace's and Cramér's Transforms . . . 240
9.1.2 The Large Deviation Rate Function . . . 243
9.2 A Relationship Between Large Deviation Probabilities for Sums of Random Variables and Those for Sums of Their Cramér Transforms. The Probabilistic Meaning of the Rate Function . . . 250
9.2.1 A Relationship Between Large Deviation Probabilities for Sums of Random Variables and Those for Sums of Their Cramér Transforms . . . 250
9.2.2 The Probabilistic Meaning of the Rate Function . . . 251
9.2.3 The Large Deviations Principle . . . 254
9.3 Integro-Local, Integral and Local Theorems on Large Deviation Probabilities in the Cramér Range . . . 256
9.3.1 Integro-Local and Integral Theorems . . . 256
9.3.2 Local Theorems . . . 261
9.4 Integro-Local Theorems at the Boundary of the Cramér Range . . . 264
9.4.1 Introduction . . . 264
9.4.2 The Probabilities of Large Deviations of S_n in an $o(n)$-Vicinity of the Point $\alpha_+ n$; the Case $\psi''(\lambda_+) < \infty$ . . . 264
9.4.3 The Class of Distributions $\mathcal{ER}$. The Probability of Large Deviations of S_n in an $o(n)$-Vicinity of the Point $\alpha_+ n$ for Distributions $\mathbf{F}$ from the Class $\mathcal{ER}$ in Case $\psi''(\lambda_+)=\infty$ . . 266
9.4.4 On the Large Deviation Probabilities in the Range $\alpha > \alpha_+$ for Distributions from the Class $\mathcal{ER}$ . . . 269
9.5 Integral and Integro-Local Theorems on Large Deviation Probabilities for Sums S_n when the Cramér Condition Is not Met . 269
9.5.1 Integral Theorems . . . 270
9.5.2 Integro-Local Theorems . . . 271
9.6 Integro-Local Theorems on the Probabilities of Large Deviations of S_n Outside the Cramér Range (Under the Cramér Condition) . . 274

10 Renewal Processes . . . 277
10.1 Renewal Processes. Renewal Functions . . . 277
10.1.1 Introduction . . . 277
10.1.2 The Integral Renewal Theorem for Non-identically Distributed Summands . . . 280
10.2 The Key Renewal Theorem in the Arithmetic Case . . . 285
10.3 The Excess and Defect of a Random Walk. Their Limiting Distribution in the Arithmetic Case . . . 290
10.4 The Renewal Theorem and the Limiting Behaviour of the Excess and Defect in the Non-arithmetic Case . . . 293
10.5 The Law of Large Numbers and the Central Limit Theorem for Renewal Processes . . . 298
10.5.1 The Law of Large Numbers . . . 298
10.5.2 The Central Limit Theorem . . . 299

10.5.3 A Theorem on the Finiteness of the Infimum of the Cumulative Sums . . . 300
10.5.4 Stochastic Inequalities. The Law of Large Numbers and the Central Limit Theorem for the Maximum of Sums of Non-identically Distributed Random Variables Taking Values of Both Signs . . . 302
10.5.5 Extension of Theorems 10.5.1 and 10.5.2 to Random Variables Assuming Values of Both Signs . . . 304
10.5.6 The Local Limit Theorem . . . 306
10.6 Generalised Renewal Processes . . . 307
10.6.1 Definition and Some Properties . . . 307
10.6.2 The Central Limit Theorem . . . 309
10.6.3 The Integro-Local Theorem . . . 311

11 Properties of the Trajectories of Random Walks. Zero-One Laws . . . 315
11.1 Zero-One Laws. Upper and Lower Functions . . . 315
11.1.1 Zero-One Laws . . . 315
11.1.2 Lower and Upper Functions . . . 318
11.2 Convergence of Series of Independent Random Variables . . . 320
11.3 The Strong Law of Large Numbers . . . 323
11.4 The Strong Law of Large Numbers for Arbitrary Independent Variables . . . 326
11.5 The Strong Law of Large Numbers for Generalised Renewal Processes . . . 330
11.5.1 The Strong Law of Large Numbers for Renewal Processes . 330
11.5.2 The Strong Law of Large Numbers for Generalised Renewal Processes . . . 331

12 Random Walks and Factorisation Identities . . . 333
12.1 Factorisation Identities . . . 333
12.1.1 Factorisation . . . 333
12.1.2 The Canonical Factorisation of the Function $\mathfrak{f}_z(\lambda) = 1 - z\varphi(\lambda)$ . . . 335
12.1.3 The Second Factorisation Identity . . . 336
12.2 Some Consequences of Theorems 12.1.1–12.1.3 . . . 340
12.2.1 Direct Consequences . . . 340
12.2.2 A Generalisation of the Strong Law of Large Numbers . . . 343
12.3 Pollaczek–Spitzer's Identity. An Identity for $S = \sup_{k\geq 0} S_k$ . . . 344
12.3.1 Pollaczek–Spitzer's Identity . . . 345
12.3.2 An Identity for $S = \sup_{k\geq 0} S_k$ . . . 347
12.4 The Distribution of S in Insurance Problems and Queueing Theory . . . 348
12.4.1 Random Walks in Risk Theory . . . 348
12.4.2 Queueing Systems . . . 349
12.4.3 Stochastic Models in Continuous Time . . . 350

