Springer Texts in Statistics

Kenneth Lange

Applied Probability

Second Edition

Springer

Kenneth Lange

Applied Probability

Second Edition

常州大学山书馆
藏书章

Springer

Prof. Kenneth Lange
University of California, Los Angeles
Departments of Biomathematics,
Human Genetics, and Statistics
90095-1766 Los Angeles
California
USA
klange@ucla.edu

STS Editorial Board
George Casella
Department of Statistics
University of Florida
Gainesville, FL 32611-8545
USA

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburg, PA 15213-3890
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

ISSN 1431-875X
ISBN 978-1-4419-7164-7 e-ISBN 978-1-4419-7165-4
DOI 10.1007/978-1-4419-7165-4
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010933109

© Springer Science+Business Media, LLC 2003, 2010
All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media, LLC, 233 Spring Street, New York, NY 10013, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks, and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Springer Texts in Statistics

Series Editors:
G. Casella
S. Fienberg
I. Olkin

For other titles published in this series, go to
http://www.springer.com/series/417

Preface to the Second Edition

My original intent in writing *Applied Probability* was to strike a balance between theory and applications. Theory divorced from applications runs the risk of alienating many potential practitioners of the art of stochastic modeling. Applications without a clear statement of relevant theory drift in a sea of confusion. To a lesser degree I was also motivated by a desire to promote the nascent field of computational probability. Current students of the mathematical sciences are more computer savvy than ever. Putting the right computational tools in their hands is bound to advance probability and the broader good of science.

The second edition of *Applied Probability* remains true to these aims. I have added two new chapters on asymptotic and numerical methods and an appendix that separates some of the more delicate mathematical theory from the steady flow of examples in the main text. In addition to these major changes, there is now a much more extensive list of exercises. Some of these are trivial, but others will challenge even the best students. Finally, many errors, both large and small, have been corrected.

Chapter 4 on combinatorics includes new sections on bijections, Catalan numbers, and Faà di Bruno's formula. The proof of the inclusion-exclusion formula has been clarified. Chapter 7 on Markov chains contains new material on rates of convergence to equilibrium in reversible finite-state chains. This discussion draws on students' previous exposure to eigenvalues and eigenvectors in linear algebra. Chapter 9 on branching processes features a new section on basic reproduction numbers. Here the idea is to devise easy algebraic tests for deciding when a process is subcritical, critical, or supercritical. Chapter 11 on diffusion processes gives better coverage of

此为试读,需要完整PDF请访问: www.ertongbook.com

Brownian motion. The last two sections of the chapter have been moved to the new Chapter 13 on numerical methods. The orphan material on convergent sequences of random variables in Chapter 1 has been moved to the new Chapter 12 on asymptotic methods.

Once again I would like to thank the students of my UCLA biomathematics classes for their help. Particularly noteworthy are David Alexander, Kristin Ayers, Forrest Crawford, Kate Crespi, Gabriela Cybis, Lewis Lee, Sarah Nowak, John Ranola, Mary Sehl, Tongtong Wu, and Jin Zhou. I owe an especially heavy debt to Hua Zhou, my former postdoctoral fellow, for suggesting many problems and lecturing in my absence. I also thank my editor, John Kimmel, for his kind support. Finally, I am glad to report that my mother, to whom both editions of this book are dedicated, is alive and well. If I can spread even a fraction of the cheer she has spread, then I will be able to look back over a life well lived.

Preface

Despite the fears of university mathematics departments, mathematics education is growing rather than declining. But the truth of the matter is that the increases are occurring outside departments of mathematics. Engineers, computer scientists, physicists, chemists, economists, statisticians, biologists, and even philosophers teach and learn a great deal of mathematics. The teaching is not always terribly rigorous, but it tends to be better motivated and better adapted to the needs of students. In my own experience teaching students of biostatistics and mathematical biology, I attempt to convey both the beauty and utility of probability. This is a tall order, partially because probability theory has its own vocabulary and habits of thought. The axiomatic presentation of advanced probability typically proceeds via measure theory. This approach has the advantage of rigor, but it inevitably misses most of the interesting applications, and many applied scientists rebel against the onslaught of technicalities. In the current book, I endeavor to achieve a balance between theory and applications in a rather short compass. While the combination of brevity and balance sacrifices many of the proofs of a rigorous course, it is still consistent with supplying students with many of the relevant theoretical tools. In my opinion, it is better to present the mathematical facts without proof rather than omit them altogether.

