International Statistical Review
Revue Internationale de Statistique

International Statistical Review (2012), 80, 1, 176–204 doi:10.1111/j.1751-5823.2012.00179.x

Short Book Reviews

Editor: Simo Puntanen

Introduction to Statistical Data Analysis for the Life Sciences
Claus Thorn Ekstrøm, Helle Sørensen
Chapman & Hall/CRC, 2011, xi + 415 pages, £31.99/$69.95, softcover
ISBN: 978-1-4398-2555-6

Table of Contents

Readership: Students and practitioners of biology and life sciences.

This book can be valuable assistance for students of life sciences and the other biological faculties and it can be treated both as a first handbook to statistical methods as well as a suitable tool to systematize earlier experiences. The authors put the emphasis not on mathematical formulas, definitions and theorems but—as they write in the Preface—on "providing the readers a feeling of being able to model and analyze the data".

The book is divided into 13 chapters and 3 appendixes. The material introduced in chapters from 1 to 12 is illustrated with many examples and biological problems. Moreover, at the end of each chapter the reader can find a lot of exercises. Some of them are supposed to be done by hand, whereas the remaining should be solved with R-package. The authors prepared one appendix as an introduction to R for readers unfamiliar with this software. The nice advantage of this book is chapter 13 containing 10 case exercises which are more complicated than previous ones and require the compilation of the methods presented before. All the data sets used in the exercises are available from the supporting web site.

The book is written in a clear and engaging style. The authors put much emphasis on the modelling part of statistical analysis and on biological interpretation of obtained results. It could be recommended for students but also other readers looking for a handbook of "practical" statistics.

Ewa Skotarczak: *efalsa@au.poznan.pl*
Department of Mathematical and Statistical Methods
Poznań University of Life Sciences
Wojska Polskiego 28, 60-637 Poznań, Poland

Handbook of Fitting Statistical Distributions with R
Zaven A. Karian, Edward J. Dudewicz
Chapman & Hall/CRC, 2011, xlv + 1672 pages, £95.00/$149.95, hardcover
ISBN: 978-1-58488-711-9

Table of Contents (parts)

Readership: Rather topical and restricted. Not advisable for students.

The title of the book, *Handbook of Fitting Statistical Distributions with R*, is misleading in several respects: this book is not a handbook, i.e. a reference book that can be easily consulted, it does not cover standard statistical distributions but rather a restricted class of parametrized distributions, the R content is at best marginal, and the two authors only wrote part of the book, with (a) a rather large portion being written by other authors and (b) a large overlap with the earlier Karian & Dudewicz (2000). Most of the novelty in this version is due to the inclusion of chapters written by/with additional authors.

The main purpose of *Handbook of Fitting Statistical Distributions with R* is to promote a class of quantile distributions, the generalized lambda distributions (GLD), first introduced by Ramberg & Schmeiser (1974). Those distributions are defined via their quantile function, which is a location-scale transform of

$$Q(y|\lambda_3, \lambda_4) = y^{\lambda_3} - (1-y)^{\lambda_4}$$

(under the constraint on (λ_3, λ_4) that the above function of y is monotonous). There is nothing wrong *per se* with those distributions, but neither is there a particular reason to prefer them to the standard parametric distributions: The first part of the book spends a large amount of space on the fact that GLDs approximate reasonably well (in the L1 or L2 norm sense) those standard distributions, but it does not explain why the substitution is of interest. Furthermore, the estimation of the parameters in GLDs (i.e. the fitting part) is quite involved. Since the likelihood function is not available in closed form, alternatives to maximum likelihood or Bayesian inference have to be devised. (Su (2007) proposes an approximate maximum likelihood resolution based on a crude plug-in inversion of the quantile function.) The book concentrates on moment and percentile estimators as the central estimation tool, with no clear message on which method to favour (see, e.g. Section 5.5). The variability of those estimators is evaluated by parametric bootstrap later in the book.

The R aspect of the book is quite limited: the attached CD-ROM contains R codes that are mentioned within *Handbook of Fitting Statistical Distributions with R*. While related to an earlier version of the book (Karian & Dudewicz, 2000), or to contributed chapters, Matlab and even Maple codes are still to be found in this edition.

There are many finer issues I could criticize about *Handbook of Fitting Statistical Distributions with R*, from the inclusion of numerous tables to the inconclusive assessment of estimation procedures. However I see little point in engaging into this: the book does not suit a purpose other than presenting a collection of papers on the state-of-the art research in the branch of

GLDs. In my opinion, the book cannot serve as a reference for practitioners, nor as a support in an academic course and, given its unusual price, I see no reason in recommending it.

Christian P. Robert: *christian.robert@ceremade.dauphine.fr*
Ceremade—Université Paris-Dauphine, Bureau B638
Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France

References

Karian, Z. & Dudewicz, E. (2000). *Fitting Statistical Distributions: The Generalized Lambda Distribution and Generalized Bootstrap Methods*. New York: Chapman & Hall.

Ramberg, J. & Schmeiser, B. (1974). An approximate method for generating asymmetric random variables. *Commun. ACM*, **17**, 78–82.

Su, S. (2007). Fitting single and mixture of generalized lambda distributions to data via discretized and maximum likelihood methods: GLDEX in R. *J. Statist. Software*, **21**, 9.

The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy

Sharon Bertsch McGrayne
Yale University Press, 2011, xiii + 320 pages, $27.50, hardcover
ISBN: 978-0-300-16969-0

Table of Contents

Readership: General public, students, and researchers alike.

The subtitle of the book, "How Bayes' rule cracked the Enigma code, hunted down Russian submarines and emerged triumphant from two centuries of controversy", tells it all: this is a collection of stories about the surprising survival of Bayes' theorem as a basis for statistical inference. As indicated above, this is primarily a general public book, hence it restrains from getting into technicalities, neither mathematical nor philosophical. Nonetheless, it is a highly enjoyable and entertaining book for statisticians and non-statisticians alike, even though the former may object at some interpretations. Some may also object to such a personification of science, but the book shows in the most lively manner how individuals contributed to the survival then success of the field. I stress that it is based on both documents and correspondence with major statisticians, thus highly trustworthy in its account.

While the book inevitably starts with the (incomplete) story of Bayes' life, the second chapter is about "the man who did everything", namely Laplace. How he attacked the issue of astronomical errors is brilliantly depicted, rooting the man within statistics and explaining why this "French Newton" would soon move to the "probability of causes", rediscovering Bayes' theorem. In the chapter about the dark ages of Bayesian statistics, I particularly liked the French connections, learning that Henri Poincaré himself testified at Dreyfus' trial with a Bayesian argument. The description of the fights of Fisher against Bayesians and non-Bayesians alike is as always both entertaining and sad. Sad also is the fact that Jeffreys (1939) ended up getting so little recognition at the time, despite its foundational features.

The following part is about Bayesian contributions to the war (WWII). The story of Turing (and Good) at Bletchley Park is well-told and, as always, I cannot but be moved by the waste of such a superb intellect, thanks to the stupidity of bureaucracies. The role of Madansky in the safety assessment of nuclear weapons is also striking, stressing the inevitability of a Bayesian assessment of a potential one-time event. The involvement of Tukey into military research (and a potential reason for his avowed anti-Bayesian views) was also a novelty for me, but not as much as his use of Bayesian small area methods for NBC election night previsions. Not so paradoxically, I ended up appreciating Chapter 15 even for the part about the search for the missing Palomares H-bomb, as it exposes the plusses a Bayesian analysis would have brought.

