Graduate Texts in Mathematics

A. N. Shiryaev

Probability

Second Edition

概率论

第2版

Springer-Verlag
世界图书出版公司

A. N. Shiryaev

Probability

Second Edition

Translated by R. P. Boas

With 54 Illustrations

Springer
世界图书出版公司

书　　名: Probability　Second Edition
作　　者: A.N. Shiryaev
中 译 名: 概率论 第 2 版
出 版 者: 世界图书出版公司北京公司
印 刷 者: 北京世图印刷厂
发　　行: 世界图书出版公司北京公司 (北京朝内大街 137 号　100010)
联系电话: 010-64015659, 64038347
电子信箱: kjsk@vip.sina.com
开　　本: 24　　　印　张: 27
出版年代: 2004 年 11 月
书　　号: 7-5062-7188-5/O · 506
版权登记: 图字: 01-2004-5050
定　　价: 65.00 元

A. N. Shiryaev
Steklov Mathematical Institute
Vavilova 42
GSP-1 117966 Moscow
Russia

R. P. Boas (*Translator*)
Department of Mathematics
Northwestern University
Evanston, IL 60201
USA

Mathematics Subject Classification (1991): 60-01

Library of Congress Cataloging-in-Publication Data
Shiriaev, Al'bert Nikolaevich.
[Veroiatnost'. English]
Probability / A.N. Shiryaev; translated by R.P. Boas.—2nd ed.
p. cm.—(Graduate texts in mathematics; 95)
Includes bibliographical references (p. –) and index.
ISBN 0-387-94549-0 (hardcover: alk. paper)
1. Probabilities. I. Title. II. Series.
QA273.S54413 1995
519.5—dc20 95-19033

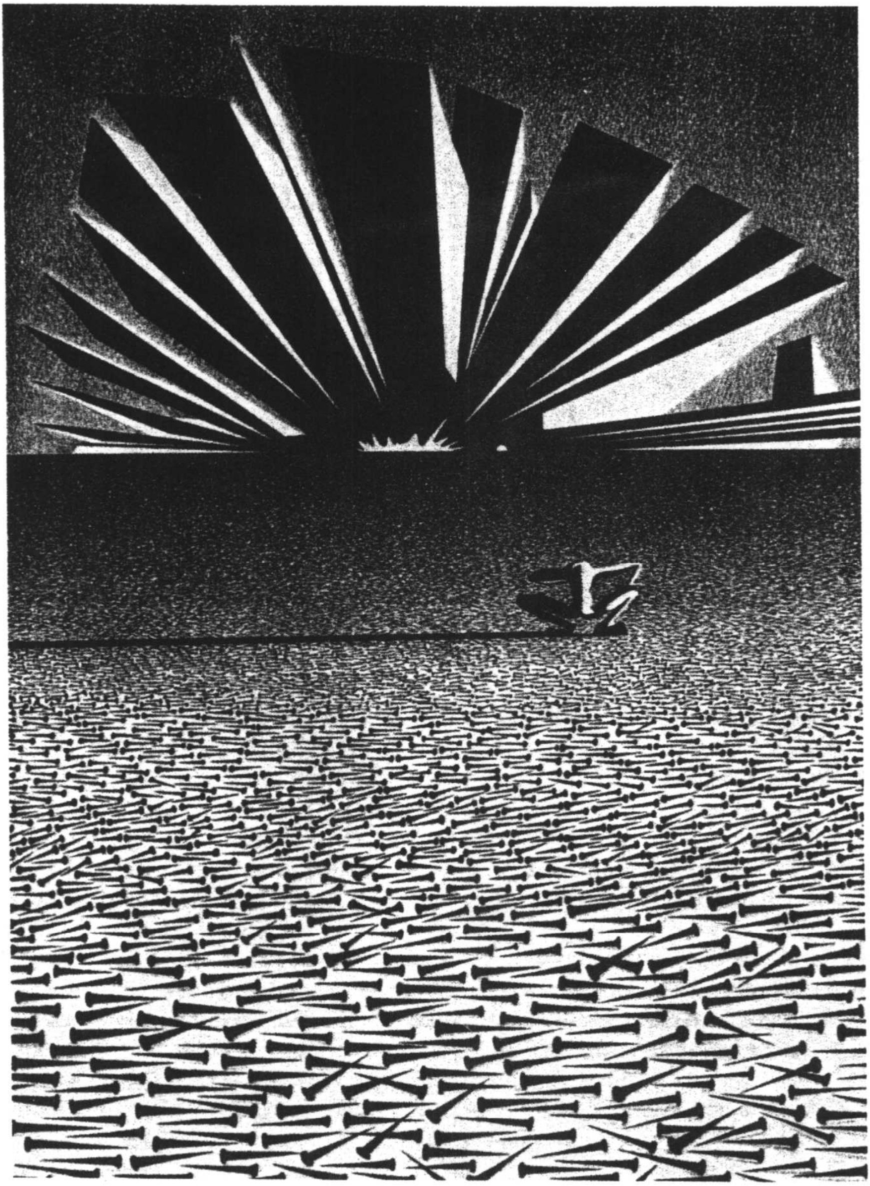
Original Russian edition: *Veroiatnost'*. Moscow: Nauka, 1980 (first edition); second edition: 1989.

