Universitext

Springer-Science+Business Media, LLC

Universitext

Editors (North America): S. Axler, F.W. Gehring, and P.R. Halmos

Aksoy/Khamsi: Nonstandard Methods in Fixed Point Theory
Aupetit: A Primer on Spectral Theory
Booss/Bleecker: Topology and Analysis
Borkar: Probability Theory; An Advanced Course
Carleson/Gamelin: Complex Dynamics
Cecil: Lie Sphere Geometry: With Applications to Submanifolds
Chae: Lebesgue Integration (2nd ed.)
Charlap: Bieberbach Groups and Flat Manifolds
Chern: Complex Manifolds Without Potential Theory
Cohn: A Classical Invitation to Algebraic Numbers and Class Fields
Curtis: Abstract Linear Algebra
Curtis: Matrix Groups
DiBenedetto: Degenerate Parabolic Equations
Dimca: Singularities and Topology of Hypersurfaces
Edwards: A Formal Background to Mathematics I a/b
Edwards: A Formal Background to Mathematics II a/b
Foulds: Graph Theory Applications
Gardiner: A First Course in Group Theory
Gårding/Tambour: Algebra for Computer Science
Goldblatt: Orthogonality and Spacetime Geometry
Hahn: Quadratic Algebras, Clifford Algebras, and Arithmetic Witt Groups
Holmgren: A First Course in Discrete Dynamical Systems
Howe/Tan: Non-Abelian Harmonic Analysis: Applications of $SL(2, R)$
Howes: Modern Analysis and Topology
Humi/Miller: Second Course in Ordinary Differential Equations
Hurwitz/Kritikos: Lectures on Number Theory
Jennings: Modern Geometry with Applications
Jones/Morris/Pearson: Abstract Algebra and Famous Impossibilities
Kelly/Matthews: The Non-Euclidean Hyperbolic Plane
Kostrikin: Introduction to Algebra
Luecking/Rubel: Complex Analysis: A Functional Analysis Approach
MacLane/Moerdijk: Sheaves in Geometry and Logic
Marcus: Number Fields
McCarthy: Introduction to Arithmetical Functions
Meyer: Essential Mathematics for Applied Fields
Mines/Richman/Ruitenburg: A Course in Constructive Algebra
Moise: Introductory Problems Course in Analysis and Topology
Morris: Introduction to Game Theory
Porter/Woods: Extensions and Absolutes of Hausdorff Spaces
Ramsay/Richtmyer: Introduction to Hyperbolic Geometry
Reisel: Elementary Theory of Metric Spaces
Rickart: Natural Function Algebras
Rotman: Galois Theory
Sagan: Space-Filling Curves
Samelson: Notes on Lie Algebras

(continued after index)

Vivek S. Borkar

Probability Theory

An Advanced Course

Springer

Vivek S. Borkar
Department of Electrical Engineering
Indian Institute of Science
Bangalore 560012
India

Mathematics Subject Classification (1991): 60-02

Library of Congress Cataloging-in-Publication Data
Borkar, Vivek S.
Probability theory : an advanced course / Vivek S. Borkar.
p. cm. – (Universitext)
Includes bibliographical references (p. -) and index.
ISBN 978-0-387-94558-3 ISBN 978-1-4612-0791-7 (eBook)
DOI 10.1007/978-1-4612-0791-7
1. Probabilities. I. Title.
QA273.B823 1995
519.2–dc20 95-34388

Printed on acid-free paper.

Production managed by Robert Wexler; manufacturing supervised by Jacqui Ashri.
Photocomposed copy prepared from the author's TeX files.

9 8 7 6 5 4 3 2 1

Preface

This book presents a selection of topics from probability theory. Essentially, the topics chosen are those that are likely to be the most useful to someone planning to pursue research in the modern theory of stochastic processes. The prospective reader is assumed to have good mathematical maturity. In particular, he should have prior exposure to basic probability theory at the level of, say, K.L. Chung's 'Elementary probability theory with stochastic processes' (Springer-Verlag, 1974) and real and functional analysis at the level of Royden's 'Real analysis' (Macmillan, 1968).

