Book Reviews

This section will review those books whose content and level reflect the general editorial policy of *Technometrics*. Publishers should send books for review to Ejaz Ahmed, Department of Mathematics and Statistics, University of Windsor, Windsor, ON N9B 3P4 (*techeditor@uwindsor.ca*).

The opinions expressed in this section are those of the reviewers. These opinions do not represent positions of the reviewer's organization and may not reflect those of the editors or the sponsoring societies. Listed prices reflect information provided by the publisher and may not be current.

The book purchase programs of the American Society for Quality can provide some of these books at reduced prices for members. For information, contact the American Society for Quality at 1-800-248-1946.

Editor Reports on New Editions, Proceedings, Collections, and Other Books

---

Introduction to Stochastic Calculus Applied to Finance (2nd ed.), by Damien LAMBERTON and Bernard LAPEYRE, Boca Raton, FL: Chapman & Hall/CRC, 2008, ISBN 978-1-58488-626-6, 253 pp., $69.95.

The second edition of this book provides a concise and accessible introduction to the probabilistic techniques needed to understand the most widely used financial models. This edition incorporates many new techniques and concepts to be used to describe the behavior of financial markets. It features fully updated materials on stochastic volatility models, option pricing, and credit risk modeling. The authors provide many numerical experiments and real-world examples from their own experience. They also implement some algorithms using SciLab, free software available for download from *http://www.scilab.org*. They provide some computer experiments to illustrate some of the book's main ideas; the solutions obtained using SciLab for these computer experiments are available at *http://cermics.enpc.fr/~bl/scilab*. These experiments were well designed by the authors based on their teaching and research experience and were found to be effective in communicating these concepts and ideas and enhancing the understanding of readers.

The book comprises 9 chapters:

1. Discrete-Time Models
2. Optimal Stopping Problem and American Options

TECHNOMETRICS, MAY 2009, VOL. 51, NO. 2
DOI 10.1198/TECH.2009.0021

3. Brownian Motion and Stochastic Differential Equations
4. The Black–Scholes Model
5. Option Pricing and Partial Differential Equations
6. Interest Rate Models
7. Asset Models With Jumps
8. Credit Risk Models
9. Simulation and Algorithm for Financial Models.

Chapter 1 presents the main ideas related to option theory within the very simple mathematical framework of discrete-time models. Chapter 2 introduces the pricing and hedging of American options to establish the link between these questions and the optimal stopping problem. Chapter 3 introduces the mathematical tools needed to model financial assets and to price options in a continuous-time framework. Chapter 4 introduces the Black–Scholes model, which tackles the problem of pricing and hedging a European option (call or put) on a nondividend-paying stock. The case of the American option is investigated, and some extensions of the model are presented. Chapter 5 discusses the connection between the diffusions of the option pricing and the partial differential equations and presents some numerical methods based on this connection. Chapter 6 presents the main features of interest rate modeling, which is widely used to price and hedge bonds and interest rate options. The concept of forward measures is discussed, and some of the most widely used models are reviewed. Chapter 7 introduces the asset models with jumps. To describe these models, the main properties of the Poisson process are reviewed. Then the dynamics of the risky asset are studied, and the computation of European option prices is discussed. Chapter 8 introduces the basic concepts of credit risk modeling methods, including structural models and intensity models. Chapter 9 discusses the methods that can be used to simulate financial models and compute prices.

Overall, *Introduction to Stochastic Calculus Applied to Finance* provides a solid introduction to stochastic approaches used in the financial world. The authors cover many key finance topics, including martingales, arbitrage, option pricing, American and European options, the Black–Scholes model, optimal hedging, and computer simulation of financial models. The book can be used as a reference text by researchers and graduate students in financial mathematics. It also is ideal reading material for practicing financial analysts and consultants using mathematical models for finance.

Weiming KE
South Dakota State University

Linear Mixed Models: A Practical Guide Using Statistical Software, by Brady T. WEST, Kathleen B. WELCH, Andrzej T. GALECKI, and Brenda W. GILLESPIE, Boca Raton, FL: Chapman & Hall/CRC, 2007, ISBN 1-58488-480-0, 354 pp., $79.95.

In this book the authors take on the herculean task of demonstrating how to perform complex linear mixed models (LMM) analyses with five programs: HLM, R/S–PLUS, SAS, SPSS, and Stata. It is much more than a software manual; through the use of excellent introductory material and details given throughout, it provides a solid introduction to LMM analysis.

The book's website is available at *http://www.umich.edu/~bwest/almmussp.html*. In addition to the data sets, code, and errata list in the book, the website provides updates and extra material. Software-oriented books can quickly become dated, but the website appears to keep up with new developments; for example, it includes material on the use of the new `lmer` package in R, including examples of how to use the `mcmcsamp(·)` function to obtain the highest posterior density intervals for the model parameters.

The first two chapters provide motivation and a thorough description of linear mixed models. Besides specifying a LMM as a linear equation consisting of fixed and random terms, a useful way to conceptualize the structure of LMMs is to use the HLM-based concept of *levels*. This multilevel framework is introduced immediately and used throughout the book. A highlight in Chapter 1 is the detailed and nearly complete history of LMM analysis from the perspectives of theory and software. Chapter 2 summarizes the theoretical background of LMM analysis, with descriptions of the types of data structures amenable to LMM analysis, a matrix-oriented description of the model, estimation, inference, computation, model building processes, and diagnostics.

The remaining chapters use case studies to demonstrate how to analyze and interpret a wide variety of experiments amenable to LMM. Each is centered on a specific (nontrivial) data set. Chapters 3 and 4 describe the analysis of two- and three-level clustered data. The two-level example data set involves the weights of rat pups, where the rat pups are at level 1 and litters are at level 2. Covariates are involved with each level. Three-level clustered data are from an educational achievement data collection with students at level 1, classrooms at level 2, and schools at level 3. In this example, each level has covariates. (Without any covariates, the model would be a threefold nested random-effects model.) Chapter 5 illustrates repeated-measures analysis with an example in which measurements are obtained in several brain locations in a group of rats. This topic continues in Chapter 6, which considers two-level longitudinal models with random coefficients, and Chapter 7, which covers three-level longitudinal models.

Chapters 3–7 follow a common format. The application is thoroughly explained and placed into a multilevel context alongside similar examples from other settings. Next, the possible parameters involved (intercepts, regression coefficients, and variance components) are carefully detailed, along with assumptions or hypotheses about independence, heterogeneous variances, an so on. Often graphical methods are essential for formulating these details, and the authors do a fine job. They then specify a model-building plan, involving multiple modeling steps to make the final model as parsimonious as possible. One of two plans is used. The *top–down* procedure involves using a model with as many fixed-effects terms as possible, including interactions, developing an appropriate model for the random effects, and finally trimming away insignificant fixed effects. The *step-up* method starts with development of an appropriate model for the random effects using no fixed effects terms other than the intercept. Then fixed effects are entered (or dropped) from the model one level at a time. Either way, many models are fit, providing ample opportunity to explain concepts and recommendations. The p value is used as the vehicle for determining what terms to drop and which to keep. (Given the number of hypotheses tested, we must "look the other way" when it comes to multiple testing issues.) The plan is then executed with each of the software packages, with a focus on one of them. The final model is critiqued using graphical diagnostics, including the analysis of residuals and random-effect predictions (i.e., best linear unbiased predictors). Parameter estimates from the different software packages are compared, and discrepancies are explained.

