Graduate Texts in Mathematics

95

Probability

Springer-Verlag

A. N. Shiryayev

Probability

Translated by R. P. Boas

With 54 Illustrations

Springer-Verlag
New York Berlin Heidelberg Tokyo
World Publishing Corporation, Beijing, China

Preface

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.

A. N. Shiryayev
Steklov Mathematical Institute
Vavilova 42
GSP-1 117333 Moscow
U.S.S.R.

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

AMS Classification: 60-01

Library of Congress Cataloging in Publication Data
Shirîaev, Al'bert Nikolaevich.
Probability.
(Graduate texts in mathematics; 95)
Translation of: Veroîatnost'.
Bibliography: p.
Includes index.
1. Probabilities. I. Title. II. Series.
QA273.S54413 1984 519 83-14813

Original Russian edition: *Veroîatnost'*. Moscow: Nauka, 1979.

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

Typeset by Composition House Ltd., Salisbury, England.
Printed and bound by R. R. Donnelly & Sons, Harrisonburg, Virginia.
Reprinted in China by World Publishing Corporation
For distribution and sale in the People's Republic of China only
只限在中华人民共和国发行

ISBN 0-387-90898-6 Springer-Verlag New York Berlin Heidelberg Tokyo
ISBN 3-540-90898-6 Springer-Verlag Berlin Heidelberg New York Tokyo
ISBN 7-5062-0124-0 World Publishing Corporation China

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. Prohorov, 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 A. N. SHIRYAYEV
Steklov Mathematical Institute

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.

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 historical 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

Contents

Introduction

The subject matter of probability theory is the mathematical analysis of random events, i.e. of those empirical phenomena which—under certain circumstances—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 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 branches of mathematics and to a wide extension of the ideas and methods of Chebyshev."
>
> (S. N. Bernstein [3])

To prove the central limit theorem of probability theory (the theorem on convergence to the normal distribution), Chebyshev and Markov used what is known as the method of moments. With more general hypotheses and a simpler method, the method of characteristic functions, the theorem was obtained by Lyapunov. The subsequent development of the theory has shown that the method of characteristic functions is a powerful analytic tool for establishing the most diverse limit theorems.

The modern period in the development of probability theory begins with its axiomatization. The first work in this direction was done by S. N. Bernstein (1880–1968), R. von Mises (1883–1953), and E. Borel (1871–1956). A. N. Kolmogorov's book *Foundations of the Theory of Probability* appeared in 1933. Here he presented the axiomatic theory that has become generally accepted and is not only applicable to all the classical branches of probability theory, but also provides a firm foundation for the development of new branches that have arisen from questions in the sciences and involve infinite-dimensional distributions.

The treatment in the present book is based on Kolmogorov's axiomatic approach. However, to prevent formalities and logical subtleties from obscuring the intuitive ideas, our exposition begins with the elementary theory of probability, whose elementariness is merely that in the corresponding probabilistic models we consider only experiments with finitely many outcomes. Thereafter we present the foundations of probability theory in their most general form.

The 1920s and '30s saw a rapid development of one of the new branches of probability theory, the theory of stochastic processes, which studies families of random variables that evolve with time. We have seen the creation of theories of Markov processes, stationary processes, martingales, and limit theorems for stochastic processes. Information theory is a recent addition.

The present book is principally concerned with stochastic processes with discrete parameters: random sequences. However, the material presented in the second chapter provides a solid foundation (particularly of a logical nature) for the study of the general theory of stochastic processes.

It was also in the 1920s and '30s that mathematical statistics became a separate mathematical discipline. In a certain sense mathematical statistics deals with inverses of the problems of probability: If the basic aim of probability theory is to calculate the probabilities of complicated events under a given probabilistic model, mathematical statistics sets itself the inverse problem: to clarify the structure of probabilistic–statistical models by means of observations of various complicated events.

