Numerical Methods of Statistics
Second Edition

This book explains how computer software is designed to perform the tasks required for sophisticated statistical analysis. For statisticians, it examines the nitty-gritty computational problems behind statistical methods. For mathematicians and computer scientists, it looks at the application of mathematical tools to statistical problems. The first half of the book offers a basic background in numerical analysis that emphasizes issues important to statisticians. The next several chapters cover a broad array of statistical tools, such as maximum likelihood and nonlinear regression. The author also treats the application of numerical tools; numerical integration and random number generation are explained in a unified manner reflecting complementary views of Monte Carlo methods. Each chapter contains exercises that range from simple questions to research problems. Most of the examples are accompanied by demonstration and source code available on the author's Web site. New in this second edition are demonstrations coded in R, as well as new sections on linear programming and the Nelder-Mead search algorithm.

John F. Monahan is a Professor of Statistics at North Carolina State University, where he joined the faculty in 1978 and has been a professor since 1990. His research has appeared in numerous computational as well as statistical journals. He is also the author of *A Primer on Linear Models* (2008).

CAMBRIDGE SERIES IN STATISTICAL AND PROBABILISTIC MATHEMATICS

This series of high-quality upper-division textbooks and expository monographs covers all aspects of stochastic applicable mathematics. The topics range from pure and applied statistics to probability theory, operations research, optimization, and mathematical programming. The books contain clear presentations of new developments in the field and also of the state of the art in classical methods. While emphasizing rigorous treatment of theoretical methods, the books also contain applications and discussions of new techniques made possible by advances in computational practice.

1. *Bootstrap Methods and Their Application,* by A. C. Davison and D. V. Hinkley
2. *Markov Chains,* by J. Norris
3. *Asymptotic Statistics,* by A. W. van der Vaart
4. *Wavelet Methods for Time Series Analysis,* by Donald B. Percival and Andrew T. Walden
5. *Bayesian Methods,* by Thomas Leonard and John S. J. Hsu
6. *Empirical Processes in M-Estimation,* by Sara van de Geer
7. *Numerical Methods of Statistics,* by John F. Monahan
8. *A User's Guide to Measure Theoretic Probability,* by David Pollard
9. *The Estimation and Tracking of Frequency,* by B. G. Quinn and E. J. Hannan
10. *Data Analysis and Graphics Using R,* by John Maindonald and John Braun
11. *Statistical Models,* by A. C. Davison
12. *Semiparametric Regression,* by D. Ruppert, M. P. Wand, and R. J. Carroll
13. *Exercise in Probability,* by Loic Chaumont and Marc Yor
14. *Statistical Analysis of Stochastic Processes in Time,* by J. K. Lindsey
15. *Measure Theory and Filtering,* by Lakhdar Aggoun and Robert Elliott
16. *Essentials of Statistical Inference,* by G. A. Young and R. L. Smith
17. *Elements of Distribution Theory,* by Thomas A. Severini
18. *Statistical Mechanics of Disordered Systems,* by Anton Bovier
19. *The Coordinate-Free Approach to Linear Models,* by Michael J. Wichura
20. *Random Graph Dynamics,* by Rick Durrett
21. *Networks,* by Peter Whittle
22. *Saddlepoint Approximations with Applications,* by Ronald W. Butler
23. *Applied Asymptotics,* by A. R. Brazzale, A. C. Davison, and N. Reid
24. *Random Networks for Communication,* by Massimo Franceschetti and Ronald Meester
25. *Design of Comparative Experiments,* by R. A. Bailey
26. *Symmetry Studies,* by Marlos A. G. Viana
27. *Model Selection and Model Averaging,* by Gerda Claeskens and Nils Lid Hjort
28. *Bayesian Nonparametrics,* by Nils Lid Hjort, Peter Müller, Stephen G. Walker
29. *From Finite Sample to Asymptotic Methods in Statistics,* by Pranab K. Sen, Julio M. Singer, and Antonio C. Pedroso de Lima
30. *Brownian Motion,* by Peter Mörters and Yuval Peres
31. *Probability: Theory and Examples,* fourth edition, by Rick Durrett

Numerical Methods of Statistics

Second Edition

JOHN F. MONAHAN
North Carolina State University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314-321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi - 110025, India

103 Penang Road, #05-06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521139519

First published 2001
Second edition published 2011

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

Library of Congress Cataloging in Publication data
Monahan, John F.
Numerical methods of statistics / John F. Monahan.
p. cm. – (Cambridge series in statistical and probabilistic mathematics)
Includes bibliographical references.
ISBN 978-0-521-19158-6
1. Mathematical statistics – Data processing. 2. Numerical analysis. I. Title.
II. Cambridge series in statistical and probabilistic mathematics.
QA276.4.M65 2001
519.5–dc21 00-031269

ISBN 978-0-521-19158-6 Hardback
ISBN 978-0-521-13951-9 Paperback

Contents

Preface to the Second Edition

In the ten years since the first edition of this book went to press, the field of statistical computing has exploded with innovations in many directions. At one time my goal was to write a comprehensive book on the subject. At this moment, however, my goals for a second edition must be more modest. Because the field has grown so much, the scope of this book has now become the core for a subset of this field. To fill in some gaps in this new core, a few sections have been added (e.g., linear programming) and others have been expanded. Many corrections have been made; I can only hope that just a few errors remain.

