J. R. Statist. Soc. A (2004)
167, *Part* 1, *pp.* 183–198

Book reviews

Books for review

If you would like to review a book, and thereby to retain it for your collection, please contact the Book Reviews Editor, whose details can be found by clicking on 'books currently available' in the information on the Royal Statistical Society's Web site:

`http://www.rss.org.uk/publications/`

Reviews in this issue

 0964–1998/04/167183

Bioinformatics: the Machine Learning Approach, 2nd edn
P. BALDI AND S. BRUNAK, 2001
Cambridge, MIT Press
xxii + 452 pp., £36.95
ISBN 0-262-02506-X

This hardback book is a key resource for those who are involved in the application of computer science to biological data. In the bioinformatics evolution, the rapidly expanding databases are coupled with the increasingly sophisticated computer systems, and framed in a probabilistic perspective. The aim of the book is to provide the methods of machine learning algorithms, which are based on Bayesian models and inference, so that interpretation can be enhanced in areas of biology where there is little theory but plenty of data. Neural networks and hidden Markov models are the machine learning techniques that are most often employed. Among the materials which are new to this edition are additional neural network applications, and a new chapter on the emerging high throughput technologies of microarrays and gene expression.

This book is divided into 13 chapters and six appendices. Chapter One covers an introduction to biological data, genes, proteins and databases. Chapter Two covers the probabilistic framework of machine learning, with Chapter Three giving specific examples. Chapter Four addresses machine learning algorithms. Chapters Five and Six cover neural network theory and applications, whereas Chapters Seven and Eight cover the theory and application of hidden Markov models. Chapter Nine discusses probabilistic graphic models, including combining neural networks and hidden Markov models to form hybrid models. Probabilistic models of evolution, expressed as phylogenetic trees constructed from sequence data, is the topic of Chapter Ten. Chapter Eleven explores stochastic grammars and linguistics. Deoxyribonucleic acid microarrays and gene expression are discussed in Chapter Twelve, and 18 pages of Internet resources and public databases are presented in the book's last chapter.

The index is six pages long and was found to be fairly useful. The bibliography is extensive with 587 entries. The technical prerequisites for this book are a little demanding in terms of requiring a somewhat extensive mathematics background. The strength of this book is that it has a broad scope and covers diverse applications. The main weakness is that because it is so technical it can be difficult to follow at times, though, for some, this could be an attractive feature. Also the six appendices provide information for increasing one's understanding. This book is of most interest to someone with a background in computer science or computational biology who is working with medical or biological databases involving deoxyribonucleic acid, ribonucleic acid or proteins. This book is highly recommended for anyone with an interest in bioinformatics.

Mark A. Best
St Matthew's University
Grand Cayman

Statistical Rules of Thumb
G. VAN BELLE, 2002
New York, Wiley
xviii + 222 pp., £38.95
ISBN 0-471-40227-3

I guess that we all promulgate to students the standard rules of thumb that a sample of at least 30 justifies appeal to the central limit theorem, that $np > 5$ is satisfactory for a normal approximation to the binomial distribution, that expected frequencies of at least 5 should govern applications of χ^2 goodness-of-fit tests and so on. van Belle has accumulated a whole book-full of such rules, some familiar, some obscure—at least to this reader—and mainly rather more substantial. The book is organized into eight chapters with about a dozen sections in each. It is the sections that give the rules. Each has a descriptive title, e.g.

> 'Use at least twelve observations in constructing a confidence interval',

always followed by the same five subsections: 'Introduction'—a general background description; 'Rule of thumb'—the bare statement of the rule; 'Illustration'—a hypothetical or genuine case-study; 'Basis of the rule'—a description of the theoretical underpinning of the rule; 'Discussion'; 'Extensions'.

The first chapter, entitled 'The basics', looks at fundamental issues in statistical modelling, hypothesis testing and estimation. Sample rules include the above one on confidence intervals,

> 'Beware of linear models',

> 'Use p-values to determine sample size, confidence intervals to report results'

and

> 'Be eclectic and ecumenical in inference',

this last one acknowledging that an applied statistician may use methods from each of Neyman–Pearson, likelihood and Bayesian approaches as appropriate. Chapter 2, on 'Sample size', includes rules such as

> 'Calculating sample size using the coefficient of variation',

'Sample size calculation for the binomial distribution'

and

'When unequal sample sizes matter; when they don't'.

The third chapter on 'Covariation' covers rules such as

'Do not correlate rates or ratios indiscriminately'

and

'Pairing data is not always a good idea'.

The next two chapters, on 'Epidemiology' and 'Environmental studies' respectively, are rather more specialized than the rest of the book and reflect the author's own long association with these areas. However, I found these sections interesting precisely because of my unfamiliarity with much of the material. Chapter 6 covers the design and analysis of experiments and clinical trials, and features much good advice for the practitioner, including pieces on

'Blocking is the key to reducing variability',

'Distinguish between design structure and treatment structure'

and

'Address multiple comparisons before starting the study'.

The next chapter I particularly liked. It presents rules of thumb for communicating numerical information by using words, tables and graphs. Indeed, the first rule is to use text for a few numbers, tables for many numbers and graphs for complex relationships. In my current work environment, I cannot resist taking the opportunity to point colleagues—and students—to

'Never use a pie chart',

'Stacked bargraphs are worse than bargraphs'

and

'Three-dimensional bargraphs constitute misdirected artistry'!

The final chapter, on 'Consulting', is one that can usefully be read by anyone who is called on to give statistical or indeed any sort of 'expert' advice. The beginning practitioner in particular will find this chapter invaluable, with sections including

'Structure a consultation session to have a beginning, a middle, and an end',

'Tailor advice to the level of the investigator',

'Ethics precedes, guides and follows consultation'

and

'Listen to, and heed the advice of experts in the field'.

I thoroughly recommend this book. It is well written and presented and, although it is true that many of the rules are formalized common sense that we would all apply, many have a less obvious technical justification and I would be surprised if any reader found no surprises. By its structure, it is the kind of book that one can happily dip into and find something of interest but it is also well indexed by topic and author and there is an excellent list of references. As the author acknowledges, you may look for a favourite rule and fail to find it or you may find a rule with which you would take issue. In both cases you are invited to contact the author. In any case I would urge you to visit the Web site (`www.vanbelle.org`) from where you can download Chapter 2 as a taster as well as consult the growing 'Rule of the month' list and the disturbingly lengthy list of corrections, to which I am about to make some additions!

Quentin L. Burrell
Isle of Man International Business School
Douglas

Information Gap Decision Theory: Decisions under Severe Uncertainty
Y. Ben-Haim, 2001
San Diego, Academic Press
x + 330 pp., £66.95
ISBN 0-12-088251-5

This (hardback) book presents a new theory of decision-making under severe uncertainty. The author argues that using information gap theory allows the decision maker to deal with situations where there is a disparity between what is known and what needs to be known in order to make a well-founded decision. Although information gap theory deals with various problems of classical decision analysis such as risk assessment, gambling and value of information, the character of information gap uncertainty gives new insights into the situations that are analysed.

The book considers several applications such as medical decisions, strategic planning and environmental management, and it is in these applications that the book's strength lies. Although it is not an easy book to understand, it is well worth the effort that is required to understand how to apply information gap theory to decision-making under severe uncertainty. I would certainly recommend this book for practising decision analysts or for use in a post-

graduate course on decision theory.