When approaching near recent times and contemporaries, Sharon McGrayne gives a very detailed coverage of the coming-of-age of Bayesians like Savage and Lindley, along with the important impact of Raiffa and Schlaifer, both on business schools and on modelling prior beliefs [via conjugate priors]: Raiffa & Schlaifer (1961) is a beautiful book that prefigures and inspired both DeGroot (1970) and Berger (1980). Similarly, while I'd read detailed scientific accounts of Mosteller's and Wallace's superb Federalist Papers study, the story is well-rendered in Chapter 12.

The final part, entitled Eureka!, is about the computer revolution we witnessed in the 1980s, culminating with the (re)discovery of MCMC methods covered, e.g. in Robert & Casella (2011). Because it contains stories that are closer and closer to today's time, it inevitably scatters into shorter and shorter accounts. However, *the theory that would not die* conveys the essential message that Bayes' rule has at last become operational, with its own computer language and objects like graphical models and Bayesian networks that could tackle huge amounts of data and real-time constraints, attracting companies like Microsoft and Google.

Obviously, I highly enjoyed reading this book. The only missing entry is in my opinion Jaynes (2003) and the specificity of the MaxEnt community, which could have deserved a chapter on its own. The book is written, edited, and printed with care, the references are numerous and relevant. A perfect Christmas gift for your fellow statisticians, and your non-statistician relatives as well!

Christian P. Robert: *christian.robert@ceremade.dauphine.fr*
Ceremade—Université Paris-Dauphine, Bureau B638
Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France

References

Berger, J. (1980). *Statistical Decision Theory and Bayesian Analysis*. New York: Springer.
DeGroot, M. (1970). *Optimal Statistical Decisions*. New York: McGraw-Hill.
Jaynes, E. (2003). *Probability Theory*. Cambridge: Cambridge University Press.
Jeffreys, H. (1939). *Theory of Probability, 1st ed*. Oxford: The Clarendon Press.
Raiffa, H. & Schlaifer, R. (1961). *Applied Statistical Decision Theory*. Cambridge, MA: MIT Press.
Robert, C. & Casella, G. (2011). A history of Markov chain Monte Carlo: subjective recollections from incomplete data. *Statist. Science*, **26**, 102–115.

A Historian Looks Back: The Calculus as Algebra and Selected Writings
Judith V. Grabiner
Mathematical Association of America, 2010, xv + 287 pages, $62.95, hardcover
ISBN: 978-0-88385-527-0

Table of Contents

Readership: Mathematicians and teachers of mathematics.

There is a fundamental difference between mathematics and science. This is that, once a mathematical truth has been discovered, it remains discovered. In contrast, scientific truths are always at risk of refutation. Having said that, the history of mathematics is full of conjectures and refutations, so that there are certainly similarities between the development of mathematics and the development of science. And it is because of this similarity that Judith Grabiner describes the history of mathematics as "not rationally reconstructible", adding "it must be the subject of empirical investigation". Furthermore, Grabiner claims that students often have difficulties similar to the mathematicians who originally elucidated key ideas and that "knowledge of the history of the relevant mathematical concepts can help the teacher understand what is troubling the student".

The first half of this book is an exploration of the argument that the work of Lagrange was central to making calculus rigorous by reducing it to algebra. The second half is a collection of Grabiner's writings, several of which have won awards, about the history of mathematics, certainly, but also about how and why mathematical concepts come to be, and how they are refined. The question posed by a fictional student at the start of Chapter 2 nicely captures the flavour of this aim: the student begins by asking what a speed of 50 miles per hour means, and, on being given an epsilon/delta definition as answer, asks "how in the world did anybody ever think of such an answer"? That last is the question on which this book seeks to shed some light.

This is a rather beautiful book (I meant in intellectual content, but it is also very nicely produced). It would make excellent reading for any serious student of mathematics, to show them the context in which the mathematics they are learning has developed – and how that development has influenced its form. The style and scope of the book are different from, but I put it alongside Courant and Robbins's *What is Mathematics?*, MacLane's *Mathematics Form and Function*, and Hersh's *What Is Mathematics, Really?* in showing the nature of mathematics.

It also shows how "[mathematicians] make false starts or go in the wrong direction. They can be right for the wrong reasons, like Lagrange. And the 'right' definitions often come only at the end". That is, it shows how even brilliant mathematicians struggle when getting to grips with new ideas. It is as inspiring as Sir David Cox's prefatory comments to the papers in *Selected Statistical Papers of Sir David Cox* (Cambridge University Press, 2005), in which he described how the ideas for the papers arose, and the struggles he had in clarifying them.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Knowledge Discovery from Data Streams
João Gama
Chapman & Hall/CRC, 2010, xix + 237 pages, £53.99/$83.95, hardcover
ISBN: 978-1-4398-2611-9

Table of Contents

Readership: Students and researchers new to streaming data mining, and researchers in any application domain which involves streaming data.

Streaming data are data which keep on coming, often described as "like water from a fire hose". Such data are a consequence of modern measurement technologies, in which electronic measuring systems capture the data automatically, requiring no human intervention. And the analysis of such data has become a hot topic. Streaming data present special challenges for data mining, including such issues as the need to process each data point only once, the need for adaptive updating, the need for fast and efficient algorithms to cope with a data rate which is often huge, and the need to cope with non-stationarity, both in the form of population drift and concept drift. The range of application domains in which the data are characterized by the above is wide, spanning network monitoring, web mining, sensor networks, telecomms, the financial markets, credit card transaction data, biomedical telemetry, etc.

As far as I am aware, this book is the first authored text (that is, not an edited collection) about the area, though four of the sections were contributed by authors other than Gama. It covers standard data mining methods (clustering, change detection, frequent pattern mining, anomaly detection, supervised classification, and so on), showing how they need to be adapted to cope with streaming data. A final chapter covers what the author calls "ubiquitous data mining", which arises when processors are embedded in devices and equipment all around us (mobile phones, intelligent meters, vehicles, even refrigerators and coffee machines). These constitute large scale distributed systems which interact beneficially with users. The book covers a lot of ground in just 200 pages, including discussion of relatively advanced methods such as wavelets, bagging, boosting, dynamic time warping, and symbolic representation of time series.

There is also, I was pleased to see, a chapter on evaluating streaming algorithms, which includes discussion of the design of evaluation experiments and evaluation metrics. Evaluation, in general, deserves more attention than it generally receives, so I was delighted to see the focus on it here.

The flavour is very much the mix of statistics, machine learning, computer science, and other areas which is characteristic of data mining.

Overall, I would describe it as providing a good introduction to an area of data analysis which is going to be very important indeed.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Train Your Brain: A Year's Worth of Puzzles
George Grätzer
A K Peters/CRC, 2011, xviii + 235 pages, $29.95/£18.99, softcover
ISBN: 978-1-56881-710-1

Table of Contents

Readership: Those who love to do mathematical puzzles.