The first chapter is a rapid overview of the basics. Each subsequent chapter deals with a separate topic in detail. There is clearly some selection involved and therefore many omissions, but that cannot be helped in a book of this size. The style is deliberately terse to enforce active learning. Thus several tidbits of deduction are left to the reader as labelled exercises in the main text of each chapter. In addition, there are supplementary exercises at the end.

In the preface to his classic text on probability ('Probability', Addison-Wesley, 1968), Leo Breiman speaks of the right and left hands of probability. To quote him: "On the right is the rigorous foundational work using the tools of measure theory. The left hand 'thinks probabilistically', reduces problems to gambling situations, coin-tossing, motions of a physical particle." This is a right-handed book, though a brief "prologue" has been inserted to give an inkling about the missing left hand. An ambidextrous book would be huge and also very difficult to write (at least for the present author). The reader is warned of this shortcoming and is strongly advised to acquire the "left hand" on his own through supplementary reading. Not having it is a major handicap. Also, it's no fun.

Needless to say, I did not invent this subject. Thus barring some novelty of organization and occasional variations on usual proofs, the material is standard. My rendition of it has been heavily influenced by two factors. The first is a beautiful stream of probability courses I took at Berkeley, taught variously by Professors David Blackwell, Michael Klass, Aram Thomasian, J.W. Pitman and David Aldous. (In particular, Chapter VI owes a lot to a course I took from Prof. Aldous.) The second is the several texts I "grew up with" during the formative years as a graduate student. These include, in addition to Breiman's book mentioned above, 'A first course in probability theory' by K.L. Chung (Academic, 1974), 'Probability theory — independence, interchangeability and martingales' by Y.S. Chow and H. Teicher (Springer-Verlag, 1978), 'Discrete parameter martingales', by J. Neveu (North Holland, 1975), 'Convergence of probability measures' by P. Billingsley (Wiley, 1968) and 'Probabilities and potential' by C. Dellacherie and P.A. Meyer (North Holland, 1978). I have also used 'Probability and measure' by P. Billingsley (Wiley, 1979) and 'Introduction to probability and measure' by K.R. Parthasarathy (Macmillan (India), 1977).

Several people have contributed to bringing this book about. A major credit goes to Dr. V.V. Phansalkar who contributed a lot to the "clean up" operations at various stages. Dr. P.G. Babu, Dr. P.S. Sastry and Mr. G. Santharam also helped in several ways. The financial burden was borne by a generous grant from the Curriculum Development Cell of the Centre for Continuing Education, Indian Institute of Science, for which I am extremely grateful.

Writing this book has been my pet project for quite a while. I hope that in its final execution I have done justice to myself and to its potential readers.

This book is dedicated to the memory of my father-in-law, the late Shri Manohar N. Budkule.

Prologue

"The equanimity of your average tosser of coins depends upon the law, or rather a tendency, or let us say a probability, or at any rate a mathematically calculable chance, which ensures that he will not upset himself by losing too much nor upset his opponent by winning too often. This made for a kind of harmony and a kind of confidence. It related the fortuitous and the ordained into a reassuring union which we recognized as nature." — Guildenstern in Tom Stoppard's *Rosencrantz and Guildenstern Are Dead* (Faber and Faber Ltd, London, 1967).

What is probability? It is not easy to answer this question. At the level of gut feeling, one can hardly better the above quote from Stoppard. Trying to go any deeper into the definition of probability will quickly get us into the realms of logic and philosophy. (See, e.g., [24].) We shall evade this issue altogether by taking a phenomenological viewpoint. Take, for instance, the simplest and the most quoted probabilistic phenomenon — the tossing of a coin. Consider the pair of "cause" (tossing a coin) and "effect" (coin drops on the floor) that goes with it. This is as clear an instance of a cause-effect relationship as one would wish to have. But refine the "effect" a little, say, by replacing it by "coin drops with head up" and we are already in trouble. This is only one of the possible effects since the coin could have equally well settled down with tail up. Which of these possibilities will occur? We cannot say a priori. But our intuition suggests that they are equally likely. Of course, another person with a different kind of intuition (or a better knowledge of the coin's composition) could think otherwise. But even after granting that it is a subjective judgement, how do we justify it? One way to do so would be to say that in absence of any extra information, there is no reason to prefer one outcome over the other ("the principle of insufficient reason"). Alternatively, we may take a purely empirical viewpoint, viz., toss the coin several times and verify that both head and tail come up in approximately half the instances. The latter already suggests a way to quantify our judgement — we say that head and tail have equal probability of one half each.

