Springer Texts in Statistics

Advisors:
George Casella Stephen Fienberg Ingram Olkin

F.M. Dekking C. Kraaikamp
H.P. Lopuhaä L.E. Meester

A Modern Introduction to Probability and Statistics

Understanding Why and How

With 120 Figures

Springer

Frederik Michel Dekking
Cornelis Kraaikamp
Hendrik Paul Lopuhaä
Ludolf Erwin Meester
Delft Institute of Applied Mathematics
Delft University of Technology
Mekelweg 4
2628 CD Delft
The Netherlands

Whilst we have made considerable efforts to contact all holders of copyright material contained in this book, we may have failed to locate some of them. Should holders wish to contact the Publisher, we will be happy to come to some arrangement with them.

British Library Cataloguing in Publication Data
A modern introduction to probability and statistics. —
(Springer texts in statistics)
1. Probabilities 2. Mathematical statistics
I. Dekking, F. M.
519.2
ISBN 1852338962

Library of Congress Cataloging-in-Publication Data
A modern introduction to probability and statistics : understanding why and how / F.M. Dekking ... [et al.].
p. cm. — (Springer texts in statistics)
Includes bibliographical references and index.
ISBN 1-85233-896-2
1. Probabilities—Textbooks. 2. Mathematical statistics—Textbooks. I. Dekking, F.M. II. Series.
QA273.M645 2005
519.2—dc22 2004057700

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

ISBN-10: 1-85233-896-2
ISBN-13: 978-1-85233-896-1

Springer Science+Business Media
springeronline.com

© Springer-Verlag London Limited 2005

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant laws and regulations and therefore free for general use.

The publisher makes no representation, express or implied, with regard to the accuracy of the information contained in this book and cannot accept any legal responsibility or liability for any errors or omissions that may be made.

Printed in the United States of America
12/3830/543210 Printed on acid-free paper SPIN 10943403

Preface

Probability and statistics are fascinating subjects on the interface between mathematics and applied sciences that help us understand and solve practical problems. We believe that you, by learning how stochastic methods come about and why they work, will be able to understand the meaning of statistical statements as well as judge the quality of their content, when facing such problems on your own. Our philosophy is one of *how* and *why*: instead of just presenting stochastic methods as cookbook recipes, we prefer to explain the principles behind them.

In this book you will find the basics of probability theory and statistics. In addition, there are several topics that go somewhat beyond the basics but that ought to be present in an introductory course: simulation, the Poisson process, the law of large numbers, and the central limit theorem. Computers have brought many changes in statistics. In particular, the bootstrap has earned its place. It provides the possibility to derive confidence intervals and perform tests of hypotheses where traditional (normal approximation or large sample) methods are inappropriate. It is a modern useful tool one should learn about, we believe.

Examples and datasets in this book are mostly from real-life situations, at least that is what we looked for in illustrations of the material. Anybody who has inspected datasets with the purpose of using them as elementary examples knows that this is hard: on the one hand, you do not want to boldly state assumptions that are clearly not satisfied; on the other hand, long explanations concerning side issues distract from the main points. We hope that we found a good middle way.

A first course in calculus is needed as a prerequisite for this book. In addition to high-school algebra, some infinite series are used (exponential, geometric). Integration and differentiation are the most important skills, mainly concerning one variable (the exceptions, two dimensional integrals, are encountered in Chapters 9–11). Although the mathematics is kept to a minimum, we strived

to be mathematically correct throughout the book. With respect to probability and statistics the book is self-contained.

The book is aimed at undergraduate engineering students, and students from more business-oriented studies (who may gloss over some of the more mathematically oriented parts). At our own university we also use it for students in applied mathematics (where we put a little more emphasis on the math and add topics like combinatorics, conditional expectations, and generating functions). It is designed for a one-semester course: on average two hours in class per chapter, the first for a lecture, the second doing exercises. The material is also well-suited for self-study, as we know from experience.

We have divided attention about evenly between probability and statistics. The very first chapter is a sampler with differently flavored introductory examples, ranging from scientific success stories to a controversial puzzle. Topics that follow are elementary probability theory, simulation, joint distributions, the law of large numbers, the central limit theorem, statistical modeling (informal: why and how we can draw inference from data), data analysis, the bootstrap, estimation, simple linear regression, confidence intervals, and hypothesis testing. Instead of a few chapters with a long list of discrete and continuous distributions, with an enumeration of the important attributes of each, we introduce a few distributions when presenting the concepts and the others where they arise (more) naturally. A list of distributions and their characteristics is found in Appendix A.

With the exception of the first one, chapters in this book consist of three main parts. First, about four sections discussing new material, interspersed with a handful of so-called Quick exercises. Working these—two-or-three-minute—exercises should help to master the material and provide a break from reading to do something more active. On about two dozen occasions you will find indented paragraphs labeled *Remark*, where we felt the need to discuss more mathematical details or background material. These remarks can be skipped without loss of continuity; in most cases they require a bit more mathematical maturity. Whenever persons are introduced in examples we have determined their sex by looking at the chapter number and applying the rule "He is odd, she is even." Solutions to the quick exercises are found in the second to last section of each chapter.

The last section of each chapter is devoted to exercises, on average thirteen per chapter. For about half of the exercises, answers are given in Appendix C, and for half of these, full solutions in Appendix D. Exercises with both a short answer and a full solution are marked with ⊞ and those with only a short answer are marked with ⊡ (when more appropriate, for example, in "Show that ..." exercises, the short answer provides a hint to the key step). Typically, the section starts with some easy exercises and the order of the material in the chapter is more or less respected. More challenging exercises are found at the end.

Much of the material in this book would benefit from illustration with a computer using statistical software. A complete course should also involve computer exercises. Topics like simulation, the law of large numbers, the central limit theorem, and the bootstrap loudly call for this kind of experience. For this purpose, all the datasets discussed in the book are available at http://www.springeronline.com/1-85233-896-2. The same Web site also provides access, for instructors, to a complete set of solutions to the exercises; go to the Springer online catalog or contact textbooks@springer-sbm.com to apply for your password.

Delft, The Netherlands
January 2005

F. M. Dekking
C. Kraaikamp
H. P. Lopuhaä
L. E. Meester

Contents