The book is divided into 12 chapters. Chapters One and Ten give an overview of what information gap theory can do, without any details about how to do it. Chapter One gives a basic overview whereas Chapter Ten is entitled a 'Retrospective essay: risk assessment in project management' and considers value judgments, robustness and quantitative decision support systems. Chapter Two introduces some of the information gap models and gives four axioms of information gap uncertainty. Chapter Three covers robustness and opportunity, and gives three components of information gap models, namely the system model, performance requirements and the uncertainty model. This is followed by eight examples including engineering design and project management. Chapter Four considers value judgments whereas Chapter Five looks at antagonistic and sympathetic immunities. Basically, if the opportunity function increases with robustness against failure, then the two immunities are said to be sympathetic. If, however, an increase in one causes a decrease in the other, then the two immunities are said to be antagonistic.

Chapter Six covers gambling and risk sensitivity, whereas Chapter Seven covers value of information. Both chapters give examples in project management. In particular, Chapter Seven considers the Allais and the Ellsberg paradoxes. Chapter Eight considers information gap supervision of a classifier where a decision problem exists for selecting one from a number of classes by use of a quantitative measured vector. As the author states,

> 'While strategic structural modification of the decision process may entail revision of the classes themselves, the operative decision algorithm assumes that any measured vector arises from one and only one class'.

Coherent uncertainties and consensus are the topics that are covered in Chapter Nine, with hybrid uncertainties in Chapter Eleven. Finally, the book concludes with a chapter on the implications of information gap uncertainty.

As stated earlier in this review, the book's strength lies in the number and diversity of applications that it covers within the chapters. They assist the reader in understanding and applying the theory as do the problems given at the end of each chapter for the reader to work on. It would, however, be helpful if the solutions to the problems were given in an appendix rather than giving problems with no solutions available. However, I would still recommend this book for people who are currently working in the field of decision analysis or risk analysis, or those with a particular interest in this area. I would not recommend it for anyone who does not have a good grounding in mathematics and a reasonable understanding of decision analysis.

Lesley F. Wright
Liverpool John Moores University

Components of Variance
D. R. Cox and P. J. Solomon, 2002
London, Chapman and Hall–CRC
x + 170 pp., £39.99
ISBN 1-58488-354-5

This is, in my opinion, a splendid book. Its title might, at first, sound narrow, boring and old fashioned; far from it, the content of the book ranges far and wide over much of mainstream statistics, it is short and sufficiently concise not to be boring and it is 'afashionable' (of which more below), ranging from pre World War II foundations to microarrays and data mining.

There are six chapters. The first, 'Key models and concepts', illustrates by example some 'simple situations involving more than one level of variation' and discusses two key distinctions: that between nesting and cross-classification and that between fixed and random effects. Chapter 2, 'One-way balanced case', discusses balanced one-way analysis of variance with random effects, in considerable detail. Unusually, perhaps, this reviewer can concur wholeheartedly with the cherry-picked prepublication review comments on the back cover of the book (thanks to Vern Farewell):

> 'A particular strength is attention given to first principles as a prelude to the treatment of many of the technical topics'.

Chapter 3 goes on to 'More general balanced arrangements', covering components of covariance and regression, time effects, empirical Bayes methods, measurement error in regression and design issues. Chapter 4 treats 'Unbalanced situations', first in the one-way case and then in more complex situations, again in considerable detail and clarity; meta-analysis is treated in this chapter. 'Nonnormal problems' is the title of Chapter 5, firstly concerning Poisson and binomial distributions and then survival data, before considering various generalities, including generalized linear mixed models. Finally, Chapter 6, 'Model extensions and criticism', ranges over a variety of topics from outliers to transformations and much more.

This book is not a cookery-book. Vern Farewell again accurately observes that

> 'What distinguishes this book from other material is the depth of the discussion combined with

the use of only essential technical details'.

There is a depth of understanding of the basics that the authors try to get across here, together with a consequent clarity of explanation of where the many more advanced techniques fit in. For a statistician like me who has spent too little of his career in the statistical mainstream, the book is hugely useful in 'reminding' me how all those fancy techniques, so often treated separately, fit together. If you then feel inspired or required to investigate particular methodology in more spelled-out detail, alternative reading will be needed; the authors' bibliographic notes seem very good and provide a starting-point.

And what I also really like about this book is what I called its afashionable nature above. Techniques from all eras of the past 70 years or more that remain relevant are treated in their proper ongoing context and no particular emphasis is given to 2003 fashions. And so the current fads that I mentioned in my first paragraph sit side by side with Tippett's and Daniels's work in the cotton and wool industries respectively in the 1930s and with huge tracts of statistical methodology from the period in between, e.g. multilevel or hierarchical models, clinical trials, longitudinal data analysis, industrial design, meta-analysis and even a little bit of Bayes.

At a presentational level, other nice things include some substantial examples from the authors' experience and a summary of contents at the start of each chapter, denoted 'Preamble'. Four 'Computational/software notes' sections give a good idea of what software can handle which situations. There are also 'Further results and exercises', but these seem somewhat an afterthought and it is unclear whom the exercises (without solutions) are for; short 'Further results' sections without the exercise format might have been better. That is my only (minor) criticism of the authors. A criticism of the publishers is that my copy of the book, at least, has been printed in rather faint type, bringing home to this 40-something reviewer that reading glasses are becoming necessary.

Professor Cox has certainly not lost his book writing touch and Professor Solomon has clearly contributed mightily also. The result is a book that every statistical modeller and many serious users of statistics should have on their shelves. To quote from Vern Farewell one more time:

> 'This is a superb book on a topic of central importance in a wide variety of areas of research'.

I strongly agree!

M. C. Jones
The Open University
Milton Keynes

Time Series Analysis by State Space Methods
J. Durbin and S. J. Koopman, 2001
New York, Oxford University Press
xviii + 254 pp., £35.00
ISBN 0-19-852354-8

State space models are widely used in time series analysis and forecasting. This book provides an up-to-date exposition and comprehensive treatment of state space models in time series analysis. State space models are quite flexible in nature and can handle a much wider range of problems compared with the traditional Box–Jenkins autoregressive integrated moving average models. This book will be helpful to graduate students and applied statisticians working in the area of econometric modelling as well as researchers in the areas of engineering, medicine and biology where state space models are used.

This book is divided into two parts. The first part discusses various techniques for the analysis of the linear Gaussian state space model, consisting of traditional techniques and some recent developments in this area. The second part is more interesting and discusses non-Gaussian and non-linear state space models; these are extensions of the models that are discussed in the first part and are more recently discussed in the literature.

This book is organized into 14 chapters. Chapter 1 provides the basic ideas of state space analysis and gives the background to linear and non-linear as well as non-Gaussian models. After the first chapter the book is divided into two parts. Chapters 2–9 comprise the first part; they are on the linear Gaussian state space model. Chapters 10–14 form the second part. Chapter 2 introduces the key concepts of state space analysis by giving a simple example of the state space model: the local level model. Basic techniques like filtering, smoothing, initialization and forecasting are explained in a very simple way. There are several examples to demonstrate the various concepts. In Chapter 3, the linear Gaussian state space model is introduced and it is shown that Box–Jenkins models are special cases of state space models. The link between exponential smoothing and state space modelling is also discussed in this chapter. Chapters 4–7 discuss various inference problems concerning the linear Gaussian state space model. In Chapter 4, derivations of the Kalman filter and smoothing recursion for the estimation of the state vector are given. Chapter 5 discusses the initialization of the Kalman filter in cases where some or all of the elements of the initial state space vector are unknown. In Chapter 6 various computational aspects of filtering and smoothing are described and in Chapter 7 maximum likelihood estimation of unknown parame-

ters is considered. The use of a score vector and the EM algorithm is also discussed in this chapter. Chapter 8 discusses Bayesian analysis of the linear Gaussian model. Chapter 9 illustrates the use of the whole methodology by applying the techniques to real data sets.