One may say all this to a *random* man on the street and in all *probability,* he won't find anything spurious with this argument. But wait — there is a *chance* that he is a sceptic and may say that the "probability" in the coin-tossing experiment is only apparent. If one finds out exactly all the parameters affecting the coin's motion (its initial coordinates, orientation, direction and magnitude of the thrust, details of the surface it falls upon, etc.) one can exactly predict the outcome. This is certainly a valid objection, but this kind of reasoning applied to probabilistic phenomena has its limitations. There may be limitations to measurement of these parameters, either practical or fundamental (e.g., quantum mechanical or "computational complexity" based). Even when they are not there, it may be easier (quicker, cheaper, etc.) to model and analyze the phenomenon as probabilistic. Without getting into the details of these issues, we shall accept probability as a tried and tested paradigm and tool for modelling and analysis of certain phenomena, viz., those in which a cause can lead to one of many outcomes (for whatever reasons) to each of which a quantitative measure of comparative likelihood can be assigned (in whatever manner). Essentially, the reader hereby is being asked to accept the "gut feeling" we started with in lieu of a definition. Having done so, let us try to mathematicize the concept.

Going back to coin-tossing, consider the set of all possible values of the parameters that determine the coin's motion. We shall call this the "sample space", denoted by Ω. Our act of tossing the coin at a particular time in a particular surrounding in a particular manner picks ("samples") a point from this space. This may either fall into the set $A = \{$points of Ω that lead to head$\}$ or $A^c = \{$points of Ω that lead to tail$\}$. (Note that I am already discounting the "improbable" events such as the coin standing on its edge.) This partitions Ω into two subsets, to each of which our "gut feeling" assigns a probability $\frac{1}{2}$. Next, consider a slightly more complicated situation, say, the rolling of a die. The new sample space Ω now partitions into six subsets $A_i, 1 \leq i \leq 6$, with $A_i = \{$sample points that lead to number $i\}$ for each i. Again, our intuition assigns a probability of $\frac{1}{6}$ to each A_i. Moreover, it makes sense to talk of the probability of getting an even number = the probability of $A_2 \bigcup A_4 \bigcup A_6 = (\frac{1}{6}) + (\frac{1}{6}) + (\frac{1}{6}) = \frac{1}{2}$, and so on.

More generally, one has a set Ω called the sample space and a collection $\mathcal{F}$ of its subsets called events. Elements of Ω are called sample points. To each event A is assigned a number between zero and one, called its probability and denoted by $P(A)$. Now we expect ("gut feeling" again) that $\mathcal{F}$ and $P(.)$ as a map from $\mathcal{F}$ to [0,1] should satisfy certain requirements. For example, "nothing happens" (empty set ϕ) and "something happens" (Ω) should be events. If A is an event, A^c(= "A does not occur") should also be one. If $A_1, A_2, A_3, \cdots$ are events, "at least one of the A_i's occurs" ($\bigcup A_i$) and "all of $\{A_i\}$ occur" ($\bigcap A_i$) should also be events. In short, $\mathcal{F}$ is a σ-field. As for $P(.), P(\Omega) = 1$ by convention. (This can also be rationalized via the "relative frequency of occurrence" interpretation.) Of course, $P(\phi)$ has to be zero and $P(A^c) = 1 - P(A)$. Moreover, if $A_1, A_2, A_3, \cdots$, are disjoint, it makes sense to demand that $P(\bigcup A_i) = \Sigma P(A_i)$. In other words, P is a countably additive nonnegative measure on the measurable space $(\Omega, \mathcal{F})$ with total mass 1. We call such a measure a probability measure.