Although the book covers a huge amount of important material, a few things are missing. Generalized mixed-effects models for dealing with responses, such as categorical and ordinal categorical responses, are not covered. The book sticks to continuous responses and normally distributed random effects. Moreover, confidence interval procedures are not described.

Inference using the results given by the software packages is covered thoroughly. But I believe that the authors should have included more information on what is not known (or at least not settled) about LMM inference, especially when sample sizes are not large and the data are unbalanced. Or instance, F tests for fixed effects with unbalanced designs are usually approximate, and the issue of "how approximate is approximate" may be important. Applied statisticians must use the tools that are available to them, but should be aware of the limitations of these tools. Slightly more emphasis on this would have been helpful.

I am familiar with R and SAS but cannot comment on the use of the other programs. But the use of SAS and R is accurate and well coded. Missing is some guidance on preparing the data for analysis, however. The data sets used in the book are all conveniently in the "correct" format. Restructuring data sets from the "wide" format (e.g., one variable per time point in a repeated-models data set) to the "long" format (e.g., a time variable and single response variable), a requirement for many applications, is tricky. Some advice on this would have been welcome.

Overall, this book is a tremendous contribution to the field of applied mixed modeling. It is much more than a software manual. It is well organized, has minimal typographical errors, and contains a complete index. It could easily serve as a reference guide. Anyone working with LMMs should seriously consider obtaining this book.

Charles HECKLER
University of Rochester

Bayesian Process Monitoring, Control and Optimization, edited by Bianca M. COLOSIMO and Enrique DEL CASTILLO, New York: Chapman & Hall/CRC, 2007, ISBN 10: 1-58488-544-0, 13: 978-1-58488-544-3, 336 pp., $97.95.

As stated in the Preface, the aim of this edited volume is to "provide a state-of-the-art survey of applications of Bayesian statistics in three specific fields of industrial engineering and applied or industrial statistics, namely, process monitoring, process control (or adjustment), and process optimization." The volume comprises four parts. Part I consists of two chapters, covering an introduction to Bayesian methods and Bayesian computational simulation techniques. This part is very well written, stimulating interest in applying Bayesian methods to industrial data analysis. Part II consists of five chapters describing Bayesian methods for process control, empirical Bayes process monitoring methods, Bayesian multivariate quality control procedures, Bayesian two-sided control chart design, and the Bayes rule of information and monitoring in manufacturing integrated circuits. Part III consists of two chapters presenting Bayesian methods for analyzing pulse trains corrupted by noise and missing pulses at unknown locations, as well as Bayesian algorithms for detecting a persistent process mean shift and for adjusting the process back to target. Part IV consists of three chapters reviewing Bayesian predictive methods for multiple response surface optimization, Bayesian methods for sequential empirical optimization, and Bayesian methods for analyzing the data from saturated factorial designs.

Although several outstanding books on Bayesian statistics are available, this volume is a special collection of informative and valuable articles in industrial statistics, particularly in the areas of process monitoring, control/adjustment, and optimization. The volume includes contributors from different parts of the world in both academia and industry sharing their research knowledge, experience, and wisdom in this particular area. In addition, this volume demonstrates the great effort being made to reach out to researchers in this area from both industry and academia. The publisher has kept the price of the volume reasonable for these researchers.

Subir GHOSH
University of California Riverside

Applied Nonparametric Statistical Methods (4th ed.), by Peter SPRENT and Nigel C. SMEETON, Boca Raton, FL: Chapman & Hall/CRC Press, 2007, ISBN 978-1-58488-701-0, 544 pp., $83.95.

Most fourth editions look surprisingly similar to the third editions. *Applied Nonparametric Statistical Methods* is an exception. Sprent and Smeeton have taken an accessible and well-regarded work and expanded, reorganized, and improved on it.

The reorganization is as follows. Where the authors previously dove right into nonparametric concepts in Chapter 1, they now offer a quick overview of such statistical topics as hypothesis testing, estimation, and ethical issues. Nonparametric fundamentals are covered in Chapter 2. The next five chapters cover single-sample location inference, other single-sample inference, paired samples, two independent samples, and three or more samples methodologies.

The new edition has new chapters on structured and survival data. Sections on correlation and concordance, regression, categorical data, association in categorical data, and robust estimation follow the new sections, whereas they immediately succeeded the section on three or more samples methodology in the third edition. The book concludes with a new section on modern nonparametrics.

The book's intended audience is undergraduate students with some exposure to statistics, and the authors really mean it! One need not have spent countless hours poring over dense mathematical texts to enjoy the fruits of the authors' labor. Sacrificing intense derivations in favor of lucid and concise prose, Sprent and Smeeton offer a strong connection with respect to the how and why of the techniques. Of course, this treatment likely will fail to resonate with a reader who prefers more rigorous methodological development. (Most undergraduate students do not suffer from this particular malady, however.)

The book's major strength is its prioritization of coverage. The authors take painstaking care to inculcate an understanding of the appropriate use of nonparametric methods, as well as an appreciation for their application over a wide range of fields. The examples are well chosen, and the variety should ensure that every reader finds at least some of the problems interesting. Good luck finding errors in grammar or syntax; the book is pretty tight (unless one believes that "normalise" should be spelled "normalize," but that is well beyond the scope of this review).

A search for weaknesses in this book is a bit of a gnat-straining excursion for the most part. One could argue that the "ethical issues" presented are too myopic, related primarily to British laws. The discussion of bootstrapping is inadequate, never covering resampling in any detail. Anyone desiring a more complete knowledge of bootstrapping will doubtlessly look elsewhere, however. The lack of tables will find its share of detractors, as will the comparative dearth of exercise problems in the chapters. But the most glaring shortcoming is the lack of computer applications. Although the authors discuss StatXact and other packages, they provide no direct link or disk with the text.

As a competitor to the texts by Conover (1999), Gibbons and Chakraborti (2004), Higgins (2004), and Wasserman (2006), *Applied Nonparametric Statistical Methods* more than holds its own. The combination of clear writing and comprehensive coverage make it an excellent introductory text. As with many other topics, faculty likely will prefer to use multiple references for their introductory classes. From the students' perspective, this text would be a well-appreciated choice.

Each text offers its own strengths. That of Gibbons and Chakraborti (2004) is highly readable, that of Higgins (2004) provides more comprehensive coverage of design and resampling-based techniques, that of Conover (1999) offers a more theoretical treatment, and that of Wasserman (2006) is balanced and notable for its breadth. It is likely that each user will have personal reasons for cherishing one of these texts above the other. (I personally learned nonparametric statistics from Jean Gibbons and Subhabrata Chakraborti and will always feel loyal to their work.)