From p. 232: "My collection of brain teasers relies heavily on the beautiful book *Mathematical Recreations* by M. Kraitchik (London, George Allen and Unwin Ltd, 1955. I've also borrowed a number of problems from the Hungarian *Mathematical and Physical Journal for Secondary Schools*". There are 140 puzzles in all, presented as 3 a week for 36 weeks and then 2 harder ones ("Black belt") for 16 weeks. There are 18 pages of hints and 131 pages of solutions. Here are three example teasers:

(p. 6) An Arab left 19 horses to 3 sons, the eldest to get a half, the middle one to get one quarter and the youngest one fifth. Naturally, the sons didn't want to kill horses. What did the "wise Quadi" suggest?

(p. 17) "Three jealous husbands want to cross a river with their wives in a two-person boat. None of the men, though, dares leave his wife alone with another man. Arrange the river crossing".

(p. 51) "The numbers 190,246,849 and 190,302,025 are the squares of two consecutive odd numbers. Without the use of a calculator or a computer, figure out the square of the even number that lies between the two odd numbers".

The puzzles have been re-scrutinized by two math experts Mária Halmos and Erika Kuczmann; these two have "refined nearly every puzzle and solution and corrected many errors".

This is a perfectly delightful book which will provide many hours of pleasure to the mathematically inclined child or adult. It would be a nice birthday present, for example.

Norman R. Draper: *draper@stat.wisc.edu*
Department of Statistics, University of Wisconsin—Madison
1300 University Avenue, Madison, WI 53706-1532, USA

A Guide to Plane Algebraic Curves
Keith Kendig
Mathematical Association of America, 2011, xv + 193 pages, $49.95, hardcover
ISBN: 978-0-88385-353-5

Table of Contents

Readership: Those who wish to study algebraic curves in the two-dimensional plane.

This attractive little book shows us how a number of interesting mathematical shapes in two dimensions are expressed in algebraic terms. Requirements for readers are "some basic complex analysis, including Cauchy-Riemann equations, complex analytic functions, meromorphic functions and Laurent expansions" and "definitions of field, field isomorphism, algebraic extension of a field, integral domain, ideal and prime ideal". (p. viii). The tone is conversational and interesting and the explanations are concise and clear. There are numerous fascinating diagrams. The small index of about 2.1 pages is unsatisfactory, however. For example, references to "rose" are listed from p. 96 on, omitting references in Chapter 1, on pages 10 and 27. Bernoulli's lemniscate (p. 131) isn't listed. The Witch of Agnesi is mentioned on p. 10, without further explanation. The many well-drawn figures (good) have no captions (bad). All in all then, this is a very nice, interesting book with a few technical flaws.

Norman R. Draper: *draper@stat.wisc.edu*
Department of Statistics, University of Wisconsin—Madison
1300 University Avenue, Madison, WI 53706-1532, USA

Introduction to General and Generalized Linear Models
Henrik Madsen, Poul Thyregod
Chapman & Hall/CRC, 2011, xiii + 302 pages, £39.99/$83.95, hardcover
ISBN: 978-1-4200-9155-7

Table of Contents

Readership: Students and researchers in statistics and various other areas where data analysis and statistical modelling are used.

This book presents a well-structured introduction to both general linear models and generalized linear models. The modelling approach used is based on the likelihood techniques, therefore it provides a flexible framework of the analysis and model building for problems based on Gaussian data and those linked to, for example, binary, positive, integer, ordinal and qualitative data, as the author aim. The book facilitates a useful comparison between the two classes of models, followed by mixed effects models and hierarchical models for both Gaussian and other data. A

number of real-world examples involving various types of data are included in an independent chapter. For most chapters, extensive illustrated examples, annotated remarks and R codes, following the definitions, methods and results, are given. Exercises are put at the end of the chapter. An author book webpage http://www2.imm.dtu.dk/~hm/GLM/ is created to offer the reader lecture notes, data and solutions. I would recommend the book as a suitable text for senior undergraduate or postgraduate students studying statistics or reference for researchers in areas of statistics and its applications.

Shuangzhe Liu: *shuangzhe.liu@canberra.edu.au*
Faculty of Information Sciences and Engineering
University of Canberra, ACT 2601, Australia

Handbook of Markov Chain Monte Carlo
Steve Brooks, Andrew Gelman, Galin Jones, Xiao-Li Meng (Editors)
Chapman & Hall/CRC, 2011, xxv + 592 pages, £63.99/$99.95, hardcover
ISBN: 978-1-4200-7941-8

Table of Contents

Readership: All who are interested in Bayesian inference with MCMC.

The handbook of Markov Chain Monte Carlo (MCMC) becomes the third volume in the recently created handbooks series of Chapman & Hall/CRC on Modern Statistical Methods. The editors

have designed 24 chapters, half of them focusing on theory, half on applications. The book gives an overview of current developments, but clearly not all interesting areas are covered uniformly. Topics I missed in the handbook are econometrics, Bayesian networks, (multivariate) time series and spatial statistics

MCMC is a new technology that has revolutionized empirical model fitting in statistics. Researchers at all levels of familiarity with MCMC will find novel morsels of material to chew on. In general, I found this to be a remarkable book on the current state of MCMC methods in statistics. Any newcomer to the field will appreciate the thoughtful collection of articles, all written by well-known people in the field (including some pioneers of MCMC), but also experts will find new aspects and the book as a valuable reference book.

The second part of the book (12 chapters) is mainly on applied topics and the last part of the book contains 3 contributions on applications: one on ecology (fisheries), one on rural-urban migration and one on hierarchical models in educational research. This might be the only weak part of the handbook, since despite the efforts of the editors to give a balanced cross-section of the field, because any short selection from constantly growing literature on MCMC applications is a rather arbitrary one.

A handbook on MCMC is a good idea, mainly because the field is expanding so fast and nobody has a complete understanding of all the current developments. Thus, to find a group of experts that contributes to a handbook is, in principle, a good idea. What speaks against the handbook is the high price and the most likely expectation that some parts of the book might be outdated soon. Thus, it would be much more desirable that handbooks like this one on MCMC is available on the internet and could be updated by the author or being commented. MCMC is still a work in progress, and as in any new field there exists the possibility that more is under way.

Besides a new algorithmic understanding, MCMC estimation of complex models requires a good knowledge on computational skills. “Can I simulate it then I understand it”. Thus, I would have liked to see some hints in some chapters, e.g. as what methods are likely to work and what aspects could be improved or are still an open problem.

After reading some of the articles application part I was surprised to find out how diversified the application of MCMC method have become in the two decades since the computational revolution started. Each of the fields chosen, like ecology (capture-recapture models), pedigree research, genomics, astronomy shows their own developments of MCMC toolboxes, so that it is rather difficult to see where the common computation techniques are, except that they are called MCMC: consisting of a Markov chain and random number generators. If the development continues with this tremendous speed we will see a new evolutionary branch of statistics in near future.

Impressive I found the following quote on p. 379 by B. Caffo *et al.*: “While our applications possessed only tens of thousands of parameters, current MRI and genomic technology puts the relevant number closer to millions”. A sentence being unthinkable in a statistics handbook some short time ago. The future of MCMC has just begun.

My expectation was that the articles in the handbook are a summary of certain acknowledged facts of modern MCMC technology. Alternatively, I expected that contributions of a handbook are reflecting the current state of the art. Both expectations were met plus some interesting insights on top of it.

