

BOOK REVIEWS

EDITOR:
THOMAS M. LOUGHIN

- | | | | |
|--|--------------------------|--|--------------------------|
| Applied Spatial Data Analysis with R
(R. S. Bivand, E. J. Pebesma,
and V. Gomez-Rubio) | <i>Jay M. Ver Hoef</i> | Adaptive Design Theory and Implementation Using SAS and R
(M. Chang) | <i>Feifang Hu</i> |
| Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology
(A. B. Lawson) | <i>J. Law</i> | Ecological Models and Data in R
(B. M. Bolker) | <i>Rachel M. Fewster</i> |
| Disease Surveillance: A Public Health Informatics Approach
(J. S. Lombardo and
D. L. Buckeridge, editors) | <i>Iris Pigeot</i> | Applied Multiway Data Analysis
(P. M. Kroonenberg) | <i>Renato Coppi</i> |
| Survival Analysis for Epidemiologic and Medical Research
(S. Selvin) | <i>M. G. Valsecchi</i> | <i>Brief Reports by the Editor</i> | |
| Survival and Event History Analysis: A Process Point of View
(O. O. Aalen, O. Borgan,
and H. K. Gjessing) | <i>Patricia Grambsch</i> | Sampling of Populations: Methods and Applications, 4th edition
(P. S. Levy and S. Lemeshow) | |
| Nonlinear Dimensionality Reduction
(J. A. Lee and M. Verleysen) | <i>Haonan Wang</i> | Applied Survival Analysis: Regression Modeling of Time-to-Event Data, 2nd edition
(D. W. Hosmer, S. Lemeshow, and S. May) | |
| Model Selection and Model Averaging
(G. Claeskens and N. L. Hjort) | <i>Thomas M. Loughin</i> | SAS for Data Analysis: Intermediate Statistical Methods
(M. G. Marasinghe and W. J. Kennedy) | |
| Meta-Analysis of Binary Data Using Profile Likelihood
(D. Böhning, R. Kuhnert,
and S. Rattanasiri) | <i>Eloise Kaizar</i> | Advances in Mathematical and Statistical Modeling
(B. C. Arnold, N. Balakrishnan, J. M. Sarabia,
and R. Mínguez, editors) | |
| Wavelet Methods in Statistics with R
(G. P. Nason) | <i>Jeffrey S. Morris</i> | An Introduction to Generalized Linear Models, 3rd edition
(A. J. Dobson and A. G. Barnett) | |
| | | Design and Analysis of Bioavailability and Bioequivalence Studies, 3rd edition
(S.-C. Chow and J.-P. Liu) | |

BIVAND, R. S., PEBESMA, E. J., and GOMEZ-RUBIO, V. **Applied Spatial Data Analysis with R**. Springer, New York, 2008. xiv+378 pp/\$59.95/€48.10/ ISBN 9780387781709.

Most authors struggle to pick a title that adequately conveys all of the material covered in a book. When I first saw *Applied Spatial Data Analysis with R*, I expected a review of spatial statistical models and their applications in packages (libraries) from the CRAN site of R. The authors' title is not misleading, but I was very pleasantly surprised by how deep the word "applied" is here. The first half of the book essentially covers how R handles spatial data. To some

statisticians this may be boring. Do you want, or *need*, to know the difference between S3 and S4 classes, how spatial objects in R are organized, and how various methods work on the spatial objects? A few years ago I would have said "no," especially to the "want" part. Just let me slap my EXCEL spreadsheet into R and run some spatial functions on it. Unfortunately, the world is not so simple, and ultimately we want to minimize effort to get all of our spatial analyses accomplished. The first half of this book certainly convinced me that some extra effort in organizing my data into certain spatial class structures makes the analysis easier and less subject to mistakes. I also admit that I found it very interesting and I learned a lot.

The book begins with an introductory chapter that leans toward computing and data. It lists four main types of spatial data: points, lines, polygons, and grids (essentially very small polygons that emulate a continuous surface). Chapter 2 begins the description of classes and methods for spatial data in R. Not surprisingly, there is a class for each type of spatial data, and an extra sparse matrix representation of a grid. All of these classes inherit a certain underlying structure from a foundational spatial class. The payoff for taking the extra energy and time to structure your data as an object in Chapter 2 is immediately apparent from Chapter 3, which covers visualizing spatial data. Chapter 4 covers spatial data import and export. Several years ago, I struggled on a project that required managing various spatial data with different projections and support using the `rgdal` package in R; I really wish I had this book at that time! Chapters 4 and 5 go into greater detail on handling spatial data, including checking and manipulating topologies, and even creating your own spatial data classes.

The second half of the book covers statistical analysis, divided into the three main areas of spatial statistics; spatial point patterns (Chapter 7), geostatistics (Chapter 8), and lattice models (called areal data modeling and disease mapping in this book, Chapters 9–11). As the authors admit, these are fairly thin and eclectic treatments of these subjects, which are designed to “showcase” the use of spatial models using some of the packages in R; they are not designed as an introduction to spatial statistics. For example, Chapter 7 uses sophisticated examples of epidemiology from point locations of disease. The geostatistics chapter jumps rather quickly to multivariable models. Modeling areal data includes topics such as Moran eigenvectors and Marshall’s global EB estimator.

In summary, this is an excellent book that should be on the shelf of any applied statistician who is analyzing spatial data using R. It is not appropriate as a text on spatial statistics, but it would be a valuable companion to any course that uses spatial packages in R.

JAY M. VER HOEF
NOAA National Marine Mammal Lab
University of Alaska, Fairbanks
Fairbanks, Alaska, U.S.A.

LAWSON, A. B. **Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology**. CRC Press, Boca Raton, Florida, 2008. xvii + 344 pp. US\$79.95/£44.99. ISBN: 9781584888406.

The author of this book has contributed a series of books on methodologies in spatial analysis of disease data for various levels of readers. His co-authored book *An Introductory Guide to Disease Mapping* (Lawson and Williams, 2001) is still a good choice for novice readers of disease mapping. This introductory book, however, is limited on the practical aspects of disease mapping (McNally, 2002). It lacks technical details and is far from comprehensive for readers who are learning to apply statistical methodologies in disease mapping. His other book, *Statistical Methods in Spatial Epidemiology* (Lawson, 2006), which covers comprehensively statistical techniques in the field of spatial epidemiology (Bailey, 2007), is a valuable reference for statisticians. Spatial epidemiology, also referred

to as disease mapping, analyzes the spatial distribution of diseases taking into account spatial autocorrelation of disease outcomes. Nowadays, a Bayesian approach is becoming more and more popular, not just among biostatisticians, but many other disciplines including medical geographers, public health professionals, environmental health workers, planners, and social scientists. A book is needed for nonstatisticians in these fields who want to apply Bayesian disease mapping but find it challenging. In this regard, the author should be congratulated for writing this new book that is most welcome and timely. This book fills the gap between an introductory and advanced publication on spatial epidemiology with a modern focus on Bayesian methodologies.

This book is an excellent reference for intermediate learners of Bayesian disease mapping. It is divided into two main parts: background and themes. The part on background provides a clear introduction to basic ideas and concepts of Bayesian methods, including Bayesian inference and modeling, computational issues, residuals, and goodness-of-fit. This part contains exercises at the end of the chapters, except for Chapter 1 where datasets are introduced. However, there are no answers provided to the exercises. The second part of this book covers a wide range of applications of Bayesian methodologies in disease mapping. It includes common applications of disease map construction, relative risk estimation, disease cluster detection, ecological analysis, as well as some advanced applications that have received increased attention in the past few years. These relatively advanced applications include multiple scale analysis (also referred to as the problem of modifiable areal unit, change of support, or spatial misalignment, where spatial data at different aggregation levels are used within an analysis), multivariate disease analysis (where multiple diseases are jointly modeled spatially), spatial survival and longitudinal analysis (where both temporal variation and spatial effects are considered in survival and longitudinal analysis), and spatiotemporal disease mapping (where disease maps are analyzed with an associated temporal dimension as applied in the modeling of spatiotemporal spread of infectious disease). Discussions on applications are separated into those for case event (point) and count (aggregated) data. Examples and datasets for illustration are mostly from the United States and United Kingdom. Materials in the book are clearly presented and well laid out in general making it more attractive to read than many other books on Bayesian methodologies, in part because it uses short thick lines and plenty of spacing in between sections. That said, definitions or explanations of abbreviations used (e.g., CH, UH, AKA) are often not given on the same page. This might create difficulties for some readers in following the associated discussions, especially for those who do not read chapter by chapter from the beginning of the book.