In the second part of the book (Chapters 10–14) the theory of the state space model is extended to non-linear and non-Gaussian models. Chapter 10 discusses the various non-linear and non-Gaussian models which include the exponential family models, heavy-tailed distributions and mixtures of normal densities. Apparently these models can be analysed by the methodology in this part of the book. Chapter 11 gives simulation techniques for non-Gaussian models based on importance sampling and antithetic variables. Chapter 12 discusses the classical inference model, based on sampling, for analysing data from non-Gaussian and non-linear models by estimating the conditional means and variances of functions of elements of the state space vector via simulation and antithetic variables. Chapter 13 studies the analysis of non-Gaussian and non-linear state space models from the Bayesian inference point of view. Finally, in Chapter 14 the authors have demonstrated the whole of the methodology discussed in part 2 of the book by using real data sets.

The authors have given a Web site for the book, for data, code etc. 36 corrections are also mentioned on this Web site which should be seen before reading the book. Overall the book is well organized and clearly written. It is a good mixture of theory and practical applications. It may not be suitable for beginners but graduate and research students will definitely enjoy this book. Also practitioners will find the book quite useful. I would also recommend it for library purchase.

Kuldeep Kumar
Bond University
Gold Coast

A Course in Financial Calculus
A. ETHERIDGE, 2002
Cambridge, Cambridge University Press
viii + 196 pp., £18.78
ISBN 0-521-89077-2

Mathematical Finance and Probability: a Discrete Introduction
P. K. MEDINA AND S. MERINO, 2003
Basel, Birkhäuser
x + 328 pp., £31.27
ISBN 3-7643-6921-3

Financial theory as a discipline parted from econometrics during the first half of the 20th century. By the 1980s, Kolody and Copeland, in volume 3 of the *Encyclopedia of Statistical Sciences*, could write

> 'the emphasis in financial research has shifted rapidly from descriptive to analytical'.

They commented, however, that

> 'at the time of writing there is no comprehensive source on this subject'.

Over the last 20 years we have seen an increasing flow of books that are aimed at those bright young mathematicians, statisticians and physicists who would like the intellectual and financial advantages of joining the financial world. Many universities now offer tailor-made courses to advanced undergraduates and new graduates; companion texts have been written.

Etheridge's book has its origins in lectures given in 1997–1998 to advanced undergraduates at Oxford University; some of her course material was, I suspect, taken from Baxter and Rennie's *Financial Calculus*. Her own book is much more mathematically demanding and has many exercises which she considers to be essential for the learning process (solutions are available to *bona fide* teachers from `solutions@cambridge.org`).

The requisite background in probability theory is at the level of Grimmett and Welsh's *Probability: an Introduction*. Some familiarity with the kind of material in David Williams's *Probability with Martingales* would be very useful.

Her book has seven chapters, beginning with basic definitions and material on binomial trees, martingales and change of measure in discrete time. This leads into discussions on Brownian motion in Chapter 3 and stochastic calculus in Chapter 4.

The reader can now master the continuous time world of the Black–Scholes model in Chapter 5. The fundamental theorem of asset pricing is proved and an explicit expression for stock and bond holding in the replicating portfolio is obtained.

The last two chapters deal with more complicated models. Chapter 6 considers exotic options such as packages, European options and Asian options; there is also a brief look at American options. Chapter 7 examines some of the more general market models, e.g. stock price models with jumps and stochastic volatility.

Typographically this is a very well-organized text that makes it easy to learn. However, the rate of introduction of new concepts would raise eyebrows among educational technologists. It is definitely not a book for the weak student.

Medina and Merino's *Mathematical Finance and Probability* is also not a book for the faint hearted, though here the prerequisites are allegedly more

modest—three semesters of linear algebra and calculus at university level.

> 'All additional material is developed from the beginning as needed',

including an introduction to modern probability and measure theory

> 'albeit mostly within the context of finite sample spaces'.

A nodding acquaintanceship with, for example, Billingsley's *Probability and Measure* and Williams's *Probability and Martingales* would speed a student's progress considerably.

Medina and Merino's aim is to impart an understanding of the pricing of financial derivatives in securities markets at an elementary but rigorous level to

(a) financial economics students,
(b) mathematics and physics students who are interested in studying a modern applied topic and
(c) mathematicians and others who are already working in the financial world who want to master the main ideas underlying option pricing.

There are 16 chapters; as their titles show, they tend to alternate between finance and probability. They are 1, 'Introduction', 2, 'A short primer on finance', 3, 'Positive linear functionals', 4, 'Finite probability spaces', 5, 'Random variables', 6, 'General one-period models', 7, 'Information and randomness', 8, 'Independence', 9, 'Multi-period models: the main issues', 10, 'Conditioning and martingales', 11, 'The fundamental theorems of asset pricing', 12, 'The Cox–Ross–Rubinstein model', 13, 'The central limit theorem', 14, 'The Black–Scholes formula', 15, 'Optimal stopping', and 16, 'American claims'. Each chapter ends with suggestions for further reading, but there are no exercises. There are also two supplementary appendices (A, 'Euclidean space and linear algebra', and B, 'Proof of the theorem of de Moivre–Laplace'), a four-page bibliography and a short subject index. A notation glossary would be a useful addition in any future edition.

How far do the two books complement one another? Etheridge has had in mind an Oxbridge-type high-flyer—one who might find Medina and Merino's book a little repetitive, particularly if the student already has a detailed understanding of probability theory. (Medina and Merino have deliberately been willing to sacrifice conciseness for clarity, even when that has involved some redundancy.) Lecturers with less brilliant students would probably prefer to use Medina and Merino's book, supplemented by exercises from Etheridge. For either book familiarity with advanced mathematical notation is a necessity. Neither book is particularly suitable for self-tuition. If you are a student on a course that is based on some other book on mathematical finance, then I recommend that you buy one of these as well—both contain the fundamentals of the subject and a second writer's way of presenting a difficult topic can be very helpful.

Freda Kemp
St Andrews University

Modern Medical Statistics: a Practical Guide
B. S. EVERITT, 2003
London, Arnold
xiv + 236 pp., £40.00
ISBN 0-340-80869-1

Topics covered are generalized linear models including those for longitudinal data, generalized additive models, classification and regression trees, survival analysis, Bayesian methods, meta-analysis, exact inferences for categorical data and finite mixture models. There is also a useful chapter on missing values, drop-outs, compliance and intention to treat, as well as a glossary and an appendix on statistical graphics in medical investigation.

The emphasis is on examples—mostly medical applications. There are exercises for students on applied medical statistics courses, some with solutions. In a prologue the author states that his aim is

> 'to introduce numerate medical researchers with a reasonable background in statistics to a range of newer methods, concentrating not on their full technical detail, but more on their practical possibilities'.