First edition published by Springer-Verlag Berlin Heidelberg.

This book is part of the Springer Series in Soviet Mathematics.

9 8 7 6 5 4 3 2 1

ISBN 0-387-94549-0 Springer-Verlag New York Berlin Heidelberg



"Order out of chaos"
(*Courtesy of Professor A. T. Fomenko of the Moscow State University*)

Preface to the Second Edition

In the Preface to the first edition, originally published in 1980, we mentioned that this book was based on the author's lectures in the Department of Mechanics and Mathematics of the Lomonosov University in Moscow, which were issued, in part, in mimeographed form under the title "Probability, Statistics, and Stochastic Processors, I, II" and published by that University. Our original intention in writing the first edition of this book was to divide the contents into three parts: probability, mathematical statistics, and theory of stochastic processes, which corresponds to an outline of a three-semester course of lectures for university students of mathematics. However, in the course of preparing the book, it turned out to be impossible to realize this intention completely, since a full exposition would have required too much space. In this connection, we stated in the Preface to the first edition that only *probability theory* and the *theory of random processes with discrete time* were really adequately presented.

Essentially all of the first edition is reproduced in this second edition. Changes and corrections are, as a rule, editorial, taking into account comments made by both Russian and foreign readers of the Russian original and of the English and German translations [S11]. The author is grateful to all of these readers for their attention, advice, and helpful criticisms.

In this second English edition, new material also has been added, as follows: in Chapter III, §5, §§7–12; in Chapter IV, §5; in Chapter VII, §§8–10. The most important addition is the third chapter. There the reader will find expositions of a number of problems connected with a deeper study of themes such as the distance between probability measures, metrization of weak convergence, and contiguity of probability measures. In the same chapter, we have added proofs of a number of important results on the rapidity of convergence in the central limit theorem and in Poisson's theorem on the approximation of the binomial by the Poisson distribution. These were merely stated in the first edition.

We also call attention to the new material on the probability of large deviations (Chapter IV, §5), on the central limit theorem for sums of dependent random variables (Chapter VII, §8), and on §§9 and 10 of Chapter VII.

During the last few years, the literature on probability published in Russia by Nauka has been extended by Sevastyanov [S10], 1982; Rozanov [R6], 1985; Borovkov [B4], 1986; and Gnedenko [G4], 1988. It appears that these publications, together with the present volume, being quite different and complementing each other, cover an extensive amount of material that is essentially broad enough to satisfy contemporary demands by students in various branches of mathematics and physics for instruction in topics in probability theory.