In the preface to his lovely recent textbook [209], David Williams writes, "Probability and Statistics used to be married; then they separated; then they got divorced; now they hardly see each other." Although this split is doubtless irreversible, at least we ought to be concerned with properly bringing up their children, applied probability and computational statistics.

If we fail, then science as a whole will suffer. You see before you my attempt to give applied probability the attention it deserves. My other recent book [122] covers computational statistics and aspects of computational probability glossed over here.

This graduate-level textbook presupposes knowledge of multivariate calculus, linear algebra, and ordinary differential equations. In probability theory, students should be comfortable with elementary combinatorics, generating functions, probability densities and distributions, expectations, and conditioning arguments. My intended audience includes graduate students in applied mathematics, biostatistics, computational biology, computer science, physics, and statistics. Because of the diversity of needs, instructors are encouraged to exercise their own judgment in deciding what chapters and topics to cover.

Chapter 1 reviews elementary probability while striving to give a brief survey of relevant results from measure theory. Poorly prepared students should supplement this material with outside reading. Well-prepared students can skim Chapter 1 until they reach the less well-known material of the final two sections. Section 1.8 develops properties of the multivariate normal distribution of special interest to students in biostatistics and statistics. This material is applied to optimization theory in Section 3.3 and to diffusion processes in Chapter 11.

We get down to serious business in Chapter 2, which is an extended essay on calculating expectations. Students often complain that probability is nothing more than a bag of tricks. For better or worse, they are confronted here with some of those tricks. Readers may want to skip the final two sections of the chapter on surface area distributions on a first pass through the book.

Chapter 3 touches on advanced topics from convexity, inequalities, and optimization. Besides the obvious applications to computational statistics, part of the motivation for this material is its applicability in calculating bounds on probabilities and moments.

Combinatorics has the odd reputation of being difficult in spite of relying on elementary methods. Chapters 4 and 5 are my stab at making the subject accessible and interesting. There is no doubt in my mind of combinatorics' practical importance. More and more we live in a world dominated by discrete bits of information. The stress on algorithms in Chapter 5 is intended to appeal to computer scientists.

Chapters 6 through 11 cover core material on stochastic processes that I have taught to students in mathematical biology over a span of many years. If supplemented with appropriate sections from Chapters 1 and 2, there is sufficient material here for a traditional semester-long course in stochastic processes. Although my examples are weighted toward biology, particularly genetics, I have tried to achieve variety. The fortunes of this book doubtless will hinge on how compelling readers find these examples.

You can leaf through the table of contents to get a better idea of the topics covered in these chapters.

In the final two chapters, on Poisson approximation and number theory, the applications of probability to other branches of mathematics come to the fore. These chapters are hardly in the mainstream of stochastic processes and are meant for independent reading as much as for classroom presentation.

All chapters come with exercises. (In this second printing, some additional exercises are included at the end of the book.) These are not graded by difficulty, but hints are provided for some of the more difficult ones. My own practice is to require one problem for each hour and a half of lecture. Students are allowed to choose among the problems within each chapter and are graded on the best of the solutions they present. This strategy provides an incentive for the students to attempt more than the minimum number of problems.

I would like to thank my former and current UCLA and University of Michigan students for their help in debugging this text. In retrospect, there were far more contributing students than I can possibly credit. At the risk of offending the many, let me single out Brian Dolan, Ruzong Fan, David Hunter, Wei-hsun Liao, Ben Redelings, Eric Schadt, Marc Suchard, Janet Sinsheimer, and Andy Ming-Ham Yip. I also thank John Kimmel of Springer–Verlag for his editorial assistance.

Finally, I dedicate this book to my mother, Alma Lange, on the occasion of her 80th birthday. Thanks, Mom, for your cheerfulness and generosity in raising me. You were, and always will be, an inspiration to the whole family.