Applied Nonparametric Statistical Methods's positives far outweigh the negatives. The text is a viable choice for an undergraduate course on nonparametric methodology. For a course that emphasizes application, it is definitely a good choice. It also could serve a reference text.

Many instructors could benefit from the authors' writing style. It is engaging without being proselytizing (or here, "proselytising"), something that most authors cannot seem to grasp. Nevertheless, I would encourage keeping at least three or four texts on nonparametric methods in one's library. No single text can suffice for theory, application, coverage, exercises, and writing style. Unfortunately, all books cannot be all things to all people (even statisticians).

J. Douglas BARRETT
University of North Alabama

REFERENCES

Conover, W. J. (1999), *Practical Nonparametric Statistics* (3rd ed.), New York: Wiley.

Gibbons, J. D., and Chakraborti, S. (2004), *Nonparametric Statistical Inference* (4th ed.), New York: Marcel Dekker.

Higgins, J. J. (2004), *Introduction to Modern Nonparametric Statistics*, Belmont, CA: Duxbury.

Wasserman, L. (2006), *All of Nonparametric Statistics*, New York: Springer.

Linear Models in Statistics (2nd ed.), by Alvin C. RENCHER and G. Bruce SCHAALJE, Hoboken, NJ: Wiley, 2008, ISBN 978-0-471-75498-5, xvi + 672 pp., $ 135.00.

The first edition of this book was published in 2000 and reviewed by Burdick (2001) in *Technometrics* and by Sengupta (2001) in *JASA*. This second edition still has the merits cited in these reviews. The relatively transparent organization, clear prose, and large number of problems with solutions or hints make it a good choice as an introduction to the basic theory of linear models in a class or for individual study.

Criticisms of the first edition (i.e., neglected topics and lacking rigor) have been addressed. With an increase in length of nearly 100 pages, the book includes two new chapters, one "on Bayesian inference in linear models (including Gibbs sampling)" and another on linear mixed models, and the authors have "upgraded the material in all other chapters" through the additions of sections and proofs.

Although the following comments might be picking nits, they point out some apparent editorial concerns that could detract from the otherwise exceptional clarity of exposition. These examples were gleaned from a detailed study of only selected portions of the book; whether these misprints are representative is unclear.

- In the Preface (pp. xiv–xv), the authors state that "the datasets and SAS command files for all numerical examples and problems in the text are available on the Intertet; see Appendix B." The copy of the book available for review contained no Appendix B. A search of the Internet including all the listings of the book found no site containing the datasets and command files.
- Although equation numbering within the text generally was used in the standard manner to reference equations, the equations between (2.61) and (2.68) did not have numbers. This disparity was first uncovered when

searching for "(62)" referenced on page 38. Similarly, "(63)" is referenced on page 47.

- On pages 46–73, the scalar, λ, is sometimes in bold type, giving the confusing implication that it is a matrix.
- In a simple slip contradicted in the same second line of page 95, the "x" in "$f(\mathbf{y}|x)$" should be bolded.
- Although the omitted number is obviously 15.3.2, the third sentence in Section 15.3.1 (p. 421) states that "the constrained model is discussed in Section."

This second edition can be highly recommended as an introduction to the general theory of linear models. An errata listing and website should be made available to readers. Statisticians might disagree about the inclusion or coverage of appropriate topics. This book introduces the core theory in a very understandable way and provides guideposts to the plethora of texts on the more popular current branches and augmentations.

Jim RUTHERFORD
Chevron Oronite Company, LLC

REFERENCES

Burdick, R. K. (2001), Review of *Linear Models in Statistics*, by Alvin C. Rencher, *Technometrics*, 43, 234–235.
Rencher, A. C. (2000), *Linear Models in Statistics*, New York: Wiley.
Sengupta, D. (2001), Review of *Linear Models in Statistics*, by Alvin C. Rencher, *Journal of the American Statistical Association*, 96, 1138.

Vertical Density Representation and Its Applications, by Marvin D. TROUTT, W. K. PANG, and S. H. HOU, Singapore: World Scientific Publishing, 2004, ISBN 981-238-693-9, 268 pp. $69.00.

This book presents a representation of a probability density function, f, on R^n in terms of the density function induced by a mapping from R^n to R. (The mapping may be the probability density function itself.) This representation is motivated by considerations that result from scoring random variables/vectors (e.g., the distribution of the resulting scores) and the Box–Muller method for generating random numbers with the normal distribution. By virtue of the applications given, the book appears to be aimed at a management science audience. It should be accessible to readers with a background in multivariate calculus and calculus-based statistics and familiarity with the Lebesgue measure.

The book comprises nine chapters. In Chapters 1 and 2, the authors introduce what they call the "vertical density representation" and give applications of a somewhat general nature. In Chapters 3 and 4, they present what they call the "multivariate vertical density representation" with applications that include the generation of random numbers with a given distribution. Chapters 5–8 are devoted to applications that are somewhat more specialized: the generation of uniformly distributed random numbers by chaotic generator functions in Chapter 5 and the solution of certain management science problems in Chapters 6–8. In Chapter 9, the authors pose a number of open questions and identify areas for future research. The book is not self-contained; it refers to certain key topics, such as the rejection-acceptance method of generating random numbers, mathematical chaos, and certain estimation methods from management science, without giving good definitions or introductions. (A certain familiarity with these terms is assumed.)

In Chapter 1, the authors present some of the book's key results. For example, given a probability density function, f, defined on R^n, an expression for the probability density function induced on R by f when f, in the second instance, is viewed as a mapping from R^n to R. (The authors call this induced probability density function the vertical density.) Conversely, an expression for recreating the probability density function on R^n when the density induced on R by a mapping from R^n to R is given, and the conditional distribution of the original values in R^n given the mapped value is uniform. (The authors call this result the vertical density representation.)

Applications of these results include the construction of probability density functions and a two-step process for generating random numbers from a probability density function: Randomly generate the ordinate of the density, and then randomly generate a point on the probability density function contour associated with that ordinate. Chapter 1 ends with a generalization of the two foregoing results; the first result is extended to mappings that are more general, and for the second result, the uniform distribution condition is relaxed. Chapter 2 gives applications to univariate probability density functions, including measuring a probability density's tail behavior by its normalized vertical density, creating a new class of very thick-tailed probability density functions, and decomposing correlation into vertical and contour components. In Chapter 3 the authors introduce what they call the multivariate vertical density representation and use it to devise a method for generating random vectors that follow a given multivariate distribution and also to study the tail behavior of certain classes of multivariate distributions. Continuing in Chapter 4, the authors give another application of the multivariate vertical density representation, proposing a method for generating nonuniform random numbers based on the rejection–acceptance method that they call the "vertical strip method."