Gnedenko's textbook [G4] contains many well-chosen examples, including applications, together with pedagogical material and extensive surveys of the history of probability theory. Borovkov's textbook [B4] is perhaps the most like the present book in the style of exposition. Chapters 9 (Elements of Renewal Theory), 11 (Factorization of the Identity) and 17 (Functional Limit Theorems), which distinguish [B4] from this book and from [G4] and [R6], deserve special mention. Rozanov's textbook contains a great deal of material on a variety of mathematical models which the theory of probability and mathematical statistics provides for describing random phenomena and their evolution. The textbook by Sevastyanov is based on his two-semester course at the Moscow State University. The material in its last four chapters covers the minimum amount of probability and mathematical statistics required in a one-year university program. In our text, perhaps to a greater extent than in those mentioned above, a significant amount of space is given to set-theoretic aspects and mathematical foundations of probability theory.

Exercises and problems are given in the books by Gnedenko and Sevastyanov at the ends of chapters, and in the present textbook at the end of each section. These, together with, for example, the problem sets by A. V. Prokhorov and V. G. and N. G. Ushakov (*Problems in Probability Theory*, Nauka, Moscow, 1986) and by Zubkov, Sevastyanov, and Chistyakov (*Collected Problems in Probability Theory*, Nauka, Moscow, 1988) can be used by readers for independent study, and by teachers as a basis for seminars for students.

Special thanks to Harold Boas, who kindly translated the revisions from Russian to English for this new edition.

Moscow A. Shiryaev

Preface to the First Edition

This textbook is based on a three-semester course of lectures given by the author in recent years in the Mechanics–Mathematics Faculty of Moscow State University and issued, in part, in mimeographed form under the title *Probability, Statistics, Stochastic Processes, I, II* by the Moscow State University Press.

We follow tradition by devoting the first part of the course (roughly one semester) to the elementary theory of probability (Chapter I). This begins with the construction of probabilistic models with finitely many outcomes and introduces such fundamental probabilistic concepts as sample spaces, events, probability, independence, random variables, expectation, correlation, conditional probabilities, and so on.

Many probabilistic and statistical regularities are effectively illustrated even by the simplest random walk generated by Bernoulli trials. In this connection we study both classical results (law of large numbers, local and integral De Moivre and Laplace theorems) and more modern results (for example, the arc sine law).

The first chapter concludes with a discussion of dependent random variables generated by martingales and by Markov chains.

Chapters II–IV form an expanded version of the second part of the course (second semester). Here we present (Chapter II) Kolmogorov's generally accepted axiomatization of probability theory and the mathematical methods that constitute the tools of modern probability theory (σ-algebras, measures and their representations, the Lebesgue integral, random variables and random elements, characteristic functions, conditional expectation with respect to a σ-algebra, Gaussian systems, and so on). Note that two measure-theoretical results—Carathéodory's theorem on the extension of measures and the Radon–Nikodým theorem—are quoted without proof.

The third chapter is devoted to problems about weak convergence of probability distributions and the method of characteristic functions for proving limit theorems. We introduce the concepts of relative compactness and tightness of families of probability distributions, and prove (for the real line) Prohorov's theorem on the equivalence of these concepts.

The same part of the course discusses properties "with probability 1" for sequences and sums of independent random variables (Chapter IV). We give proofs of the "zero or one laws" of Kolmogorov and of Hewitt and Savage, tests for the convergence of series, and conditions for the strong law of large numbers. The law of the iterated logarithm is stated for arbitrary sequences of independent identically distributed random variables with finite second moments, and proved under the assumption that the variables have Gaussian distributions.

Finally, the third part of the book (Chapters V–VIII) is devoted to random processes with discrete parameters (random sequences). Chapters V and VI are devoted to the theory of stationary random sequences, where "stationary" is interpreted either in the strict or the wide sense. The theory of random sequences that are stationary in the strict sense is based on the ideas of ergodic theory: measure preserving transformations, ergodicity, mixing, etc. We reproduce a simple proof (by A. Garsia) of the maximal ergodic theorem; this also lets us give a simple proof of the Birkhoff–Khinchin ergodic theorem.