Contents

Preface to the Second Edition v

Preface vii

1 Basic Notions of Probability Theory 1
1.1 Introduction . . . 1
1.2 Probability and Expectation . . . 1
1.3 Conditional Probability . . . 6
1.4 Independence . . . 8
1.5 Distributions, Densities, and Moments . . . 9
1.6 Convolution . . . 13
1.7 Random Vectors . . . 14
1.8 Multivariate Normal Random Vectors . . . 17
1.9 Problems . . . 20

2 Calculation of Expectations 25
2.1 Introduction . . . 25
2.2 Indicator Random Variables and Symmetry . . . 25
2.3 Conditioning . . . 29
2.4 Moment Transforms . . . 31
2.5 Tail Probability Methods . . . 36
2.6 Moments of Reciprocals and Ratios . . . 38
2.7 Reduction of Degree . . . 40
2.8 Spherical Surface Measure . . . 42

2.9 Dirichlet Distribution 44
2.10 Problems 46

3 Convexity, Optimization, and Inequalities **55**
3.1 Introduction 55
3.2 Convex Functions 56
3.3 Minimization of Convex Functions 61
3.4 The MM Algorithm 63
3.5 Moment Inequalities 66
3.6 Problems 70

4 Combinatorics **75**
4.1 Introduction 75
4.2 Bijections 75
4.3 Inclusion-Exclusion 78
4.4 Applications to Order Statistics 83
4.5 Catalan Numbers 84
4.6 Stirling Numbers 86
4.7 Application to an Urn Model 89
4.8 Application to Faà di Bruno's Formula 91
4.9 Pigeonhole Principle 93
4.10 Problems 94

5 Combinatorial Optimization **103**
5.1 Introduction 103
5.2 Quick Sort 104
5.3 Data Compression and Huffman Coding 106
5.4 Graph Coloring 108
5.5 Point Sets with Only Acute Angles 112
5.6 Sperner's Theorem 113
5.7 Subadditivity and Expectations 114
5.8 Problems 118

6 Poisson Processes **123**
6.1 Introduction 123
6.2 The Poisson Distribution 124
6.3 Characterization and Construction 124
6.4 One-Dimensional Processes 127
6.5 Transmission Tomography 131
6.6 Mathematical Applications 134
6.7 Transformations 136
6.8 Marking and Coloring 138
6.9 Campbell's Moment Formulas 139
6.10 Problems 142

此为试读,需要完整PDF请访问: www.ertongbook.com

7 Discrete-Time Markov Chains **151**
7.1 Introduction . . . 151
7.2 Definitions and Elementary Theory . . . 151
7.3 Examples . . . 155
7.4 Coupling . . . 158
7.5 Convergence Rates for Reversible Chains . . . 163
7.6 Hitting Probabilities and Hitting Times . . . 165
7.7 Markov Chain Monte Carlo . . . 168
7.7.1 The Hastings-Metropolis Algorithm . . . 168
7.7.2 Gibbs Sampling . . . 170
7.7.3 Convergence of the Independence Sampler . . . 171
7.8 Simulated Annealing . . . 173
7.9 Problems . . . 174

8 Continuous-Time Markov Chains **187**
8.1 Introduction . . . 187
8.2 Finite-Time Transition Probabilities . . . 187
8.3 Derivation of the Backward Equations . . . 189
8.4 Equilibrium Distributions and Reversibility . . . 190
8.5 Examples . . . 193
8.6 Calculation of Matrix Exponentials . . . 197
8.7 Kendall's Birth-Death-Immigration Process . . . 200
8.8 Solution of Kendall's Equation . . . 203
8.9 Problems . . . 206

9 Branching Processes **217**
9.1 Introduction . . . 217
9.2 Examples of Branching Processes . . . 218
9.3 Elementary Theory . . . 219
9.4 Extinction . . . 221
9.5 Immigration . . . 225
9.6 Multitype Branching Processes . . . 229
9.7 Viral Reproduction in HIV . . . 231
9.8 Basic Reproduction Numbers . . . 232
9.9 Problems . . . 235

10 Martingales **247**
10.1 Introduction . . . 247
10.2 Definition and Examples . . . 247
10.3 Martingale Convergence . . . 251
10.4 Optional Stopping . . . 255
10.5 Large Deviation Bounds . . . 260
10.6 Problems . . . 264

11 Diffusion Processes **269**
11.1 Introduction 269
11.2 Basic Definitions and Properties 270
11.3 Examples Involving Brownian Motion 272
11.4 Other Examples of Diffusion Processes 276
11.5 Process Moments 280
11.6 First Passage Problems 282
11.7 The Reflection Principle 287
11.8 Equilibrium Distributions 289
11.9 Problems 291

