Probability Theory

R. G. LAHA and V. K. ROHATGI

Probability Theory

R. G. LAHA and V. K. ROHATGI
Bowling Green State University

JOHN WILEY & SONS
New York Chichester Brisbane Toronto

Library of Congress Cataloging in Publication Data:

Laha, R. G.
Probability theory.

(Wiley series in probability and mathematical statistics)
Includes bibliographical references and indexes.
1. Probabilities. 2. Measure theory. I. Rohatgi, V. K., 1939– joint author. II. Title.

QA273.L18 519.2 78-24431
ISBN 0-471-03262-X

Printed in the United States of America

10 9 8 7 6 5 4 3 2

Preface

This book is designed for a beginning or an intermediate graduate course in probability theory. It is intended for a serious student of probability, whether a mathematics and/or a statistics major. A working knowledge of real analysis at the level of Royden, complex analysis at the level of Hille, and abstract measure theory at the level of Halmos is assumed. For students who wish to specialize in probability theory it is expected that this course will be followed by one in stochastic processes and/or one in probability measures on abstract spaces for research preparation. For others, a course based on this book will accord adequate preparation.

The objective of this book is to make the basic concepts of probability theory easily accessible to both students and research workers in a comprehensive manner. We believe that a student of probability theory at graduate level should have a comprehensive knowledge of both the measure-theoretic foundations and the analytical tools of probability theory. This should include, for example, the axiomatic foundations of probability theory as developed by Kolmogorov and the analytical tools as developed, amongst others, by Lévy, Cramér, Feller, and Prokhorov. Some of the existing books on the subject do not concentrate long enough on the central topics such as laws of large numbers, the law of the iterated logarithms, infinitely divisible laws, and the central limit theory, while others do not treat the analytical tools in a comprehensive manner. This book is written to overcome not only these drawbacks but also to meet the requirements set out above.

Some special features of this book are as follows:

1. The strong interrelationship between probability theory and mathematical analysis is emphasized. For example, a detailed discussion of the properties of characteristic functions and $\mathscr{L}_p$ spaces, and a chapter on random variables taking values in normed linear spaces, are included. We emphasize classical as well as modern methods.

2. Special stress is placed on probability that is applicable rather than probability as analysis. Applications of probability, in particular, to statistics and analysis are emphasized.
3. Some recent developments in probability theory are included. For example, a detailed proof of Prokhorov's theorem and its applications (Section 3.9) are given. Section 3.8 deals with semigroups of probability distributions and their infinitesimal generators. In Chapter 7 we prove the Minlos–Sazonov theorem and derive an analogue of the Lévy–Khintchine representation of the Fourier transform of an infinitely divisible probability measure on a Hilbert space. We also discuss in detail the general central limit problem in a Hilbert space.
4. Every attempt has been made to make the book self-contained. Only well-known results from analysis and measure theory have been used. We have avoided using results by quotation from sources such as monographs or research papers.
5. A large number of examples and remarks elucidate the text.
6. An adequate number of problems at the end of each chapter (subdivided by sections) supplement the text.
7. Notes and comments at the end of each chapter include references to sources and to additional reading material.
8. An extensive list of references is included.

A few words are in order about the selection of topics and applications. The choice of topics is somewhat traditional, but the reader will find here some material that is available only in specialized monographs. The ordering of the chapters is for ease in presentation. However, the reader need not follow this order. For example, most of Chapter 6 can easily be understood after Chapter 3. Similarly parts of Chapter 7 can also be followed after Chapter 3 has been read. As for the choice of applications, only those are included for which little or no preparatory work is needed. In view of our intended audience, applications to statistics and analysis figure prominently in our selection, whereas applications to number theory, stochastic processes, and the like are given less attention.

We do not claim any originality in methods of proofs or their presentation. However, a special attempt has been made to present complete proofs in a lucid and precise manner. Most of the results that are included in this book are fairly well known. For this reason we have avoided overburdening the text with credit references. Rather we have cited, wherever possible, monographs and books where such references can be found. Similarly we have referred to only the sources where the results are stated or proved, in the formulation best suited to our needs. For the organization and presentation of the material we have relied heavily on the well known works of Loève, Lukacs, and Gnedenko and Kolmogorov. To these authors we express our indebtedness.

The numbering of chapters, sections, subsections, theorems, and so forth is traditional. Each chapter has been subdivided into several sections. Each section has been further subdivided into subsections wherever necessary. Definitions, theorems, equations, and so on are numbered consecutively within each section. Thus equation $(i.j.k)$ stands for the kth equation in Section j of Chapter i. Section $i.j$ stands for Section j in Chapter i; Section $i.j.k$ stands for Subsection k of Section j in Chapter i; and so on.