In Chapter 5 the authors develop a class of uniform random number generators based on chaos functions using ideas and results from the vertical density representation. Chapter 6, the first of three chapters devoted to applications in management science, introduces what the authors call "Tolstoy's law of the mode," which states that within a population, members that score high on a certain criterion will tend to be closer to one another than those that score low on the same criterion. The authors derive conditions for this property to hold. Also included in this chapter are discussions of normal-like performance on finite intervals, unimodality, and the use of the general vertical density representation in solving an inverse linear programming problem. Chapter 7 examines the aggregate production planning problem. The authors propose a method that they call "minimum decisional regret" to estimate cost parameters and derive a target-mode agreement criterion for model validation by applying results derived previously. (The minimum decisional regret principle says that past managerial decisions or plans minimize the decisional regret, that is, the cost for past plans minus the cost for optimal plans.) In Chapter 8 the authors discuss an activity-based costing problem where for each of several comparable operational units, a number of factors drive one or more cost categories. They define the benchmark (i.e., lowest attainable) cost per unit for a given factor and propose the "maximum performance efficiency" principle as a basis for estimating the benchmark costs. (The maximum performance efficiency principle says to estimate the unknown vector of benchmark costs by that vector for which the average cost efficiency across the operational units is greatest.) They use the vertical density representation to specify a distributional characteristic (normal-like-or-better) that indicates consistency with the goal of 100% efficiency, a basic underlying assumption of the maximum performance efficiency principle's applicability. This mirrors the requirement of $\mathrm{N}(0, \sigma)$ residuals in ordinary least squares regression. Finally, in Chapter 9 the authors raise some open questions and identify areas for future research.

In summary, the book's main ideas—the vertical density representation of probability density functions—are contained in Chapters 1 and 3, with the rest of the book essentially devoted to applications. (The applications with the broadest appeal are those on random number generation.) In presenting the material, the authors sometimes do not introduce key concepts (e.g., the term "chaotic function" is presented without first being defined). In some instances, the mathematics should have been tightened up; for example, in discussing recursions, the authors start with a continuous function, f, from an interval $[a, b]$ to itself and a starting point x_0 yielding a sequence $(x_n, n \geq 0)$ in $[a, b]$ defined recursively by $x_{n+1} = f(x_n)$. They then go on to refer to $P(x_n \in A)$ in defining a probability density function ϕ, when it exists, by $P(x_n \in A) = \int_A \phi(x)\,dx$, where A is a measurable subset of $[a, b]$. But the x_n's are not random variables, and thus the expression $P(x_n \in A)$ is inappropriate. What the authors may have meant was the limit of the relative frequency of points from the sequence being in A. In addition, the book's notation is not consistent with standard notation (i.e., upper case for random variables; lower case for data points), and is not consistent throughout (e.g., random variables are sometimes in upper case and sometimes in lower case). These shortcomings, together with typos/errors, increase the effort needed to read and understand the material. Overall, this book would seem to be best appreciated by a reader who is already familiar with the various topics covered.

Errol CABY
AT&T Labs

Asymptotic Theory of Statistics and Probability, by A. DASGUPTA, New York: Springer, 2008, ISBN 978-0387759708, xxvii + 724 pp., $89.95.

This book covers large-sample theory at a doctoral level. Large-sample theory, or asymptotic theory, is useful for approximating probabilities and for finding cutoffs for large-sample tests and confidence intervals. The prerequisite for the text is a year of probability and measure at the level of the text by Billingsley (1995).

Although the author specifies gives recommended chapters for 10 different one-semester courses (p. xi), the text may be difficult to use in a course because of the lack of proofs. Sources for the proofs are given, but only two proofs are provided in the book's first 300 pages. Alternative texts, roughly in increasing order of difficulty, include those of Lehmann (1999), Ferguson (1996), Sen and Singer (1993), and Serfling (1980).

Asymptotic Theory of Statistics and Probability is an excellent reference for asymptotic theory, covering an unusually large number of topics with excellent motivation for why the theory is useful for the particular area of research. Chapter 1 covers basic convergence theorems, such as the central limit theorem. Chapter 2 reviews metrics and convergence, and Chapter 3 considers weak and strong laws and the delta method. Chapter 4 considers transformations, such as variance stabilizing transformations. Chapter 5 covers general central limit theorems. Chapter 6 covers moment convergence and uniform integrability. Chapter 7 considers sample percentiles and order statistics, and Chapter 8 covers sample extremes. Chapters 9 and 10 cover central limit theorems for dependent sequences and Markov chains. Chapter 11 considers the accuracy of the central limit theorem, with, for example, the Berry–Esseen inequality. Chapter 12 considers invariance principles. Chapters 13 and 14 consider Edgeworth expansions and saddlepoint approximations. Chapters 15–19 cover U-statistics, maximum likelihood estimators, M-estimators, trimmed means and estimators of multivariate location. Chapter 20 considers Bayes procedures and posterior distributions. Chapters 21 and 22 consider testing problems, including asymptotic efficiency. Chapter 23 covers some general large deviation results. Chapter 24 covers the asymptotic theory of nonparametric estimators. Chapter 25 considers the two-sample problems. Chapters 26–28 cover goodness of fit. Chapters 29–31 consider the bootstrap, jackknife and permutation tests. Chapter 32 covers density estimation, Chapter 33 covers mixture models and nonparametric Deconvolution, and Chapter 34 covers high-dimensional inference and false discovery rates. Finally, Chapter 35 gives a collection of inequalities from probability, linear algebra, and analysis.

David J. OLIVE
Southern Illinois University

REFERENCES

Billingsley, P. (1995), *Probability and Measure* (3rd ed.), New York: Wiley.

Ferguson, T. S. (1996), *A Course in Large Sample Theory*, New York: Chapman & Hall.

Lehmann, E. L. (1999), *Elements of Large-Sample Theory*, New York: Springer.

Sen, P. K., and Singer, J. M. (1993), *Large-Sample Methods in Statistics: An Introduction With Applications*, New York: Chapman & Hall.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Applied Multiway Data Analysis, by Pieter M. KROONENBERG, Hoboken, NJ: Wiley, 2008, ISBN 978-0-470-16497-6, xxi + 579 pp., $115.00.

This book is focused primarily toward graduate students in the areas of chemistry, social and behavioral sciences, and environmental sciences, although the techniques and methods used can be more broadly used in other areas, such as finance and engineering, as well. About one-third of the book is dedicated to some basic multiway algorithms and how they are extended from two-way data to multiway data, whereas the remainder covers some interesting areas of interest to practitioners. The book covers various topics as they are related to multiway data, including modeling issues (e.g., PCA, Tucker, Parafac), model selection, handling of missing data, residual analysis, longitudinal analysis, clustering, and categorical data.

Considering the complexity of the topics covered in the book, along with the great amount of research being done in this field to resolve various issues, the book is very nicely written. At times, it was difficult to discern the distinction between multiway and multimode, even though the author spends some energy defining their differences as he perceives them. Using just one of the two terms might have been a better approach.

The book starts by examining two-way data, then expands the discussion to three-way data and then, near the end of the book, to N-way data. It might have been better to present the concept of N-way data earlier in the book, along with a discussion of how the different issues relate to N-way data. The author presents the concepts of nested design and multisample data very nicely, once again demonstrating the book's practical nature.

