

Cambridge University Press
0521773393 - Statistical Models
A. C. Davison
Frontmatter
[More information](#)

Statistical models

CAMBRIDGE SERIES IN STATISTICAL AND PROBABILISTIC
MATHEMATICS

Editorial Board:

- R. Gill, *Department of Mathematics, Utrecht University*
B.D. Ripley, *Department of Statistics, University of Oxford*
S. Ross, *Department of Industrial Engineering, University of California, Berkeley*
M. Stein, *Department of Statistics, University of Chicago*
D. Williams, *School of Mathematical Sciences, University of Bath*

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
0521773393 - Statistical Models
A. C. Davison
Frontmatter
[More information](#)

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>

© Cambridge University Press 2003

This book is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without
the written permission of Cambridge University Press.

First published 2003

Printed in the USA

Typeface Times 10/13 pt *System* L^AT_EX 2_ε [TB]

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

ISBN 0 521 77339 3 hardback

Contents

<i>Preface</i>	ix
1 Introduction	1
2 Variation	15
2.1 Statistics and Sampling Variation	15
2.2 Convergence	28
2.3 Order Statistics	37
2.4 Moments and Cumulants	44
2.5 Bibliographic Notes	48
2.6 Problems	49
3 Uncertainty	52
3.1 Confidence Intervals	52
3.2 Normal Model	62
3.3 Simulation	77
3.4 Bibliographic Notes	90
3.5 Problems	90
4 Likelihood	94
4.1 Likelihood	94
4.2 Summaries	101
4.3 Information	109
4.4 Maximum Likelihood Estimator	115
4.5 Likelihood Ratio Statistic	126
4.6 Non-Regular Models	140

vi	<i>Contents</i>
4.7	Model Selection 150
4.8	Bibliographic Notes 156
4.9	Problems 156
5	Models 161
5.1	Straight-Line Regression 161
5.2	Exponential Family Models 166
5.3	Group Transformation Models 183
5.4	Survival Data 188
5.5	Missing Data 203
5.6	Bibliographic Notes 218
5.7	Problems 219
6	Stochastic Models 225
6.1	Markov Chains 225
6.2	Markov Random Fields 244
6.3	Multivariate Normal Data 255
6.4	Time Series 266
6.5	Point Processes 274
6.6	Bibliographic Notes 292
6.7	Problems 293
7	Estimation and Hypothesis Testing 300
7.1	Estimation 300
7.2	Estimating Functions 315
7.3	Hypothesis Tests 325
7.4	Bibliographic Notes 348
7.5	Problems 349
8	Linear Regression Models 353
8.1	Introduction 353
8.2	Normal Linear Model 359
8.3	Normal Distribution Theory 370
8.4	Least Squares and Robustness 374
8.5	Analysis of Variance 378
8.6	Model Checking 386
8.7	Model Building 397
8.8	Bibliographic Notes 409
8.9	Problems 409

<i>Contents</i>	vii
9 Designed Experiments	417
9.1 Randomization	417
9.2 Some Standard Designs	426
9.3 Further Notions	439
9.4 Components of Variance	449
9.5 Bibliographic Notes	463
9.6 Problems	464
10 Nonlinear Regression Models	468
10.1 Introduction	468
10.2 Inference and Estimation	471
10.3 Generalized Linear Models	480
10.4 Proportion Data	487
10.5 Count Data	498
10.6 Overdispersion	511
10.7 Semiparametric Regression	518
10.8 Survival Data	540
10.9 Bibliographic Notes	554
10.10 Problems	555
11 Bayesian Models	565
11.1 Introduction	565
11.2 Inference	578
11.3 Bayesian Computation	596
11.4 Bayesian Hierarchical Models	619
11.5 Empirical Bayes Inference	627
11.6 Bibliographic Notes	637
11.7 Problems	639
12 Conditional and Marginal Inference	645
12.1 Ancillary Statistics	646
12.2 Marginal Likelihood	656
12.3 Conditional Inference	665
12.4 Modified Profile Likelihood	680
12.5 Bibliographic Notes	691
12.6 Problems	692

Appendix A. Practicals

696

Bibliography

699

Name Index

712

Example Index

716

Index

718

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