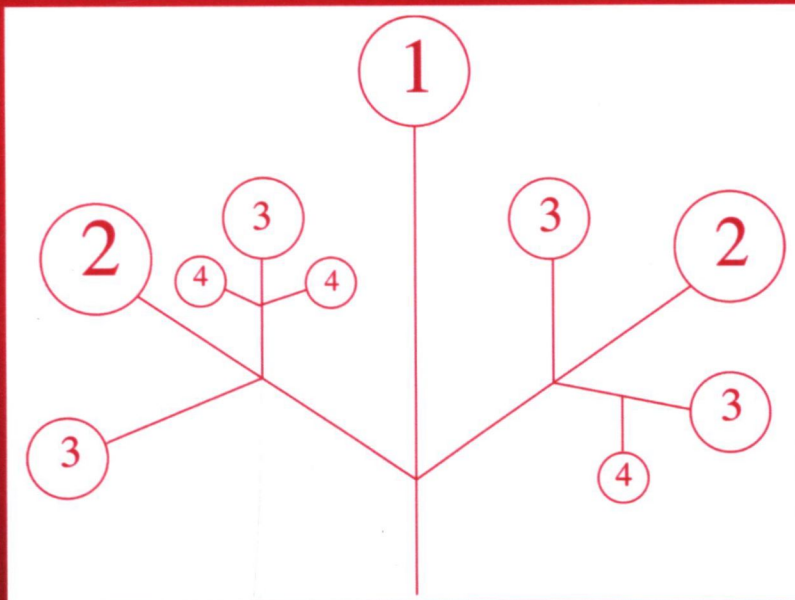


Texts in Statistical Science

Applied Stochastic Modelling

Second Edition



Byron J. T. Morgan



CRC Press
Taylor & Francis Group

A CHAPMAN & HALL BOOK

Statistics

Highlighting modern computational methods, **Applied Stochastic Modelling, Second Edition** provides you with the practical experience of scientific computing in applied statistics through a range of interesting real-world applications. It also successfully revises standard probability and statistical theory. Along with an updated bibliography and improved figures, this edition offers numerous updates throughout.

New to the Second Edition

- An extended discussion on Bayesian methods
- A large number of new exercises
- A new appendix on computational methods

The book covers both contemporary and classical aspects of statistics, including survival analysis, Kernel density estimation, Markov chain Monte Carlo, hypothesis testing, regression, bootstrap, and generalised linear models. Although the book can be used without reference to computational programs, the author provides the option of using powerful computational tools for stochastic modelling. All of the data sets and MATLAB® programs found in the text are also available online.

Continuing in the bestselling tradition of its predecessor, this book remains an excellent resource for understanding how to fit stochastic models to data.

Features

- Covers important advances in the theory and practice of statistics
- Explains how to construct, fit, and evaluate statistical and stochastic models for use in diverse areas, such as geology, sociology, biology, and economics
- Incorporates traditional statistical and probability theory with modern computational methods
- Offers the data sets from the book, the latest MATLAB computer programs, and corresponding versions in R on a supplementary Web site

Second
Edition

Applied Stochastic Modelling

Morgan



Texts in Statistical Science

Applied Stochastic Modelling

Second Edition

Byron J. T. Morgan

University of Kent
UK



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

First issued in hardback 2017

© 2009 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

ISBN-13: 978-1-5848-8666-2 (pbk)
ISBN-13: 978-1-1384-6969-3 (hbk)

This book contains information obtained from authentic and highly regarded sources. While all reasonable efforts have been made to publish reliable data and information, neither the author[s] nor the publisher can accept any legal responsibility or liability for any errors or omissions that may be made. The publishers wish to make clear that any views or opinions expressed in this book by individual editors, authors or contributors are personal to them and do not necessarily reflect the views/opinions of the publishers. The information or guidance contained in this book is intended for use by medical, scientific or health-care professionals and is provided strictly as a supplement to the medical or other professional's own judgement, their knowledge of the patient's medical history, relevant manufacturer's instructions and the appropriate best practice guidelines. Because of the rapid advances in medical science, any information or advice on dosages, procedures or diagnoses should be independently verified. The reader is strongly urged to consult the relevant national drug formulary and the drug companies' and device or material manufacturers' printed instructions, and their websites, before administering or utilizing any of the drugs, devices or materials mentioned in this book. This book does not indicate whether a particular treatment is appropriate or suitable for a particular individual. Ultimately it is the sole responsibility of the medical professional to make his or her own professional judgements, so as to advise and treat patients appropriately. The authors and publishers have also attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com/ or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

