Introduction to probability theory

KIYOSI ITÔ

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

CAMBRIDGE UNIVERSITY PRESS
Cambridge
London New York New Rochelle
Melbourne Sydney

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
10 Stamford Road, Oakleigh, Melbourne 3166, Australia

Introduction to Probability Theory by Kiyosi Itô
Originally published in Japanese by Iwanami-Shoten, Publishers,
Tokyo, 1978

Translated by Kiyosi Itô

First published in English, 1984
Reprinted 1986

Library of Congress Cataloging in Publication Data

Itô, Kiyosi, 1915–

Introduction to probability theory.

Translation of: Kakuritsuron. 1-4-shō.

Includes index.

1. Probabilities.
QA273.I822513 1984   519.2   83-23187
ISBN 0 521 26418 9 hardcover
ISBN 0 521 26960 1 paperback

Transferred to digital printing 1999

Contents

Preface

Because a point in space can be represented by a triple of real numbers, all geometric properties of spatial figures can be expressed in terms of real numbers. Hence one can theoretically understand geometry solely through analysis. But a true appreciation of geometry requires not only analytical technique but also intuition of geometric objects. The same holds for probability theory. The modern theory of probability is formulated in terms of measures and integrals and so is part of modern analysis from the logical viewpoint. But to really enjoy probability theory, one should grasp the orientation of development of the theory with intuitive insight into random phenomena. The purpose of this book is to explain basic probabilistic concepts rigorously as well as intuitively.

In Chapter 1 we restrict ourselves to trials with a finite number of outcomes. The concepts discussed here are those of elementary probability theory but are dealt with from the advanced standpoint. We hope that the reader appreciates how random phenomena are discussed mathematically without being annoyed with measure-theoretic complications.

In the subsequent chapters we expect the reader to be more or less familiar with basic facts in measure theory.

In Chapter 2 we discuss the properties of those probability measures that appear in this book.

In Chapter 3 we explain the fundamental concepts in probability theory such as events, random variables, independence, conditioning, and so on. We formulate these concepts on a perfect separable complete probability space. The additional conditions "perfectness" and "separability" are imposed to construct the theory in a more natural way. The reader will see that such conditions are satisfied in all problems appearing in applications.

In the standard textbook conditional probability is defined with respect to σ-algebras of subsets of the sample space (Doob's definition). Here we first define it with respect to decompositions of the sample space

(Kolmogorov's definition) to make it easier for the reader to understand its intuitive meaning and then explain Doob's definition and the relation between these two definitions.

In Chapter 4 we discuss the properties of infinite sums of independent real random variables.

At the end of each section several problems are presented with hints for solution to help the reader understand the material in the section.

This book is the English version of the first four chapters of my book *Probability Theory* (in Japanese, Iwanami-Shoten, Tokyo, 1978), based on my course of probability theory at Kyoto University (Japan), at Aarhus University (Denmark), and at Cornell University (U.S.A.). I am grateful to my colleagues and students for their valuable comments. Also I thank David Tranah at Cambridge University Press for his kind cooperation and Mrs. H. Shinohara for her painstaking job of typing.

Kiyosi Itô
Tokyo

Notation and abbreviations

Symbols and terms marked (∗) may not be in universal usage and so should be given special attention.

General

(∗) topological space = T_1-space
(∗) increasing = nondecreasing
(∗) positive-definite = non–negative-definite
(∗) countable = at most countable

Set theory

$\varnothing$ = the empty set
(∗) 2^A = the power set of A
(∗) $\#A$ = the cardinal number of A
A^c = the complement of A
$\cup$ = union
(∗) $+, \Sigma$ = disjoint union
(∗) $-$ = proper difference
(∗) Δ = symmetric difference
$\times, \Pi$ = Cartesian product
(∗) π_k = projection to the kth component space
$f = A \to B, x \to f(x)$ = the map f from A into B carrying x to $f(x)$
$\mathscr{D}(f)$ = the domain of definition of f
$f(C)$ = the image of C under f
$f^{-1}(D)$ = the inverse image of D under f

Analysis

N = the natural numbers
Q = the rational numbers

$\mathbf{R}$ = the real numbers
$\mathbf{C}$ = the complex numbers
$[a, b) = \{x \in \mathbf{R} | a \leq x < b\}$
$(a, b), (a, b], [a, b]$: defined similarly
$\mathbf{R}^n$ = the n-space
(*) $\mathbf{R}^\infty$ = the sequence space
(*) $\binom{n}{r} = {}_nC_r = n!/r!(n-r)!$
$[x] = \max\{n \in N | n \leq x\}$ (Gauss bracket)
$x \vee y = \max(x, y)$
$x \wedge y = \min(x, y)$
$x^+ = x \vee 0$
$x^- = (-x) \vee 0$
Re = real part
Im = imaginary part
ess. sup. = essential supremum
sup = supremum (least upper bound)
inf = infimum (greatest lower bound)
lim sup = limit superior (upper limit)
lim inf = limit inferior (lower limit)
(*) $f(a+) = f(a+0) = \lim_{\varepsilon \downarrow 0} f(a+\delta)$
(*) $f(a-) = f(a-0) = \lim_{\varepsilon \downarrow 0} f(a-\varepsilon)$
(*) $1_A = \chi_A$ = the indicator (or characteristic function) of A
supp = support (of a function)
(*) $f(x)|_{x=a}$ = the value of $f(x)$ evaluated at $x = a$
$f|_A$ = the restriction of f to A
$\mathscr{F}$ = Fourier transform
a.e. = almost everywhere
$\| \, \|_p$ = pth-order norm
$L^p(X, m)$ = the L^p-space over (X, m)
(*) $mf^{-1} = fm$ = the image measure of m by f
(*) f is measurable $\mathscr{B}_1/\mathscr{B}_2 \Leftrightarrow f^{-1}(\mathscr{B}_2) \subset \mathscr{B}_1$