Overall, I enjoyed reading the book, but I would feel much more comfortable if the (chapter) contributions of the book would be available in the Internet or could be found by search machines.

Wolfgang Polasek: *polasek@ihs.ac.at*
Department of Economics and Finance, Institute for Advanced Studies
Stumpergasse 56, 1060 Vienna, Austria

Monte Carlo Simulation for the Pharmaceutical Industry: Concepts, Algorithms, and Case Studies
Mark Chang
Chapman & Hall/CRC, 2011, xxiii + 539 pages, £57.99/$89.95, hardcover
ISBN: 978-1-4398-3592-0

Table of Contents

Readership: As specified by the author, the primary readers of this book are "industry statisticians, scientists and software engineers/programmers". I would add teachers looking for simulation exercises and motivating examples (though restricted here to the pharmaceutical context) and curious people interested in an overview of the drug production process.

First of all, I would like to say that this book is well-written and easy to read.

As is obvious from the table of contents, the author first motivates reader's interest on simulation techniques by listing problems as varied as possible which can be solved by a Monte Carlo simulation tool. Chapter 2 gives a quick introduction to simulation techniques while Chapter 3 is an overview on drug development. Every following chapter (from 4 to 12) deals with one particular aspect of the drug development (from pharmacoeconomic to pharmacokinetic and pharmacodynamic).

As stressed by the author, this book is original (and maybe unique) in the sense that in order to convince the pharmaceutical industry of the importance of Monte Carlo simulation, every single step in the drug development is considered from that perspective.

From an occasional medicine consumer's point of view, I found this book quite interesting, in particular Chapter 3 which describes the whole drug development process in a very instructive way. All the aspects seem to have been covered and this concern of comprehensiveness has to be stressed.

From a teacher's point of view, I see this book as a source of possible exercises and motivating problems. Chapter 2 supplies an enumeration of algorithm to sample from standard distributions. Note that the Figure 2.2 on page 76 gives a known, unreadable but always impressive synthetic view of the relations between the standard probability distributions.

From a statistician's point of view, I wonder if the objective of this book was not too ambitious which prevents the author from proposing a practical tool.

Indeed, no comments are given on the prerequisite mathematical skills. I think clarifying this point would have helped understanding the objective of the book.

On the one hand, the methods for sampling standard distributions presented in Chapter 2 are good exercises and require a low level in probabilities. However, they are useless in practice since statistical software such as R provide optimized random number generators. Moreover, some of the proposed methods are really naive: page 19, in order to model a competitive phenomenon, a minimum of exponential distributions is simulated without using the fact that it results into an exponential distribution with the sum of the respective parameters.

On the other hand, the readers are introduced to complex algorithms such as Monte Carlo Markov Chain in only 4 pages: a novice in Monte Carlo simulation will not measure their importance whereas an initiate won't need these pages.

Finally, I would say that this book is worthwhile reading as a long (540 pages) introduction to Monte Carlo simulation and its eventual application in pharmaceutical industry. It can convince people to consider this methodology but a real application would need a true implication and reading specialized books.

Sophie Donnet: *donnet@ceremade.dauphine.fr*
Ceremade—Université Paris-Dauphine, Bureau B627
Place du Maréchal de Lattre de Tassigny, 75775 Paris Cedex 16, France

Theoretical Statistics: Topics for a Core Course

Robert W. Keener
Springer, 2010, xvii + 538 pages, €89.95/£81.00/$99.00, hardcover
ISBN: 978-0-387-93838-7

Table of Contents

Readership: Graduate students in statistics.

This volume provides an excellent course in the mathematical theory underlying statistical ideas and methods, for advanced (meaning postgraduate) students. The amount of material covered is indicated by the fact that it evolved from a three-semester sequence of courses given by the author. Its suitability as a course text is materially aided by very extensive exercises, along with solutions to selected exercises. Anyone who works through this book will end up with a first class understanding of the mathematical ideas underlying modern statistical concepts and methods.

However, . . . I used the adjective "mathematical" above, and I think that adding that to the title would more properly represent the content than its current title. There are aspects of statistics which have equal justification to be called "core theoretical statistics" to those it does cover—such as the challenges of observational data, issues of causality, the relative properties and merits of different schools of inference, issues concerned with modern application areas such as small n, large p problems, and (okay, perhaps understandably, since one cannot cover everything) experimental design. Neither does it have much to say about modern computation methods—the bootstrap and empirical Bayes are covered, but what about ensemble methods and permutation tests for example. Likewise, unless I missed it in the 538 pages, it does not distinguish hypothesis testing from significance testing (indeed, the word "significance" does

not appear in the index—though this is probably an editorial oversight, since I did spot it in the text itself). There remains considerable confusion in the scientific (even statistical) literature about how *p*-values should be interpreted, and one might have expected a book on "theoretical statistics" to cover this.

So I do have a criticism, but this is really merely about the extent to which the title describes the content. For what it is—a solid grounding in the mathematical basis of theoretical statistics—this volume is first class.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Testing Statistical Hypotheses of Equivalence and Noninferiority, Second Edition
Stefan Wellek
Chapman & Hall/CRC, 2010, xvi + 415 pages, £66.99/$104.95, hardcover
ISBN: 978-1-4398-0818-4

Table of Contents

Readership: Researchers or graduate students in pharmaceutical or clinical trials.

Equivalence and noninferiority trials have become increasingly important in recent decades: the authors note that "nowadays, at least half of the prescription drug units sold in the leading industrial countries are generic drugs that have been approved to be marketed on the basis of some bioequivalence trial". Equivalence and noninferiority trials are experiments which explore whether a proposed new treatment is of similar effectiveness to or no worse than an existing treatment. The use of such trials avoids the ethical issues of exposing patients to a placebo treatment when there is an existing effective treatment. Subtleties arise from the fact that in an equivalence trial the aim is one of verifying rather than rejecting the hypothesis of no difference—or, more accurately, of no difference of clinically significant size. It is not correct to infer equivalence of treatments from a nonsignificant traditional two-sided test because, as Altman and Bland put it in 1995, "absence of evidence is not evidence of absence".

Chapters 1, 2, and 3 lay the foundations of such tests, and later chapters describe examples of equivalence and noninferiority tests in a large number of different contexts and scenarios, so that researchers with a particular type of problem can jump straight to a discussion of that type. This second edition expands the first by including more discussion of noninferiority testing. The book is liberally illustrated with real examples, and includes an appendix describing software routines (in R, SAS, etc.) where programs for the different calculations can be found.

This book will clearly be an extremely valuable practical guide for researchers faced with such problems, and would also provide a good introduction to anyone new to the special challenges of equivalence and noninferiority testing.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Music Data Mining
Tao Li, Mitsunori Ogihara, George Tzanetakis (Editors)
Chapman & Hall/CRC, 2011, xxiv + 360 pages, £57.99/$89.95, hardcover
ISBN: 978-1-4398-3552-4

Table of Contents

Readership: Researchers and graduate students in acoustics, computer science, electrical engineering, music, statistics, and other related fields who are interested in music data mining and information retrieval.