I recommend this book as background reading for students on an advanced undergraduate or postgraduate course in medical statistics, but I have some reservations about its likely appeal to *numerate medical researchers with a reasonable background in statistics*. Much depends on how one interprets 'reasonable background'. Clinicians who are qualified to take major responsibility for the statistical aspects of their trials should at least be aware of many of the tools that are discussed in this book because, as the author points out, the application of some of these is increasingly common in medical research.

Many routine statistical analyses are done by those with a basic knowledge of the subject but who are not professional statisticians. However, there are no constraints governing the use of statistical techniques by those with partial statistical training so it

is inevitable that advanced concepts may be misunderstood and incorrectly used as a result of limited statistical know-how.

This is why I have reservations about the usefulness of this book to the author's stated target audience. For some topics the treatment is unnecessarily detailed for numerate clinicians who only need to be aware of these techniques, but at the same time there is often insufficient information for anyone who is not previously familiar with the basic ideas to apply them safely without further professional statistical advice. As Everitt points out, relevant computer software is essential for the routine application of many of the techniques. However, the versatility of many such programs makes it easy to use them inappropriately in the absence of expert statistical guidance. The book does not give clear warnings of this danger.

There is inconsistency in the level of coverage, this ranging from superficial to detailed. This to some extent reflects a choice of topics that was guided by the author's own interests, some being more straightforward than others. Nevertheless, anyone who is unfamiliar with the material in the brief elementary introduction to Bayesian methods is unlikely to be comfortable with the discussions of overdispersion or of some of the complexities in application to longitudinal data for generalized linear models.

A well-balanced chapter is that on missing values, drop-outs, compliance and intention to treat. It gives valuable information about important features of clinical trials for a statistician coming afresh to medical statistics, while explaining the statistical implications of these concepts for clinicians.

Many difficult points are clearly explained. However, some minor blemishes will soon be detected by the more critical reader. One such occurs in the very first example on page 4 where the author claims that reference to a scatterplot matrix indicates that blood pressure is most closely related to age and weight, whereas in fact the matrix (or the correlation matrix) indicates that blood pressure is most closely related to weight and to body surface area. Because body surface area is highly correlated with weight this turns out to be irrelevant so far as the formal analysis is concerned, but this is not explained. Many important points are relegated to exercises.

The author attempts to overcome the problem about technical terms which may or may not be familiar to all readers by providing a glossary of 24 terms ranging from *Yates's correction* to *quasi-likelihood*. The choice of terms is somewhat unbalanced and many of the explanations are sketchy. That for Yates's correction does not say that it is generally only appropriate for 2×2 tables.

On balance the good points in the book outweigh any deficiencies. The emphasis is strongly on data analysis and it is unfortunate that the role of statistics in the design of clinical trials is not covered.

In summary, the book is good background reading for an advanced course on medical statistics and a useful introduction to the topics covered for highly numerate clinicians with a good statistical background, but many of the latter will probably find parts of it heavy going whereas other parts will already be familiar to them.

P. Sprent
Wormit

Data Mining
D. Hand, H. Mannila and P. Smyth, 2001
Cambridge, MIT Press
xxxii + 546 pp., £34.50
ISBN 0-262-08290-X

Data mining

> 'is a new discipline, lying at the intersection of statistics, machine learning, data management and databases, pattern recognition, artificial intelligence, and other areas'.

It uses a very wide variety of techniques and comprises many different topics. David Hand and his co-authors have not tried to include everything that could be regarded as data mining in this book but have concentrated on the aspects that they consider to be most fundamental. A major purpose of the book is to unify the computational and statistical approaches via an understanding of the underlying theory and algorithms.

The book is written in the form of a text for final year undergraduate and Master degree students. It will also appeal to researchers and those who are already engaged in data mining. It falls naturally into three parts: 'Fundamentals' (Chapters 1–4), 'Data mining components' (Chapters 5–8) and 'Data mining tasks and algorithms' (Chapters 9–14). Suggestions are made concerning essential reading and also subsets of the book that are suitable for students with differing backgrounds and lengths of courses.

Chapter 1 is introductory, focusing on the basic elements such as data mining tasks and approaches. The nature of data sets and their components is studied in Chapter 2. Summarizing and displaying data sets form the subject-matter of Chapter 3 (this includes the use of principal components analysis and multidimensional scaling). An appreciation of the ramifications of uncertainty in Chapter 4 leads into a tutorial on the elements of statistical methodology (random variables, samples and samp-

ling, statistical inference, estimation and hypothesis testing).

Chapter 5 gives a systematic overview of data mining algorithms. Chapter 6 swings back to the statistical approach by considering the fundamentals of stochastic modelling. Discussions about score functions for data mining algorithms in Chapter 7 and search and optimization methods in Chapter 8 end the second part of the book. The organization of data and the management of databases are considered later on, in Chapter 12.

The third part of the book is much the longest. It includes material on descriptive modelling (Chapter 9), predictive modelling for classification (Chapter 10), predictive modelling for regression (Chapter 11) and finding patterns and rules (Chapter 13). The final chapter concentrates on the technical problems that are associated with the retrieval of information such as the retrieval of text, images and sequences. The difficulties when evaluating the performance of retrieval are stressed.

This is a handsome book—the text is well organized and it is supported by very many examples and diagrams. The typography is excellent (though some of the diagrams are printed too faintly for clarity).

I hope that this will form the basis for a 'bible' on data mining, with update editions and companion volumes as the subject develops and modifies. It deserves to be in every computational science and statistical science department, as well as in university libraries generally.

Freda Kemp
St Andrews University

Applied Quantitative Finance: Theory and Computational Tools
W. Härdle, T. Kleinow and G. Stahl, 2002
Berlin, Springer
xxii + 402 pp., £49.00
ISBN 3-540-43460-7

The book under review is based on the Center for Applied Statistics and Economics course at Humboldt University in Berlin and is designed for researchers who wish to develop professional skill in modern quantitative applications in finance. This is an electronic book and is designed as an interactive document with a stream of text and information with various hints and links to additional tools and features. The reader can access practical examples without purchasing additional software or material to be downloaded.

Prerequisite for reading the book is a sound knowledge of probability, statistics and some software applications. The book contains four main parts, each consisting of articles by 30 contributors including the three authors. The contributors are either from universities or from financial institutions. Part I is about value at risk. The three chapters in this part deal with statistical approximations of value at risk in conditional Gaussian models, the calculation of value at risk by using copulas and the quantification of spread risk by means of simulation.