The book concentrates on explaining the different approaches rather than getting too involved in the mathematical detail. This allows the author to point out the properties and characteristics of the different methods instead of concentrating on things like proofs and algorithmic issues. The author mitigates this by listing appropriate references for the reader to consult if interested in more details. Once again, this allows the book to be very practitioner-focused.

Personally, I found the book too wordy and lacking sufficient diagrams to illustrate the points. One of the main issues when dealing with multiway data is the multidimensional structure of the data. Here diagrams and illustrations can be very useful in helping the reader understand the author's point of view. I believe the book would have benefited from additional diagrams and illustrations.

It was nice to see an entire chapter devoted to missing data and ways to handle it, particularly because most of the approaches to handling missing data are based on PCA applied to two-mode data. In the case of multiway data, one must either extend the two-mode approach of PCA or devise new methods to take advantage of the data's multiway structure.

In a field that is actively evolving, with multiple possible methods of dealing with a problem, the author has covered the different perspectives fairly and has done a fine job pointing out the limitations of the different methods. Overall, *Applied Multiway Data Analysis* is a great book for practitioners in this field.

Ali ESMAILI
Air Products and Chemical Inc.

Statistical Principles and Techniques in Scientific and Social Investigations, by Woitek J. KRZANOWSKI, New York: Oxford University Press, 2007, ISBN 978-0-19-921309-2, 241 pp., $85.00.

According to the author, this book's purpose is to express the guiding principles of statistics as a discipline used by others to collect analyze and interpret the meaning of data. The author expects to be both a teacher and a missionary to readers who are avid users of statistical principles as well as those who are skeptics. In the twenty-first century, this latter group should be quite small. The author notes the many changes in the analysis of data over the past half or full century, and one would tend to believe that skeptics are rare. The author's objective is to fulfill the desire to explore the pertinent statistical topics necessary for all types of readers in scientific and social research. I doubt that a true scientific and social researcher could ever be a skeptic; however, skepticism is a common characteristic of scientific and social critics. We need not examine history very long to see much groundless skepticism of good and careful scientific and social research; for example, there are still those who disbelieve that government-run social programs "never" achieve any goals of increasing social welfare.

On page xiv of the Preface, the author states, "I tried to keep them (being algebraic and mathematical notation) at a rudimentary algebraic level, and have not strayed into any areas of advanced mathematics." This is true; there is no use of differential and integral calculus, and differential equations of highly complex formulas are used only rarely in the solutions to probability questions. On the other hand, the author does not consider such devices as the use of tables in the teaching of probability rules (e.g., addition and multiplication rules), nor does he use tables when explaining Bayes theorem calculations. The formulas for the normal, beta, and Weibull distributions, as examples, also would be not understood easily or perhaps not at all by those with only a rudimentary knowledge of college algebra. My 40-plus of teaching statistics at the level of rudimentary algebra has shown me that few such people understand the uses of or the need for statistics. In this way, we achieve the creation of skepticism, because most readers would not understand what the author is trying to convey.

Based on the premise that is now broken, we can evaluate the book's contribution knowing that a major goal is not satisfied. Do the illustrations on pages 2 and 3 convey to the reader the early notions of inference? More specifically, are they good examples? If a health practitioner informed me that there is a 1 in 100 chance that I contracted a debilitating disease, should I be happy or still very worried? I would want additional testing done to determine what would be the best diagnosis and procedures to follow. The cost of an error is very high in this situation. In essence, there are many and better examples than the one used by the author on pages 2 and 3.

Given the book's purpose and intended audience, the probability example on page 7 concerning the lottery engenders more confusion in the reader. The author does not solve this problem in its entirety, and does not address the more difficult probability question. According to the author, one must resort to "the topics of *permutations* and *combinations* in a specialist text" to solve this difficult problem. Why the author brings up this illustration at all, considering that the result is beyond the scope of the text, is unclear.

The definition of "at *random*" on page 8 is wrong. Random does not mean that each result of a statistical experiment is *equally likely.* This error is repeated on page 29, where random is defined a second time.

In the development of probability theory, the author defines the *multiplication* rule before he defines the *addition* rule. This is unusual; books from the noncalculus approach to that of a professional mathematician would define the simpler rule (addition) first. Why on page 9 does the author define conditional probability so early in the development of probability is beyond me as well as probably the great majority of my colleagues in education. The approach to Bayes theorem again is not well written. Although the use of the formula is correct, students can understand this theorem better when teachers use tables to illustrate the calculations, which eventually leads to a clearer understand of the theorem and its applications.

In Chapter 2, a simple introduction to boxplots, histograms, and the like would help the reader understand the concepts associated with analyzing data. The notion of symmetry could be explained, as could normality, skewness, and kurtosis. For example, the author should define the five number solutions for describing boxplots, which could allow the reader to see how data sets differ in both location and dispersion. Nothing like this is developed or well elucidated in this chapter.

In Chapter 3 the author does point out the need to identify the population model for identifying the parameters to be estimated. The geometric distribution is not necessary at this point. The extensive formulas used to describe distributions are not very illuminating at best. The conventional reader would benefit from charts, diagrams, and plots of these distributions with adequate references to useful applications.

Chapter 4 introduces *inferences* about μ and σ, but does not relate this to the discussion of population models in the preceding chapter. Furthermore, there is no explicit definition of degrees of freedom. All of this leaves one to believe that the author is not doing a complete job in these early chapters. I am surprised that an editor did not notice these omissions before publication. Chapter 5 continues the discussion of the Bayes theorem and its approach to inference. The discussion of the controversy on pp. 94–95 is misplaced; the author should have fully illustrated the Bayesian approach first. Then he should have pointed out the differences in the thought process with reference to the "frequentist" approach and the conclusions to be drawn. At that point, the author could discuss the "great debate" among statisticians as to their usefulness.

In the latter half of the book, the author describes the essential topics of regression (linear, nonlinear, and multiple), correlation, scaling, and multivariate methods. Of course, these are essential topics; however, it is doubtful that these discussions would be fruitful for one with no or limited experience with these topics. At best, these chapters are perhaps useful as reference material, but not as learning material. There is no discussion of computer software for the huge computational problems. The author does not warn against the use of simpler presentation software with some statistical algorithms (i.e., EXCEL), which may provide inaccurate and often wrong results. Too often, I have found miscalculations with this software, leading to inappropriate and often false conclusions or decision signals. Finally, better application applications should illustrate the author's essential points.

The reviewer thanks Ernest Kurnow, Professor Emeritus, New York University for helping with this review.

Jeffrey E. JARRETT
University of Rhode Island

Probability and Statistics With R, by María Dolores UGARTE, Ana F. MILITINO, and Alan T. ARNHOLT, Boca Raton, FL: Chapman & Hall/CRC, 2008, ISBN 978-1-58488-891-8, 728 pp., $89.95.