Probability theory

(Ω, P) = the base probability space
$\mathscr{D}(P)$ = the domain of definition of P = the P-measurable sets
a.s. = almost surely
(*) i.o. = infinitely often
(*) f.e. = with a finite number of exceptions
i.p. = in probability

$A, B, C, \ldots$ = classes of subsets of Ω, i.e., subsets of 2^{Ω}
$\mathbf{2} = \{A \in \mathcal{D}(P)\colon P(A) = 0 \text{ or } 1\}$ = the trivial σ-algebra
(*) $\delta[A]$ = the Dynkin class generated by A
(*) $\sigma[A]$ = the σ-algebra generated by A
$\mathcal{B}(S)$ = the Borel system (topological σ-algebra) on S
$\bigwedge_\alpha \mathcal{B}_\alpha$ = the greatest σ-algebra that is included in every $\mathcal{B}_\alpha = \bigcap_\alpha B_\alpha$
$\bigvee_\alpha B_\alpha$ = the least σ-algebra that includes every $\mathcal{B}_\alpha = \sigma[\bigcup_\alpha \mathcal{B}_\alpha]$
$\Delta, \Delta', \ldots$ = decompositions of Ω
(*) $\Delta_{\mathcal{A}}$ = the decomposition generated by A
(*) $X_\Delta(\omega)$ = the element of Δ that contains ω
(*) $\mathcal{B}_\Delta$ = the σ-algebra generated by Δ
$\Delta \succ \Delta'$ = Δ is finer than Δ'
$\bigvee_n \Delta_n$ = the least upper bound of $\{\Delta_n\}$
$\bigwedge_n \Delta_n$ = the greatest lower bound of $\{\Delta_n\}$
$\lambda, \mu, \nu, \ldots$ = n-dimensional distributions $(n = 1, 2, \ldots, \infty)$
$\mathcal{P}^n$ = the n-dimensional distributions
$\mu * \nu$ = the convolution of μ and ν
$\check{\mu}$ = the reflection of μ
$\tilde{\mu} = \check{\mu} * \mu$
C_μ = the continuity points of μ
D_μ = the discontinuity points of μ
$M^p(\mu)$ = the pth-order moment of μ
$|M|^p(\mu)$ = the pth-order absolute moment of μ
$M(\mu) = M^1(\mu)$ = the mean value of μ
$V(\mu)$ = the variance of μ
$\gamma(\mu)$ = the central value of μ
$\delta(\mu)$ = the dispersion of μ
$\varphi_\mu = \mathcal{F}\mu$ = the characteristic function of μ
$N_{m,v}$ = Gauss distribution
$C_{m,c}$ = Cauchy distribution
p_λ = Poisson distribution
δ = delta distribution
$X, Y, Z, \ldots$ = random variables
(*) $P^X = PX^{-1}$ = the probability distribution of X
$E(X)$ = the expectation of $X = M(P^X)$
$V(X)$ = the variance of $X = V(P^X)$
$\sigma(X)$ = the standard deviation of $X = \sigma(P^X)$
(*) $V(X, Y)$ = the covariance of X and Y
$R(X, Y)$ = the correlation coefficient between X and Y

(∗) $\gamma(X)$ = the central value of $X = \gamma(P^X)$
(∗) $\delta(X)$ = the dispersion of $x = \delta(P^X)$
φ_X = the characteristic function of $X = \mathscr{F}P^X$
$M_p(X)$ = the pth-order moment of $X = M_p(P^X)$
$|M|_p(X)$ = the pth order absolute moment of $X = |M|_p(P^X)$
Δ_X = the decomposition generated by X
$\sigma[X]$ = the σ-algebra generated by X
(∗) $\bar{\sigma}[X] = B_{\Delta_X} = X^{-1}(\mathscr{D}(P^X))$ $(\sigma[X] \subset \bar{\sigma}[X] = \subset \sigma[X] \vee \mathbf{2})$
(∗) $P_{\Delta_X}(P_{\mathscr{B}})$ = conditional probability measure under $\Delta(\mathscr{B})$
(∗) $P_{\Delta}(P^X_{\mathscr{B}})$ = conditional probability distribution under $\Delta(\mathscr{B})$
(∗) $E_{\Delta}(E_{\mathscr{B}})$ = conditional expectation under $\Delta(\mathscr{B})$
(∗) $L^0(\Omega, P)$ = the real random variables on (Ω, P)
(∗) $\|X\|_0 = E(|X| \wedge 1)$
(∗) $L^0_{\mathscr{B}}(\Omega, P)$ = the $\mathscr{B}$-measurable random variables on (Ω, P)
(∗) $L^p_{\mathscr{B}}(\Omega, P) = \{X \in L^0_{\mathscr{B}}(\Omega, P) \mid \|X\|_p < \infty\}$