The discussion of sequences of random variables that are stationary in the wide sense begins with a proof of the spectral representation of the covariance fuction. Then we introduce orthogonal stochastic measures, and integrals with respect to these, and establish the spectral representation of the sequences themselves. We also discuss a number of statistical problems: estimating the covariance function and the spectral density, extrapolation, interpolation and filtering. The chapter includes material on the Kalman–Bucy filter and its generalizations.

The seventh chapter discusses the basic results of the theory of martingales and related ideas. This material has only rarely been included in traditional courses in probability theory. In the last chapter, which is devoted to Markov chains, the greatest attention is given to problems on the asymptotic behavior of Markov chains with countably many states.

Each section ends with problems of various kinds: some of them ask for proofs of statements made but not proved in the text, some consist of propositions that will be used later, some are intended to give additional information about the circle of ideas that is under discussion, and finally, some are simple exercises.

In designing the course and preparing this text, the author has used a variety of sources on probability theory. The Historical and Bibliographical Notes indicate both the historial sources of the results and supplementary references for the material under consideration.

The numbering system and form of references is the following. Each section has its own enumeration of theorems, lemmas and formulas (with no indication of chapter or section). For a reference to a result from a different section of the same chapter, we use double numbering, with the first number indicating the number of the section (thus, (2.10) means formula (10) of §2). For references to a different chapter we use triple numbering (thus, formula (II.4.3) means formula (3) of §4 of Chapter II). Works listed in the References at the end of the book have the form [L n], where L is a letter and *n* is a numeral.

The author takes this opportunity to thank his teacher A. N. Kolmogorov, and B. V. Gnedenko and Yu. V. Prokhorov, from whom he learned probability theory and under whose direction he had the opportunity of using it. For discussions and advice, the author also thanks his colleagues in the Departments of Probability Theory and Mathematical Statistics at the Moscow State University, and his colleagues in the Section on probability theory of the Steklov Mathematical Institute of the Academy of Sciences of the U.S.S.R.

Moscow
Steklov Mathematical Institute

A. N. SHIRYAEV

Translator's acknowledgement. I am grateful both to the author and to my colleague, C. T. Ionescu Tulcea, for advice about terminology.

R. P. B.

Contents

Introduction

The subject matter of probability theory is the mathematical analysis of random events, i.e., of those empirical phenomena which—under certain circumstance—can be described by saying that:

They do not have *deterministic regularity* (observations of them do not yield the same outcome);

whereas at the same time

They possess some *statistical regularity* (indicated by the statistical stability of their frequency).

We illustrate with the classical example of a "fair" toss of an "unbiased" coin. It is clearly impossible to predict with certainty the outcome of each toss. The results of successive experiments are very irregular (now "head," now "tail") and we seem to have no possibility of discovering any regularity in such experiments. However, if we carry out a large number of "independent" experiments with an "unbiased" coin we can observe a very definite statistical regularity, namely that "head" appears with a frequency that is "close" to $\frac{1}{2}$.

Statistical stability of a frequency is very likely to suggest a hypothesis about a possible quantitative estimate of the "randomness" of some event A connected with the results of the experiments. With this starting point, probability theory postulates that corresponding to an event A there is a definite number $\mathsf{P}(A)$, called the probability of the event, whose intrinsic property is that as the number of "independent" trials (experiments) increases the frequency of event A is approximated by $\mathsf{P}(A)$.

Applied to our example, this means that it is natural to assign the probability $\frac{1}{2}$ to the event A that consists of obtaining "head" in a toss of an "unbiased" coin.

There is no difficulty in multiplying examples in which it is very easy to obtain numerical values intuitively for the probabilities of one or another event. However, these examples are all of a similar nature and involve (so far) undefined concepts such as "fair" toss, "unbiased" coin, "independence," etc.

Having been invented to investigate the quantitative aspects of "randomness," probability theory, like every exact science, became such a science only at the point when the concept of a probabilistic model had been clearly formulated and axiomatized. In this connection it is natural for us to discuss, although only briefly, the fundamental steps in the development of probability theory.