12.5 Cases Where Factorisation Components Can Be Found in an Explicit Form. The Non-lattice Case 351
12.5.1 Preliminary Notes on the Uniqueness of Factorisation . . . 351
12.5.2 Classes of Distributions on the Positive Half-Line with Rational Ch.F.s 354
12.5.3 Explicit Canonical Factorisation of the Function $\mathfrak{v}(\lambda)$ in the Case when the Right Tail of the Distribution **F** Is an Exponential Polynomial 355
12.5.4 Explicit Factorisation of the Function $\mathfrak{v}(\lambda)$ when the Left Tail of the Distribution **F** Is an Exponential Polynomial . . 361
12.5.5 Explicit Canonical Factorisation for the Function $\mathfrak{v}^0(\lambda)$ . . 362
12.6 Explicit Form of Factorisation in the Arithmetic Case 364
12.6.1 Preliminary Remarks on the Uniqueness of Factorisation . 365
12.6.2 The Classes of Distributions on the Positive Half-Line with Rational Generating Functions 366
12.6.3 Explicit Canonical Factorisation of the Function $\mathfrak{v}(z)$ in the Case when the Right Tail of the Distribution **F** Is an Exponential Polynomial 367
12.6.4 Explicit Canonical Factorisation of the Function $\mathfrak{v}(z)$ when the Left Tail of the Distribution **F** Is an Exponential Polynomial 370
12.6.5 Explicit Factorisation of the Function $\mathfrak{v}^0(z)$ 371
12.7 Asymptotic Properties of the Distributions of $\chi_\pm$ and S 372
12.7.1 The Asymptotics of $\mathbf{P}(\chi_+ > x \mid \eta_+ < \infty)$ and $\mathbf{P}(\chi_-^0 < -x)$ in the Case $\mathbf{E}\xi \le 0$ 373
12.7.2 The Asymptotics of $\mathbf{P}(S > x)$ 376
12.7.3 The Distribution of the Maximal Values of Generalised Renewal Processes 380
12.8 On the Distribution of the First Passage Time 381
12.8.1 The Properties of the Distributions of the Times $\eta_\pm$ 381
12.8.2 The Distribution of the First Passage Time of an Arbitrary Level x by Arithmetic Skip-Free Walks 384

13 Sequences of Dependent Trials. Markov Chains 389
13.1 Countable Markov Chains. Definitions and Examples. Classification of States 389
13.1.1 Definition and Examples 389
13.1.2 Classification of States 392
13.2 Necessary and Sufficient Conditions for Recurrence of States. Types of States in an Irreducible Chain. The Structure of a Periodic Chain 395
13.3 Theorems on Random Walks on a Lattice 398
13.3.1 Symmetric Random Walks in $\mathbb{R}^k$, $k \ge 2$ 400
13.3.2 Arbitrary Symmetric Random Walks on the Line 401
13.4 Limit Theorems for Countable Homogeneous Chains 404

13.4.1 Ergodic Theorems 404
13.4.2 The Law of Large Numbers and the Central Limit Theorem for the Number of Visits to a Given State 412
13.5 The Behaviour of Transition Probabilities for Reducible Chains . . 412
13.6 Markov Chains with Arbitrary State Spaces. Ergodicity of Chains with Positive Atoms 414
13.6.1 Markov Chains with Arbitrary State Spaces 414
13.6.2 Markov Chains Having a Positive Atom 420
13.7 Ergodicity of Harris Markov Chains 423
13.7.1 The Ergodic Theorem 423
13.7.2 On Conditions (I) and (II) 429
13.8 Laws of Large Numbers and the Central Limit Theorem for Sums of Random Variables Defined on a Markov Chain 436
13.8.1 Random Variables Defined on a Markov Chain 436
13.8.2 Laws of Large Numbers 437
13.8.3 The Central Limit Theorem 443

14 Information and Entropy 447
14.1 The Definitions and Properties of Information and Entropy 447
14.2 The Entropy of a Finite Markov Chain. A Theorem on the Asymptotic Behaviour of the Information Contained in a Long Message; Its Applications 452
14.2.1 The Entropy of a Sequence of Trials Forming a Stationary Markov Chain 452
14.2.2 The Law of Large Numbers for the Amount of Information Contained in a Message 453
14.2.3 The Asymptotic Behaviour of the Number of the Most Common Outcomes in a Sequence of Trials 454