Music information retrieval (MIR) is the small but growing scientific study of retrieving and analyzing information from music. Being still a young field of research, most studies are only presented in conference proceedings and, more sparsely, in a variety of journals. This book, as a collection of papers, brings together some of the leading scholars of the field to tackle a number of data mining techniques aiming mainly at data classification.

Readers new to the elementary music data mining and audio feature extraction techniques are encouraged to gear up for the subsequent chapters by browsing through the two introductory chapters. They summarize some previous research related to the subject and introduce some key concepts and terminology every modern MIR researcher should be aware of. The remaining nine chapters are on more specific MIR topics. Each chapter being somewhat self-contained with separate bibliographies at the end, it is unnecessary to read these nine chapters in strict order. Instead, one may choose to read any chapter one finds appealing without a constant need to look for explanations elsewhere.

There is a bunch of musical notes and symbols on the cover of the book, which might lead the reader to think this is a book on symbolic MIR. However, the majority of the eleven chapters deal with the auditory domain of MIR with virtually only two of them bringing some aspects of the symbolic domain into question. Personally, I find this lack of eclecticism a little disappointing because, according to my way of thinking, the essence of music lies in its high-level features such

as pitch and meter rather than in the low-level ones such as timbre. I understand the availability of musical data in the form of digital audio file collections together with the increased computing power has facilitated the analysis of these low-level features. Nevertheless, a book of this kind should not leave the amount of exploration of basic MIDI file collections to as modest level as this book does.

The book is not technically adroit, meaning that a basic knowledge of mathematics, statistics, or signal processing is sufficient to follow the book. Naturally, this is also a drawback: the presentation, although being direct and mostly engaging and coherent, is not as detailed as a researcher or graduate student would like it to be. For a researcher already working on the field, there are some novel results such as those in the sixth chapter, yet they are not that substantial contributions to the field in my opinion.

It is worth mentioning that there are many similar paper collections being published in the field every year. Only few MIR monographs, though, have been published so far, so it would be nice to see one on the book's theme to be published soon.

Joonas Kauppinen: *joonas.kauppinen@uta.fi*
School of Information Sciences
FI-33104 University of Tampere, Finland

Principles of Uncertainty
Joseph B. Kadane
Chapman & Hall/CRC, 2011, xxvii + 475 pages, £57.99/$89.95, hardcover
ISBN: 978-1-4398-6161-5

Table of Contents

Readership: Graduate students and researchers of statistics, mathematics and economics.

This text provides a unique blend of theory, methods, philosophy and applications that is suitable for a course in Bayesian probability and statistics. An elementary course might use Chapters 1, 2, Sections 1, 2, and 6–10 of Chapter 3, and Sections 1–6 of Chapter 4. The book provides also enough material for an advanced probability course or a Bayesian course. This book introduces a student to the relevant parts of mathematics, computing and philosophy, as they bear on statistics. The book is useful to scholars of all persuasions who may find there alternative interpretations of probability and interesting examples. Each of the first few chapters begins by introducing a new concept or assumption. The rest of those chapters explores the consequences of that new assumption and provides thought-provoking material for teaching. Some few sections are mathematically highly technical or specialized. The text contains a summary and exercises at the end of each section.

Erkki P. Liski: *erkki.liski@uta.fi*
School of Information Sciences
FI-33104 University of Tampere, Finland

Modeling and Reasoning with Bayesian Networks
Adnan Darwiche
Cambridge University Press, 2009, xii + 548 pages, £58.00/$99.00, hardcover
ISBN: 978-0-521-88438-9

Table of Contents

Readership: Advanced or graduate students in statistics, artificial intelligence, and related areas. Professional statisticians, and researchers in other disciplines who would like to model their data and make inferences using such structures.

The field of Bayesian networks is a hybrid one, representing a confluence of ideas from statistics and artificial intelligence. Such networks are a very important subclass of more general graphical model structures, permitting inference in situations involving many, sometimes very many, variables. From a statistical perspective, the network structure is a way of representing and understanding joint probability distributions. From a computational perspective, so that inferences can be made using this structure, it is necessary to develop practical algorithms, both for construction of the networks, and for propagating probabilities through them. Recent years have seen significant advances in these areas, to the extent that highly effective methods now exist. Given the continuing growth of large data sets and increasingly complex problems (and this is not going to ease up), Bayesian networks should be a potential tool in every modern statistician's toolbox.

The book assumes no prior knowledge; it outlines any necessary ideas of logic as well as probability from a perspective best matched to the ideas to be presented. With just one or two exceptions, discussion is restricted to networks of discrete variables with a finite number of values, and to directed graphs. I particularly enjoyed the opening chapter, which provides a concise and clear historical summary, setting the development of such networks in a broader context.

The book is clearly written. Although the authors do say that it is intended to be a focused thorough treatment rather than encyclopaedic, to me it seemed fairly comprehensive. Proofs are included, but are placed at the end of the chapters so as not to disrupt the flow of ideas, and exercises are also given at the end of each chapter.

In all, the clarity, continuity, and depth of the presentation mean that his would make a first class course text, as well as serving as a very useful reference work. I shall certainly recommend it for teaching purposes, and doubtless refer to it to remind myself about particular aspects of such models.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Empirical Model Building: Data, Models, and Reality, Second Edition
James R. Thompson
Wiley, 2011, xvii + 430 pages, £83.50/€100.20/$125.00, hardcover
ISBN: 978-0-470-46703-9

Table of Contents

Readership: Postgraduate or advanced undergraduate statisticians and modellers. Applied statisticians.

This is a curious book. It is not a text in the traditional sense, in that it does not lead the reader through a progressive development of empirical model building. Rather it discusses a series of areas in which the author has worked, to provide examples of and insight into how to build mathematical and statistical models. It appears to be a very substantial extension of the (1989, 242 pages) first edition, with fourteen chapters taking the place of the previous five. Some of the chapters in the new edition appear to be expanded and split versions of earlier chapters, but others are entirely new additions.

I can see two main uses for this volume (in addition to the traditional one of propping open a door of course). The first is that it would provide really excellent supplementary material for a more traditional course on modelling, both for supporting reading and to provide a source of case studies and examples. The breadth of areas of application will demonstrate clearly how ubiquitous are statistical and mathematical modelling, and the clear expositions will enable students to work through it with little assistance. There are also problems at the end of each chapter, so that students can be asked to do more than merely read.

As for the second use, I once wrote a review in which I suggested that the volume then in question would make a superb birthday present for a young statistician, since it was full of stimulating illustrations—like a more advanced version of a science popularization book. I was going to say the same about this volume, but hesitated when I recalled that a colleague, on reading the earlier review, had remarked that they felt sorry for the people I sent presents to. Nonetheless, I do think a young (or indeed, even old) statistician or mathematical modeller would really enjoy this book.

It is a pity that the first typo occurs as early as the ninth word of the preface—though this is presumably the editor's fault rather than the author's. But I suppose I cannot criticise, having published a book which includes a typo in the title on the spine.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Biosurveillance: Methods and Case Studies
Taha Kass-Hout, Xiaohui Zhang (Editors)
Chapman & Hall/CRC, 2011, xv + 363 pages, £59.99/$89.95, hardcover
ISBN: 978-1-4398-0046-1

Table of Contents

Readership: Biosurveillance researchers or readers wishing to see examples of cutting edge applications in the area.