References are given at the end of the book and are denoted by numbers enclosed in brackets: [ ].

The set of lecture notes on which this book is based has been used by both of us over the last 10 years. The first-named author class-tested a major portion of the present version during the year 1977–1978.

We are indebted to Professors J. Sethuraman of Florida State University, A. J. Terzuoli of the Polytechnic Institute of New York, and H. Braun of Princeton University who read the first draft and made many suggestions. We are specially grateful to Professors R. J. Tompkins and C. C. Heyde for making numerous suggestions. Thanks are also due to D. Borowiak for his assistance in proofreading.

Kae Lea Main deserves our special thanks for her efficient and error-free typing. Finally we express our indebtedness to Ms. Beatrice Shube of John Wiley & Sons for her efficient editorial guidance.

R. G. LAHA
V. K. ROHATGI

Bowling Green, Ohio
January 1979

Contents

CHAPTER 1

Basic Concepts of Probability Theory

The fundamental concepts of probability theory have roots in measure theory. Like any branch of mathematics, probability theory has its own terminology and its own tools. In this chapter we introduce some of this terminology and study some basic concepts of probability theory. It is assumed that the reader has a working knowledge of measure theory at the level of Halmos [31], and of real analysis at the level of Royden [71]. We will frequently use results from measure theory and real analysis without reference to their source. Results which are of particular importance in probability theory, however, are proved in some detail. These include for example, Theorems 1.1.2 and 1.1.5.

1.1 PROBABILITY SPACES AND RANDOM VARIABLES

1.1.1 Notation and Probability Terminology

We denote by Ω a nonempty set. The elements of Ω will be called *points* and be denoted generically by ω. The following set-theoretic notation will be used:

Points	ω
Sets	capital letters E, F, G, etc.
Union	$E \cup F$, $\bigcup_\alpha E_\alpha$
Intersection	$E \cap F$, $\bigcap_\alpha E_\alpha$
Complement	E^c
Difference	$E - F = E \cap F^c$
Singleton set	$\{\omega\}$
Set inclusion	$E \subset F$ (not excluding $E = F$)

Classes	script capital letters $\mathscr{A}$, $\mathscr{B}$, $\mathscr{S}$, etc.
Inclusion	$\mathscr{A} \subset \mathscr{B}$ (not excluding $\mathscr{A} = \mathscr{B}$)
Belonging to	$\omega \in E$, $E \in \mathscr{S}$
Empty set	$\varnothing$

In the following list we give the correspondences between the probability and measure-theoretic terms which are frequently used:

Sample space	Measurable space
Probability	Normed measure
Probability space	Normed measure space
Elementary event	Singleton set
Event	Measurable set
Sure event	Whole space Ω
Impossible event	Empty set $\varnothing$
Almost sure, almost surely (a.s.) (with probability 1)	Almost everywhere (a.e.)
Random variable	(Finite-valued numerical) measurable function
Expectation	Integral

We summarize below in probability language some results which are specializations of corresponding results in measure theory.

1.1.2 Probability Space

Let Ω be a nonempty set. Let $\mathscr{S}$ be a σ-field of subsets of Ω, that is, a nonempty class of subsets of Ω which contains Ω and is closed under countable union and complementation.

Let P be a measure defined on $\mathscr{S}$ satisfying $P(\Omega) = 1$. Then the triple $(\Omega, \mathscr{S}, P)$ is called a *probability space*, and P, a *probability measure*. The set Ω is the *sure event*, and elements of $\mathscr{S}$ are called *events*. Singleton sets $\{\omega\}$ are called *elementary events*. The symbol $\varnothing$ denotes the empty set and is known as the *null* or *impossible event*. Unless otherwise stated, the probability space $(\Omega, \mathscr{S}, P)$ is fixed, and $A, B, C, \ldots$, with or without subscripts, represent events.

We note that, if $A_n \in \mathscr{S}$, $n = 1, 2, \ldots$, then A_n^c, $\bigcup_{n=1}^{\infty} A_n$, $\bigcap_{n=1}^{\infty} A_n$, $\liminf_{n\to\infty} A_n$, $\limsup_{n\to\infty} A_n$, and $\lim_{n\to\infty} A_n$ (if it exists) are events. Also, the probability measure P is defined on $\mathscr{S}$, and for all events A, A_n

$$P(A) \geq 0, \qquad P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n) (A_n\text{'s disjoint}), \qquad P(\Omega) = 1.$$

It follows that

$$P(\emptyset) = 0, \qquad P(A) \le P(B) \quad \text{for } A \subset B, \qquad P\left(\bigcup_{n=1}^{\infty} A_n\right) \le \sum_{n=1}^{\infty} P(A_n).$$

Moreover,

$$P\left(\liminf_{n\to\infty} A_n\right) \le \liminf_{n\to\infty} P(A_n) \le \limsup_{n\to\infty} P(A_n) \le P\left(\limsup_{n\to\infty} A_n\right),$$

and, if $\lim_{n\to\infty} A_n$ exists, then

$$P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n).$$

The last result is known as the continuity property of probability measures.