Like many other textbooks on Bayesian methodologies, this book provides readers the datasets and `WinBUGS` codes for some of the examples discussed. It would be better if `WinBUGS` codes for all models fitted were provided to guide readers to implement any of the models. For example, in the analysis of Georgia low birth weight on pages 262–264, seven models are compared but the `WinBUGS` code for only one of the models is provided. At the time when this review was written, those datasets, `WinBUGS` file, and R codes from the book that are supposed to be acquirable from the author’s website

(<http://www.sph.sc.edu/alawson/default.htm>) were inaccessible and therefore could not be evaluated for this review.

In conclusion, my opinion is that readers who do not have a fairly strong background in statistics yet are trying to learn how to apply Bayesian methodologies in analyzing spatial data will benefit most from reading this book. In fact, many of the methodologies discussed in this book are applicable not only to spatial epidemiology but also to many other fields that utilize spatial data. As far as spatial epidemiology is concerned, this book is useful for Biometrics readership. Its work of focus is spatial biostatistics, as described precisely by the author.

REFERENCES

- Bailey, T. C. (2007). Book review of Lawson A. B., *Statistical Methods in Spatial Epidemiology* (second edition), John Wiley and Sons Ltd., Chichester, 2006. *Biometrics* **63**, 968–969.
- Lawson, A. B. (2006). *Statistical Methods for Spatial Epidemiology*, 2nd edition. New York: Wiley.
- Lawson, A. B. and Williams, F. L. R. (2001). *An Introductory Guide to Disease Mapping*. New York: Wiley.
- McNally, R. J. Q. (2002). Book review of Lawson A. B. and Williams, F. L. R., *An introductory guide to disease mapping*, John Wiley and Sons Ltd., 2001. *Biometrics* **58**, 693–694.

J. LAW

Department of Health Studies and Gerontology
University of Waterloo
Ontario, Canada

LOMBARDO, J. S. and BUCKERIDGE, D. L. (eds). **Disease Surveillance: A Public Health Informatics Approach**. Wiley, Hoboken, New Jersey, 2007. xxiii + 458 pp. US\$120.00/€85.80. ISBN 9780470068120.

In the face of an increasing threat to public health—on the one hand due to the rapid spread of partially unknown infectious diseases and on the other hand due to bioterrorism—advanced disease surveillance systems are needed. These should be able to automatically acquire data, archive, document, and process them to eventually present the data in a comprehensive format to the user such that decisions can be quickly made. The development and maintenance of such systems require an interdisciplinary approach bringing together skilled personnel from information technology, public health, medicine, epidemiology, and biostatistics. A fast adaptation of disease surveillance systems to a changing environment is a major and challenging task in their implementation.

The objective of this book is to address this broad perspective and thus to cover all components of a modern and effective disease surveillance system. For this purpose, the book is subdivided into three major parts where the single chapters are all written by experts with experience in the specific aspects of disease surveillance systems that are of interest in the respective part. Each chapter contains several real examples to illustrate its topic and ends with a brief summary. In addition, the chapters are all interrelated by briefly reviewing the topic of the previous chapter in the overview of the current one.

The first chapter serves as an introduction to disease surveillance where not only a historic perspective is given, but also the components of an early recognition disease

surveillance system are described presenting some examples of systems in use today.

Chapters 2–5 are more devoted to the technological and methodological aspects of the design and implementation of modern disease surveillance systems. Chapter 2 introduces several potential data sources for disease surveillance purposes that may contain early health indicators. It also discusses possibilities for evaluation and comparison of different data sources. These data sources come into play again in Chapter 3, where it is described how the data can be acquired and stored. Standards in health communication are reviewed, including issues related to data protection. Chapter 4 examines analytic processes that can be applied to data streams to achieve timely recognition of abnormalities with minimum false alarms. Chapter 5 focuses on the processes and techniques to visualize and display data to the users. It also discusses various system architectures and Web-based applications and ends with a description of skills that are required to operate and maintain such disease surveillance systems.

The second part of the book comprises Chapters 6–10 that present various case studies. Chapter 6 reviews the requirements that are posed by health departments to monitor the health of the respective population and describes systems that are in operation, also providing examples where an electronic surveillance system supported epidemiologists to identify a public health issue and to react accordingly. While Chapter 6 focuses on systems in the United States the remaining chapters summarize experiences with advanced surveillance systems outside the U.S. Chapter 7 describes an initiative within the Public Health Agency of Canada, whereas Chapter 8 describes the use of tele-health data in England and Wales. Chapter 9 is devoted to good practice examples for the surveillance of emerging infectious diseases in developing countries.

In the third part of the book, Chapter 10 covers aspects that are related to the evaluation of disease surveillance systems, such as the context in which the evaluation takes place, the components of the evaluation process, and methods to measure the performance of surveillance systems. Chapter 11 reviews current training practices in surveillance such as distance-education-based opportunities. Potential future advances and the current progress in the field of automated disease surveillance are the topic of Chapter 12.

Although some of the chapters could be more concise, the book is especially valuable for anyone interested in automated disease surveillance because of its broad scope addressing all issues related to developing and operating automated disease surveillance systems. As the authors themselves state, the book is intended as a textbook for public health informatics students. The study questions given at the end of each chapter help to review the contents of each chapter and to get the take-home message. The book is also intended to serve as a reference for health departments and as teaching material for workshops on disease surveillance.

IRIS PIGEOT

Bremen Institute for Prevention Research
and Social Medicine
University Bremen
Bremen, Germany

SELVIN, S. **Survival Analysis for Epidemiologic and Medical Research**. Cambridge University Press, New York, 2008. xiii + 282 pp. US\$41.95/£22.99. ISBN: 9780521719377.

This book explains survival analysis to readers with a minimal background in statistics, with emphasis on intuitive explanation of the methods, rather than on statistical theory, and on the interpretation of results. Data from the San Francisco Men's Health Study, a population-based prospective investigation of the epidemiology and natural history of AIDS conducted at the time the epidemic was emerging, are used throughout the book to illustrate the statistical methods presented. In the spirit of the series (Practical Guides to Biostatistics and Epidemiology), it introduces the computational tools for running survival data analysis in R. The book covers in nine chapters the basics of survival analysis, presenting the analytic approaches most commonly used because of their simplicity in application and interpretation, albeit sometimes without highlighting sufficiently the limitations of these methods.

Chapter 1 is a nice introduction to survival probability, average rates, and hazard rates, which could be even more efficacious if the part on the binomial transformation for normal approximation would be moved to the appendix. The difference between probability and rate (inverse of time dimension) is not easy to teach without mathematical tools but here the topic is nicely covered and almost always the author is able to avoid the use of the word "rate" to indicate both (which is allowed in English), which is often a matter of confusion to nonstatistically oriented readers. Censoring is not even mentioned in the first chapter and is not needed in the first part of Chapter 2, where the construction of a current complete life table is presented. Not until the second part of Chapter 2 is the topic of noninformative right censoring introduced, when the author discusses an example of what he calls "Clinical life table analysis" with "incomplete observations" of individual follow-up data. The actuarial estimators of the yearly conditional probability of death and hazard rate are presented distinguishing the different information that these two quantities give on the survival experience. One inconsistency is present here between the definition of the hazard rate given in the previous chapter and the estimator that does not adjust the effective person-time at risk by subtracting half the number of deaths (as done in the standard actuarial estimator of the average death rate over an interval). Chapter 3 presents the principles of maximum-likelihood estimation and likelihood-based testing (Wald test, likelihood-ratio test, with the concept of nested models, but not the score test) in a clear and concise way, without any use of mathematics. Chapter 4 nicely introduces the product limit estimator (with some imprecision in the definitions given for the different types of censoring) and the calculation of pointwise confidence bounds based on appropriate transformation of the cumulative probability of survival. By showing the exponential model, Chapter 5 introduces with simplicity and clarity the crucial aspect of estimation and goodness of fit in parametric modeling. Nice features of this chapter are that the parameter estimate is derived by using the memory-less property of the model and that a final example on U.S. lung cancer mortality data shows the use of the model to derive age-adjusted population survival probabilities. Chapter 6 introduces the Weibull model,

discusses the interpretation of its parameters and goodness of fit by means of an application. It also introduces very simply the driving concept of simulated survival times from a given distribution, without details (and without censoring). The exponential and Weibull models are extended to include regressors in Chapter 7. This covers the parametric comparison of two samples, after showing the nonparametric approach with the log-rank test only (other tests are not mentioned). The model-based approach deals with comparison under proportional hazards only, but the parameterization chosen for the Weibull model also allows showing quite simply the possibility to interpret the parameters in the context of accelerated time models. Chapter 8 is on the potentialities and limitations of applying proportional hazards regression models. It reviews how to include covariates in a regression model (continuous versus categorical variables—without explaining dummy variables—and first-order interaction terms) and how to use residuals for assessing goodness of fit (Cox and Snell residuals). This chapter is general enough to lead into the following one on the Cox model (Chapter 9), where these concepts are used again to explain the interpretation of the estimated regression coefficients and to assess goodness of fit (with some inconsistencies and typos in the formulas for the modified Cox and Snell residuals). Chapter 10 gives the R code related to SURFIT, SURVREG, SURVDIFF, COXPH, which are fundamental functions for any statistical analysis of survival data. A nice problem set that does not require computer computation is given at the end.