Now an experiment such as tossing a coin or rolling a die picks a point ω from the sample space and maps it into an element of another space E. (E = {head, tail} and {1,2,3,4,5,6} resp. in the two examples.) Thus it is a map $X : \Omega \to E$. Since our idea is to have sets of the type $\{\omega \mid X(\omega) \in B\}$ to be events (i.e., elements of $\mathcal{F}$) for a suitable collection of subsets $B \subset E$, considerations analogous to the above for $\mathcal{F}$ suggest that we equip E with a σ-field ξ and require the map $X : (\Omega, \mathcal{F}) \to (E, \xi)$ to be measurable. Such a map will be called an (E-valued) random variable. Our mathematical formalism for probability theory is now ready.

Having glibly said all this, let me hasten to add that none of it is obvious. The mathematical formulation of probability theory has a long history and was, in fact, among the major open issues in mathematics early this century. Its eventual settlement via the measure theoretic framework as sketched above is due to Kolmogorov and followed a lot of early work by several others such as Markov and Borel. There were other contenders too, such as the "relative frequency" approach of Von Mises. By now, Kolmogorov's formulation is the most widely accepted one in the mathematics community and we shall stick to it. The foundational issues, however, are by no means dead, one of the most contested issues being the hypothesis of countable additivity. See [43] for a recent debate on these matters. Readers interested in foundational issues should look up [20, 41]. For historical details, see [10, 14, 33].

Note that in reality one observes only a single "realization" of a random variable X, i.e., $X(\omega)$ for a particular $\omega \in \Omega$. Thus the probability space $(\Omega, \mathcal{F}, P)$ in the background is a hypothetical entity and its choice is by no means unique. There is also the problem of choosing P. Some methods thereof are listed below.

(i) *Principle of insufficient reason* — In absence of any reason to favour one outcome over the other, we may deem them equally likely. This can be made a building block for deriving more complicated P. For example, the observable outcome may be a known function of more basic variables to which this principle applies. This function need not be one-one and thus the possible values of the observed outcome need not be equally likely. This is the basis of the Darwin–Fowler approach to statistical mechanics [30, Chapter 5].

(ii) *Subjective probability* — P may simply be a quantification of one's subjective beliefs regarding the relative likelihoods of various events and thus liable to change from person to person. These considerations are important in economic applications [41].

(iii) *Physical reasoning* — Some knowledge of the underlying physical phenomena coupled with simplifying assumptions and probabilistic reasoning can lead to a natural choice of P in some cases. For example, noise in electric circuits is often the cumulative effect of many small and essentially independent phenomena. The central limit theorem (Chapter IV) then suggests that it may be taken to be Gaussian.

(iv) *Worst case analysis* — When a random effect is unwanted (e.g. noise), the "worst case analysis" approach suggests that we take the P that is the worst in an appropriate sense. The "maximum entropy" method in statistics is based on such considerations [26].

(v) *Mathematical simplicity* — One often hypothesizes a P for its mathematical simplicity. The ensuing analysis should in principle be justified by suitable "robustness" results which show that the deductions remain essentially valid even when the P is perturbed.

(vi) *Measurements* — In repeatable random phenomena, one may estimate P from measurements. This is what "Statistics" is all about.

Probability theory has had a symbiotic relationship with several other disciplines and no essay on probability would be complete without at least a mention of these. They are:

(1) Engineering ("noise" in communication engineering, random vibrations in structural engineering),

(2) Operations research and computer science (queuing models, stochastic search algorithms),

(3) Physics (statistical and quantum mechanics, astrophysics),

(4) Biology (genetics, population dynamics),

(5) Economics (econometrics, information economics),

(6) Other branches of mathematics (ergodic theory, partial differential equations),

and so on.

With this we conclude our preamble and move on to the mathematical theory of probability, which begins with $(\Omega, \mathcal{F}, P)$.

Contents