The second part is on credit risk and has two chapters, one on the analysis of rating migration probabilities and the other on sensitivity analysis of credit portfolio models where two widely used approaches—the factor structure and the direct specification of a copula within the framework of a default-based credit risk model—are studied. Part III is on implied volatility and has four chapters. In the first of these chapters, the Black–Scholes formula is solved for the constant volatility parameter by using observed option prices to derive implied volatilities. Then the implied volatilities are analysed with two variants of principal components and the results are interpreted in the context of risk management. In the next chapter the construction of implied binomial trees is described and they are applied to estimate state price densities. In the third chapter in part III, a local polynomial model of the implied volatility and its derivatives is used to estimate state price densities. In the fourth chapter of this part, the methods proposed are used to analyse the German DAX index and the performance of skewness and kurtosis trades are investigated. Part IV of the book is on econometrics wherein some recently developed econometric methods are discussed. This part has nine chapters. In Chapter 10 on multivariate volatility models, multivariate generalized autoregressive conditional heteroscedastic models are introduced and are applied to a study of exchange rates. Statistical process control methods that are used to monitor sequential data are discussed in Chapter 11. In Chapter 12, the empirical likelihood technique is used to construct test procedures for the goodness of fit of a diffusion model. Combining hedonic regression with Kalman filtering, a simple state space model is developed in Chapter 13 which is used to infer the common component in the movement of prices of single-family homes sold in a district of Berlin. The influence of long memory effects on financial time series is the subject of Chapter 14, wherein Hurst analysis is applied to a study of a trading strategy for German voting and non-voting stocks. An adaptive estimation algorithm for time series is discussed in Chapter 15 with applications to data for two financial time series. This method helps to detect time intervals where the model approximately holds. Chapter

16 introduces Monte Carlo and quasi-Monte-Carlo techniques for pricing exotic options. Using deconvolution kernel estimates, a nonparametric estimation approach of generalized autoregressive conditional heteroscedastic models is discussed in Chapter 17. The last chapter presents and demonstrates a net-based spreadsheet solution for modern statistical and econometric analysis using which one can have access to the XploRe programs that are developed in the book with standard software. Each chapter in the book ends with a bibliography which includes the list of references in that chapter.

The book will definitely be useful for beginners who are curious to know what is happening in computational mathematical finance. To acquire a better understanding of the underlying principles the reader should have sound knowledge of probability and statistics. There are many misprints and errors; a list of some of these is available from the reviewer.

This electronic book will be welcomed by all who are interested in computational mathematical finance and students may derive much practical training working with the data sets and the XploRe programs. This book can be recommended to undergraduate libraries in statistics, econometrics and finance.

Ravi Sreenivasan
University of Mysore

The Elements of Statistical Learning
T. HASTIE, R. TIBSHIRANI AND J. FRIEDMAN, 2001
New York, Springer
xvi + 534 pp., £56.00
ISBN 0-387-95284-5

During the past decade, there has been an explosion in computation and information technology. With it has come vast amounts of data in a variety of fields such as medicine, biology, finance and marketing. The challenge of understanding these data has led to the development of new tools in the field of statistics and has spawned new areas such as data mining, machine learning and bioinformatics. Many of these tools have common underpinnings but are often expressed with different terminology.

This book describes the important ideas in these areas in a common conceptual framework. Its coverage is broad, from supervised learning where the classes are unknown *a priori* and need to be 'discovered' from the data (class prediction) to unsupervised learning where the classes are predefined and the task is to understand the basis for the classification from a set of labelled objects (class discovery). The many topics include neural networks, support vector machines, classification trees and boosting. The emphasis is on concepts rather than mathematics, and several examples are given as illustration. The book consists of 14 chapters.

Chapter 1 ('Introduction') briefly explains the role of statistical learning and presents some real life learning problems that are discussed throughout the book. Chapter 2 ('Overview of supervised learning') presents an overview of supervised learning problems, whereas Chapter 3 ('Linear methods for regression') and Chapter 4 ('Linear methods for classification') discuss some linear methods for regression and classification. Chapter 5 ('Basis expansions and regularization') presents splines, wavelets and regularization or penalization methods for a single predictor, whereas Chapter 6 ('Kernel methods') discusses kernel methods and local regression. Chapter 7 ('Model assessment and selection') studies the concepts of bias and variance, overfitting and methods like cross-validation for choosing models, whereas Chapter 8 ('Model inference and averaging') presents an overview of maximum likelihood, Bayesian inference, bootstrapping, the EM algorithm, Gibbs sampling and bagging. A related procedure called boosting is the focus of Chapter 10 ('Boosting and additive trees'); this is the first comprehensive treatment of the topic in any book so far. Chapter 9 ('Additive models, trees, and related methods'), Chapter 10 ('Boosting and additive trees'), Chapter 11 ('Neural networks'), Chapter 12 ('Support vector machines and flexible discriminants') and Chapter 13 ('Prototype methods and nearest-neighbors') describe a series of structured methods for supervised learning for regression and classification problems. Finally, Chapter 14 ('Unsupervised learning') describes some methods for unsupervised learning. The book also contains about 14 pages of relevant references. Author and subject indexes are also provided.

This book is designed for researchers and students in a broad variety of fields such as statistics, artificial intelligence, engineering and finance. It should be a valuable resource for those who are interested in data mining in science or industry. I believe that it will be a very useful addition to any scholarly library.

Theofanis Sapatinas
University of Cyprus
Nicosia

Logistic Regression: a Self-learning Text, 2nd edn
D. G. KLEINBAUM AND M. KLEIN, 2002
New York: Springer
xiv + 514 pp., £56
ISBN 0-387-95397-3

Ideal preparation for this book is claimed to be

a course on quantitative methods in epidemiology and a course in applied multiple regression. So, imagine that you are sitting attentively in a lecture, and the formula

$$P(X) = \frac{1}{1 + e^{-(\alpha+\beta_1 E)}}$$

goes onto the board. Then, just in case you were not quite on the ball (notwithstanding the courses in regression and epidemiology that you have attended), the lecturer helpfully writes underneath:

> '... the logistic model $P(X)$ equals 1 over 1 plus e to minus the quantity α plus β_1 times E ...'.

Well, that was helpful was it not? The next formula swiftly follows:

$$E = 1: \quad \frac{1}{1 + e^{-(\alpha+[\beta_1 \times 1])}} = \frac{1}{1 + e^{-(\alpha+\beta_1)}},$$

and then more helpful text,

> '... by substituting the value E equals 1 into the model formula ... we then obtain 1 over 1 plus e to minus the quantity α plus β_1 times 1, or simply 1 over 1 plus e to minus the quantity α plus β_1'.

Is this not clear yet? We continue:

$$E = 0: \frac{1}{1 + e^{-(\alpha+[\beta_1 \times 0])}} = \frac{1}{1 + e^{-\alpha}},$$

and the text

> '... For E equal to zero ... substitute E equal to zero into the model formula and we obtain 1 over 1 plus e to minus α'.

But, by now, you are asleep and so am I. Insomniacs need look no further. This text is repetitious beyond belief, tedious beyond hope, unimaginatively detailed beyond caring. As your eyes glaze over reading this review, just keep in mind how lucky you are to have been spared the full narcotic verbiage of the original.

The material is mostly of very basic logistic regression, very thoroughly covered. Readers staying awake through the first seven chapters (to page 226) should be able to grasp the mechanics of logistic regression modelling. However, the authors manage to conceal the joy and excitement of statistical exploration: the single running example is used only to illustrate the nuts and bolts and not to motivate one model or another.

Only three or four data sets are used. These are described again and again unnecessarily in the same or similar words wherever they appear. The data sets are available on an accompanying disc but are not listed in the book: a pity. About half-way through the book, I had a sudden shocking thought, a thought which should have occurred to me sooner, but then absence is so less stimulating than presence: I could not remember seeing a single graph. Well, I checked and there are fewer than a handful, including these: a drawing on page 5 to show us what the logistic function looks like and two scatterplots on page 339 of 10 (presumably faked) pairs of data to show positive and negative correlation. Need I say more? I shall. We teach our students to graph data, first and last, as probably the most important thing to do with a data set. Many of us will have made some stupid blunder that could have been avoided by plotting the data (I certainly have), and this is all the more important when the models start to become complicated and our natural intuitions desert us. It seems to me ludicrous to offer a text-book on logistic regression which offers not a single genuine data plot, either to help interpretation or to diagnose modelling problems. Really, those interested in learning about logistic regression would acquire a much better feel for the subject by reading chapter 6 of Friendly (2001), which contributes far more to motivation and understanding than this text does.