12 Asymptotic Methods **297**
12.1 Introduction 297
12.2 Asymptotic Expansions 298
12.2.1 Order Relations 298
12.2.2 Finite Taylor Expansions 299
12.2.3 Exploitation of Nearby Exact Results 301
12.2.4 Expansions via Integration by Parts 302
12.3 Laplace's Method 304
12.3.1 Stirling's Formula 306
12.3.2 Watson's Lemma 307
12.4 Euler-Maclaurin Summation Formula 308
12.5 Asymptotics and Generating Functions 311
12.6 Stochastic Forms of Convergence 314
12.7 Problems 318

13 Numerical Methods **327**
13.1 Introduction 327
13.2 Computation of Equilibrium Distributions 328
13.3 Applications of the Finite Fourier Transform 331
13.4 Counting Jumps in a Markov Chain 336
13.5 Stochastic Simulation and Intensity Leaping 339
13.6 A Numerical Method for Diffusion Processes 343
13.7 Application to the Wright-Fisher Process 347
13.8 Problems 350

14 Poisson Approximation **355**
14.1 Introduction 355
14.2 Applications of the Coupling Method 356
14.3 Applications of the Neighborhood Method 360
14.4 Proof of the Chen-Stein Estimates 363
14.5 Problems 368

15 Number Theory **373**
15.1 Introduction 373

15.2 Zipf's Distribution and Euler's Theorem 374
15.3 Dirichlet Products and Möbius Inversion 378
15.4 Averages of Arithmetic Functions 382
15.5 The Prime Number Theorem 386
15.6 Problems 391

Appendix: Mathematical Review **395**
A.1 Elementary Number Theory 395
A.2 Nonnegative Matrices 397
A.3 The Finite Fourier Transform 401
A.4 The Fourier Transform 403
A.5 Fourier Series 406
A.6 Laplace's Method and Watson's Lemma 410
A.7 A Tauberian Theorem 412

References **415**

Index **429**

1
Basic Notions of Probability Theory

1.1 Introduction

This initial chapter covers background material that every serious student of applied probability should master. In no sense is the chapter meant as a substitute for a previous course in applied probability or for a future course in measure-theoretic probability. Our comments are merely meant as reminders and as a bridge. Many mathematical facts will be stated without proof. This is unsatisfactory, but it is even more unsatisfactory to deny students the most powerful tools in the probabilist's toolkit. Quite apart from specific tools, the language and intellectual perspective of modern probability theory also furnish an intuitive setting for solving practical problems. Probability involves modes of thought that are unique within mathematics. As a brief illustration of the material reviewed, we derive properties of the multivariate normal distribution in the final section of this chapter. Later chapters will build on the facts and vocabulary mentioned here and provide more elaborate applications.

1.2 Probability and Expectation

The layman's definition of probability is the long-run frequency of success over a sequence of independent, identically constructed trials. Although this law of large numbers perspective is important, mathematicians have found it helpful to put probability theory on an axiomatic basis [24, 53, 60, 80,

K. Lange, *Applied Probability*,
DOI 10.1007/978-1-4419-7165-4_1, © Springer Science+Business Media, LLC 2010

166, 171, 208]. The modern theory begins with the notion of a sample space Ω and a collection $\mathcal{F}$ of subsets from Ω subject to the following conventions:

(1.2a) $\Omega \in \mathcal{F}$.

(1.2b) If $A \in \mathcal{F}$, then its complement $A^c \in \mathcal{F}$.

(1.2c) If $A_1, A_2, \ldots$ is a finite or countably infinite sequence of subsets from $\mathcal{F}$, then $\bigcup_i A_i \in \mathcal{F}$.

Any collection $\mathcal{F}$ satisfying these postulates is termed a σ-field or σ-algebra. Two immediate consequences of the definitions are that the empty set $\emptyset \in \mathcal{F}$ and that if $A_1, A_2, \ldots$ is a finite or countably infinite sequence of subsets from $\mathcal{F}$, then $\bigcap_i A_i = (\bigcup_i A_i^c)^c \in \mathcal{F}$. In probability theory, we usually substitute everyday language for set theory language. Table 1.1 provides a short dictionary for translating equivalent terms.