Some of the problems and methods of mathematical statistics are also discussed in this book. However, all that is presented in detail here is probability theory and the theory of stochastic processes with discrete parameters.

CHAPTER I

Elementary Probability Theory

§1. Probabilistic Model of an Experiment with a Finite Number of Outcomes

1. Let us consider an experiment of which all possible results are included in a finite number of outcomes $\omega_1, \ldots, \omega_N$. We do not need to know the nature of these outcomes, only that there are a finite number N of them.

We call $\omega_1, \ldots, \omega_N$ *elementary events*, or *sample points*, and the finite set

$$\Omega = \{\omega_1, \ldots, \omega_N\},$$

the *space of elementary events* or the *sample space*.

The choice of the space of elementary events is the *first step* in formulating a probabilistic model for an experiment. Let us consider some examples of sample spaces.

EXAMPLE 1. For a single toss of a coin the sample space Ω consists of two points:

$$\Omega = \{\mathrm{H}, \mathrm{T}\},$$

where H = "head" and T = "tail". (We exclude possibilities like "the coin stands on edge," "the coin disappears," etc.)

EXAMPLE 2. For n tosses of a coin the sample space is

$$\Omega = \{\omega : \omega = (a_1, \ldots, a_n), a_i = \mathrm{H} \text{ or } \mathrm{T}\}$$

and the general number $N(\Omega)$ of outcomes is 2^n.

EXAMPLE 3. First toss a coin. If it falls "head" then toss a die (with six faces numbered 1, 2, 3, 4, 5, 6); if it falls "tail", toss the coin again. The sample space for this experiment is

$$\Omega = \{H1, H2, H3, H4, H5, H6, TH, TT\}.$$

We now consider some more complicated examples involving the selection of n balls from an urn containing M distinguishable balls.

2. EXAMPLE 4 (Sampling with replacement). This is an experiment in which after each step the selected ball is returned again. In this case each sample of n balls can be presented in the form $(a_1, \ldots, a_n)$, where a_i is the label of the ball selected at the ith step. It is clear that in sampling with replacement each a_i can have any of the M values $1, 2, \ldots, M$. The description of the sample space depends in an essential way on whether we consider samples like, for example, (4, 1, 2, 1) and (1, 4, 2, 1) as different or the same. It is customary to distinguish two cases: *ordered* samples and *unordered* samples. In the first case samples containing the same elements, but arranged differently, are considered to be different. In the second case the order of the elements is disregarded and the two samples are considered to be the same. To emphasize which kind of sample we are considering, we use the notation $(a_1, \ldots, a_n)$ for ordered samples and $[a_1, \ldots, a_n]$ for unordered samples.

Thus for ordered samples the sample space has the form

$$\Omega = \{\omega : \omega = (a_1, \ldots, a_n), a_i = 1, \ldots, M\}$$

and the number of (different) outcomes is

$$N(\Omega) = M^n. \tag{1}$$

If, however, we consider unordered samples, then

$$\Omega = \{\omega : \omega = [a_1, \ldots, a_n], a_i = 1, \ldots, M\}.$$

Clearly the number $N(\Omega)$ of (different) unordered samples is smaller than the number of ordered samples. Let us show that in the present case

$$N(\Omega) = C^n_{M+n-1}, \tag{2}$$

where $C^l_k \equiv k!/[l!(k-l)!]$ is the number of combinations of l elements, taken k at a time.