Example 1.1.1. Let $\Omega = \{\omega_j : j \ge 1\}$, and let $\mathscr{S}$ be the σ-field of all subsets of Ω. Let $\{p_j, j \ge 1\}$ be any sequence of nonnegative real numbers satisfying $\sum_{j=1}^{\infty} p_j = 1$. Define P on $\mathscr{S}$ by

$$P(E) = \sum_{\omega_j \in E} p_j, \qquad E \in \mathscr{S}.$$

Then P defines a probability measure on $(\Omega, \mathscr{S})$, and $(\Omega, \mathscr{S}, P)$ is a probability space.

Example 1.1.2. Let $\Omega = (0, 1]$ and $\mathscr{S} = \mathscr{B}$ be the σ-field of Borel sets on Ω. Let λ be the Lebesgue measure on $\mathscr{B}$. Then $(\Omega, \mathscr{S}, \lambda)$ is a probability space.

Definition 1.1.1. Let $(\Omega, \mathscr{S}, P)$ be a probability space. A real-valued function X defined on Ω is said to be a random variable if

$$X^{-1}(E) = \{\omega \in \Omega : X(\omega) \in E\} \in \mathscr{S} \qquad \text{for all } E \in \mathscr{B},$$

where $\mathscr{B}$ is the σ-field of Borel sets in $\mathbb{R} = (-\infty, \infty)$; that is, a random variable X is a measurable transformation of $(\Omega, \mathscr{S}, P)$ into $(\mathbb{R}, \mathscr{B})$.

We note that it suffices to require that $X^{-1}(I) \in \mathscr{S}$ for all intervals I in $\mathbb{R}$, or for all semiclosed intervals $I = (a, b]$, or for all intervals $I = (-\infty, b]$, and so on. Unless otherwise specified, $X, Y, \ldots$, with or without subscripts, will represent random variables.

We note that a random variable X defined on $(\Omega, \mathscr{S}, P)$ induces a measure P_X on $\mathscr{B}$ defined by the relation

$$P_X(E) = P\{X^{-1}(E)\} \qquad (E \in \mathscr{B}).$$

Clearly P_X is a probability measure on $\mathscr{B}$ and is called the *probability distribution* or, simply, the *distribution* of X. We note that P_X is a Lebesgue–Stieltjes measure on $\mathscr{B}$.

Definition 1.1.2. For every $x \in \mathbb{R}$ set

$$F_X(x) = P_X(-\infty, x] = P\{\omega \in \Omega: X(\omega) \leq x\}. \tag{1.1.1}$$

We call $F_X = F$ the distribution function of the random variable X.

In the following we write $\{X \leq x\}$ for the event $\{\omega \in \Omega: X(\omega) \leq x\}$. We first prove the following elementary property of a distribution function.

Theorem 1.1.1. The distribution function F of a random variable X is a nondecreasing, right-continuous function on $\mathbb{R}$ which satisfies

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0$$

and

$$F(+\infty) = \lim_{x \to \infty} F(x) = 1.$$

Proof. Note that for every $x \in \mathbb{R}$ and $h > 0$

$$F(x + h) - F(x) = P\{x < X \leq x + h\} \geq 0,$$

so that F is nondecreasing.

Next, let $\{h_n\}$ be a sequence of real numbers such that $0 < h_n \downarrow 0$, as $n \to \infty$. Then, for every $n \geq 1$,

$$F(x + h_n) - F(x) = P\{x < X \leq x + h_n\}.$$

It follows from the continuity property of P that

$$\lim_{n \to \infty} [F(x + h_n) - F(x)] = 0,$$

and hence that F is right-continuous.

Finally, for every $N \geq 1$ we have

$$F(N) - F(-N) = P\{-N < X \leq N\}.$$

Taking the limit of both sides as $N \to \infty$ and using the continuity property once again, we conclude that

$$F(+\infty) - F(-\infty) = 1.$$

But since $0 \leq F(x) \leq 1$ for every $x \in \mathbb{R}$, it follows that $F(-\infty) = 0$ and $F(+\infty) = 1$. ■

Corollary. A distribution function F is continuous at $x \in \mathbb{R}$ if and only if $P\{\omega: X(\omega) = x\} = 0$.