**Applied
Stochastic
Modelling**
Second Edition

CHAPMAN & HALL/CRC
Texts in Statistical Science Series

Series Editors

Bradley P. Carlin, *University of Minnesota, USA*

Julian J. Faraway, *University of Bath, UK*

Martin Tanner, *Northwestern University, USA*

Jim Zidek, *University of British Columbia, Canada*

Analysis of Failure and Survival Data

P.J. Smith

**The Analysis of Time Series—
An Introduction, Sixth Edition**

C. Chatfield

**Applied Bayesian Forecasting and Time Series
Analysis**

A. Pole, M. West and J. Harrison

**Applied Nonparametric Statistical Methods, Fourth
Edition**

P. Sprent and N.C. Smeeton

**Applied Statistics — Handbook of GENSTAT
Analysis**

E.J. Snell and H. Simpson

Applied Statistics — Principles and Examples

D.R. Cox and E.J. Snell

Applied Stochastic Modelling, Second Edition

B.J.T. Morgan

Bayesian Data Analysis, Second Edition

A. Gelman, J.B. Carlin, H.S. Stern
and D.B. Rubin

**Bayesian Methods for Data Analysis,
Third Edition**

B.P. Carlin and T.A. Louis

Beyond ANOVA — Basics of Applied Statistics

R.G. Miller, Jr.