We prove this by induction. Let $N(M, n)$ be the number of outcomes of interest. It is clear that when $k \leq M$ we have

$$N(k, 1) = k = C^1_k.$$

Now suppose that $N(k, n) = C_{k+n-1}^{k}$ for $k \le M$; we show that this formula continues to hold when n is replaced by $n + 1$. For the unordered samples $[a_1, \ldots, a_{n+1}]$ that we are considering, we may suppose that the elements are arranged in nondecreasing order: $a_1 \le a_2 \le \cdots \le a_n$. It is clear that the number of unordered samples with $a_1 = 1$ is $N(M, n)$, the number with $a_1 = 2$ is $N(M - 1, n)$, etc. Consequently

$$\begin{aligned} N(M, n+1) &= N(M, n) + N(M-1, n) + \cdots + N(1, n) \\ &= C_{M+n-1}^{n} + C_{M-1+n-1}^{n} + \cdots C_{n}^{n} \\ &= (C_{M+n}^{n+1} - C_{M+n-1}^{n+1}) + (C_{M-1+n}^{n+1} - C_{M-1+n-1}^{n+1}) \\ &\quad + \cdots + (C_{n+1}^{n+1} - C_{n}^{n}) = C_{M+n}^{n+1}; \end{aligned}$$

here we have used the easily verified property

$$C_k^{l-1} + C_k^l = C_{k+1}^l$$

of the binomial coefficients.

Example 5 (Sampling without replacement). Suppose that $n \le M$ and that the selected balls are not returned. In this case we again consider two possibilities, namely ordered and unordered samples.

For ordered samples without replacement the sample space is

$$\Omega = \{\omega : \omega = (a_1, \ldots, a_n), a_k \ne a_l, k \ne l, a_i = 1, \ldots, M\},$$

and the number of elements of this set (called *permutations*) is $M(M-1)\cdots(M-n+1)$. We denote this by $(M)_n$ or A_M^n and call it "the number of permutations of M things, n at a time").

For unordered samples (called *combinations*) the sample space

$$\Omega = \{\omega : \omega = [a_1, \ldots, a_n], a_k \ne a_l, k \ne l, a_i = 1, \ldots, M\}$$

consists of

$$N(\Omega) = C_M^n \tag{3}$$

elements. In fact, from each unordered sample $[a_1, \ldots, a_n]$ consisting of distinct elements we can obtain $n!$ ordered samples. Consequently

$$N(\Omega) \cdot n! = (M)_n$$

and therefore

$$N(\Omega) = \frac{(M)_n}{n!} = C_M^n.$$

The results on the numbers of samples of n from an urn with M balls are presented in Table 1.

Table 1

M^n	C^n_{M+n-1}	With replacement
$(M)_n$	C^n_M	Without replacement
Ordered	Unordered	Type \ Sample

For the case $M = 3$ and $n = 2$, the corresponding sample spaces are displayed in Table 2.

EXAMPLE 6 (Distribution of objects in cells). We consider the structure of the sample space in the problem of placing n objects (balls, etc.) in M cells (boxes, etc.). For example, such problems arise in statistical physics in studying the distribution of n particles (which might be protons, electrons, ...) among M states (which might be energy levels).

Let the cells be numbered $1, 2, \ldots, M$, and suppose first that the objects are distinguishable (numbered $1, 2, \ldots, n$). Then a distribution of the n objects among the M cells is completely described by an ordered set $(a_1, \ldots, a_n)$, where a_i is the index of the cell containing object i. However, if the objects are indistinguishable their distribution among the M cells is completely determined by the unordered set $[a_1, \ldots, a_n]$, where a_i is the index of the cell into which an object is put at the ith step.

Comparing this situation with Examples 4 and 5, we have the following correspondences:

$$\text{(ordered samples)} \leftrightarrow \text{(distinguishable objects)},$$

$$\text{(unordered samples)} \leftrightarrow \text{(indistinguishable objects)},$$

Table 2

(1, 1) (1, 2) (1, 3) (2, 1) (2, 2) (2, 3) (3, 1) (3, 2) (3, 3)	[1, 1] [2, 2] [3, 3] [1, 2] [1, 3] [2, 3]	With replacement
(1, 2) (1, 3) (2, 1) (2, 3) (3, 1) (3, 2)	[1, 2] [1, 3] [2, 3]	Without replacement
Ordered	Unordered	Type \ Sample