From Finite Sample to Asymptotic Methods in Statistics

Exact statistical inference may be employed in diverse fields of science and technology. As problems become more complex and sample sizes become larger, mathematical and computational difficulties can arise that require the use of approximate statistical methods. Such methods are justified by asymptotic arguments but are still based on the concepts and principles that underlie exact statistical inference. With this in perspective, this book presents a broad view of exact statistical inference and the development of asymptotic statistical inference, providing a justification for the use of asymptotic methods for large samples. Methodological results are developed on a concrete and yet rigorous mathematical level and are applied to a variety of problems that include categorical data, regression, and survival analyses.

This book is designed as a textbook for advanced undergraduate or beginning graduate students in statistics, biostatistics, or applied statistics but may also be used as a reference for academic researchers.

Pranab K. Sen is the Cary C. Boshamer Professor of Biostatistics and Professor of Statistics and Operations Research at the University of North Carolina at Chapel Hill. He is the author or coauthor of numerous textbooks in statistics and biostatistics and editor or coeditor of numerous volumes in the same field. He has more than 600 publications in leading statistics journals and has supervised 83 doctoral students. Sen is a Fellow of both the Institute of Mathematical Statistics and the American Statistical Association. In 2002, he was the Senior Noether Awardee for his lifelong contributions to nonparametrics and received the Commemoration Medal from the Czech Union of Physicists and Mathematicians in 1998.

Julio M. Singer is a Professor in the Department of Statistics, University of São Paulo, Brazil, and is the codirector of the university's Center for Applied Statistics. Professor Singer is the coauthor of books on categorical data and large sample theory and has publications in both methodological and applications-oriented journals. He was the 1993 James E. Grizzle Distinguished Alumnus in Biostatistics from the University of North Carolina at Chapel Hill. He supervised several graduate students and contributed to the development of the doctoral program in statistics at the University of São Paulo, Brazil.

Antonio C. Pedroso de Lima is an Associate Professor at the Department of Statistics, University of São Paulo, Brazil, and is the codirector of the university's Center for Applied Statistics. He received his doctoral degree in biostatistics from the University of North Carolina at Chapel Hill. He is the coauthor of a book on introductory statistics, and his research has been published in theoretical, methodological, and applications-oriented journals. Professor Pedroso de Lima has advised a number of master's degree and doctoral students in the graduate program in statistics at the University of São Paulo.

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.

Already Published

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

From Finite Sample to Asymptotic Methods in Statistics

PRANAB K. SEN
University of North Carolina

JULIO M. SINGER
Universidade de São Paulo

ANTONIO C. PEDROSO DE LIMA
Universidade de São Paulo

CAMBRIDGE
UNIVERSITY PRESS

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

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521877220

First published 2010

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Sen, Pranab Kumar, 1937–
From finite sample to asymptotic methods in statistics / Pranab K. Sen, Julio M. Singer, Antonio C. Pedroso de Lima.
p. cm. – (Cambridge series in statistical and probabilistic mathematics)
Includes bibliographical references and index.
ISBN 978-0-521-87722-0 (hardback)
1. Mathematical statistics. 2. Probabilities. 3. Estimation theory. 4. Asymptotic expansions.
I. Singer, Julio da Motta, 1950– II. Lima, Antonio C. Pedroso de, 1961– III. Title. IV. Series.
QA276.S358 2009
519.5–dc22 2009034794

ISBN 978-0-521-87722-0 Hardback

To our beloved parents,

Nagendra and Kalyani
José and Edith
Manoel and Neusa

Cambridge University Press
978-0-521-87722-0 - From Finite Sample to Asymptotic Methods in Statistics
Pranab K. Sen, Julio M. Singer and Antonio C. Pedroso de Lima
Frontmatter
More information