In conclusion, this book meets its goals of presenting efficaciously, using only elementary mathematics, those methods that should be known by everybody with interest in understanding survival data analysis in clinical and epidemiological studies. The book makes the ideas of survival analysis accessible to nonstatisticians involved in public health and clinical research or to graduate students with basic statistical background. It can be a very useful book to provide them with training on the fundamental techniques used in most applications.

M. G. VALSECCHI

Center of Biostatistics—Dept. of Clinical Medicine
and Prevention
University of Milano-Bicocca
Monza, Italy

AALEN, O. O., BORGAN, O., and GJESSING, H. K. **Survival and Event History Analysis: A Process Point of View**. Springer, New York, 2008. xviii + 554 pp. US\$84.95/€69.50. ISBN 9780387202877.

Recently, there has been a spate of outstanding books on survival: Cook and Lawless (2007), Duchateau and Janssen (2008), Martinussen and Schieke (2006), O'Quigley (2008), and Pintilie (2006). The book under review is unique in this group; it focuses on stochastic processes as the lenses through which to view survival data. These include counting processes with related martingales and submartingales, Poisson, renewal, Markov, Weiner and related diffusion processes, and Levy processes. They are used to model not only standard biomedical data—survival, multistate and event history data, recurrent events, and multivariate (clustered) survival—but

also the underlying hazard and dynamic processes, both frailty and causal. The authors emphasize concepts over mathematical technicalities, but without “dumbing down” the presentation. As they point out (p. 42) “. . . the basic ideas and results for stochastic processes are mostly relatively simple, and that they can be understood at an ‘intuitive technical’ level without going into mathematical details. . . Very often, the intuitive content in stochastic processes tends to drown in complex mathematical presentations, which is probably the reason this material is not so much used in applied statistics. We want to contribute to a demystification of martingales, stochastic integrals, and other stochastic process concepts.” The authors have succeeded admirably. Measure theory is neither presented nor required. The most mathematical chapter, Chapter two, which presents counting processes martingale theory, is a delight. The authors use discrete time processes to explain the mysterious concept of a predictable process and the key theorem, the Doob–Meyer decomposition. By focusing on martingales with continuous compensators living on finite-time intervals, referring to conditioning on the past (filtration is not mentioned in the comprehensive index) and eschewing local processes, they let the elegance of martingale theory reveal itself. For other processes, an appendix gives the basic material. When a process appears in the text, clarity is ensured by plentiful graphics picturing its behavior and thorough discussion and derivation of relevant formulae, always relating them back to the simplest cases.

Middle chapters cover applications of martingale theory: nonparametric analysis, semiparametric regression for both relative risk (Cox model and excess relative risk) and additive models, and parametric likelihood. Each chapter gives thorough descriptions of several data analyses, most of which had direct involvement of one or more of the authors, postponing theoretical discussion to the end. Highlights include nonstandard data examples, such as the Nelson–Aalen estimator for mating flies, a discussion of practical aspects of the product integral for Markov process transition matrices, the clearest derivation of the likelihood for parametric counting processes that this reviewer has ever read, and the counting process formulation for nested simple and counter-matched case–control sampling. None of this is surprising in view of the authors’ expertise. Aalen introduced counting process/martingale theory to statisticians in his thesis and invented the additive model for counting process intensities. Borgan developed this field for parametric likelihood, did fundamental work on martingale theory for nonparametric tests, and was the main contributor to nested case–control sampling. Gjessing’s strengths in genetic/family studies and diffusion and Levy processes come into play in the latter half of the book where new and/or controversial areas are discussed.

The last half of the book focuses on three intertwined themes: frailty, dynamic analysis, and causality. Much of the material is new, all is exciting. The tone is different from a standard statistics textbook. It is almost philosophical; not in the derogatory sense used by the late Dr Richard Feynman of vacuous verbiage substituting for content, but in the sense of deeply thoughtful discussions of big concepts, drawing on a variety of sources and viewpoints. The mathematics is all there, clearly explained, but subordinate to ideas and applications. Real examples are drawn from recurrent event and mul-

tivariate (clustered) survival data. Relative risk and likelihood models are considered but the emphasis is on additive models. Chapter 6 discusses univariate frailty, showing how ignoring unobserved heterogeneity may mislead, with many striking examples, including a parallel drawn between crossing population hazards and Simpson’s paradox. Shared frailty models follow, particularly for parametric likelihood, but the highlight is the new development of a multivariate frailty forming a hierarchy based on Levy distributions. An application to testicular cancer data with familial dependence is given. Dynamic analyses use time-varying covariates based on the past history of the process. They are of key importance in causal models. Aalen additive linear models can separate direct and indirect effects of variables of interest, such as interventions and portray them via dynamic path diagrams, directed acyclic graphs whose vertices are partitioned into covariate processes and the outcome process. The edges represent effects of some covariates on others, ultimately impinging on the outcome (indirect effects) and direct effects of covariates on the outcome. Each event in the outcome process is portrayed by a potentially different set of edges in the graph and each vertex including the outcome is modeled as a linear function of the covariates acting on it. Plotting the regression coefficients as functions of time gives a dynamic picture of the size of potentially causal indirect and direct effects.

Dynamic path analysis gives the beginnings of a causal picture. The authors provide a caveat: “. . . any kind of dependence observed in a process can . . . always be interpreted as effects of the past; that is, it can always be given a dynamic interpretation” due to the Doob–Meyer decomposition. “On the other hand, any kind of dependence can also be given a frailty type interpretation with no effects of the past involved. . . generally, any dynamic effects may reflect unobserved underlying variables, or they may represent real causal effects. Without information besides the processes themselves, we don’t know.” Nonetheless, the authors state that statisticians should be more concerned with causality. They describe several historic and current concepts of causality but prefer considering it as the dynamic influence of stochastic processes on each other.

The final chapters explore dynamic models for hazards and frailties. Hazards may be modeled by the time to absorption of Markov chains with a single absorbing state or by hitting times of diffusion processes. This allows for two different covariate influences: the distance between starting point and barrier and the strength of the drift toward that barrier. An R program for this will be appearing on the book Web site shortly. The book concludes with the ultimate use of stochastic processes: functions of diffusion processes model hazards and functions of time-transformed Levy processes model frailties. Despite the somewhat speculative nature of this material, the authors derive interpretable models that can be fit to data, including a dynamic frailty model fit to Norwegian divorce data for three different decades, providing another interpretation to compete with the well-worn “seven year itch.”

The book is intended as a text for biostatistics graduate students. It will fill that role excellently. It will expose them to ideas they are unlikely to encounter in depth in a standard curriculum and is precisely the sort of book to inspire theses and other research projects. Prerequisites include exposure

to stochastic processes and basic survival analysis, as well as the mathematical statistics that the standard graduate program provides. Each chapter contains relevant probability theory and data analyses and concludes with a set of exercises. Although solutions are not given, hints are often provided. With its comprehensive and up-to-date bibliography, and extensive index, it is also ideal for self-study. There are only a few errata. The book has Springer's high quality with pleasing typesetting and good margins. The book Web site at Springer contains an author's page, still under construction, with corrections and the data and descriptions for seven of the datasets used in the book. My only complaints are the dearth of practical information on computer implementation to replicate analyses from the book. There is very occasional mention of \mathbf{R} , but nothing further. Finally, why "prognostic index" when "risk score" is so much simpler?

