Applied Asymptotics: Case Studies in Small-Sample Statistics

In fields such as biology, medical sciences, sociology and economics researchers often face the situation where the number of available observations, or the amount of available information, is sufficiently small that approximations based on the normal distribution may be unreliable. Theoretical work over the last quarter-century has led to new likelihood-based methods that yield very accurate approximations in finite samples, but this work has had limited impact on statistical practice. This book illustrates by means of realistic examples and case studies how to use the new theory, and investigates how and when it makes a difference to the resulting inference. The treatment is oriented towards practice and is accompanied by code in the R language which enables the methods to be applied in a range of situations of interest to practitioners. The analysis includes some comparisons of higher order likelihood inference with bootstrap and Bayesian methods.

ALESSANDRA BRAZZALE is a Professor of Statistics at the Università degli Studi di Modena e Reggio Emilia.

ANTHONY DAVISON is a Professor of Statistics at the Ecole Polytechnique Fédérale de Lausanne.

NANCY REID is a University Professor of Statistics at the University of Toronto.

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. *Exercises 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
20. *Random Graph Dynamics*, by Rick Durrett
21. *Networks*, by Peter Whittle
22. *Saddlepoint Approximations with Applications*, by Ronald W. Butler

Applied Asymptotics Case Studies in Small-Sample Statistics

A. R. Brazzale
Università degli Studi di Modena e Reggio Emilia

A. C. Davison
Ecole Polytechnique Fédérale de Lausanne

N. Reid
University of Toronto

CAMBRIDGE UNIVERSITY PRESS

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

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/9780521847032

First published 2007

Printed in the United Kingdom at the University Press, Cambridge

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

ISBN 978-0-521-84703-2 hardback

Contents

Preface

The likelihood function plays a central role in both statistical theory and practice. Basic results about likelihood inference, which we call first order asymptotics, were developed in fundamental work by R. A. Fisher during the 1920s, and now form an essential and widely taught part of both elementary and advanced courses in statistics. It is less well known that Fisher later proposed a more refined approach, which has been developed over the past three decades into a theory of higher order asymptotics. While this theory leads to some extremely accurate methods for parametric inference, accounts of the theory can appear forbidding, and the results may be thought to have little importance for statistical practice.

The purpose of this book is dispel this view, showing how higher order asymptotics may be applied in realistic examples with very little more effort than is needed for first order procedures, and to compare the resulting improved inferences with those from other approaches. To do this we have collected a range of examples and case studies, provided details on the implementation of higher order approximations, and compared the resulting inference to that based on other methods; usually first order likelihood theory, but where appropriate also methods based on simulation. Our examples are nearly all derived from regression models for discrete or continuous data, but range quite widely over the types of models and inference problems where likelihood methods are applied.

In order to make higher order methods accessible, we have striven for as simple an exposition as we thought feasible, aiming for heuristic explanation rather than full mathematical rigour. We do not presuppose previous knowledge of higher order asymptotics, key aspects of which are explained early in the book. The reader is assumed to have knowledge of basic statistics including some central classes of models, and some experience of standard likelihood methods in practice. We intend that the book be useful for students of statistics, practising statisticians, and data analysts, as well as researchers interested in a more applied account of the methods than has so far been available. Our effort has been made practicable by software developed by Alessandra Brazzale and Ruggero Bellio over many years, of which the `hoa` package bundle now available in R is the culmination. This software is extensively used throughout the book, and the ideas behind the `hoa` packages, described in Chapter 9, formed the basis for our approaches to programming

when new software was needed for some of the examples. The `hoa` package bundle and other materials may be obtained from the book's web page

`http://statwww.epfl.ch/AA`

This book could not have been written without the help of many colleagues. We thank particularly Sven Bacher, Douglas Bates, Ruggero Bellio, Nathalie Chèvre, David Cox, Don Fraser, Torsten Hothorn, Alessandra Salvan, Ana-Maria Staicu, Charlotte Vandenberghe, and members of the R Development Core Team, for access to data, fruitful collaboration, helpful discussion, valuable comments on the material, and help with computational aspects of our work. David Tranah and Diana Gillooly of Cambridge University Press have been supportive and patient editors. We thank also the following institutions for financial and material support: the EPFL; the University of Padova; the University of Toronto; the Italian National Research Council, and in particular its Institute of Biomedical Engineering, where much of Alessandra Brazzale's work was performed; the Italian Ministry of Education, University and Research; the Canadian Natural Sciences and Engineering Research Council; and the Swiss National Science Foundation. We thank also our friends and families for their enduring encouragement and support.

A. R. Brazzale, A. C. Davison and N. Reid
Reggio Emilia, Lausanne and Toronto