On the positive side, the authors clearly know what they are talking about and, if you do not mind ploughing through the text, you will obtain a fairly good idea of the basic recipes for fitting logistic regression models. This includes a treatment of generalized estimating equation models in rather more detail than in Hosmer and Lemeshow (2000), brief coverage of generalized linear mixed models and a useful appendix on software: SAS appears to be the authors' choice, but there are examples for SPSS and STATA.

The bibliography is short: 21 references. McCullagh and Nelder (1989) is included, but there is no room for the three texts that I consult most for such models: Collett (1991), Agresti (1990) and Hosmer and Lemeshow (2000) (which has 191 references, to save you the trouble). However, just in case you had missed your ideal preparation in epidemiology and applied multiple regression, you will find references to ... oh!, well ... two texts by D. G. Kleinbaum: happy reading.

References

Agresti, A. (1990) *Categorical Data Analysis*. New York: Wiley.

Collett, D. (1991) *Modelling Binary Data*. Boca Raton: CRC.

Friendly, M. (2001) *Visualizing Categorical Data*. Cary: SAS Institute.

Hosmer, D. W. and Lemeshow, S. (2000) *Applied Logistic Regression*. New York: Wiley.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. London: Chapman and Hall.

David Wooff
University of Durham

Hidden Markov Models for Bioinformatics
T. Koski, 2001
Dordrecht, Kluwer
xvi + 392 pp., £94.00
ISBN 1-4020-0135-5

This text on probabilistic modelling for bioinformatical applications is another welcome volume that has graduated from humble beginnings as undergraduate lecture notes (in mathematics, computer science and biochemistry) into a polished and instructive course. Here, the 'bioinformatics' in the title is taken in the more restricted sense of probabilistic sequence analysis (as opposed to the very broad interpretation that it has since adopted). It is concerned with the numerous methods of assigning probabilities to, and performing operations on, strings of alphabet elements of the type that are encountered in genome analysis (e.g. A, T, G, C). It presents the methods and model families of probabilistic learning in increasing gradients: from multinomial processes to Markov chains, and from there to hidden Markov models. Particular sophisticated modifications of these basic frameworks which have proved fruitful in biological sequence analysis are also investigated.

The first third of the book is concerned with providing a detailed account of the prerequisite ideas that are required at later stages. Beyond basic probability this amounts to discussions on information theory and coding theorems, the learning of probabilistic models via Bayes theorem and a detailed examination of the EM algorithm for finite mixtures of distributions (special cases of hidden Markov models). Chapters on sequence alignment, scoring and profiles are complemented with the associated dynamic programming search algorithms (e.g. the Smith–Waterman algorithm).

The middle section introduces, firstly, Markov chains and how they can be trained from default prior beliefs. A chapter is then given to a summary of the restrictions that models based on Markov chains impose when they are straightforwardly applied to biological analysis. Secondly an overview of hidden Markov models is presented and similarly applied in the context of sequences of deoxyribonucleic acid. Here the modularity of hidden Markov models is emphasized, along with discussions on motif-based and profile hidden Markov models and the ideas behind modelling protein sequence families and multiple alignments. The final chapters in the book are given to various recursion and learning algorithms. Derin's backward recursive formula for smoothing posterior probabilities, the forward–backward algorithm and the Viterbi algorithm are all presented in considerable detail. Discussion of the Baum–Welch learning algorithm merits two whole chapters.

The adopted lecture notes style of presentation throughout the discussion could be considered a little terse for casual reading; the focus is strongly on the mathematical development of the subject-matter, although numerous illustrative and instructive examples and exercises are usefully provided throughout, and some detail is postponed to chapter-specific appendices. However, the strength of this book is precisely in the detailed development that is often passed over in other texts; the reader will be rewarded with a solid foundation in the core material, and a grounded appreciation of why such models have achieved success in the given applications.

One novelty is the referencing structure; deriving from the original format of the text in which references were handed out weekly with the original class notes, citations are included at the end of each chapter. Within each bibliography, which includes a 'further reading' section, references are further subpartitioned according to within-chapter subject-matter. Thus locating a hazily recalled reference requires identification via subject rather than author if a laborious search is to be avoided; this may prove an annoyance to some. A detailed reading will also require many fingers in previous pages, as the references to equations in considerably earlier chapters are frequent, although this is probably more a reflection of the comprehensive nature of the course rather than an editorial oversight on readability.

According to the author this book is aimed at advanced undergraduate and graduate students with good mathematical ability and a basic familiarity with the ideas behind algorithmic sequence analysis. This is perhaps a fair assessment; the mathematical content is strict throughout, and there is no deeply detailed motivation of the biological application, although this can easily be located in numerous other sources. The value to the researcher is more likely to be as a very useful reference for the detail of arguments and proofs, in addition to being a valuable teaching aid.

In summary, this is a comprehensive and mathematically strict text that provides a systematic, lecture style introduction and presentation of probabilistic modelling as applied to bioinformatics. This text should be considered for library purchase for

any institution with a serious bioinformatics (or informatics) programme.

Scott Sisson
University of Puerto Rico

Statistical Analysis of Data from Marked Bird Populations
B. J. T. Morgan and D. L. Thomson (eds), 2002
Bristol, Routledge, Taylor and Francis
668 pp.
ISSN 0266-4763

Volume 29 of the *Journal of Applied Statistics* comprises the proceedings from the international EURING conference held in October 2000 at the Point Reyes Bird Laboratory, California, which was dedicated to George Seber, who published many key papers within the analysis of bird or animal populations. The conference is generally composed of population biologists and statisticians and aims to develop new methodological advances within the study of capture–recapture data, particularly applied to bird populations, in addition to discussing recent advances in this field and identifying important areas for future research. In total there are more than 30 papers and nine discussion contributions within this volume, authored by many of the leading researchers in the field.

The initial preface of the book gives a very brief but interesting overview of capture–recapture studies and outlines areas of current research, with plenty of citations supplied for further reference. The main collection of papers within the book are then classified into seven sections and covers a wide range of topics within the analysis of bird populations. At the beginning of the majority of sections there are very good overviews of the main ideas which also discuss the limitations and difficulties that are associated with the areas.

The first section entitled 'Evolutionary biology' broadly covers the issue of heterogeneity in the population, often modelled via the use of covariate information. Included within these papers are both good overviews of the area and the application of many of these techniques to particular sets of data where the results are not only reported but also interpreted. Alternative approaches to modelling heterogeneity and overdispersion in the population are presented in the section 'Random effects models' which has been an area of recent research. Random effects have also been introduced in relation to the reduction of the number of parameters, with the temporal variation in the parameters assumed to be random around some 'average' value. The corresponding issue of model selection is also addressed within this context.

There are a series of papers considering the extension of general capture–recapture data to the multistate case. In particular, there is a very balanced and concise overview of multistate models, primarily applied to geographical locations but also to more general individual covariates. The majority of papers in this section discuss real data but also recognize the need for further work in this area, discussing the limitations of the methods presented, not least in terms of the statistical difficulties relating to the complexity of the models and the issue of model choice.