REFERENCES

- Cook, R. J. and Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Events*. New York: Springer.
 Duchateau, L. and Janssen, P. (2008). *The Frailty Model*. New York: Springer.
 Martinussen, T. and Schieke, T. H. (2006). *Dynamic Regression Models for Survival Data*. New York: Springer.
 O'Quigley, J. (2008). *Proportional Hazards Regression*. New York: Springer.
 Pintilie, M. (2006). *Competing Risks: A Practical Perspective*. New York: Wiley.

PATRICIA GRAMBSCH

Division of Biostatistics, School of Public Health
 University of Minnesota
 Minneapolis, Minnesota, U.S.A.

LEE, J. A. and VERLEYSEN, M. **Nonlinear Dimensionality Reduction**. Springer, New York, 2007. XVIII, 310 pp. US\$79.95/€64.95. ISBN: 978-0-387-39350-6.

With the emergence of modern data-gathering technology and data storage devices, large and increasingly complex data that are being collected are often from high-dimensional spaces with unknown low-dimensional nonlinear underlying structures. In another words, these data often lie close to low-dimensional manifolds. Such low-dimensional representations (manifolds) of high-dimensional data often contain valuable information, which requires advanced statistical techniques for its extraction. For example, nonlinear structures are often found in the studies of medical imaging, computer visions, motor control, speech processing, and a much larger range of other physical and biological sciences. Currently, there is still a shortage of practical statistical methods to understand the geometric properties of such manifold structures. As a result, a text that bridges nonlinear dimension reduction techniques with high-dimensional data analysis will provide a useful contribution to the researchers and practitioners from various scientific fields.

This beautifully produced book covers various innovative topics in nonlinear dimensionality reduction, such as Isomap, locally linear embedding, and Laplacian eigenmaps, etc. Those topics are usually not covered by existing texts on multivariate statistical techniques. Moreover, the text offers an excellent overview of the concept of intrinsic dimension.

Special attention is devoted to the topic of estimation of the intrinsic dimension, which has been previously overlooked by many researchers.

The book includes seven chapters. In Chapter 1, several examples are used for both practical and theoretical motivations. Two benchmark manifolds, the "Swiss roll" and "open box," are described for comparison in later chapters. Chapter 2 summarizes some possible features as well as properties that are used to characterize high-dimensional data analysis methods. In Chapter 3, the authors give a brief overview of the concept of the intrinsic dimension, including topological dimension, fractal dimension, and other topics. Several dimension estimators are discussed with comparisons based on synthetic data. Chapters 4 and 5 focus on various dimensionality reduction techniques using distance preservation criterion and topology preservation criterion, respectively. Those techniques include multidimensional scaling, Isomap, self-organizing map, locally linear embedding, and Laplacian eigenmaps. For each method, the authors have developed mathematical details and outlined algorithmic implementation and also pointed out the advantages as well as drawbacks. Those techniques are carefully compared through extensive numerical studies in Chapter 6. In Chapter 7, some concluding remarks and directions for future development are provided. Overall, the topics selected by the authors are very attractive and impressive. It is not perfect, of course. For instance, there is no discussion of confidence intervals of those dimension estimators. Also, recent developments in factor analysis for nonlinear factors can also be very useful and competitive methods for dimensionality reduction.

A strong feature of the book is the style of presentation. The book is clearly written, and typographic errors are very few. A large number of examples and graphical displays in color help the reader in understanding the ideas. For each method discussed, the authors do a credible job by starting from motivating examples and intuitive ideas, introducing rigorous mathematical notation without being cumbersome, and ending with discussion and conclusion remarks.

All in all, this is an interesting book, and I would recommend this text to those researchers who want to learn quickly about this new field of manifold learning. This book will serve as a useful and necessary resource to several advanced statistics courses in machine learning and data mining. In practice, it would be a good choice as supplemental reading material alongside other textbooks on machine learning, such as Hastie, Tibshirani, and Friedman (2001) and Bishop (2006). For effective student use, a companion chapter in the book or website on the issue of computational implementation is crucial. In addition, the `Matlab` and `R` packages will surely enhance the learning resources and increase the accessibility of this book to data analysts.

REFERENCES

- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.
 Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning*. New York: Springer.

HAONAN WANG

Department of Statistics
 Colorado State University
 Fort Collins, Colorado, U.S.A.

CLAESKENS, G. and HJORT, N. L. **Model Selection and Model Averaging**. Cambridge University Press, New York, 2008. xvii + 312 pp. US\$70.00/£40.00. ISBN 9780521852258.

Model selection has seen renewed attention in the statistics literature lately as researchers attempt to undo the damage that has been done over the years by our reliance on stepwise selection procedures. Recent books spanning the gamut of technical levels and interests include Konishi and Kitagawa (2008), Anderson (2008), and the second edition of its popular predecessor, Burnham and Anderson (2002). Where the first of these books is almost entirely mathematical and the latter two are almost entirely applied, this new offering by Claeskens and Hjort is somewhere in the middle, although somewhat more mathematical than applied. Its target audience is reportedly master's students with a background in regression and statistics-savvy researchers in other disciplines. As will be outlined below, the former group may be well-served by the book, but the latter would likely be hard-pressed to follow along unless they were mathematical-statistics savvy.

The primary focus of this book is model selection based on information criteria. This topic takes up about two-thirds of the book, including separate chapters on the Akaike (AIC), Bayesian (BIC), and Focussed (FIC) information criteria, the latter of which is the authors' own development. After starting with a chapter introducing some examples, the chapter on AIC jumps quickly into vector notation without warning. While this chapter is not mathematically dense, it does deal primarily with derivations and mathematics-based explanations, and it demands a certain level of sophistication of its readers (it refers without explanation to Lévy processes and influence functions, for example). Descriptions of developments relating to AIC are given, as well as some applications to problems. The third chapter is a similar discussion of BIC, while Chapter 4 provides theoretical comparisons of AIC, BIC, and their kin.

The next two chapters turn to discussions related to some of the authors' own work, and represent the main area in which this book is unique in the literature. Chapter 5 digs into some mathematical aspects of full- vs. reduced-model comparisons. Chief among these is the concept of a "tolerance region" of values for a parameter, where incorrectly choosing a reduced-model value for the parameter nonetheless provides better estimation in terms of a bias-for-variance trade-off. This then lays the groundwork for the introduction and analysis of the FIC, which ranks models not on their closeness to some ideal, but rather on their ability to produce estimates that have good properties (e.g., low mean squared errors) for specific quantities of interest. Many examples are given, including some numerical examples.

There is one chapter on model averaging, which includes a very pointed discussion of the inappropriateness of standard inference methods when applied post-model-selection. A theorem is given unifying results on sampling distributions for model-averaged and post-model-selection estimators. This chapter also contains one small section on shrinkage-based selections such as LASSO and LARS. Research on these methods has led much of the charge behind the renewed interest in model selection that I mentioned above. It would have made this a much more complete book to give these subjects the attention that they deserve, but I also understand that the

authors' main interest and expertise is in areas relating to their own work. Readers should be aware that they will not learn about these other modern topics in this book.

Next, the book gives an interesting perspective on testing lack of fit by embedding the null model into a parameter-indexed family of expansions using basis functions. The decision rule is then whether the null model is selected from out of potentially infinitely many members of this family. A penultimate chapter gives several additional numerical examples. The final chapter gives some attention to further topics, such as missing data and boundary-value problems.

Examples presented throughout the book tend to be a mix of derivations for special parametric problems and analyses of actual numerical data sets. Each chapter begins with an abstract, which I think is a nice feature, and ends with notes on the literature and some exercises. Like the examples, the exercises are a combination of mathematical derivations and numerical problems. Rarely in the text, but more commonly in the exercises, reference to R functions is made or small snippets of R code are given.

The book is structured in such a way that reading it from front to back is really beneficial. The chapters are not meant to be self-contained: too often, reference is made to notation in previous chapters, leaving readers to locate and understand previous discussions before delving into whatever they really wanted to learn. Also, because much of the material on FIC is not yet well known, Chapter 6 would have been vastly improved from an application standpoint by the inclusion of detailed computational algorithms and/or R code for the examples.