R (R Development Core Team 2008) is an independent, open-source, and free implementation of the S programming language that provides an environment for statistical computing and graphics. R allows users to increase its functionality by defining new functions. This has allowed numerous users to contribute packages and libraries to expand the capabilities of R in recent years. For this reason, R not only has received much attention from statisticians and scientists as a tool in research, but also has been used as a teaching media for statistics. This book introduces R as teaching software for probability and statistics and uses R to solve problems in statistical inference and data analysis.

This book covers a wide range of topics in both theoretical and applied statistics, including exploratory data analysis, univariate and multivariate distribution theory, sampling theory and sampling distributions, point and interval estimation, hypothesis testing, nonparametric methods, experimental design, and regression analysis. Each chapter contains numerous examples with detailed steps as well as R codes to illustrate the theories and methodologies. Problems are included in every chapter for practice.

Although the book's title demonstrates that its focus is on R, the authors list both R and S–PLUS (Insightful Corp 2008) commands and clearly note when a command is applicable only in either S–PLUS or R. Therefore, S–PLUS users also should find this book useful. Detailed executable codes and codes to generate the figures in each chapter are available online at *http://www1.appstate.edu/~arnholta/PASWR/front.htm*.

The book begins in Chapter 1, the book starts with a concise introduction of the basics in S that includes an overview of S language and syntax, structure, data manipulation, and user-written functions in R and S–PLUS. Chapter 2 discusses exploratory data analysis and presents some important graphical and quantitative techniques. This chapter provides a good coverage of the commands and syntax in R for graphics.

After studying the first two chapters, a reader should be able to manage the basic techniques in R for data manipulation and graphics that are needed in the subsequent chapters. Chapters 3–5 discuss general probability, random variables, and distribution theories, along with the use of R to compute and display probability mass/density and distribution functions, compute numerical integrations, and generate pseudo–random numbers.

Chapters 6–9 cover topics on statistical inference. Chapter 6 introduces sampling theory, parameters, estimators, and sampling distributions. Chapters 7 and 8 cover point and interval estimation. Chapter 9 discusses hypothesis testing. In these chapters R is used to illustrate the ideas in sampling distributions, obtain point and interval estimators, compute critical values for hypothesis testing, and perform the test procedures.

Chapter 10 continues the discussion on statistical inference by encompassing various nonparametric methods and nonparametric statistical tests, including two-sample problems, goodness-of-fit tests, categorical data analysis, nonparametric bootstrapping, and permutation techniques.

The book's last two chapters are devoted to statistical modeling. Chapter 11 discusses experimental design, and Chapter 12 covers simple and multiple regression analysis. These chapters contain three case studies based on real-world examples that serve as exercises in applying the procedures discussed.

Appendixes covering S commands and random vectors and matrixes are included. It is noteworthy that the appendix on S commands provides a quick reference list of the commands used throughout the book with clear descriptions, which will be valuable for many readers.

Today the capability to manage at least one of the statistical computing environments becomes an essential component in undergraduate and graduate studies in the field of statistical science. This book serves this purpose well by nicely blending mathematical statistics, statistical inference, statistical methods, and computational statistics using S language together. Students or self-learners can learn some basic techniques for using R in statistical analysis on their way to learning about various topics in probability and statistics. This book also could serve as a wonderful stand-alone textbook in probability and statistics if the computational statistics portions are skipped.

Hon Keung Tony NG
Southern Methodist University

REFERENCES

Insightful Corp (2008), *S–PLUS® Version 8.0 for Windows*, Seattle, WA: Insightful Corp.

R Development Core Team (2008), *R: A Language and Environment for Statistical Computing*, Vienna, Austria: The R Foundation for Statistical Computing.

Introduction to Probability With R, by K. BACLAWSKI, New York: Chapman & Hall, 2008, ISBN 1-4200-6521-1, 384 pp., $89.95.

Established introductory texts on probability can be roughly divided into two types. Most titles are mathematical introductions that treat probability as applied finite-measure theory. Although this level of understanding is essential to those using stochastic calculus and advanced continuous-time processes, these books require far more mathematical prerequisites than most students possess. Examples include the books by Billingsley (1995), Chung (2001), Stroock (1993), Bhattacharya and Waymire (2007), and Athreya and Lahiri (2006). Applied introductions are fewer in number and assume a mathematical prerequisite that any engineer, physicist, or applied math major would have. They emphasize the application of probabilistic reasoning rather than the mathematical details. Examples include books by Ross (2002) and Meester (2008). Both types of book assume that the reader has a good knowledge of multivariate calculus. Often, contemporary students have a shaky grasp of the prerequisites and have more skill in computing than in abstract mathematical reasoning. *Introduction to Probability in R* is aimed at this kind of student, combining elements of both types of text.

Backlawski presents the standard progression of basic probabilistic ideas, starting from samples spaces and events and then moving on to define probability, random variables, expected values, limit laws, conditional distributions, and transformations of random variables. This is followed by short chapters on the Poisson process, entropy, and stochastic processes. There are many exercises at the end of each chapter, followed by detailed solutions to some of the questions. These solutions serve as back-up examples, and a small random sample of them were found to be error-free. This is impressive, given the tendency of probability text solution pages to be corrupted with errors. The book also makes extensive use of R to illustrate various aspects of probability. Limiting behavior of random variables is demonstrated through animations, and the difference between Normal and Cauchy data is nicely illustrated. Some basics of R are provided, although the reader would benefit from some familiarity with R before trying to follow some of the more advanced programs. The code used to produce all graphs in the text and the code in all examples can be downloaded from the author's website.

If this were a solid applied probability text with additional R exercises, it would be an ideal book for students with shaky backgrounds in mathematics. Unfortunately, the author has made a number of highly unusual choices in terms of content and style that counteract the positive features of his book.

Introduction to Probability in R does not resemble a classic text that includes definitions, theorems, examples, and discussion arranged in a straightforward fashion. Instead, it reads like a set of augmented class notes from a very interesting and unusual class. Each chapter and each section must be read from beginning to end, because there is no way to isolate and focus on specific ideas. This makes the book very difficult to use as a reference. The chatty and unthreatening prose might be very helpful to a nervous beginner, but it can be maddeningly vague to anyone who knows the subject.

Although the themes of the chapters are conventional, their contents are not. The book was based on a course taught by Gian-Carlo Rota at MIT, and it betrays its highly mathematical ancestry. Almost all examples are motivated by issues that would intrigue a mathematician but not motivate a working probabilist or statistician. The section on continuous distributions briefly mentions the standard continuous models but gives almost no examples of their application in nonmathematical contexts. Instead, almost all examples are based on the distribution of gap lengths between the points in a random sample from a uniform distribution on a finite interval. Although many of these examples will be interesting to those who already understand probability, they replace more conventional examples that most students would find useful. Although the choice of examples makes the book difficult to teach from, some of the examples and presentations are very unusual and clearly written. There are very interesting sections on generating functions and on the use of entropy to characterize 'most random' distributions in various contexts that leave an experienced reader wanting to learn more, because these topics are seldom included in probability texts. Unfortunately, the book includes almost no references. The eight references included at the end are not cited in the text, and the index must be used to determine where they are cited. None of them refer to any of the major, interesting, and unusual ideas presented in the book, and there are no suggestions of advanced works for further reading on any topic.