The origins of biosurveillance can be traced back to John Snow's 1854 London cholera investigation and beyond, but recent advances in computer and electronic data capture technology are revolutionizing things. This book provides a collection of examples of such advances, ranging from simulating biosurveillance data, through the role of animals and zoos, via prediction markets, to aberration detection using R software. Every chapter is based around real cases with which the various authors have been concerned, so that this is not an abstract technical volume but one which is solidly grounded in biosurveillance practice. With the exception of a section in Chapter 3, it is not concerned with statistical methodological details of biosurveillance (such as are described in Lawson and Kleinman's *Spatial and Syndromic Surveillance*, for example) but rather with higher level issues of building and managing surveillance systems.

Different scientific domains have different attitudes to making data publicly available. The second chapter of this volume notes that biosurveillance data are difficult to gain access to, unless one works in a biosurveillance research group, and describes the simulation of such data. But other chapters describe some of the exciting new sources of data, including SMS text messaging, remote sensing, and even rumour-based information sources.

Given the high level and case-based nature of this volume, it would make excellent background or motivational reading for advanced students entering the area. It provides up-to-date illustrations of where this fast-developing field is now.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Stochastic Processes
Richard F. Bass
Cambridge University Press, 2011, xv + 390 pages, £45.00/$75.00, hardcover
ISBN: 978-1-107-00800-7

Table of Contents

Readership: Beginning graduate students and researchers from applied disciplines.

This book is pitched at a slightly more advanced level than, for example, Ross's *Stochastic Processes* or Grimmett and Stirzaker's *Probability and Random Processes*—as is illustrated by the fact that it assumes, as prerequisites, "a sound knowledge of basic measure theory and a course in the classical aspects of probability".

The first eight chapters of the book introduce basic notions, including Brownian motion and martingales. Later groups of chapters then outline stochastic calculus, describe jump processes, Markov processes, stochastic differential equations, financial and filtering applications, probability measures on metric spaces, and give important examples.

The amount of material is illustrated by the fact that the authors describe it as containing "far too much material . . . to cover in a single semester, and even too much for a full year". However,

the flip side of that is that it does provide a deep, comprehensive, and detailed exposition of the fundamentals of stochastic processes. There are exercises following each chapter.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Controversial Statistical Issues in Clinical Trials
Shein-Chung Chow
Chapman & Hall/CRC, 2011, xix + 591 pages, £63.99/$99.95, hardcover
ISBN: 978-1-4398-4961-3

Table of Contents

Readership: Scientists in the pharmaceutical industry, medical/statistical reviewers in government regulatory agencies, and researchers and students.

This volume deals with an important area—issues in clinical trials research which are perhaps not fully resolved. While some of the topics certainly are controversial (traditional Chinese medicine, for example, the subject of Chapter 23, of which more below), I think that "controversial" might be too strong a word to describe many of the topics dealt with, and "unresolved" or even simply "misused" might be more appropriate alternatives.

As the contents list shows, it is wide ranging, covering all aspects of clinical trials, and has excellent links and references to regulatory aspects. It will therefore provide a useful reference work for clinical trials researchers. However, the style is at times rather unhelpful (e.g. page 24: "Statistical inference on a parameter of interest of a population under study is usually derived under the probability structure of the parameter", "It is suggested that randomization be performed using an appropriate randomization method under a valid randomization model according to the study design to ensure the validity, accuracy, and reliability of the derived statistical inference") and even misleading (e.g. page 203: "a statistical difference means that the difference is not by chance alone and it is reproducible").

For at least some of the topics dealt with it would be necessary to follow up the description here with a more detailed and careful reading elsewhere. Indeed, the author explicitly states that the aim of the book is to "post commonly seen controversial issues rather than provide resolutions".

In Chapter 23, noting that it can take more than twelve years to bring a promising compound to the market, and that only a tiny fraction of proposed compounds finally achieve regulatory approval, the author suggests (page 443) that "as a result, an alternative approach for drug discovery is necessary" and that this has led to the study of traditional Chinese medicines. These are defined as "a Chinese herbal medicine developed for treating patients with certain diseases as diagnosed by the four major Chinese diagnostic techniques of inspection, auscultation and olfaction, interrogation, and pulse taking and palpation, based on the traditional Chinese medical theory of global dynamic balance among the functions/activities of all the organs of the body". The chapter goes on to say "unlike evidence-based clinical research and development of a Western medicine, clinical research and development of a [traditional Chinese medicine] is usually experience-based with anticipated variability due to a subjective evaluation of the disease under study. The use of [traditional Chinese medicine] in humans for treating various diseases has a history of more than 5000 years and yet no scientific documentation is available regarding clinical evidence of safety and efficacy . . . ". This topic will certainly conform to the title's "controversial issues".

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

Modern Issues and Methods in Biostatistics
Mark Chang
Springer, 2011, xiv + 307 pages, €79.95/£72.00/$89.95, hardcover
ISBN: 978-1-4419-9841-5

Table of Contents

Readership: Researchers, biostatisticians, statisticians, and scientists who are interested in quantitative analyses.

This is a first class book. It discusses a wide range of deep issues in statistics, and although focused on topics arising in biostatistics, pharmaceuticals, and clinical trials it would make stimulating and thought-provoking reading for any statistician.

It is lucid and yet comprehensive in the issues it discusses. For example, the opening chapter on multiple hypothesis testing strategies is the clearest concise (less than thirty pages) description I have seen. Despite the chapter's brevity it manages to cover the main principles (such as union-intersection and intersection-union testing, the closure and partitioning principle, coherence and consonance, and others) as well as all the major procedures (too many to list here), and various kind of error rate (for example, familywise error rate, generalized familywise error rate, false discovery rate, and so on). That it does all this in a clear and straightforward manner in so short a space is nothing short of remarkable.

The other chapters are similarly impressive. Most chapters include a section on controversies and challenges.

The only aspect I could find to criticise is that I was misled by the section heading in each chapter of "Further reading and exercises" since these are merely lists of references. Perhaps a future edition could expand "further reading" into a few lines of guidance, so that readers do not have to guess which of the references will be enlightening.

There are exercises at the end of each chapter, and I am certainly tempted to use the book as the basis for a short course for beginning postgraduate students (in statistics, not merely biostatistics), since it would open their eyes to some challenging and indeed fascinating aspects of modern statistics.

David J. Hand: *d.j.hand@imperial.ac.uk*
Mathematics Department, Imperial College
London SW7 2AZ, UK

The R Primer
Claus Thorn Ekstrøm
Chapman & Hall/CRC, 2012, xii + 287 pages, £25.99/$39.95, softcover
ISBN: 978-1-4398-6206-3

Table of Contents

Readership: Statisticians at all levels.

The programming language R, tailored for statistical analysis and graphics, is widely used, immensely useful through the availability of a vast number of specialized packages, and can be downloaded for free. The present book introduces R by means of a collection of very concrete examples, rather than attempting to explain each feature in full generality. In my experience, starting from concrete examples is the easiest and preferable way to learn a new programming language. This makes the book a nice starting point for learning R, and suitable for self-study provided the reader has some background in statistics. There are, however, several other introductory texts on the market, and it is not clear to me whether Ekstrøm's book is preferable to, e.g. his Danish compatriot Peter Dalgaard's book "Introductory Statistics with R" (2nd ed., Springer, 2008) which is a bit more generous with discussions of the statistical theory involved. In the end it boils down to a matter of taste.