In summary, this is a very fine book from which to teach a graduate level topics class on information-based model selection. It is also a must-read for researchers in model selection, especially those just starting off, even though it is not especially comprehensive. It is *not* suited as a reference for practitioners, as the applied content is minimal. Finally, its heavy notational reliance on previous chapters would make it difficult to use as a general reference for researchers in other areas wanting to connect quickly with model selection issues.

REFERENCES

- Anderson, D. A. (2008). *Model Based Inference in the Life Sciences: A Primer on Evidence*. New York: Springer.
- Burnham, K. P. and Anderson, D. A. (2002). *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*, 2nd edition. New York: Springer.
- Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling*. New York: Springer.

THOMAS M. LOUGHIN

Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC, Canada

BÖHNING, D., KUHNERT, R., and RATTANASIRI, S. **Meta-Analysis of Binary Data Using Profile Likelihood**. CRC Press, Boca Raton, Florida, 2008. 208 pp. US\$79.95/£44.99. ISBN 9781584886303.

Three broad approaches are predominantly used to combine information from collections of 2×2 tables resulting from clinical trials. These are (1) weighted combinations of

effect sizes, (2) likelihood-based approaches that approximate the distribution of the observed effect sizes, and (3) complete likelihood-based approaches that directly model event counts and so must involve nuisance parameters related to the baseline or control group risk. In using the complete likelihood approach to estimation, one might maximize over all the parameters jointly or average over a prior distribution for the nuisance parameters as in Bayesian modeling. Another way to work with the complete likelihood is to replace the nuisance parameters with their maximum-likelihood estimates to create the *profile likelihood* (PL). The PL can then be used as if it were a usual likelihood to estimate the parameters of interest. While PL methodology has been developed in other settings, this book presents its application to meta-analysis. The authors argue that the PL approach is superior to jointly estimating all the parameters because it allows the focus to be solely on the parameters of interest and results in estimates with smaller errors.

In particular, this book focuses on profile maximum-likelihood (PML) estimates of risk ratios derived from independent Poisson models for event counts in the two arms of clinical trials. Binomial distribution-based PLs and the estimation of odds ratios are also presented, but in less detail. After introducing meta-analysis in broad strokes, the authors present a thorough discussion of the use of the PL to estimate the risk ratio in the case of no covariates or between-study effect variation. Although the discussion is in places technical, the authors' presentation includes enough detail that a dedicated reader unfamiliar with this topic can follow the arguments. Because closed-form solutions for these PML estimates are not available, the authors also present and compare numerical algorithms. This simple PL is then extended to include covariates as well as study effect clustering. Some technicalities of estimation are included but the discussion is not as detailed, and readers not familiar with mixture modeling will likely need to learn the basics of this approach before reading those sections. The authors also propose new measures of heterogeneity that compare favorably to conventional measures.

This book soundly presents the foundational ideas of the use of the PL in meta-analysis and includes much material that has not been published elsewhere. Its descriptions are greatly enhanced by the authors' consistent use of real examples. I also appreciated the inclusion of other currently used methods alongside the presentation of PL methods for meta-analysis. The details of these alternate methods are appropriately brief with relevant references. The authors wisely recognize that the PL-based estimates they present in this book are not simple to compute, and so have provided freely downloadable Windows-based software: Computer Assisted Analysis of Meta-Analysis using the PL Model (CAMAP). CAMAP was easy to download, install, and use. It computes estimates, standard errors, and measures of heterogeneity for all the Poisson-based methods for relative risk PML estimation presented in the book.

Although in general the presentation of PML estimation was good, several topics may have benefited from a more thorough discussion. For example, there were inconsistencies in the authors' recommendations regarding likelihood-ratio tests for model selection. In light of PL methodology's reliance on asymptotic arguments, it would have been useful to

describe the performance of PML estimates in common finite settings. I also felt that better support would have improved a few of the arguments for choosing the PL approach over other methods.

Despite these few reservations, I am not aware of a more complete source for this topic. The authors' presentation of the core ideas behind the derivation and use of PML estimates is accessible to anyone familiar with standard likelihood-based estimation. The many good examples facilitate intelligent application of these ideas, and the described software makes implementation simple. This book will be useful to those who would like to learn about using PL-based methods as an alternative to the currently available approaches to the combination of information from collections of 2×2 tables.

ELOISE KAIZAR
Department of Statistics
Ohio State University
Columbus, Ohio, U.S.A.

NASON, G. P. **Wavelet Methods in Statistics with R**. Springer, New York, 2008. x+257 pp. US\$59.95/€48.10. ISBN 9780387759609.

Wavelets are a specific class of basis functions that can parsimoniously represent many types of functional data. They are derived to possess certain properties that make them outstanding candidates for a number of quantitative tasks, including data compression, signal and image denoising, change point and peak detection, and the modeling of irregular functional data. Ignited by the seminal work of Ingrid Daubechies (1992), wavelet research has exploded in the past two decades, with wavelets becoming the dominant tool in many signal and image processing applications, and gaining increasing attention in other application areas, including statistics. Statistical interest in wavelets has also grown during this period, sparked by the work of Donoho and Johnstone (1994, 1995) demonstrating how wavelets could be used to perform adaptive denoising in nonparametric regression models. This book deals with the contributions of wavelets in statistical research, with its primary aims being to "cover the main areas in statistics where wavelets have found use or have potential" and (promote the) "use of wavelet methods as well as their description." The author has been an active researcher in wavelet methods in statistics since their early days.

This book is clearly written and well laid out. It starts with an excellent, intuitive introduction to wavelets. Here, the key properties of wavelets are built up from basic principles, using the concrete and easy-to-understand Haar wavelets as the backdrop. This approach helps readers unfamiliar with wavelets to quickly comprehend the key concepts of wavelets, which can be very tricky when one is first beginning study of this area. This presentation successfully strikes a careful balance of keeping the presentation intuitive and straightforward, yet precise and accurate. Following this introduction, the book extensively discusses wavelet shrinkage and describes how this concept is used to perform adaptive nonparametric regression of functional data. These sections provide a nearly comprehensive treatment of the subject, presenting methods from both Bayesian and frequentist perspectives. The remainder of the book then covers details regarding various

more advanced topics in wavelet research in which the author and his collaborators have made contributions, including locally stationary time series and multiscale variance stabilizing transformations.

One useful aspect of the book is that the code for generating all of the figures is made available on the Web, allowing the reader to follow along and gain first-hand experience applying the methods contained within. The code uses an R package `WaveThresh` (available at <http://www.stats.bris.ac.uk/~wavethresh>) that has been developed by the author. This package is straightforward to download and use, and contains myriad functions for performing various wavelet-based analyses. This will increase the book's applied impact, as researchers will be able to quickly move from reading about the methods described in the book to applying them to their own data sets.

This book would be a very good introductory textbook for graduate students and researchers interested in gaining access to the field, and also serve as a useful reference for wavelet-based research for more experienced researchers. There are a large number of books providing an introduction to wavelets, but only a few that focus on statistical applications, and most of those written for statisticians are several years old, and so do not contain some of the more recent material. *Wavelet Methods in Statistics with R* contains a thorough overview of the various alternative methods available for wavelet regression, and also contains more advanced material not contained in the other books, including methods for locally stationary time-series analysis and connections of wavelets with variance stabilizing transformations. *Wavelet Methods for Time Series Analysis* by Percival and Walden (2002) provides a more complete survey of wavelet approaches for time-series analysis, but does not include the locally stationary time-series methods included here.

This book does not claim to be a comprehensive treatment of wavelets in statistics, but one of its stated aims is to describe areas of statistics where wavelets have found application or show promise. While thoroughly covering certain major areas, this book does not provide any coverage of several recently emerging applications of wavelets to statistics. One primary area is the field of functional data analysis, which involves the analysis of samples involving multiple functions or images. The unique properties of wavelets that make them suitable for nonparametric regression of irregular functions in the single-function case also make them useful for modeling samples of irregular functions in functional data analysis settings. There has been recent research on function registration (e.g., Bigot, 2006), clustering (e.g., Ray and Mallick, 2006), classification (e.g., Wang, Ray, and Mallick, 2007), and functional mixed models (e.g., Morris and Carroll, 2006). There is also some recent research using wavelet-based models for change point detection (e.g., Antoniadis and Gijbels, 2002). This work is more recent and emerging, but it would have been helpful for there to be some treatment of these topics in this book.