An interesting section collects together several Bayesian analyses of real data sets. Although this is the shortest section, containing only three papers (and two discussion contributions), these papers provide a solid introduction to Bayesian methods, while also providing practical advice and discussion on the implementation of such approaches, including the specification of priors, prior sensitivity and the issue of model uncertainty. The analyses presented within the papers under the Bayesian title include the use of covariate information, random effects and multistate models, thus providing alternative approaches to the ideas and analyses that are presented in the other sections on these topics.

The remaining sections contain papers that discuss further computational issues, the physical design of the experiments that are used to collect the capture–recapture data and the estimation of population growth (or decline).

The various contributions provide a wide range of methodological approaches to many problems, analyses of real data sets and very often stimulating discussion of a wide range of issues that are seen to be important and are yet to be addressed. Descriptions of the various computer packages that are available for analysing different forms of data have also been detailed in various places (within-area overviews and the analysis of particular data sets), providing adequate reference for anyone wishing to perform similar analyses. Overall, the series of papers provides both general and more specific overviews of different areas relating to the analysis of bird populations and presents many recent advances within this area.

Ruth King
University of St Andrews

The Estimation and Tracking of Frequency
B. G. Quinn and E. J. Hannan, 2001
Cambridge, Cambridge University Press
266 pp., £40.00
ISBN 0-521-80446-9

The Estimation and Tracking of Frequency is the

product of research of over nearly two decades when the second author, Ted Hannan, who was working on the modelling of acoustic and electronic signals, became interested in estimating two frequencies in two sinusoids when the frequencies are very close together. Initially both Hannan and Quinn had the idea of writing this book and Quinn has now finished it after years of finding the time to do so.

The authors start out with a good introductory chapter revealing the foundations of periodic functions, sinusoidal regression and the periodogram; finally they talk of frequency estimation and tracking. Readers may be confused, thinking that this book is on spectral analysis, but what I could gather from the book is that it is primarily targeted at the mathematician or researcher rather than the engineer.

The book contains six chapters which talk about 'Statistical and probabilistic methods', 'The estimation of a fixed frequency', 'Techniques from ARMA modeling', 'Techniques based on phases and autocovariances', 'Estimation using Fourier coefficients' and 'Tracking in low signal to noise ratio conditions'.

The chapters on 'Statistical and probabilistic methods' and 'The estimation of a fixed frequency' essentially lay the theoretical foundations for the rest of the book. The statistical theory for maximum likelihood estimators of frequencies is examined in the latter chapter. The authors derive several theorems here which could be verified by researchers for practical applicability, but to an engineer they may seem to be of little interest. In later chapters the authors discuss various stochastic models, maximum likelihood and the Cramér–Rao bound, and limit theorems.

In 'Techniques from ARMA modeling', the authors examine the Quinn and Fernandes technique which places poles on the unit circle and iteratively achieves an estimator whose asymptotic properties are same as that of the periodogram maximiser.

In the chapters on 'Techniques based on phases and on autocovariances' and 'Estimation using Fourier coefficients', the authors expose the readers to popular frequency estimation techniques such as the autoregressive, MUSIC techniques. 'Estimation using Fourier coefficients' looks at how to obtain a closed form estimator of frequency for a single time series by using Fourier coefficients at successive frequencies.

The final chapter on 'Tracking in low signal to noise ratio conditions' may be of practical interest to engineers. Here, the author explores maximum likelihood tracking, hidden Markov models and Viterbi algorithms for the tracking of frequency.

These last three chapters along with the appendix of MATLAB programs may be of practical interest to the engineer who is working in industry since he may be faced with situations involving tracking of frequencies in low signal-to-noise ratio conditions.

A final list of MATLAB programs to supplement understanding of the algorithms that are presented in this book will prove to be very useful to the user.

Summing up, I feel that this book will help all users who are graduates taking an advanced statistics or mathematics course in conjunction with communications systems. Chapters One–Three and the last two chapters will certainly help users with such a background. As a final note, engineers or practical implementers of systems who are interested in learning 'tracking of frequency' may not learn much from this book since it is targeted at mathematicians—it contains a large amount of mathematics and it may take plenty of time for them to digest the material.

Dharmendra Lingaiah
Cranes Software International Limited
Bangalore

Handbook of Statistics, Volume 21, **Stochastic Processes: Modeling and Simulation**
D. N. SHANBHAG AND C. R. RAO (eds), 2003
Amsterdam, North-Holland
1020 pp., £125
ISBN 0-444-50013-8

North-Holland's *Handbooks of Statistics* are written in an authoritative manner by leading researchers and are important, though expensive, additions to departmental libraries. Does this new volume shape up to expectation?

It is the companion to volume 19, *Stochastic Processes: Theory and Methods* (2001), which was also edited by D. N. Shanbhag and C. R. Rao. This earlier volume dealt with the fundamentals of branching, Lévy, continuous time autoregressive moving average, Gaussian, point and diffusion processes, random walks and fluctuation theory, stochastic networks, renewal theory, characterization and identifiability, martingales, Markov chains and Itô's stochastic calculus, to mention only some of the topics. The new volume is applications orientated.

It reviews recent developments and current researches in fields as diverse as astronomy, biology, chemistry, engineering, epidemiology, genomics, manufacturing systems, queuing theory and telecommunications. I would have liked to have seen a similar treatment of the financial field. For this volume 14, *Statistical Methods in Finance* (1996), and the theoretical Chapter Twelve on stochastic pro-

cesses in insurance and finance by Embrechts, Frey and Furrer in volume 19 must suffice. Pharmacology and pharmaceutics are also areas that are thin. Some of the chapters have a welcome bent towards modelling, numerical methods and simulation techniques.

There are 23 chapters, by a galaxy of writers. The chapter headings give a good overview of the coverage of the book: 1, 'Modelling and numerical methods in manufacturing system using control theory' (Boukas and Liu); 2, 'Models of random graphs and their applications' (Cannings and Penman); 3, 'Locally self-similar processes and their wavelet analysis' (Cavanaugh, Wang and Davis); 4, 'Stochastic models for DNA replication' (Cowan); 5, 'An empirical process with applications to testing the exponential and geometric models' (Ferreira); 6, 'Patterns in sequences of random events' (Gani); 7, 'Stochastic models in telecommunications for optimal design, control, and performance evaluation' (Gautam); 8, 'Stochastic processes in epidemic modelling and simulation' (Greenhaigh); 9, 'Empirical estimators based on MCMC data' (Greenwood and Wefelmeyer); 10, 'Fractals and the modelling of self-similarity' (Hambly); 11, 'Numerical methods in queueing theory' (Heyman); 12, 'Applications of Markov chains to the distribution theory of runs and patterns' (Koutras); 13, 'Modelling image analysis problems using Markov random fields' (Li); 14, 'An introduction to semi-Markov processes with application to reliability' (Limnios and Oprisan); 15, 'Departures and related characteristics in queueing models' (Manoharan, Alamatsaz and Shanbhag); 16, 'Discrete variate time series' (McKenzie); 17, 'Extreme value theory, models and simulation' (Nadarajah); 18, 'Biological applications of branching processes' (Pakes); 19, 'Markov chain approaches to damage models' (C. R. Rao, Albassam, M. B. Rao and Shanbhag); 20, 'Point processes in astronomy: exciting events in the universe' (Scargle and Babu); 21, 'On the theory of discrete and continuous bilinear time series models' (Subba Rao and Terdik); 22, 'Non-linear and non-Gaussian state-space modeling with Monte Carlo techniques: a survey and comparative study' (Tanizaki); 23, 'Markov modelling of burst behaviour in ion channels' (Yeo, Milne, Madsen, Li and Edeson).