Probability theory, as a science, originated in the middle of the seventeenth century with Pascal (1623–1662), Fermat (1601–1655) and Huygens (1629–1695). Although special calculations of probabilities in games of chance had been made earlier, in the fifteenth and sixteenth centuries, by Italian mathematicians (Cardano, Pacioli, Tartaglia, etc.), the first general methods for solving such problems were apparently given in the famous correspondence between Pascal and Fermat, begun in 1654, and in the first book on probability theory, *De Ratiociniis in Aleae Ludo* (*On Calculations in Games of Chance*), published by Huygens in 1657. It was at this time that the fundamental concept of "mathematical expectation" was developed and theorems on the addition and multiplication of probabilities were established.

The real history of probability theory begins with the work of James Bernoulli (1654–1705), *Ars Conjectandi* (*The Art of Guessing*) published in 1713, in which he proved (quite rigorously) the first limit theorem of probability theory, the law of large numbers; and of De Moivre (1667–1754), *Miscellanea Analytica Supplementum* (a rough translation might be *The Analytic Method* or *Analytic Miscellany*, 1730), in which the central limit theorem was stated and proved for the first time (for symmetric Bernoulli trials).

Bernoulli deserves the credit for introducing the "classical" definition of the concept of the *probability* of an event as the *ratio* of the number of possible outcomes of an experiment, that are favorable to the event, to the number of possible outcomes.

Bernoulli was probably the first to realize the importance of considering infinite sequences of random trials and to make a clear distinction between the probability of an event and the frequency of its realization.

De Moivre deserves the credit for defining such concepts as independence, mathematical expectation, and conditional probability.

In 1812 there appeared Laplace's (1749–1827) great treatise *Théorie Analytique des Probabilitiés* (*Analytic Theory of Probability*) in which he presented his own results in probability theory as well as those of his predecessors. In particular, he generalized De Moivre's theorem to the general (unsymmetric) case of Bernoulli trials, and at the same time presented De Moivre's results in a more complete form.

Laplace's most important contribution was the application of probabilistic methods to errors of observation. He formulated the idea of considering errors of observation as the cumulative results of adding a large number of independent elementary errors. From this it followed that under rather general conditions the distribution of errors of observation must be at least approximately normal.

The work of Poisson (1781–1840) and Gauss (1777–1855) belongs to the same epoch in the development of probability theory, when the center of the stage was held by limit theorems. In contemporary probability theory we think of Poisson in connection with the distribution and the process that bear his name. Gauss is credited with originating the theory of errors and, in particular, with creating the fundamental method of least squares.

The next important period in the development of probability theory is connected with the names of P. L. Chebyshev (1821–1894), A. A. Markov (1856–1922), and A. M. Lyapunov (1857–1918), who developed effective methods for proving limit theorems for sums of independent but arbitrarily distributed random variables.

The number of Chebyshev's publications in probability theory is not large—four in all—but it would be hard to overestimate their role in probability theory and in the development of the classical Russian school of that subject.

> "On the methodological side, the revolution brought about by Chebyshev was not only his insistence for the first time on complete rigor in the proofs of limit theorems, . . . but also, and principally, that Chebyshev always tried to obtain precise estimates for the deviations from the limiting regularities that are available for large but finite numbers of trials, in the form of inequalities that are valid unconditionally for any number of trials."
>
> (A. N. Kolmogorov [30])

Before Chebyshev the main interest in probability theory had been in the calculation of the probabilities of random events. He, however, was the first to realize clearly and exploit the full strength of the concepts of random variables and their mathematical expectations.

The leading exponent of Chebyshev's ideas was his devoted student Markov, to whom there belongs the indisputable credit of presenting his teacher's results with complete clarity. Among Markov's own significant contributions to probability theory were his pioneering investigations of limit theorems for sums of independent random variables and the creation of a new branch of probability theory, the theory of dependent random variables that form what we now call a Markov chain.

> "Markov's classical course in the calculus of probability and his original papers, which are models of precision and clarity, contributed to the greatest extent to the transformation of probability theory into one of the most significant