Statistical models

CAMBRIDGE SERIES IN STATISTICAL AND PROBABILISTIC MATHEMATICS

Editorial Board:

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*, A.C. Davison and D.V. Hinkley
2. *Markov Chains*, J. Norris
3. *Asymptotic Statistics*, A.W. van der Vaart
4. *Wavelet Methods for Time Series Analysis*, D.B. Percival and A.T. Walden
5. *Bayesian Methods*, T. Leonard and J.S.J. Hsu
6. *Empirical Processes in M-Estimation*, S. van de Geer
7. *Numerical Methods of Statistics*, J. Monahan
8. *A User's Guide to Measure-Theoretic Probability*, D. Pollard
9. *The Estimation and Tracking of Frequency*, B.G. Quinn and E.J. Hannan

Statistical models

A. C. Davison
Swiss Federal Institute of Technology, Lausanne

CAMBRIDGE UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2003

Printed in the USA

Typeface Times 10/13 pt *System* LaTeX 2$_\varepsilon$ [TB]

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

ISBN 0 521 77339 3 hardback

Contents

Preface

A statistical model is a probability distribution constructed to enable inferences to be drawn or decisions made from data. This idea is the basis of most tools in the statistical workshop, in which it plays a central role by providing economical and insightful summaries of the information available.

This book is intended as an integrated modern account of statistical models covering the core topics for studies up to a masters degree in statistics. It can be used for a variety of courses at this level and for reference. After outlining basic notions, it contains a treatment of likelihood that includes non-regular cases and model selection, followed by sections on topics such as Markov processes, Markov random fields, point processes, censored and missing data, and estimating functions, as well as more standard material. Simulation is introduced early to give a feel for randomness, and later used for inference. There are major chapters on linear and nonlinear regression and on Bayesian ideas, the latter sketching modern computational techniques. Each chapter has a wide range of examples intended to show the interplay of subject-matter, mathematical, and computational considerations that makes statistical work so varied, so challenging, and so fascinating.

The target audience is senior undergraduate and graduate students, but the book should also be useful for others wanting an overview of modern statistics. The reader is assumed to have a good grasp of calculus and linear algebra, and to have followed a course in probability including joint and conditional densities, moment-generating functions, elementary notions of convergence and the central limit theorem, for example using Grimmett and Welsh (1986) or Stirzaker (1994). Measure is not required. Some sections involve a basic knowledge of stochastic processes, but they are intended to be as self-contained as possible. To have included full proofs of every statement would have made the book even longer and very tedious. Instead I have tried to give arguments for simple cases, and to indicate how results generalize. Readers in search of mathematical rigour should see Knight (2000), Schervish (1995), Shao (1999), or van der Vaart (1998), amongst the many excellent books on mathematical statistics.

Solution of problems is an integral part of learning a mathematical subject. Most sections of the book finish with exercises that test or deepen knowledge of that section, and each chapter ends with problems which are generally broader or more demanding.

Real understanding of statistical methods comes from contact with data. Appendix A outlines practicals intended to give the reader this experience. The practicals themselves can be downloaded from

`http://statwww.epfl.ch/people/~davison/SM`

together with a library of functions and data to go with the book, and errata. The practicals are written in two dialects of the S language, for the freely available package `R` and for the commercial package `S-plus`, but it should not be hard for teachers to translate them for use with other packages.

Biographical sketches of some of the people mentioned in the text are given as sidenotes; the sources for many of these are Heyde and Seneta (2001) and

`http://www-groups.dcs.st-and.ac.uk/~history/`

Part of the work was performed while I was supported by an Advanced Research Fellowship from the UK Engineering and Physical Science Research Council. I am grateful to them and to my past and present employers for sabbatical leaves during which the book advanced. Many people have helped in various ways, for example by supplying data, examples, or figures, by commenting on the text, or by testing the problems. I thank Marc-Olivier Boldi, Alessandra Brazzale, Angelo Canty, Gorana Capkun, James Carpenter, Valérie Chavez, Stuart Coles, John Copas, Tom DiCiccio, Debbie Dupuis, David Firth, Christophe Girardet, David Hinkley, Wilfred Kendall, Diego Kuonen, Stephan Morgenthaler, Christophe Osinski, Brian Ripley, Gareth Roberts, Sylvain Sardy, Jamie Stafford, Trevor Sweeting, Valérie Ventura, Simon Wood, and various anonymous reviewers. Particular thanks go to Jean-Yves Le Boudec, Nancy Reid, and Alastair Young, who gave valuable comments on much of the book. David Tranah of Cambridge University Press displayed exemplary patience during the interminable wait for me to finish. Despite all their efforts, errors and obscurities doubtless remain. I take responsibility for this and would appreciate being told of them, in order to correct any future versions.

My long-suffering family deserve the most thanks. I dedicate this book to them, and particularly to Claire, without whose love and support the project would never have been finished.

Lausanne, January 2003