Olle Häggström: *olleh@chalmers.se*
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg, Sweden

Origami 5: Fifth International Meeting of Origami Science, Mathematics, and Education
Patsy Wang-Iverson, Robert J. Lang, Mark Yim (Editors)
A K Peters /CRC, 2011, xiv + 646 pages, £43.99/$69.95, softcover
ISBN: 978-1-56881-714-9

Origami Inspirations
Meenakshi Mukerji
A K Peters, 2010, xi + 120 pages, £16.99/$24.95, softcover
ISBN: 978-1-56881-584-8

(The book contains over 36 models ranging from simple to intermediate.)

Readership: The proceedings volume addresses all aspects of origami techniques that can be used potentially in science, engineering and mathematical education. The second book is for origami folders interested in geometric solids.

When I first visited Japan in 1983 I was curious to learn how origami was taught and developed in Japan. Origami is a Japanese word and means literally folding paper. My search for a connection of origami with mathematics was not successful at that time. I wished that I could have seen a book or even some articles as they are published now in this book by the annual "Origami Science" meetings.

In recent years, the theory of origami has been moving in the direction of science. When I studied origami in the Japanese classes of Akira Yoshizawa (http://en.wikipedia.org/wiki/Akira_Yoshizawa), I was told that origami is an art. But it is also seen as a symbol of peace, starting at kindergarten and a teaching device for kids, and a construction device for mathematical or architecture students. Thus, origami reveals many aspects depending on your point of view. Until recently, the scientific view of origami was rather underdeveloped and the OSME proceedings might change this.

The proceedings volume is the fifth of such meeting, 5OSME (July 13–17, 2010, Singapore Management University. The Contents lists the 4 main parts of the book:

I. Origami History, Art, and Design
II. Origami in Education
III. Origami Science, Engineering, and Technology
IV. Mathematics of Origami

There are contributions for teaching origami in mathematics classes or for using paper as a technical material. A study on (polygonal) tape knots. Learn about ideas how on micro-scale folding is done using thin metal films. Have you heard of Haga's theorem? There is a proof for trisecting a line and angles with origami papers. All to be found in part IV of the book: "Mathematics of Origami".

Part I of the book gives a good overview of what can be done today at the interface of origami and science. Creativity in this new branch of a "science between the dimensions" can be seen everywhere: Learn about Oribotics, Twirl Design, Curved Corrugations, or Snapology Technique. Part II, on "Origami in Education", is less entertaining than the other parts. Part III describes applications of Origami (in the wide sense) to micro and macro problems in technical sciences.

As mentioned, this book is a conference volume and full of papers that relate to all these aspects of origami that are listed above. You get the feeling that this is just the starting point of the Science of Origami and that its real development has not yet begun. Origami has become a catalyst for many interdisciplinary ideas and projects.

The second book, by Meenakshi Mukerji, does not cover scientific aspects but is a book on how to create geometric objects out of series of similar two-dimensional geometric origami pieces. The book starts with a "simple" task: make a cube out of 6 similar pieces that can be inter-joined into appropriate flaps at the corners of the pieces. The next chapters of the book go on to explain how more difficult geometric objects, like dodecahedras, can be put together in a similar fashion. When I tried to do this with a 10 year old boy, he lost patience very soon, even if only 6 identical pieces had to be made.

This points to an important aspect of origami: Many people are fascinated by the beauty of these complicated geometric origami object, but the completely underestimate the time and the skills that goes into such objects. (A similarly fascinating book is Gjerde, 2009). Mathematicians might be additionally wondering what theory goes behind the goal to create the three-dimensional

objects from a plain sheet of paper. Clearly such questions of meta-origami are not touched in the present origami books or the conference volume. But, eventually it might be the beginning of a new future development.

Wolfgang Polasek: *polasek@ihs.ac.at*
Department of Economics and Finance, Institute for Advanced Studies
Stumpergasse 56, 1060 Vienna, Austria

Hints and References

Interested readers can look at the website of the "inspiration" book: http://www.origami-resource-center.com/origami-inspirations.html

Gjerde, E. (2009). *Origami Tessellations: Awe-Inspiring Geometric Designs*. Wellesley, MA: A K Peters. http://www.origamitessellations.com

Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery
Graham Williams
Springer, 2011, xx + 374 pages, €54.95/£49.99/$64.95, softcover
ISBN: 978-1-4419-9889-7

Table of Contents

Readership: Data analysts whose focus is on data manipulation, discrimination and classification tasks, who require a Graphical User Interface to abilities in R packages; students with a strong computing background doing "data mining" courses.

This text is a manual for the impressive Rattle graphical user interface (GUI) for R, describing both the use of the GUI and the R code that is invoked to carry out the computations. It documents, not just the details of use of R, but the details of what happens "under the hood". An important feature of Rattle, which it shares with other R GUIs, is that the R code is available for inspection and/or adaptation to handle tasks for which Rattle does not make direct provision.

Data analysts with limited or no previous experience with the R system, and who want to use it for the types of analysis for which Rattle caters, are likely to find Rattle a helpful tool that will allow them to quickly become productive with R. Those who in due course gain some reasonable degree of comfort with use of the command line are likely, as time progresses, to move increasingly from the GUI to the command line.

There is extensive useful practical advice on data preparation and data manipulation. The language and style will in places be unfamiliar to statisticians who have not been exposed to the data mining literature. The methods ("models") discussed are those that are prominent in the data

mining literature—cluster analysis, association analysis, and classification. Classical regression, including logistic and Poisson regression, although available in Rattle, are not discussed. The training/validate/test approach to model building, with separate data performing these separate roles, is emphasized.

The Rattle package is well suited for use in intermediate level courses on regression or classification. While this text provides useful background and information on the mechanics of handling computing and data manipulation, supplementary material on relevant statistical theory would be very desirable. The challenge is to find a good marriage between statistical theory, the cross-validation and training/validate/test approaches that are emphasized in texts such as this, and computer-intensive methods more generally.

John H. Maindonald: *john.maindonald@anu.edu.au*
Centre for Mathematics & Its Applications, Building 27
Australian National University, Canberra, ACT 0200, Australia

R for SAS and SPSS Users, Second Edition
Robert A. Muenchen
Springer, 2011, xxviii + 686 pages, €89.95/£81.00/$99.00, hardcover
ISBN: 978-1-4614-0684-6

Table of Contents

Readership: Researchers using R, especially those using R in place of or in addition to SAS or SPSS.

This is a greatly expanded second edition of a text that has already proved widely popular. The explanation is careful and detailed. It uses SAS and SPSS terminology, matching it with R terminology, in the text, in the Table of Contents, and in the Index. A glossary translates R terminology into terminology that is likely to be more familiar to SAS and SPSS users. There is a comparison of major attributes of SAS and SPSS with those of R.

The focus is on running and using R, on programming in R, on data acquisition and manipulation, and on graphics. One chapter only out of eighteen is explicitly devoted to statistical methods. This is devoted mainly to statistical tests, with a short section on regression.