I am frequently asked by students or statistical researchers to recommend a book that would give them an introduction to wavelets. In the past, I struggled with which book to recommend, as I have found some presentations to be too technical and difficult to follow for a beginner, and others to

be accessible but not thorough and detailed enough. *Wavelet Methods in Statistics with R* strikes an excellent balance, motivating the concepts intuitively through concrete examples, yet providing complete and accurate coverage of the underlying ideas. Paired with the R package that implements the methods discussed in the book, this book is a useful tool for not just gaining background in the field but also equipping the reader to apply these methods to their own data sets. It is a book I will heartily recommend to statisticians looking for an entry point into the field of wavelets.

REFERENCES

- Antoniadis, A. and Gijbels, I. (2002). Detecting abrupt changes by wavelet methods. *Journal of Nonparametric Statistics* **14**, 7–29.
- Bigot, J. (2006). Landmark-based registration of curves via the continuous wavelet transform. *Journal of Computational and Graphical Statistics* **15**, 542–564.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: SIAM.
- Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81**, 425–455.
- Donoho, D. L. and Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association* **90**, 1200–1224.
- Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society, Series B* **68**, 179–199.
- Percival, D. B. and Walden, A. T. (2002). *Wavelet Methods for Time Series Analysis*. New York: Cambridge University Press.
- Ray, S. and Mallick, B. (2006). Functional clustering by Bayesian wavelet methods. *Journal of the Royal Statistical Society, Series B: Statistical Methodology* **68**, 305–332.
- Wang, X., Ray, S., and Mallick, B. K. (2007). Bayesian curve classification using wavelets. *Journal of the American Statistical Association* **102**, 962–973.

JEFFREY S. MORRIS

Department of Biostatistics
The University of Texas, M.D. Anderson Cancer Center
Houston, Texas, U.S.A.

CHANG, M. **Adaptive Design Theory and Implementation Using SAS and R**. Chapman & Hall/CRC, Boca Raton, Florida. xxii + 418 pp. US\$89.95/£57.99. ISBN 9781584889625.

Adaptive design has attracted a lot of attention in the past ten years. This book provides a unified presentation of adaptive design theories. Also, it provides computer programs in SAS and R for the design and simulation of adaptive trials. The book tries to remove barriers and to close gaps between advanced statistical methodology and applied research in clinical trials.

Chapters 1, 2, and 3 provide a good overview of clinical research and development. Chapter 1 gives a brief introduction to the commonly used adaptive designs and places emphasis on the importance of computer simulations in adaptive designs. Chapter 2 reviews the classical two-arm designs including superiority, noninferiority, and equivalence, followed by a discussion of dose–response studies, where a unified formulation and user-friendly SAS programs for sample size calculation are given. In Chapter 3, general theory of adaptive designs is introduced, which involves stopping boundaries,

formulas for power and adjusted p-values, point estimation, and derivation of confidence intervals.

Chapters 4–16 discuss the different adaptive approaches in detail. Chapters 4 and 5 introduce the family of methods called combinations of p-values. While the former focuses on methods based on a direct combination of p-values such as individual p-values, sum of p-values, and product of p-values, the latter studies methods based on linear combinations of weighted inverse-normal transformations of the stage-wise p-values (MINP). The most striking feature of MINP is that it unifies different adaptive methods including the Lan–Demets error-spending method, the Lehman–Wassmer method, and others. Chapter 6 generalizes the two-stage designs from the previous two chapters to K stages by use of simulation where normal, binary, and survival endpoints are considered.

Chapter 7 studies two-stage adaptive methods based on conditional error functions. For the Proschan–Hunsberger method, a conditional error function, as a function of the first-stage normal statistic, is pre-specified so that the unconditional error for the whole procedure equals a fixed value α . The advantage of this approach is that it allows both the treatment effect and nuisance parameters to be estimated from the data.

Chapter 8 explores the method called recursive adaptive design. It is more of a general methodology than a specific method in that it views any K -stage design as recursive two-stage designs, and thus can be adapted to all the two-stage methods. Another great benefit of recursive design is that no numerical computation is needed, since the conditional type I error required for the rest of the trial can be directly computed from the corresponding two-stage design.

Chapter 9 introduces different sample-size modification methods. Two types of sample size adjustment rules are given: one based on the effect-size ratio between the initial estimate and the observed estimate, the other based on conditional power. These rules are combined with general adaptive designs, and simulations are used to evaluate different approaches. In particular, a futility boundary is suggested to save cost. Also, the author points out that for adaptive designs, conditional power is more important than unconditional power.

Chapters 10–13 cover multiple-endpoint adaptive design, drop-lose and add-arm design, biomarker-adaptive design, and adaptive treatment switching and crossover, respectively, all of which are new and challenging areas invoked by studies such as cancer and HIV.

Chapters 14 and 15 concentrate on two types of response-adaptive designs, one for treatment allocation, and the other for dose finding, respectively. For the former, play-the-winner model, randomized play-the-winner model, and generalized urn model with various endpoints are introduced and their asymptotic properties are explored. Also, it is pointed out that these models take ethical concerns into consideration by providing a better chance of randomizing the patients to a superior treatment. For the latter, two commonly used approaches for oncology dose-escalation trials—algorithm-based escalation rules and the model-based approach—are introduced. Recently, however, many new families of response-adaptive

designs (with better properties) have been proposed in literature. These are missed in the book.

In Chapter 16, the Bayesian approach is compared to traditional frequentist methods. The differences are identified in various aspects. Bayesian rules for classic designs and adaptive designs are demonstrated. As noted by the author, the incorporation of a Bayesian approach into adaptive designs provides a powerful tool for monitoring trials, predicting outcomes, anticipating problems, and suggesting early remedies.

Chapters 17 and 18 provide concluding notes on the adaptive designs introduced in this book. In Chapter 17, practical concerns including planning, execution, analysis, and reporting are discussed. The author places emphasis on the combination of adaptive design, the Bayesian approach, and trial simulation in drug development. And in Chapter 18, controversial issues surrounding adaptive designs are discussed from both statistical and philosophical perspectives. Statistical principles such as sufficiency, conditionality, and exchangeability are examined for adaptive designs so that readers can have a whole picture of both the advantages and disadvantages of adaptive designs.

In summary, this book provides a systematic introduction to adaptive design theory. It also gives trial examples and computing code to help readers to construct a comprehensive understanding of adaptive designs. It is certainly a useful guide and reference for academic and industry statisticians alike.

FEIFANG HU

Department of Statistics

University of Virginia

Charlottesville, Virginia, U.S.A.

BOLKER, B. M. *Ecological Models and Data in R*. Princeton University Press, Princeton, New Jersey, 2008. xii + 396 pp. US\$55.00/£39.95. ISBN 9780691125220.

Over the last 20 years, statistical researchers have been having a field day with complex ecological models. Bugs and beasties have all the traits to get the modern statistician lapping for more, with their nonlinear dynamics, hierarchical processes, and flagrant heteroscedasticity. And you can bet your bottom dollar those elusive critters would not be missing-at-random. Meanwhile, vanguard ecologists have begun downloading R and BUGS, but the learning curve is steep. Benjamin Bolker's *Ecological Models and Data in R* offers a way to unite the two camps. By bringing modern statistics to ecologists using R syntax, and by encouraging a philosophy of learning by exploration and simulation, we can hope for a sparkling future in which statisticians and ecologists speak—and program in—the same language.

The Introduction suggests as prerequisites a one-semester undergraduate statistics course and “rusty” calculus, with no previous programming experience necessary. Some 400 pages later, the Afterword issues the reader with a well-deserved “Welcome to the Cutting Edge.” It's a tall order to gallop from near-scratch to the cutting edge in 400 pages, learning programming along the way, and I would suggest a firmer grounding in statistics would be needed for a full appreciation

of the book's many gems. Most students will first be in a position to engage with the material at master's and PhD level, and the text is certainly not too basic for statistics majors. For those from other backgrounds, I would recommend a firm grasp of basic probability—including the negative binomial, Poisson, and normal distributions, Bayes' theorem and conditional probability; basic data analysis such as linear regression and ANOVA; and at least some exposure to maximum likelihood for estimation and confidence intervals. Chapters 5 to 8 would be superb training for students from either statistics or ecology, to bridge the chasm between taught courses and independent research.