**Computer-Aided Multivariate Analysis,
Fourth Edition**

A.A. Afifi and V.A. Clark

A Course in Categorical Data Analysis

T. Leonard

A Course in Large Sample Theory

T.S. Ferguson

Data Driven Statistical Methods

P. Sprent

Decision Analysis — A Bayesian Approach

J.Q. Smith

**Elementary Applications of Probability Theory,
Second Edition**

H.C. Tuckwell

Elements of Simulation

B.J.T. Morgan

**Epidemiology — Study Design and
Data Analysis, Second Edition**

M. Woodward

Essential Statistics, Fourth Edition

D.A.G. Rees

**Extending the Linear Model with R: Generalized
Linear, Mixed Effects and Nonparametric Regression
Models**

J.J. Faraway

A First Course in Linear Model Theory

N. Ravishanker and D.K. Dey

**Generalized Additive Models:
An Introduction with R**

S. Wood

**Interpreting Data — A First Course
in Statistics**

A.J.B. Anderson

**An Introduction to Generalized
Linear Models, Third Edition**

A.J. Dobson and A.G. Barnett

Introduction to Multivariate Analysis

C. Chatfield and A.J. Collins

**Introduction to Optimization Methods and Their
Applications in Statistics**

B.S. Everitt

Introduction to Probability with R

K. Baclawski

**Introduction to Randomized Controlled Clinical
Trials, Second Edition**

J.N.S. Matthews

**Introduction to Statistical Methods for
Clinical Trials**

T.D. Cook and D.L. DeMets

Large Sample Methods in Statistics

P.K. Sen and J. da Motta Singer

Linear Models with R

J.J. Faraway

**Markov Chain Monte Carlo —
Stochastic Simulation for Bayesian Inference,
Second Edition**

D. Gamerman and H.F. Lopes

Mathematical Statistics

K. Knight

Modeling and Analysis of Stochastic Systems

V. Kulkarni

Modelling Binary Data, Second Edition

D. Collett

**Modelling Survival Data in Medical Research,
Second Edition**

D. Collett

**Multivariate Analysis of Variance and Repeated
Measures — A Practical Approach for Behavioural
Scientists**

D.J. Hand and C.C. Taylor

Multivariate Statistics — A Practical Approach

B. Flury and H. Riedwyl

Pólya Urn Models

H. Mahmoud

Practical Data Analysis for Designed Experiments

B.S. Yandell

Practical Longitudinal Data Analysis

D.J. Hand and M. Crowder

Practical Statistics for Medical Research

D.G. Altman

A Primer on Linear Models

J.F. Monahan

Probability — Methods and Measurement

A. O'Hagan

**Problem Solving — A Statistician's Guide, Second
Edition**

C. Chatfield

**Randomization, Bootstrap and
Monte Carlo Methods in Biology,
Third Edition**

B.F.J. Manly

Readings in Decision Analysis

S. French

Sampling Methodologies with Applications

P.S.R.S. Rao

Statistical Analysis of Reliability Data

M.J. Crowder, A.C. Kimber,
T.J. Sweeting, and R.L. Smith

Statistical Methods for Spatial Data Analysis

O. Schabenberger and C.A. Gotway

Statistical Methods for SPC and TQM

D. Bissell

**Statistical Methods in Agriculture and Experimental
Biology, Second Edition**

R. Mead, R.N. Curnow, and A.M. Hasted

**Statistical Process Control — Theory and Practice,
Third Edition**

G.B. Wetherill and D.W. Brown

Statistical Theory, Fourth Edition

B.W. Lindgren

Statistics for Accountants

S. Letchford

Statistics for Epidemiology

N.P. Jewell

**Statistics for Technology — A Course in Applied
Statistics, Third Edition**

C. Chatfield

Statistics in Engineering — A Practical Approach

A.V. Metcalfe

**Statistics in Research and Development,
Second Edition**

R. Caulcutt

**Survival Analysis Using S — Analysis of
Time-to-Event Data**

M. Tableman and J.S. Kim

The Theory of Linear Models

B. Jørgensen

Time Series Analysis

H. Madsen

Preface to the Second Edition

The structure of the second edition of this book is very similar to that of the first edition; however, there have been numerous changes throughout. In particular, a large number of new exercises have been added, there is a new appendix on computational methods, and the discussion of Bayesian methods has been extended. The bibliography has been updated, throughout figures have been improved, and, where necessary, errors have been corrected. I am grateful for the many positive comments that the first version of the book received, and to those who have written to point out mistakes, and ways in which the book might be improved. I thank especially Ted Catchpole, Rachel Fewster, Ruth King, Rachel McCrea, David Miller, Karen Palmer and Martin Ridout. I thank MATLAB[®] for providing the latest version of the package, and I am also grateful for the help and patience of Rob Calver and colleagues at CRC Chapman & Hall. The book continues to be the set text for final year under-graduate and post-graduate courses in respectively Applied Stochastic Modelling and Data Analysis, and Computational Statistics in the University of Kent. The book successfully revises and integrates the probability and statistics methods of earlier lecture courses. At the same time it brings students into contact with modern computational methods; it provides students with practical experience of scientific computing at use in applied statistics, in the context of a range of interesting real-life applications. The book Web Site contains the data sets from the book, the MATLAB computer programs, as well as corresponding versions in R. There is a solutions manual for the exercises and the computer practical sheets that are used in Kent. The book has been the basis for a course on *Applied Stochastic Modelling* held at Pfizer Central Research in Sandwich, Kent, during 2007, and also for a Continuing Professional Development course with the same name, to be held at the Royal Statistical Society, London, in 2008. The slides for these courses are also available on the book Web site.

Canterbury

Preface

This book is being completed at the end of a millenium, the last 30 years of which have seen many exciting major developments in the theory and practice of statistics. The book presents many of the most important of these advances. It has its origins in a series of talks on aspects of modern statistics for fitting stochastic models to data, commissioned by statisticians at Pfizer Central Research in Sandwich in Kent. These talks gave rise to a 30-hour lecture course given to third year undergraduates, statistics MSc students and first-year statistics PhD students at the University of Kent, and this book has grown from the notes for that lecture course. These students have found that even the most recently developed statistical methods may be readily understood and successfully applied in practice. As well as covering modern techniques, the material of the book integrates and revises standard probability and statistical theory. Much modern statistical work is implemented using a computer. Thus it is necessary in a book of this nature to include computer instructions of some kind. The integrated computer language MATLAB has been selected for this purpose, and over 50 short MATLAB programs are included throughout the book. Their distribution and purpose are described in the index of MATLAB programs. They may all be accessed from the book site on the World Wide Web. They are designed to be illustrative, rather than completely efficient. Often doubts concerning theory are dissipated when one can see computer code for the theory. Students of the book material have certainly found the MATLAB programs to be a useful aid to learning. It has been uplifting to observe the satisfaction that students have gained from running the MATLAB programs of the book. Often the students had no previous knowledge of MATLAB and also little prior experience of scientific computing. The material of Appendix B, which summarises important features of MATLAB, and tutorial assistance were all that was needed by these students. However it should be stressed that while the computer programs are included as an aid to learning, the book may be read and used without reference to the programs. S-plus versions of the programs are available on the book site on the World Wide Web. There are also some references to the use of symbolic algebra packages such as MAPLE, as these provide powerful tools for stochastic modelling.

