Cambridge University Press
978-0-521-72788-4 - Probability and Information: An Integrated Approach
David Applebaum
Frontmatter
More information

PROBABILITY AND INFORMATION

This is an updated new edition of the popular elementary introduction to probability theory and information theory, now containing additional material on Markov chains and their entropy. Suitable as a textbook for beginning students in mathematics, statistics, computer science or economics, the only prerequisite is some knowledge of basic calculus. A clear and systematic foundation to the subject is provided; the concept of probability is given particular attention via a simplified discussion of measures on Boolean algebras. The theoretical ideas are then applied to practical areas such as statistical inference, random walks, statistical mechanics and communications modelling. Topics discussed include discrete and continuous random variables, entropy and mutual information, maximum entropy methods, the central limit theorem and the coding and transmission of information. Many examples and exercises illustrate how the theory can be applied, for example to information technology. Detailed solutions to most exercises are available on the web.

DAVID APPLEBAUM is a Professor in the Department of Probability and Statistics at the University of Sheffield.

PROBABILITY AND INFORMATION

An Integrated Approach

DAVID APPLEBAUM

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521899048

First published 1996
Second edition 2008

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-0-521-89904-8 hardback
ISBN 978-0-521-72788-4 paperback

To my parents, Sadie and Robert

Cambridge University Press
978-0-521-72788-4 - Probability and Information: An Integrated Approach
David Applebaum
Frontmatter
More information

To live effectively is to live with adequate information.

Norbert Wiener *The Human Use of Human Beings*

The study of probability teaches the student that clear logical thinking is also of use in situations where one is confronted with uncertainty (which is in fact the case in almost every practical situation).

A. Renyi *Remarks on the Teaching of Probability*

Cambridge University Press
978-0-521-72788-4 - Probability and Information: An Integrated Approach
David Applebaum
Frontmatter
More information

Contents

Preface to the second edition

When I wrote the first edition of this book in the early 1990s it was designed as an undergraduate text which gave a unified introduction to the mathematics of 'chance' and 'information'. I am delighted that many courses (mainly in Australasia and the USA) have adopted the book as a core text and have been pleased to receive so much positive feedback from both students and instructors since the book first appeared. For this second edition I have resisted the temptation to expand the existing text and most of the changes to the first nine chapters are corrections of errors and typos. The main new ingredient is the addition of a further chapter (Chapter 10) which brings a third important concept, that of 'time' into play via an introduction to Markov chains and their entropy. The mathematical device for combining time and chance together is called a 'stochastic process' which is playing an increasingly important role in mathematical modelling in such diverse (and important) areas as mathematical finance and climate science. Markov chains form a highly accessible subclass of stochastic (random) processes and nowadays these often appear in first year courses (at least in British universities). From a pedagogic perspective, the early study of Markov chains also gives students an additional insight into the importance of matrices within an applied context and this theme is stressed heavily in the approach presented here, which is based on courses taught at both Nottingham Trent and Sheffield Universities.

I would like to thank all readers (too numerous to mention here) who sent me comments and corrections for the first edition. Special thanks are due to my colleagues – Paul Blackwell who patiently taught me enough S+ for me to be able to carry out the simulations in Chapter 10 and David Grey who did an excellent job on proof-reading the new chapter. Thanks are also due to staff at Cambridge University Press, particularly David Tranah and Peter Thompson for ongoing support and readily available assistance.

David Applebaum
(2007)

Cambridge University Press
978-0-521-72788-4 - Probability and Information: An Integrated Approach
David Applebaum
Frontmatter
More information

Preface to the first edition

This is designed to be an introductory text for a modern course on the fundamentals of probability and information. It has been written to address the needs of undergraduate mathematics students in the 'new' universities and much of it is based on courses developed for the Mathematical Methods for Information Technology degree at the Nottingham Trent University. Bearing in mind that such students do not often have a firm background in traditional mathematics, I have attempted to keep the development of material gently paced and user friendly – at least in the first few chapters. I hope that such an approach will also be of value to mathematics students in 'old' universities, as well as students on courses other than honours mathematics who need to understand probabilistic ideas.

I have tried to address in this volume a number of problems which I perceive in the traditional teaching of these subjects. Many students first meet probability theory as part of an introductory course in statistics. As such, they often encounter the subject as a ragbag of different techniques without the same systematic development that they might gain in a course in, say, group theory. Later on, they might have the opportunity to remedy this by taking a final-year course in rigorous measure theoretic probability, but this, if it exists at all, is likely to be an option only. Consequently, many students can graduate with degrees in mathematical sciences, but without a coherent understanding of the mathematics of probability.

Information sciences have of course seen an enormous expansion of activity over the past three decades and it has become a truism that we live in an 'information rich world'. It is perhaps a little surprising that information theory itself, the mathematical study of information, has continued to be a subject that is not widely available on university mathematics courses and again usually appears, if at all, as a final-year option. This may be because the subject is seen as being conceptually difficult, and it is certainly true that the basic concept of 'entropy' is extremely rich and subtle; nonetheless, bearing in mind that an understanding of the fundamentals

Cambridge University Press
978-0-521-72788-4 - Probability and Information: An Integrated Approach
David Applebaum
Frontmatter
More information

requires only a knowledge of elementary probability theory and familiarity with the manipulation of logarithms, there is no reason why it should not be taught earlier in the undergraduate curriculum.