15 Martingales 457
15.1 Definitions, Simplest Properties, and Examples 457
15.2 The Martingale Property and Random Change of Time. Wald's Identity 462
15.3 Inequalities 477
15.3.1 Inequalities for Martingales 477
15.3.2 Inequalities for the Number of Crossings of a Strip 481
15.4 Convergence Theorems 482
15.5 Boundedness of the Moments of Stochastic Sequences 487

16 Stationary Sequences 493
16.1 Basic Notions 493
16.2 Ergodicity (Metric Transitivity), Mixing and Weak Dependence . . 497
16.3 The Ergodic Theorem 502

17 Stochastic Recursive Sequences 507
17.1 Basic Concepts 507
17.2 Ergodicity and Renovating Events. Boundedness Conditions 508

17.2.1 Ergodicity of Stochastic Recursive Sequences 508
17.2.2 Boundedness of Random Sequences 514
17.3 Ergodicity Conditions Related to the Monotonicity of f 516
17.4 Ergodicity Conditions for Contracting in Mean Lipschitz Transformations 518

18 Continuous Time Random Processes 527
18.1 General Definitions 527
18.2 Criteria of Regularity of Processes 532

19 Processes with Independent Increments 539
19.1 General Properties 539
19.2 Wiener Processes. The Properties of Trajectories 542
19.3 The Laws of the Iterated Logarithm 545
19.4 The Poisson Process 549
19.5 Description of the Class of Processes with Independent Increments 552

20 Functional Limit Theorems 559
20.1 Convergence to the Wiener Process 559
20.2 The Law of the Iterated Logarithm 568
20.3 Convergence to the Poisson Process 572
20.3.1 Convergence of the Processes of Cumulative Sums 572
20.3.2 Convergence of Sums of Thinning Renewal Processes . . . 575

21 Markov Processes 579
21.1 Definitions and General Properties 579
21.1.1 Definition and Basic Properties 579
21.1.2 Transition Probability 581
21.2 Markov Processes with Countable State Spaces. Examples 583
21.2.1 Basic Properties of the Process 583
21.2.2 Examples 589
21.3 Branching Processes 591
21.4 Semi-Markov Processes 593
21.4.1 Semi-Markov Processes on the States of a Chain 593
21.4.2 The Ergodic Theorem 594
21.4.3 Semi-Markov Processes on Chain Transitions 597
21.5 Regenerative Processes 600
21.5.1 Regenerative Processes. The Ergodic Theorem 600
21.5.2 The Laws of Large Numbers and Central Limit Theorem for Integrals of Regenerative Processes 601
21.6 Diffusion Processes 603

22 Processes with Finite Second Moments. Gaussian Processes 611
22.1 Processes with Finite Second Moments 611
22.2 Gaussian Processes 614
22.3 Prediction Problem 616

Appendix 1 Extension of a Probability Measure 619

Appendix 2 Kolmogorov's Theorem on Consistent Distributions 625

Appendix 3 Elements of Measure Theory and Integration 629
3.1 Measure Spaces 629
3.2 The Integral with Respect to a Probability Measure 630
3.2.1 The Integrals of a Simple Function 630
3.2.2 The Integrals of an Arbitrary Function 631
3.2.3 Properties of Integrals 634
3.3 Further Properties of Integrals 635
3.3.1 Convergence Theorems 635
3.3.2 Connection to Integration with Respect to a Measure on the Real Line 636
3.3.3 Product Measures and Iterated Integrals 638
3.4 The Integral with Respect to an Arbitrary Measure 640
3.5 The Lebesgue Decomposition Theorem and the Radon–Nikodym Theorem 643
3.6 Weak Convergence and Convergence in Total Variation of Distributions in Arbitrary Spaces 649
3.6.1 Weak Convergence 649
3.6.2 Convergence in Total Variation 652

Appendix 4 The Helly and Arzelà–Ascoli Theorems 655

Appendix 5 The Proof of the Berry–Esseen Theorem 659

Appendix 6 The Basic Properties of Regularly Varying Functions and Subexponential Distributions 665
6.1 General Properties of Regularly Varying Functions 665
6.2 The Basic Asymptotic Properties 668
6.3 The Asymptotic Properties of the Transforms of R.V.F.s (Abel-Type Theorems) 672
6.4 Subexponential Distributions and Their Properties 674

Appendix 7 The Proofs of Theorems on Convergence to Stable Laws . . 687
7.1 The Integral Limit Theorem 687
7.2 The Integro-Local and Local Limit Theorems 699

Appendix 8 Upper and Lower Bounds for the Distributions of the Sums and the Maxima of the Sums of Independent Random Variables 703
8.1 Upper Bounds Under the Cramér Condition 703
8.2 Upper Bounds when the Cramér Condition Is Not Met 704
8.3 Lower Bounds 713

Appendix 9 Renewal Theorems 715

References 723

Index of Basic Notation 725

Subject Index 727