The target student audience for the book is final-year undergraduate and MSc students of mathematics and statistics. The book is also intended as a single convenient source of reference for research scientists and post-graduate students, using modern statistical methods which are currently described in

depth in a range of single-topic textbooks. Prior knowledge is assumed at the level of a typical second-year university course on probability and statistics. Appendix A summarises a number of important formulae and results from probability and statistics. A small fraction of the book sections and exercises contain more advanced material, and these are starred. Kernel density estimation is a central aspect of modern statistics, and therefore the basic ideas are summarised in Appendix C. While a limited number of exercises have solutions included in the book, a more extensive set of solutions is to be found on the World Wide Web book site.

Statistical methods are all-pervasive, contributing significantly to subjects as diverse as geology, sociology, biology and economics. The construction, fitting and evaluation of statistical and stochastic models are not only vitally important in areas such as these, but they are also great fun. It is hoped that some of the enjoyment and fascination of the subject will be gained by readers of this book.

The book is motivated by real data and problems. The examples and exercises are often chosen from my own experience and, as can be seen from the index of data sets, many have arisen from biology. The areas covered are sometimes atypical, and sometimes classical, such as survival analysis, quantal assay and capture-recapture. Several of the examples recur at various points throughout the book. The data are available on the book site on the World Wide Web.

Acknowledgments

I owe a great debt of gratitude to the many scientists and statisticians with whom I have collaborated, and with whom I continue to collaborate. I am especially grateful to my MATLAB mentor, Ted Catchpole, who with Paul Terrill, contributed some of the MATLAB programs. The original proposal for a set of lectures at Pfizer Central Research was agreed with Trevor Lewis, Group Executive Director, Biometrics, at Pfizer. The Pfizer audience was typically challenging, with Ken Martin and John Parrott in particular regularly asking probing questions. Steve Brooks, Steve Buckland, Martin Ridout, and Paul Terrill all commented helpfully on draft chapters. At Arnold, Nicki Dennis, and Kirsty Stroud were patient and encouraging. Three years of Kent students provided a sympathetic test bed and launching pad. The many drafts of the book were typed enthusiastically by Lilian Bond, Julie Snook, and Mavis Swain.

Canterbury

'For the things we have to learn before we can do them, we learn by doing them'

Aristotle

Contents

Preface to the Second Edition	xi
Preface	xiii
1 Introduction and Examples	1
1.1 Introduction	1
1.2 Examples of data sets	3
1.3 Discussion	11
1.4 Exercises	12
2 Basic Model-Fitting	17
2.1 Introduction	17
2.2 Maximum-likelihood estimation for a geometric model	17
2.3 Maximum-likelihood for the beta-geometric model	22
2.4 Modelling polyspermy	26
2.5 Which model?	31
2.6 What is a model for?	32
2.7 *Mechanistic models	32
2.8 Discussion	34
2.9 Exercises	34
3 Function Optimisation	45
3.1 Introduction	45
3.2 MATLAB; graphs and finite differences	46
3.3 Deterministic search methods	48
3.4 Stochastic search methods	60
3.5 Accuracy and a hybrid approach	67
3.6 Discussion	68
3.7 Exercises	69
4 Basic Likelihood Tools	77
4.1 Introduction	77
4.2 Estimating standard errors and correlations	80
4.3 Looking at surfaces: profile log-likelihoods	84
4.4 Confidence regions from profiles	89
4.5 Hypothesis testing in model selection	94

4.6	Score and Wald tests	101
4.7	Classical goodness of fit	106
4.8	Model selection bias	106
4.9	Discussion	107
4.10	Exercises	108
5	General Principles	123
5.1	Introduction	123
5.2	Parameterisation	123
5.3	*Parameter redundancy	130
5.4	Boundary estimates	135
5.5	Regression and influence	136
5.6	The EM algorithm	137
5.7	Alternative methods of model-fitting	148
5.8	*Non-regular problems	152
5.9	Discussion	153
5.10	Exercises	154
6	Simulation Techniques	169
6.1	Introduction	169
6.2	Simulating random variables	170
6.3	Integral estimation	175
6.4	Verification	177
6.5	*Monte Carlo inference	179
6.6	Estimating sampling distributions	180
6.7	Bootstrap	183
6.8	Monte Carlo testing	190
6.9	Discussion	192
6.10	Exercises	193
7	Bayesian Methods and MCMC	199
7.1	Basic Bayes	199
7.2	Three academic examples	200
7.3	The Gibbs sampler	201
7.4	The Metropolis-Hastings algorithm	213
7.5	A hybrid approach	218
7.6	The data augmentation algorithm	220
7.7	Model probabilities	220
7.8	Model averaging	223
7.9	Reversible jump MCMC: RJMCMC	225
7.10	Discussion	226
7.11	Exercises	228
8	General Families of Models	237
8.1	Common structure	237