TABLE 1.1. A Brief Dictionary of Set Theory and Probability Terms

Set Theory	Probability	Set Theory	Probability
set	event	null set	impossible event
union	or	universal set	certain event
intersection	and	pairwise disjoint	mutually exclusive
complement	not	inclusion	implication

The axiomatic setting of probability theory is completed by introducing a probability measure or distribution Pr on the events in $\mathcal{F}$. This function should satisfy the properties:

(1.2d) $\Pr(\Omega) = 1$.

(1.2e) $\Pr(A) \geq 0$ for any $A \in \mathcal{F}$.

(1.2f) $\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i)$ for any countably infinite sequence of mutually exclusive events $A_1, A_2, \ldots$ from $\mathcal{F}$.

A triple $(\Omega, \mathcal{F}, \Pr)$ constitutes a probability space. An event $A \in \mathcal{F}$ is said to be null when $\Pr(A) = 0$ and almost sure when $\Pr(A) = 1$.

Example 1.2.1 *Discrete Uniform Distribution*

One particularly simple sample space is the set $\Omega = \{1, \ldots, n\}$. Here the natural choice of $\mathcal{F}$ is the collection of all subsets of Ω. The uniform distribution (or normalized counting measure) attributes probability $\Pr(A) = \frac{|A|}{n}$ to a set A, where $|A|$ denotes the number of elements of A. Most of the counting arguments of combinatorics presuppose the discrete uniform distribution. ■

Example 1.2.2 *Continuous Uniform Distribution*

A continuous analog of the discrete uniform distribution is furnished by Lebesgue measure on the unit interval $[0,1]$. In this case, the best one can do is define $\mathcal{F}$ as the smallest σ-algebra containing all closed subintervals $[a,b]$ of $\Omega = [0,1]$. The events in $\mathcal{F}$ are then said to be Borel sets. Henri Lebesgue was able to show how to extend the primitive identification $\Pr([a,b]) = b - a$ of the probability of an interval with its length to all Borel sets [171]. Invoking the axiom of choice from set theory, one can prove that it is impossible to attach a probability consistently to all subsets of $[0,1]$. The existence of nonmeasurable sets makes the whole enterprise of measure-theoretic probability more delicate than mathematicians anticipated. Fortunately, one can ignore such subtleties in most practical problems. ■

The next example is designed to give readers a hint of the complexities involved in defining probability spaces.

Example 1.2.3 *Density in Number Theory*

Consider the natural numbers $\Omega = \{1, 2, \ldots\}$ equipped with the density function

$$\text{den}(A) \;=\; \lim_{n\to\infty} \frac{|A \cap \{1, 2, \ldots, n\}|}{n}.$$

Clearly, $0 \le \text{den}(A) \le 1$ whenever $\text{den}(A)$ is defined. Some typical densities include $\text{den}(\Omega) = 1$, $\text{den}(\{j\}) = 0$, and $\text{den}(\{j, 2j, 3j, 4j, \ldots\}) = 1/j$. Any σ-algebra $\mathcal{F}$ containing each of the positive integers $\{j\}$ fails the test of countable additivity stated in postulate (1.2f) above. Indeed,

$$\text{den}(\Omega) \;\neq\; 0 \;=\; \sum_{j=1}^{\infty} \text{den}(\{j\}).$$

Note that $\text{den}(A)$ does satisfy the test of finite additivity. Of course, it is possible to define many legitimate probability distributions on the positive integers. ■

In practice, most questions in probability theory revolve around random variables rather than sample spaces. Readers will doubtless recall that a random variable X is a function from a sample space Ω into the real line R. This is almost correct. To construct a consistent theory of integration, one must insist that a random variable be measurable. This technical condition requires that for every constant c, the set $\{\omega \in \Omega : X(\omega) \le c\}$ be an event in the σ-algebra $\mathcal{F}$ attached to Ω. Measurability can also be defined in terms of the Borel sets $\mathcal{B}$ of R, which comprise the smallest σ-algebra containing all intervals $[a,b]$ of R. With this definition in mind, X is measurable if and only if the inverse image $X^{-1}(B)$ of every Borel set B is an event in $\mathcal{F}$. This

此为试读,需要完整PDF请访问: www.ertongbook.com