The chapters all have clearly written introductory sections and worthwhile bibliographies; many have good concluding sections that point the way forward. This volume ought to be available to all statistical departments, especially applied ones, so that staff and research students are encouraged to realize the importance of applied stochastic processes.

Freda Kemp
St Andrews University

Handbook for Conducting Research on Human Sexuality

M. W. Wiederman and B. E. Whitley, Jr (eds), 2001
Mahweh, Erlbaum
viii + 532 pp., £95.50
ISBN 0-8058-3437-0

This multi-authored book deals with most of the methodological, conceptual, practical and ethical aspects of research into human sexuality. It has 21 chapters, split into five parts which cover theory and validity, data collection, data analysis, interpreting results and special issues. An examination of the section on data collection shows this to be a text that is more about methodology than statistics. Chapters on focus groups, diary techniques and policy capture are included as well as chapters on sampling and measurement. The data analysis section is similarly split between quantitative and qualitative approaches.

The important question to ask before buying such a book is 'is it a book about sex research for statisticians, or conversely about statistics for sex researchers?'. Many such crossover books disappoint both audiences by allowing a series of compromises that satisfy no-one. The authors of this book have, I think, achieved the right balance, if for no other reason than it is difficult to answer this question. Although many primers in statistics are available, the chapters on statistics in this book provide a creditable review of experimental design and analysis as well as regression and covariance structure modelling.

Having established that this could be a useful text for sex researchers venturing out on their first studies, the next question to pose is 'what will statisticians derive from such a text?'. If you are a social or human sciences statistician I think that the answer is a considerable amount. It is important that statisticians working in the human and social sciences maintain a sense of the context in which their work takes place. This not only means having a general grasp of the literature that is relevant to the field but also being able to recognize and acknowledge when quantitative approaches cannot provide the answers that are sought.

This can be for a variety of reasons and sexuality research provides an almost ideal illustrative case of how both quantitative and qualitative approaches have to sit together to gain an overall picture of the field.

In Chapter 18 (on cross-cultural research) the author identifies four aspects of gender: biological, psychological, social and cultural. I have taken the liberty of changing the order of these. The reason for this is that I believe that they can be set on a

continuum; each move along this continuum means that the quantitative gives ground to the qualitative. Although biological research is almost exclusively quantitative, when teleological explanations are advanced meanings are being inferred that may go beyond the data. They reflect something of the researcher. In psychological research the mass of scales and the rigid application of experimental designs are seen as a badge of scientific credibility. None-the-less, in the field of social psychology and sociology, observational and semistructured interview techniques (that can be analysed either quantitatively or qualitatively) soon give way to less structured detailed interviews and focus groups that are almost always analysed qualitatively (I use almost because of the now very occasional use of the sociogram). The only difference as far as I can see between social psychology and sociology is in subject-matter, not methodology. Though sociologists may make wider use of survey and spatial data than social psychologists, when both these disciplines turn their attention to culture their methodologies soon have to give ground to the ethnographic and participant observation techniques that are the preserve of social anthropologists.

If the world of human research was neatly divided into biologists, psychologists, sociologists and social anthropologists, then life would be simpler and much less creative and there would be no need for a book like this. That said, even when a researcher does choose a multidisciplinary area, he or she will often come from a background that predominates in either the quantitative or qualitative approach. At the clearly quantitative end of the spectrum the purely physiological aspects of sexuality can now be measured by a variety of techniques (Chapter 7). In the study of the behavioural and attitudinal aspects of sexuality there is much to be gained by the quantitative approach in terms of survey and behavioural observation techniques (Chapters 5 and 6). But it is as the research moves into enquiries about more personal and emotional issues that we start to see the influence of qualitative techniques. Some types of questionnaire will obtain the best response by being straightforward and anonymous. But sometimes an interviewer is needed who will explain to and gain the trust of the participant. A skilled interviewer may not need a rigid schedule of questions but will need to know when to continue down a certain line of questioning. It may be necessary to put an open question (one that needs more than a yes or no, or a mark on a Likert scale), and then to code the answers. It may be that not enough is known about the particular area even to use this technique and the interviews have to be directed by the respondents themselves. Or it may be that the social interaction with regard to the subject in question can only really be tapped by establishing a focus group.

In almost every human and social science there is an area where the quantitative has to start giving ground to the qualitative. It is important for the integrity of the profession of statisticians to be able to recognize when this position has been reached and to say so and not to claim for the quantitative approach things that it cannot achieve. Sexuality research, with the breadth of techniques that it has to use and integrate, is a case-study that statisticians could profit from studying before reflecting on their own areas of practice.

James Tighe
Institute of Psychiatry
London
and Bracton Centre for Substance Misuse
Bexley

Find the Right Publicity Contact. Email Book Request. Next Steps.Â First, I will be defining book influencer as anyone with a platform who is talking about and promoting books. This ranges from book bloggers to bookstagrammers, booktubers and booklrs, etc. Second, you will often see me refer to ARCs or Advance Reader Copies. You may also see these listed as AREs, DRCs, and eARCs (which refers to an electronic/eBook version).Â You can also request advance audiobook copies but the books available in that format are likely to be more limited. For this post, I will be focusing on physical and electronic ARCs, but you can apply most of the requesting information to audio as well. Electronic ARCs. A great way to get started with ARCs is to request electronic ARCs, or e-ARCs. Books and Media Reviews. The following items are currently available (listed by year of publication). There is no fee but the reviewer keeps the book, except when publishers make an ebook available for limited time.Â Sometimes, two or more books on the same topic can be reviewed together; in general, comparisons of new books with the existing literature can be most helpful. We want reviews published in our journals to read well, to be authoritative, and to be useful to the statistical community.Â The book reviews editor and/or the copy editor might edit your review mildly, mainly to put it into in â€˜house styleâ€™, but also to correct typographical errors etc; if an editor wishes to make any alteration of substance, he or she will run it past you first. In summary, this book will be found to be extremely useful for a non-statistical audience without previous knowledge of SPSS. No book can cover everything. Hence I would recommend the use of this book together with some comple-mentary reading in the areas of categorical and multivariate data and advanced SPSS program-ming. Therefore, the phrase `just another statistical book for non-statisticians' is not valid. I would deÂ®nitely recommend this book to a library. Lada Mitchell Napp Pharmaceuticals. Your book details can be found in the first section of the book setup. Your bookâ€™s information is also called your bookâ€™s metadata. This is what will show on your book's detail page. As most of these fields cannot be changed after you publish, we recommend paying close attention and double check everything before submitting. Use the links below for information on each topic: Book title & editions. Titles and edition numbers.Â Join us for an interactive, 60-minutes session where we explain how Amazon uses metadata to make your book available to readers. Register today. Book Depository is the world's most international online bookstore offering over 20 million books with free delivery worldwide.Â All prices include your local Tax. This may change dependent on your delivery location, details of which can be found in our terms and conditions. Non-book products e.g. stationery, audio books etc. have the standard rate of tax applied. Tax on pre-orders will be estimated, the actual amount paid in tax will be confirmed on the dispatch note. The price of goods will not change as a result. From 1 July 2018, as a result of changes to Australian tax legislation, Book Depository now collects and remits GST (Goods and Sales Tax) on all books with a customs value of A$1000 or less that are shipped