The text is strongly focused on examples from ecology, but most of these could be grasped without any formal ecological training. As an accessible introduction to many topics in modern statistics, the text would be a valuable asset to practitioners in any field. I would also recommend it for experienced statistical ecologists, who will appreciate the succinct coverage of a wide range of methods, and the useful R functions, including those supplied in Bolker's custom-written packages `embdbook` and `bbml`. If, for example, such a researcher were about to conduct their first GLMM or MCMC analysis, this book would be a good port of call to get the quick how-to's, without having to dredge through long-mastered theory or details. Throughout, the treatment is even-handed between Bayesian and classical approaches.

Chapters 1 to 4 cover background in modeling philosophy, R and graphics, mathematical functions and calculus, and probability distributions, respectively. While these chapters could be used to accompany taught classes as a first introduction to the material, they are probably best treated as reference sections, with the exception of Chapter 2 if the reader has not used R before. The impatient reader could happily jump straight to Chapter 5, with occasional reference backwards. Chapter 5 is a gem, showing how to use simulation in R to explore models, test code for estimation, and assess bias, variance, and power. The world would be a better place if all practitioners conducted such explorations as a matter of course, as Bolker advocates.

Chapters 6 and 7 cover likelihood and optimization, and are substantial, enjoyable, and informative, while worked examples with abundant R code are given in Chapter 8. Here, the real strengths of the book emerge: we have exactly what we need to create models, code them, fit them to data, construct confidence intervals, and conduct model selection. I particularly appreciated Bolker's openness about the difficulties inherent in model-fitting off the beaten track: occasionally his own models fail to converge, or he is caught out by a multimodal likelihood surface. Learning that these hiccups are part of the job will be an eye-opener for many grad students. Chapter 9 is a useful blast through well-trodden statistical ground, while Chapters 10 and 11 cover the more advanced topics of mixed models and dynamic models.

The style is informal and accessible, with an intuitive slant and lots of graphics to help. There is sometimes a tendency to display too much information on a single plot, but this could be used as a prompt to persuade students to recreate the plot for themselves, step by step. Above all, *Ecological Models and Data in R* is a how-to manual, and the conscientious reader will be rewarded with a much deeper understanding of their statistical models, as well as the ecological systems they represent.

RACHEL M. FEWSTER
Department of Statistics
University of Auckland
Auckland, New Zealand

KROONENBERG, P. M. **Applied Multiway Data Analysis**. Wiley, Hoboken, New Jersey, 2008. xxi + 579 pp. US\$115.00/€109.00. ISBN: 9780470164796.

Statistical reasoning is based on an appropriate processing of empirical information (statistical data) and "theoretical" information (assumptions, models), aimed at obtaining additional information for knowledge acquisition and decision making. The type and quality of empirical information is crucial for the development and utilization of suitable methods, capable to capture the informational content of the data.

Traditional two-way data ("individuals \times variables," or proximity matrices) constitute the object of the classical techniques of multivariate analysis (from generalized linear models, to principal components and multidimensional scaling). When multiway data (e.g. "individuals \times variables \times occasions" arrays) are taken into consideration, specific methods are required, enabling the researcher to grasp the new informational features stemming from this particular data structure.

Two pioneering volumes devoted to this topic preceded the publication of Kroonenberg's book (Law et al., 1984; Coppi and Bolasco, 1989). A further relevant reference is Coppi and Di Ciaccio (1994). However, *Applied Multiway Data Analysis* constitutes the first single-author in-depth and systematic presentation of methods, techniques, and applications concerning multiway data. The book is usefully divided into three parts. The first one provides a general methodological discussion of the data analysis procedures, with particular reference to three-way component models. The second part deals with such relevant operational problems as: data preprocessing, missing values, model selection, and tools for interpretation of results. The third part is mainly devoted to real-life applications of the previously illustrated models. Here a special effort is made for casting light on the more subtle aspects of model building and interpretation, as well as on how to cope with possible pitfalls that may jeopardize the uninitiated multiway data analyst when managing the complexity of both the empirical information and the mathematical models in the multiway framework.

In Part 1, the first two chapters serve as a general introduction to the topics dealt with in the book, while Chapters 3 to 5 focus, respectively, on data, models, and algorithms. In Chapters 1 and 2, the attempt to make the reader acquainted with the problems, concepts, and methods of multiway analysis is appreciated. A nice example, in this connection, is the statistical study of Chopin's preludes, in the perspective of capturing the emotions and suggestions called forth in the listener. Chapter 3 describes the various types of multiway data, pointing out the different underlying designs (fully crossed and nested designs, multidimensional scaling designs, multivariate categorical data). In Chapter 4, the primary objectives of the volume are clearly stated. It appears that the main task is to present component models for *fully crossed* three-way data (i.e., when observations exist for all the combinations of all the modes' entities). This is a specific, albeit very

important, area in the domain of multiway analysis. However, the topic is analyzed in a very deep manner, avoiding matrix algebra and other technicalities as much as possible, as it is desirable for practitioners and applied researchers. Nonetheless, references are properly added for detailed algebraic aspects. In this chapter, PCA and three-way component models are illustrated. In this respect, it is worth mentioning that the above models are introduced in the following order: PCA, Tucker3 and, finally, Parafac. This contradicts the standard way to present them (in the literature, Parafac is usually considered in between PCA and Tucker3). In any case, this is not a weakness, since the author's approach is valuable and very interesting. Chapter 5 copes with some computational aspects connected with the models introduced in Chapter 4. Note that, in this case, the standard order (Parafac, and then Tucker3) is adopted.

In Part 2 (Chapters 6–12), the problems of data handling, model selection, and interpretation are dealt with in detail. Chapters 9–11 appear to be of particular interest for the applied researcher, since they are devoted to describing useful interpretational devices for the obtained parameter matrices. Specifically, Chapter 9 concerns the scaling of both component and core arrays. Chapter 10 deals with the transformation of the component matrices for simplifying the interpretation and Chapter 11 presents some plotting procedures for multiway data.

In Part 3 (Chapters 13–19), an accurate illustration is made of how the models and procedures previously described can be applied to real-life data sets. Different types of data structures are discussed: multiway profiles, multiway rating scales, multivariate longitudinal data, multiway contingency tables, three-way binary data. In the final chapter, the perspective of managing more complex data (from four-way and beyond) is examined through a synthetic but effective description of the main related issues.

All in all, Kroonenberg's book constitutes an extremely valuable tool for applied researchers in almost all

domains of investigation (from economics to psychology, from biomedicine to technology and physical sciences), whenever they are faced with the task of analyzing complex statistical data in view of obtaining useful information in their areas of interest. The book is also a precious reference for students and scholars coping with advanced multivariate analysis. If one should find any weak aspects in the book, these might be identified in two limitations. The first one, which is also stressed by the author, consists in the choice of multiway component models as the basic tool for analyzing multiway data. This obviously cuts out a relevant part of multiway methods (e.g., those based on a multiphase strategy of analysis of a set of two-way data matrices, such as ACT-STATIS proposed by the French school of data analysis). The second limitation is somehow intrinsic in the exploratory approach chosen in the book. In fact, the sample population problem is treated in a rather superficial manner and the underlying inferential issues are not fully discussed (see, for instance, the use of the notion of "exchangeability" without reference to an explicit probabilistic set up). In spite of the above-mentioned limitations, *Applied Multiway Data Analysis* will certainly represent a milestone in the literature concerning the statistical analysis of complex data structures.

REFERENCES

- Coppi, R. and Bolasco, S. (eds). (1989). *Multiway Data Analysis*. Amsterdam: Elsevier.
- Coppi, R. and Di Ciaccio, A. (eds). (1994). Special issue on "Multiway Data Analysis: Software and Applications." *Computational Statistics and Data Analysis* **18**(1).
- Law, H. G., Snyder Jr, C. W., Hattie, J. A., and McDonald, R. P. (eds). (1984). *Research Methods for Multimode Data Analysis*. New York: Praeger.

RENATO COPPI

Department of Statistics, Probability and Applied Statistics
Sapienza University of Rome
Rome, Italy

BRIEF REPORTS BY THE EDITOR

LEVY, P. S. and LEMESHOW, S. **Sampling of Populations: Methods and Applications, 4th edition**. Wiley, New York, 2008. xxxi + 576 pp. \$120.00/€97.90. ISBN 9780470040072.