A second change in this timespan is the rapid widespread adoption of R in the field of statistics. As language and culture shape each other, my own views on computing have changed from teaching this material using R. Small changes scattered throughout reflect this change in viewpoint. Additionally, most of the demonstrations and examples – all that seemed appropriate – have been translated to R and are available on my Web site for this book (http://www4.stat.ncsu.edu/~monahan/nmos2/toc.html).

Thanks are due to Lauren Cowles of Cambridge University Press for encouraging this second edition. Karen Chiswell deserves recognition for finding numerous typos and providing other corrections. I would like to also thank Jerry Davis and Wendy Meiring for pointing out others. Bruce McCullough provided invaluable feedback, comments, questions, and suggestions. Thanks are also due to the many students who, perhaps unknowingly, provided feedback with their questions. And this second edition would not be possible without the love, support, and patience of my wife Carol.

Preface to the First Edition

This book grew out of notes for my Statistical Computing course that I have been teaching for the past 20 years at North Carolina State University. The goal of this course is to prepare doctoral students with the computing tools needed for statistical research, and I have augmented this core with related topics that through the years I have found useful for colleagues and graduate students. As a result, this book covers a wide range of computational issues, from arithmetic, numerical linear algebra, and approximation, which are typical numerical analysis topics, to optimization and nonlinear regression, to random number generation, and finally to fast algorithms. I have emphasized numerical techniques but restricted the scope to those regularly employed in the field of statistics and dropped some traditional numerical analysis topics such as differential equations. Many of the exercises in this book arose from questions posed to me by colleagues and students.

Most of the students that I have taught come with a graduate level understanding of statistics, no experience in numerical analysis, and little skill in a programming language. Consequently, I cover only about half of this material in a one-semester course. For those with a background in numerical analysis, a basic understanding of two statistical topics, regression and maximum likelihood, would be necessary.

I would advise any instructor of statistical computing not to shortchange the fundamental topic of arithmetic. I have found that most students resist the idea that computers have limited precision and employ many defense mechanisms to support that denial. Until students are comfortable with finite precision arithmetic, this psychological obstacle will cripple their understanding of scientific computation. As a result, I urge the use of single precision arithmetic in the early part of the course and introduce numerical linear algebra using a low-level language, even though students may eventually use software or languages that completely hide the calculations behind operators and double precision. These operators will continue to be mysterious black boxes until the fundamental concept of finite precision arithmetic is understood and accepted.

Early in this effort, I faced the dilemma of how to describe algorithms. The big picture is easier to present or to understand with pseudocode descriptions of algorithms. But I always felt that skipping over the details was misleading the reader, especially when the details are critical to the success of an implementation. Furthermore, there is no better challenge to one's understanding of a topic than to take a big-picture description and program it to the smallest detail. On the other hand, writing one's own implementation of an algorithm often seems like a futile reinvention of the wheel.

And so my response to this dilemma is to have it both ways: to present algorithms in pseudocode in the text, but also to supplement the pseudocode with Fortran programs and demonstrations on the accompanying disk.

These programs provide the basic tools for extending the realm of statistical techniques beyond the bounds of current statistical software. But my primary goal in providing this code is instructional. Some exercises consist of implementing a particular algorithm, and occasionally I have intentionally included my implementation for the reader to compare with, or, perhaps, improve upon. I encourage the reader to examine the details of the code and to see how the algorithms respond to changes. A secondary goal is to include as many realistic problems as practicable, having endured the frustration of failing to get code to work on anything but toy problems.

I would like to express my appreciation to the many sources of support behind this effort. First of all, three heads of the Department of Statistics have supported my work in statistical computing: Tom Gerig, Dan Solomon, and the late Dave Mason. Some of the work included here is the result of collaborations with many colleagues over the years; especially notable are Al Kinderman on random number generation and Alan Genz on numerical integration. In particular, I would like to thank Sujit Ghosh and Dave Dickey for contributing invaluable advice on Chapter 13. Dennis Boos deserves special acknowledgment as a friend, colleague, and collaborator, and most importantly, for supplying me with many interesting problems over the years. I would like to thank all of the colleagues and students who brought interesting problems to me that have become material in this book. Finally, I appreciate the feedback that students have given me each semester on earlier versions of this manuscript, including their blank stares and yawns, as well as insightful questions.