Contents

Preface

Students and investigators working in statistics, biostatistics, or applied statistics, in general, are constantly exposed to problems that involve large quantities of data. This is even more evident today, when massive datasets with an impressive amount of details are produced in novel fields such as genomics or bioinformatics at large. Because, in such a context, exact statistical inference may be computationally out of reach and in many cases not even mathematically tractable, they have to rely on approximate results. Traditionally, the justification for these approximations was based on the convergence of the first four moments of the distributions of the statistics under investigation to those of some normal distribution. Today we know that such an approach is not always theoretically adequate and that a somewhat more sophisticated set of techniques based on asymptotic considerations may provide the appropriate justification. This need for more profound mathematical theory in statistical large-sample theory is patent in areas involving dependent sequences of observations, such as longitudinal and survival data or life tables, in which the use of martingale or related structures has distinct advantages.

Unfortunately, most of the technical background for understanding such methods is dealt with in specific articles or textbooks written for a readership with such a high level of mathematical knowledge that they exclude a great portion of the potential users. We tried to bridge this gap in a previous text (Sen and Singer [1993]: *Large Sample Methods in Statistics: An Introduction with Applications*), on which our new enterprise is based. While teaching courses based on that text in the Department of Biostatistics at the University of North Carolina at Chapel Hill and in the Department of Statistics at the University of São Paulo, during the past few years, we came across a new hiatus originated in the apparent distinction between the theory covered in the exact statistical inference and the approximate approach adopted in the book. While realizing that the foundations of the large-sample statistical methods we proposed to address stem from the basic concepts and principles that underlie finite sample setups, we decided to integrate both approaches in the present text. In summary, our intent is to provide a broad view of finite-sample statistical methods, to examine their merits and caveats, and to judge how far asymptotic results eliminate some of the detected impasses, providing the basis for sound application of approximate statistical inference in large samples.

Chapter 1 describes the type of problems considered in the text along with a brief summary of some basic mathematical and statistical concepts required for a good understanding of the remaining chapters. Chapters 2 and 3 lay out the two basic building blocks of statistical inference, namely, estimation and hypothesis testing. There we address issues relating to the important concepts of likelihood, sufficiency, and invariance, among others.

In Chapter 4 we present a brief review of Decision Theory and Bayesian methods, contrasting them with those outlined in the previous chapters. Chapter 5 is devoted to stochastic processes, given their importance in the development of models for dependent random variables as well as their significance as paradigms for many problems arising in practical applications. Chapters 6 and 7 contain the essential tools needed to prove asymptotic results for independent sequences of random variables as well as an outline of the possible extensions to cover the dependent sequence case. Chapter 8 discusses some general results on the asymptotics of estimators and test statistics; their actual application to categorical data and regression analysis is illustrated in Chapters 9 and 10, respectively. Finally, Chapter 11 deals with an introductory exposition of the technical background required to deal with the asymptotic theory for statistical functionals. The objective here is to provide some motivation and the general flavor of the problems in this area, because a rigorous treatment would require a much higher level of mathematical background than the one we contemplate. The 11 chapters were initially conceived for a two-semester course for second-year students in biostatistics or applied statistics doctoral programs as well as for last-year undergraduate or first-year graduate students in statistics. A more realistic view, however, would restrict the material for such purposes to the first eight chapters, along with a glimpse into Chapter 11. Chapters 9 and 10 could be included as supplementary material in categorical data and linear models courses, respectively. Because the text includes a number of practical examples, it may be useful as a reference text for investigators in many areas requiring the use of statistics.

The authors would like to thank the numerous students who took large-sample theory courses at the Department of Biostatistics, University of North Carolina at Chapel Hill, and Department of Statistics, University of São Paulo, providing important contributions to the design of this text. Finally, we must acknowledge the Cary C. Boshamer Foundation, University of North Carolina at Chapel Hill, as well as Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil, and Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil, for providing financial support during the years of preparation of the text.

Pranab K. Sen
Julio M. Singer
Antonio C. Pedroso de Lima
Chapel Hill and São Paulo, April 2009