Nonmathematical examples are included in some sections. In the case of statistics, these examples are largely confusing and make the book unsuitable for use as a self-study text for any but the brightest and most confident students. No attempt is made to connect probability as a mathematical model with its application as a predictive tool, although there is some criticism of the relative frequency definition. There is a discussion of confidence intervals, but the author appears to confuse these intervals with the complements of rejection regions of significance tests. In the discussion of significance tests, the possible alternative hypotheses are never mentioned, and there is no discussion of how they affect the definition of the p-value. In the discussion of the central limit theorem, Baclawski states that "in the absence of better information, we *must* (sic) assume that a measurement is Normally distributed." Although the author may be stating this in reference to errors in physical measurements, the vagueness of his prose suggests to the beginner that all continuous random variables used to model nature should be assumed to be Normal. This assumption, made carelessly, can turn probabilistic analysis from a powerful tool into a pointless and potentially misleading rite. There is no suggestion that the probabilist look at data to make a choice of distribution, nor are there any examples of measurements (such as lifetimes) that are not Normal. On the cover, the book is clearly identified as belonging to Chapman & Hall's *Texts in Statistical Science* series.

The book's idiosyncrasies make it unsuitable as a general reference work or a self-study book, especially if the student is planning to use probability in statistical applications. For a beginning student of average ability, it could serve as a source of extra examples and exercises. Moreover, beginners should find the informal and nonthreatening presentation of the basic ideas very useful, especially if they had been using a more terse text. A more advanced student could use the book as an extra source of intriguing mathematical examples, as could an instructor searching for interesting items to throw into a more conventional course. There is a very interesting book struggling to emerge from the material presented here, but not one that can yet challenge the clarity and comprehensiveness of the texts by Ross (2002) and Meester (2008).

Jeffrey PICKA
University of New Brunswick

REFERENCES

Athreya, K. B., and Lahiri, S. N. (2006), *Measure Theory and Probability Theory*, New York: Springer.
Bhattacharya, R. N., and Waymire, E. C. (2007), *A Basic Course in Probability Theory*, New York: Springer.
Billingsley, P. (1995), *Probability and Measure* (3rd ed.), New York: Wiley.
Chung, K.-L. (2001), *A Course in Probability Theory* (2nd ed.), New York: Academic.
Meester, R. (2008), *A Natural Introduction to Probability Theory*, Basel: Birkhäuser.
Ross, S. (2002), *A First Course in Probability* (6th ed.), New York: Prentice-Hall.
Stroock, D. W. (1993), *Probability Theory: An Analytic View*, New York: Cambridge.

Testing 1–2–3: Experimental Design With Applications in Marketing and Service Operations, by Johannes LEDOLTER and Arthur J. SWERSEY, Stanford, CA: Stanford University Press, 2007, ISBN 978-0804756129, xii + 300 pp., $65.00.

This book covers experimental design with emphasis on business and marketing applications. Competing texts that cover factorial, fractional factorial, and Plackett–Burman designs tend to focus on industrial applications. The text is similar in level to the books by Box, Hunter, and Hunter (2005), Cobb (1998), Kirk (1982), Kuehl (1994), and Montgomery (2005).

Chapter 1 outlines the book's organization, and Chapter 2 reviews basic statistical concepts, such as the matched-pairs test. Chapter 3 considers the one-way ANOVA design and completely randomized block design. Chapters 4, 5, and 6 cover factorial, fractional factorial, and Plackett–Burman designs. These three techniques are two level orthogonal designs. Chapter 7 covers orthogonal designs in which some of the factors have more than two levels, including two-way and three-way ANOVA models. Chapter 8 gives an introduction to nonorthogonal designs, including D-optimal designs. The appendix provides 11 case studies (two with parts A and B).

The clarity of the exposition and the quality of the examples are unusually high. I was teaching design (using Box, Hunter, and Hunter 2005) when I received this book to review, and I immediately incorporated the examples into my course.

Although this book should receive strong consideration for use as the primary text in a course on design, it contains no residual plots and insufficient information on generating computer output. Model checking is extremely important, and the residual plot and the response plot (of the fitted values versus the response) are crucial tools.

David J. OLIVE
Southern Illinois University

REFERENCES

Box, G. E. P., Hunter, J. S., and Hunter, W. G. (2005), *Statistics for Experimenters* (2nd ed.), New York: Wiley.

Cobb, G. W. (1998), *Introduction to Design and Analysis of Experiments*, Emeryville, CA: Key College Publishing.

Kirk, R. E. (1982), *Experimental Design: Procedures for the Behavioral Sciences* (2nd ed.), Belmont, CA: Brooks/Cole Publishing Company.

Kuehl, R. O. (1994), *Statistical Principles of Research Design and Analysis*, Belmont, CA: Duxbury Press.

Montgomery, D. C. (2005), *Design and Analysis of Experiments* (6th ed.), New York: Wiley.

Sampling Techniques for Forest Inventories, by Daniel MANDALLAZ, Boca Raton, FL: Chapman & Hall/CRC, 2008, ISBN 978-1-58488-976-2, xv + 256 pp., $83.95.

This compact little volume is packed with important and useful ideas. It comprises 12 chapters and 3 appendixes. It is not intended for beginners, but is aimed at graduate students and professionals. The reader is assumed to have a sufficient understanding of probability theory, statistics, and linear algebra, although there is an appendix covering conditional expectations and variances. In addition, the reader is expected to have some previous knowledge of the general economic-political background of forest inventories and aspects of mensuration, as well as skill in remote sensing and GIS. The author suggests that the key ideas can be found Chapters 1 and 2, which deal with sampling finite populations in general; Chapters 4 and 5, which are specific to forest inventory; and Chapter 8, which provides a case study. Exercises are included in each chapter, and solutions to many of them are given in Appendix C. The exercises are a mix of theoretical and numerical. Many sections contain a list of "Remarks." As with any book including a lot of theory, the reader will want to have pencil and paper handy to fill in or work out details. Proofs are given for nearly all results, although in some cases these are relegated to the exercises. There are some oblique references to software but little or no explicit discussion of the importance or usefulness of particular software. The key point that the reader will want to keep in mind is that this book pertains to sampling for surveys, particularly forest surveys.

Chapter 1, *Introduction and Terminology*, is very short (two pages) and serves only to state some terminology and notation as well as to delineate principal objectives when conducting an inventory, including estimating population totals, population means, estimator variances, ratios and correlations. It is important to remember that a forest survey requires actual visitation and/or the use of some form of remote sensing. Except for a very small forest, complete tabulation usually is not possible or at least not practical, and thus some form of sampling is necessary. Finally, note that a forest can have multiple characteristics of interest.