The R system has two major packages that offer a structured approach to flat screen or hard copy graphics—*lattice* and *ggplot*2. This text has an extended and helpful account of ggplot2 graphics. The only substantial reference to lattice is a table that compares R's traditional (base) graphics with lattice and ggplot2. Arguably, this over-rates the consistency of ggplot2 while under-rating that of lattice. More seriously, it does take note of the layering features, in the separate *latticeExtra* package, that give lattice the same abilities for building plots piece by piece that are available in ggplot2. It is inevitable that a book that is so wide in its coverage will overlook a few such matters.

This is a wide-ranging and carefully compiled source of information on R. It is a strongly recommended addition to the library of anyone who comes to R from SAS or SPSS. It is supplemented by a wide-ranging and informative web site.

John H. Maindonald: *john.maindonald@anu.edu.au*
Centre for Mathematics & Its Applications, Building 27
Australian National University, Canberra, ACT 0200, Australia

Numerical Ecology with R
Daniel Borcard, François Gillet, Pierre Legendre
Springer, 2011, xi + 306 pages, €54.95/£49.99/$64.95, softcover
ISBN: 978-1-4419-7975-9

Table of Contents

Readership: Students and researchers in biostatistics and numerical ecology.

At the risk of a sweeping generalization, many ecology-oriented people are strangely reluctant when it comes to quantifying nature and using mathematical tools to help understand it. In order to improve the situation, and advance our understanding of the complexities of ecological science, it is necessary to overcome this reluctance.

The authors' text helps to redress this imbalance and is meant as a companion when working at a computer. The text is lavishly illustrated with examples throughout and relies on two main data sets that are readily available—one on fish data, the other on mite data.

The text is well written and ideal as either a course companion or for personal study. Each chapter explains clearly the objectives underlying the presentation of the material considered and each includes a final conclusion section so that the key points can be referred to subsequently. The authors' text deserves to become a standard reference for anyone working in ecological science and more specifically, in numerical ecology.

Carl M. O'Brien: *carl.obrien@cefas.co.uk*
Centre for Environment, Fisheries & Aquaculture Science
Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

Modelling and Quantitative Methods in Fisheries, Second Edition
Malcolm Haddon
Chapman & Hall/CRC, 2011, xvi + 449 pages, £39.99/$79.95, hardcover
ISBN: 978-1-58488-561-0

Table of Contents

Readership: Students of undergraduate courses in biology, marine ecology and statistics.

It is ten years since the first edition of this book appeared and I reviewed it. The text remains true to the author's initial aim of providing an introduction to the analytical methods currently being used in quantitative biology and fisheries science. It is important to remember when reading this book that there are few texts that students can truly consult on fisheries science without a detailed understanding of stock assessment and fisheries management practices—this text continues to bridge that gap.

The material has been revised and improvements made to a number of the examples. Two concerns and reservations that I commented on in my previous review have been addressed by the inclusion of two new chapters—one on characterizing uncertainty covering asymptotic errors and likelihood profiles, and the other on size-based models using abalone as an example. The book is lavishly illustrated throughout with the use of Microsoft Excel workbooks which adds to the flexibility, availability and ease of use of the text. I recommend the text both as a course companion and for private study.

Carl M. O'Brien: *carl.obrien@cefas.co.uk*
Centre for Environment, Fisheries & Aquaculture Science
Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

Composite Sampling: A Novel Method to Accomplish Observational Economy in Environmental Studies

Ganapati P. Patil, Sharad D. Gore, Charles Taillie
Springer, 2011, xiii + 275 pages, €99.95/£90.00/$129.00, hardcover
ISBN: 978-1-4419-7627-7

Table of Contents

Readership: Students and researchers in statistics, ecology and environmental science.

Composite sampling is an approach that few may encounter during their formal studies and is often employed to accomplish observational economy in a variety of environmental studies—what is desirable is not affordable, and what is affordable is not adequate.

In this monograph, the authors present the work of a 4-year initiative undertaken in the United States and discuss several previously unpublished results and applications of composite sampling in environmental studies. The examples are varied and real, and the theory presented is sufficient to understand the formal analytical approaches and the solutions presented. The text may only appeal to a limited readership but is well worth reading.

This is by no means a standard text and I found the best way to read the book was to consider the individual chapter examples in isolation first and then refer back to the relevant theory chapters. This is unlikely to become a standard student text and will appeal to those practitioners needing to balance the desire for statistical rigour in sampling design with the constraints of ever decreasing financial budgets with which to undertake such sampling.

Carl M. O'Brien: *carl.obrien@cefas.co.uk*
Centre for Environment, Fisheries & Aquaculture Science
Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

Choice-Based Conjoint Analysis: Models and Designs
Damaraju Raghavarao, James B. Wiley, Pallavi Chitturi
Chapman & Hall/CRC, 2011, xi + 180 pages, £59.99/$94.95, hardcover
ISBN: 978-1-4200-9996-6

Table of Contents

Readership: Researchers in psychology, economics, geography, marketing, tourism, human resources administration, health administration, and the political and social sciences.

It is a pleasure to review this book and to note the advances in the subject since my initial encounters with the techniques as a post-graduate student in the early 1980s. Conjoint analysis and discrete choice experimentation are tools developed since the 1960s for understanding how individuals develop preferences for alternatives; i.e. the choices that lead one person to purchase a particular brand of washing machine in preference to another, or to purchase a particular make of car when there is little to choose between their purchase prices.

In real life, people reveal their preferences through choices. The authors deal with situations for which choice alternatives may be described in terms of their components or attributes. The emphasis throughout the text is on developing practical designs for specific classes of problems. Often, if not usually, economics becomes important but may be difficult to disentangle from other factors—not independent but intimately linked.

For those already familiar with the subject, the text is well worth adding to their book collection but for others will no doubt remain a mystery until they need to consult and use the models presented by the authors.

Carl M. O'Brien: *carl.obrien@cefas.co.uk*
Centre for Environment, Fisheries & Aquaculture Science
Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK

Forest Analytics with R: An Introduction
Andrew P. Robinson, Jeff D. Hamann
Springer, 2011, xv + 339 pages, €54.95/£49.99/$64.95, softcover
ISBN: 978-1-4419-7761-8

Table of Contents

Readership: Students, researchers and forestry practitioners.

The material presented in this text is more than sufficient for a dedicated module of an applied statistics course, as the subject of forestry provides a rich environment for data analysis and poses many questions for which statistics, econometrics, and applied mathematics tools can play a constructive and informative role. The need for pragmatism in many real world problems necessitates an open and flexible approach to the use of formal numerical approaches.

The authors develop, and demonstrate, solutions to common forestry data handling and analysis challenges—drawing upon applied statistics, forest biometrics, and operations research for their inspiration and solutions. The open-source R programming language provides the platform upon which the authors demonstrate solutions to the many and varied problems presented. Whilst much of the text may be regarded as standard for the topic, the last chapter addresses an area harvest strategy which is well worth reading on its own without the remainder of the book.

The text is well written, easy to read and I recommend it to anyone interested in biometrics.

Carl M. O'Brien: *carl.obrien@cefas.co.uk*
Centre for Environment, Fisheries & Aquaculture Science
Pakefield Road, Lowestoft, Suffolk NR33 0HT, UK