In this volume, a systematic development of probability and information is presented, much of which would be suitable for a two-semester course (either semesters 1 and 2 or 2 and 3) of an undergraduate degree. This would then provide the background for further courses, both from a pure and applied point of view, in probability and statistics.

I feel that it is natural to view the mathematics of information as part of probability theory. Clearly, probability is needed to make sense of information theoretic concepts. On the other hand, these concepts, as the maximum entropy principle shows (see Chapter 6), can then help us to make 'optimal' probability assignments. It is interesting that this symbiosis between the two subjects was anticipated by two of the greatest probabilists of the twentieth century, as the following two quotations testify:

> There is no doubt that in the years to come the study of entropy will become a permanent part of probability theory.
>
> (A. I. Khinchin: *Mathematical Foundations of Information Theory*)

> Finally, I would like to emphasise that I consider *entropy* and *information* as basic concepts of probability and I strongly recommend that the teacher should spend some time in the discussion of these notions too.
>
> (A. Renyi: *Remarks on The Teaching of Probability*)

Some aspects of the subject which are particularly stressed in this volume are as follows:

(i) There is still a strong debate raging (among philosophers and statisticians, if not mathematicians) about the foundations of probability which is polarised between 'Bayesians' and 'frequentists'. Such philosophical problems are usually ignored in introductory texts, but I believe this sweeps a vital aspect of the subject under the carpet. Indeed, I believe that students' grasp of probability will benefit by their understanding this debate and being given the opportunity to formulate their own opinions. My own approach is to distinguish between the mathematical concept of probability (which is measure theoretic) and its interpretation in practice, which is where I feel the debate has relevance. These ideas are discussed further in Chapters 1 and 4, but for the record I should declare my Bayesian tendencies.

(ii) As well as the 'frequentist/Bayesian' dichotomy mentioned above, another approach to the practical determination of probabilities is the so-called classical theory (or principle of insufficient reason), much exploited by the founders of probability theory, whereby 'equally likely' events are automatically assigned equal probabilities. From a modern point of view, this 'principle of symmetry' finds a natural application in models where random effects arise through the interaction of a large number of identical particles. This is the case in many scientific applications, a paradigm case being statistical mechanics. Furthermore, thanks to the work of E. T. Jaynes, we now have a beautiful and far-reaching generalisation of this idea, namely the principle of maximum entropy, which is described in Chapter 6 and which clearly illustrates how a knowledge of information theory can broaden our understanding of probability.

(iii) The mathematical concept of probability is best formulated, as Kolmogorov taught us, in terms of measures on σ-algebras. Clearly, such an approach is too sophisticated for a book at this level; I have, however, introduced some very simple measure-theoretic concepts within the context of Boolean algebras rather than σ-algebras. This allows us to utilise many of the benefits of a measure-theoretic approach without having to worry about the complexities of σ-additivity. Since most students nowadays study Boolean algebra during their first year within courses on discrete mathematics, the jump to the concept of a measure on a Boolean algebra is not so great. (After revealing myself as a crypto-Bayesian, I should point out that this restriction to finite-additivity is made for purely pedagogical and not idealogical reasons.)

(iv) When we study vector spaces or groups for the first time we become familiar with the idea of the 'basic building blocks' out of which the whole structure can be built. In the case of a vector space, these are the basis elements and, for a group, the generators. Although there is no precise analogy in probability theory, it is important to appreciate the role of the Bernoulli random variables (i.e. those which can take only two possible values) as the 'generators' of many interesting random variables, for example a finite sum of i.i.d. Bernoulli random variables has a binomial distribution, and (depending on how you take the limit) an infinite series can give you a Poisson distribution or a normal distribution.

I have tried to present herein what I see as the 'core' of probability and information. To prevent the book becoming too large, I have postponed the development of some concepts to the exercises (such as convolution of densities and conditional expectations), especially when these are going to have a marginal application in other parts of the book.

Answers to numerical exercises, together with hints and outlines of solutions for some of the more important theoretical exercises, are given at the end of the book. Teachers can obtain fully worked solutions as a LaTeX file, available at http: www.cambridge.org/9780521727884.

Many authors nowadays, in order to be non-sexist, have dropped the traditional 'he' in favour either of the alternative 'he/she' or the ambidextrous (s)he. I intended to use the latter, but for some strange reason (feel free to analyse my unconscious motives) my word processor came out with s(h)e and I perversely decided to adopt it. I apologise if anyone is unintentionally offended by this acronym.

Finally it is a great pleasure to thank my wife, Jill Murray, for all the support she has given me in the writing of this book. I would also like to thank two of my colleagues, John Marriott for many valuable discussions and Barrie Spooner for permission to use his normal distribution tables (originally published in Nottingham Trent University *Statistical Tables*) in Appendix 4. It is a great pleasure also to thank Charles Goldie for his careful reading of part of an early draft and valuable suggestions for improvement. Last but not least my thanks to David Tranah at CUP for his enthusiasm for this project and patient responses to my many enquiries.

D. Applebaum
1995