Previous editions of this book have been reviewed in *Biometrics*, most recently the third edition in *Biometrics* **55**(3), 995. In the authors' words, this new edition is a "moderate updating and refreshing" rather than a "major change in organization or emphasis." Among the updates are: (1) *Stata* code updated to match more recent editions of the software; (2) updated examples to reflect modern times (e.g., prices were out of date); (3) including most data files on Wiley's ftp website; (4) a new chapter on construction of survey weights, guest-authored by P. P. Biemer and S. L. Christ, to fill a gap in coverage left in previous editions; and (5) a new chapter on telephone surveys, including issues raised by increasing reliance on cell phones, guest-authored by M. L. Link and M. Fahimi.

The new chapters seem to be a useful extension of the text and are written at a level similar to the rest of the book.

Plentiful numerical examples are given, with numbers plugged into formulas and occasional bits of SUDAAN or *Stata* code and output. Numerous application-oriented exercises close each chapter, with some answers supplied in the back. The book remains a very appropriately written text for classroom use, especially for students studying public health or epidemiology, or for undergraduate majors in statistics and related fields. It is not likely to serve the needs of a graduate class in statistics, nor a researcher in survey methods.

HOSMER, D. W., LEMESHOW, S., and MAY, S. **Applied Survival Analysis: Regression Modeling of Time-to-Event Data, 2nd edition**. Wiley, New York, 2008. xiii + 392 pp. \$110.00/€92.90. ISBN 9780471754992.

The first edition of this book was not reviewed in *Biometrics*, so a brief review is in order before highlighting changes from the first edition. I would describe this as an "application-leaning" book: not inherently mathematical, but definitely not a cookbook. Readers are expected to have had past course

work in applied regression and logistic regression. The general approach taken by the authors is to describe models and methods, explain how and why they are used, and then apply them to an example and discuss the results. It sounds like what every book should do, but I am impressed at how evenly the authors pursue all of these goals. For example, the math level is limited but not absent. Formulas are given for variances of many quantities; likelihoods, score functions, and information matrices are often given; and matrix notation is used where needed. This is balanced well with time spent discussing the applications, interpreting the analysis results, and offering practical advice. On the whole, this is a good, well-rounded, applied survival analysis text that could be used for a variety of teaching levels and would serve well as a practitioner's reference.

One of the changes to the second edition is that the authors have replaced the one main example that, along with occasional others, is used throughout the text. Other new highlights include variable selection via fractional polynomials, the use of time-varying covariates, and frailty models. The authors used *Stata*, primarily, for their analyses, and occasionally refer to its capabilities in the text, but no code or steps are given to guide the computing.

MARASINGHE, M. G. and KENNEDY, W. J. **SAS for Data Analysis: Intermediate Statistical Methods**. Springer, New York, 2008. xii + 557 pp. \$69.95/€53.45. ISBN 9780387773711.

The preface starts by stating, "This book is intended for use as the textbook in a second course in applied statistics..." The target audience is graduate students or advanced undergraduates in areas outside of statistics. The authors' goal is to teach both the methodology of regression and ANOVA and the use of SAS to complete such analyses. I have had a lot of experience teaching precisely these classes and never been completely satisfied with the available texts, so I was eager to evaluate this new book.

The book starts with an introduction to SAS data entry and manipulation. A second chapter covers the basic structure of an SAS program, some key programming statements, and a few essential PROCs, such as UNIVARIATE and FREQ. The third chapter deals with SAS/GRAPH, demonstrating some of the main procedures and options that are useful in making most of basic graphical displays that are likely to be used in reports and theses. These introductory chapters are welcome, as I have regularly had students ask me about books covering exactly these things. An instructor can choose either to cover much of this explicitly or to provide a few essential tools and let students refer to these chapters as needed. These three nonstatistical chapters make up about one-third of the book. The "meat" of the book is three chapters covering, respectively, regression, ANOVA, and mixed models. Most of the usual topics are covered in these chapters. Descriptions of the statistical methods are rather terse: they are presented rather than explained, accompanied by little intuition. For example, in the development of simple linear regression, the model, parameter estimates via least squares, ANOVA table, and tests and confidence intervals for parameters are all given in two

pages. This is followed by a 5½-page SAS example showing a program with several options, three figures of output and plots, and accompanying explanatory text. Programs and output are coarsely enumerated and the corresponding numbers are linked to the text. SAS options beyond those used explicitly are sometimes described.

My overall impression is mixed. As a methodological book on applied linear models, it falls far short of books like Kutner et al. (2004), but offers a much firmer link to the computation than most standard methods texts. On the other hand, as an SAS reference for linear models it falls far short of Littell, Stroup, and Freund (2002), but offers more detail on the methods. It is a reasonable compromise between the two, especially considering the very reasonable price. It might serve its audience well, especially if an instructor is willing to provide more detail and intuition on the practical application of regression and ANOVA.

REFERENCES

- Kutner, M. H., Nachtsheim, C. J., Neter, J., and Li, W. (2004). *Applied Linear Statistical Models*, 5th edition. Boston: McGraw-Hill/Irwin.
- Littell, R. C., Stroup, W. W., and Freund, R. J. (2002). *SAS[®] for Linear Models*, 4th edition. Cary, NC: SAS Institute.
- ARNOLD, B. C., BALAKRISHNAN, N., SARABIA, J. M. and MINGUEZ, R. (eds). **Advances in Mathematical and Statistical Modeling**. Birkhäuser, Boston, 2008. xv+368 pp. \$99.00/€85.49. ISBN 9780817646257.

Although not mentioned in the title, this edited volume is a tribute to the distinguished career of Professor Enrique Castillo. The title is quite clear, however, in suggesting breadth. In parallel with Castillo's publication record, the book covers a range of topics too broad to summarize meaningfully in a brief report. In particular, while the book is broken down broadly into general parts—a few of the topic headings include distribution theory, statistical methods, and order statistics—the individual papers within each part are quite varied in scope. I recommend that interested readers see the book's table of contents, listed on the publisher's website for the book. The approach taken by individual contributors varies from mathematical to applied. I would expect that few readers would be interested in everything in the book, but that very many would find something of interest here.

DOBSON, A. J. and BARNETT, A. G. **An Introduction to Generalized Linear Models, 3rd edition**. CRC Press, Boca Raton, Florida. xi + 307 pp. \$59.95/£31.99 (Paperback). ISBN 9781584889502.

Previous editions of this book have been reviewed in *Biometrics*, most recently the second edition in *Biometrics* **58(3)**, 699. This new edition includes three new chapters on Bayesian analysis and updated computer code for R and *Stata*. The book assumes familiarity with basic statistics (a solid introductory course will do), plus matrix algebra and calculus. The latter two prerequisites are a must, as there is considerable demonstration and development of the elementary theory

of generalized linear models (e.g., likelihood-based estimation and inference, including Taylor expansions). The new chapters are an appropriate-level introduction to Bayesian methods and their application, covering inference procedures, MCMC, and related diagnostics. Some `WinBUGS` code is given.

The chapters are short and concise, and the writing is clear. There are a few errors-in-fact: hypotheses for goodness-of-fit tests were curiously written in terms of the test statistics rather than some population quantities; the proof on p. 31 is botched, assuming incorrectly the independence of variables that are obviously not independent; and Exercise 10.3 asks to show that a property of the Weibull distribution is somehow not a property of its special case, the exponential. (Thanks to Rachel Altman for pointing out the last two.) Otherwise, explanations are fundamentally sound and aimed well at an upper-level undergrad or early graduate student in a statistics-related field. This is a very worthwhile book: a good class text and a practical reference for applied statisticians.

CHOW, S.-C. and LIU, J.-P. **Design and Analysis of Bioavailability and Bioequivalence Studies, 3rd edition.** CRC Press, Boca Raton, Florida, 2008. xxii+733 pp. \$99.95/£63.99. ISBN 9781584886686.

Previous editions of this book have been reviewed in *Biometrics*, most recently the second edition in *Biometrics* **56(3)**, 959–960. The reviewer described the book as a “thorough exposé of a subject about which the authors have considerable expert knowledge,” lauding its “encyclopaedic coverage” while being disappointed about its lack of attention to practical details. The new edition is an update in response to both the continual work that is being done in the titular areas and regulatory guidelines that have been published by health agencies since the last edition. The authors address the regulatory developments specifically in a new chapter near the end of the book, while methodological developments have resulted in updates made throughout the book. Additionally, two new chapters on pharmacokinetics have been added.