Proof. The proof of the corollary is an immediate consequence of the fact that

$$P\{X = x\} = F(x) - F(x - 0). \tag{1.1.2}$$

Remark 1.1.1. Let X be a random variable, and let g be a Borel-measurable function defined on $\mathbb{R}$. Then $g(X)$ is also a random variable whose distribution is determined by that of X.

We now show that a function F on $\mathbb{R}$ with the properties stated in Theorem 1.1.1 determines uniquely a probability measure P_F on $\mathscr{B}$.

Theorem 1.1.2. Let F be a nondecreasing, right-continuous function defined on $\mathbb{R}$ and satisfying

$$F(-\infty) = 0 \qquad \text{and} \qquad F(+\infty) = 1.$$

Then there exists a probability measure $P = P_F$ on $\mathscr{B}$ determined uniquely by the relation

$$P_F(-\infty, x] = F(x) \qquad \text{for every } x \in \mathbb{R}. \tag{1.1.3}$$

Proof. Let $\mathscr{P}$ be the class of all bounded left-open, right-closed intervals of the form $(a, b]$, $-\infty < a < b < \infty$. Define a set function P_F on $\mathscr{P}$ by the relation

$$P_F(a, b] = F(b) - F(a).$$

We write $P = P_F$ and note that $0 \le P(E) \le 1$ for all $E \in \mathscr{P}$. The proof of Theorem 1.1.2 is based on the following steps.

STEP 1. Let $E_0 \in \mathscr{P}$, and $E_n \in \mathscr{P}$, $n = 1, 2, \ldots$ be a sequence of disjoint sets such that $E_n \subset E_0$ for every n. We show that

$$\sum_{n=1}^{\infty} P(E_n) \le P(E_0) \tag{1.1.4}$$

holds.

Let us first consider the case where the sequence E_n consists of only a finite number of sets, say $E_1, E_2, \ldots, E_N$. Set $E_n = (a_n, b_n]$, $0 \le n \le N$. Without loss of generality, we may assume that $a_1 \le a_2 \le \cdots \le a_N$. Since the E_n are disjoint and $E_n \subset E_0$ for each n, it follows that

$$a_0 \le a_1 < b_1 \le a_2 < b_2 \le \cdots \le a_N < b_N \le b_0.$$

Then

$$\sum_{n=1}^{N} P(E_n) = \sum_{n=1}^{N} [F(b_n) - F(a_n)]$$
$$\leq \sum_{n=1}^{N} [F(b_n) - F(a_n)] + \sum_{n=1}^{N-1} [F(a_{n+1}) - F(b_n)]$$
$$= F(b_N) - F(a_1)$$
$$\leq P(E_0).$$

Next, let the sequence $\{E_n\}$ be countably infinite. Since

$$\sum_{n=1}^{N} P(E_n) \leq P(E_0) \qquad \text{for every } N,$$

letting $N \to \infty$ yields (1.1.4).

STEP 2. Let $K_0 = [a_0, b_0]$, $-\infty < a_0 < b_0 < \infty$, and let $V_n = (a_n, b_n)$, $-\infty < a_n < b_n < \infty, n = 1, 2, \ldots, N$, be such that $K_0 \subset \bigcup_{n=1}^{N} V_n$. Then we show that

$$F(b_0) - F(a_0) \leq \sum_{n=1}^{N} [F(b_n) - F(a_n)]. \tag{1.1.5}$$

Since $K_0 \subset \bigcup_{n=1}^{N} V_n$, there exists an integer k_1, $1 \leq k_1 \leq N$, such that $a_0 \in V_{k_1}$. If $b_0 \in V_{k_1}$, then clearly (1.1.5) holds. Otherwise, $b_{k_1} < b_0$. In this case there exists an integer k_2, $1 \leq k_2 \leq N$, such that $b_{k_1} \in V_{k_2}$. If $b_{k_2} < b_0$, there exists an integer k_3, $1 \leq k_3 \leq N$, such that $b_{k_2} \in V_{k_3}$, and so on. Clearly this process must terminate after a finite number of steps when we have obtained a set V_{k_m} from the sequence $\{V_n\}$ such that $b_0 \in V_{k_m}$. We may assume, without loss of generality, that $m = N$ and that $V_{k_n} = V_n$ for $1 \leq n \leq N$. We have the set of inequalities

$$a_1 < a_0 < b_1, \qquad a_2 < b_1 < b_2, \ldots, a_N < b_{N-1} < b_N. \qquad a_N < b_0 < b_N.$$

Hence

$$F(b_0) - F(a_0) \leq F(b_N) - F(a_1)$$
$$= F(b_1) - F(a_1) + \sum_{j=1}^{N-1} [F(b_{j+1}) - F(b_j)]$$
$$\leq \sum_{n=1}^{N} [F(b_n) - F(a_n)].$$