Chapter 2, *Sampling Finite Populations: The Essentials*, begins by considering a population of N individuals. Initially it is assumed that N is known and that there is a list of the individuals in the population. In many applications, particularly forest surveys, this list usually is not known. It is still possible to select a sample by using inclusion probabilities (which play a very significant role in the subsequent chapters and sections). Subsequent sections cover the Horvitz–Thompson estimator, sampling with and without replacement (both simple random and with unequal inclusion probabilities), and estimation of ratios. Multiple-stage sampling is common in many applications; this is similar to stratification but may have important differences. Suppose that a population is partitioned into *primary sampling units* (PSUs). A sample is taken from this collection; then for each PSU in the sample, a second-stage sample of secondary sampling units (SSUs) is selected. To simplify matters, two general assumptions are made. In *invariance of subsampling*, each time that the ith PSU is included in the first-stage sample, the second-stage sampling design is used. In *independence of subsampling*, the subsampling in any one PSU in the first-stage sample is carried out independently of the subsampling in any other PSU. In the forest inventory context, the volume of first-stage trees is estimated using only the DBH and the species, whereas in the second stage, the volume is estimated using the additional data on height and diameter at 7 m above the ground. Single-stage cluster sampling is a special case of two-stage sampling; at the second stage, a full census is taken of each of the selected clusters. If the first stage of two-stage sampling is a full census, then two-stage sampling is the same as stratified sampling.

Chapter 3, *Sampling Finite Populations: Advanced Topics*, extends many of the ideas covered in Chapter 2. In particular, two-stage sampling is extended to three-stage sampling (assuming invariance and independence of subsampling). When the third-stage sampling units are the elements of the population, this is called *element sampling*. The estimate of the variance of the Horvitz–Thompson estimator using three-stage sampling is seen to be an extension of the variance estimate for two-stage sampling. The estimates of the separate components of the variance also are given. These are useful in determining optimal sample sizes.

Section 3.2, "Abstract Nonsense and Elephants," begins with a discussion of the work of Cassel, Särndal, and Wretman (1977), Godambe (1955), and Basu (1991). Consider a population endowed with all possible response variables and a fixed sampling design. The Horvitz–Thompson estimator is in the class of unbiased estimators of the total, but there is no *uniformly best* member of this class (best in the sense of the variance), because it would have to have zero variance. Then the example given by Basu is presented to show that the variance of the Horvitz–Thompson might be arbitrarily large. One way to improve the precision of estimates is by using auxiliary information, for example, as in *probability proportional to size sampling* (PPS). Auxiliary variables usually are assumed to be known before sampling, that is, for each population element; however, it usually suffices that the auxiliary variable values are known for each population element in the sample as well as the sum over the population. To be useful, the auxiliary information must be related to the response variables of interest; in particular, this relationship must be modeled. But the objective is *model-assisted* estimation rather than *model-dependent* estimation. This is the subject of the last section in the chapter.

Chapter 4, *Forest Inventory: One-Phase Sampling Schemes.* Given a forest area F (a subset of the Euclidean plane), begin with a well-defined population of trees, each with a label and with associated position vectors. The response variables are observed at a fixed time. Usually the population of trees includes those with a minimum DBH, taken to be dimensionless points. The chapter begins by considering four scenarios. One of the crucial ideas is to transform the problem from a finite population to an infinite population. This is done by defining a random function such that the spatial average of the local density is the same as the true density of a response variable. Because forest inventories usually are combined inventories (i.e., information is gained directly from the forest via terrestrial plots and combined with auxiliary information gained from aerial photographs, satellite images, thematic maps or perhaps prior surveys), the survey might be in one phase or two phases. This chapter is concerned with two of the four scenarios, that is, single phase with one stage and single phase with two stages. If there are two phases, then *phase one* pertains to the gathering of the auxiliary information, and *phase two* collects terrestrial information from a subsample of the first-phase sample. The terrestrial sampling might be either single stage, in which trees are selected to collect the information, or two stage, wherein the first-stage trees are used to determine an approximation of the response variable. For each of the four scenarios, either simple random sampling or cluster random sampling might be used for the terrestrial data collection.

Chapter 5, *Forest Inventory: Two-Phase Sampling Schemes*, two-phase single-stage and two-phase two-stage scenarios are discussed by analogy with the previous chapter. Also as in the previous chapter, both simple random sampling and cluster random sampling are treated. In a two-phase scenario, there is the potential for using a regression model, possibly an external model or an internal model. A model is *internal* if fitted using terrestrial data and *external* if fitted from the first phase in a two-phase scenario. As an example of an external model, suppose that the stand structure is determined using the interpretation aerial photographs in the sample s_1, which, via preexisting yield tables, allows prediction of the volume per hectare or the number or stems per hectare. Because a two-phase sampling scheme is used, there will be a second sample (a subset of the first sample) denoted by s_2. If an external model is not available, then an internal model will be fitted using the auxiliary information. In the case of a single stage, the estimator is the sum of two averages, the average of the model values for points in s_1 and the average of the residuals (true vs. model values) for points in s_2. The mean of the residuals is not assumed to be zero. In the case of a two-phase, two-stage sampling scheme, the residuals are computed using the generalized local density in lieu of the true local density. In the case of two-phase, one-stage cluster sampling or two-phase, two-stage cluster sampling, the averages are weighted. Three different techniques are presented for estimating the coefficients in an internal model in the case of two-phase sampling.

Chapter 6, *Forest Inventory: Advanced Topics*, begins with a discussion of the difference between a design-based approach and a model-dependent approach. In the former, $Y(x)$ is random because x is the realization of a random variable. In the model-dependent approach, x is fixed and $Y(x)$ is a realization of a a spatial stochastic process. This can be considered a microscopic approach. Alternatively, a macroscopic approach might be used where the locally observed density, $Y(x)$, is modeled as a stochastic process. Microscopic models at the tree level and macroscopic models at the point level generally are not compatible in the sense that

$$E_{M_1} 1 \div \lambda(F) \sum_{i=1}^{N} Y_i = 1 \div \lambda(F) \int_F E_{M_2} Y(x)\,dx$$

may not be satisfied. In the model-dependent approach, a vector of auxiliary variables is assumed to be known at each point in space, particularly at each point in the terrestrial sample s_2. The spatial mean of this vector also is assumed known. This approach is based on a super-population model (Matern 1986) and the use of a linear model,

$$Y(x) = \mathbf{Z}(x)^t \beta + R(x).$$

The model-dependent predictor is given by

$$\widehat{Y}_{pred} = 1 \div \lambda(F) \int_F \mathbf{Z}(x)^t \widehat{\beta}_{s_2}\,dx.$$

At this point, the concept of the *anticipated mean squared error* (*ASME*) is introduced, that is, the average of the mean squared error under all possible samples generated by the design. Various asymptotic results for the ASME are given, including simple random sampling, stratified random sampling, and stratified random sampling. These all involve the covariance function of the residual term in the linear model. Next, model-assisted procedures are compared with pure model dependent procedures under different scenarios. There is a very short section on small-area estimation, followed by a somewhat longer section on modeling relationships.

Chapter 7, *Geostatistics*, is a very brief introduction to geostatistics leading up to the estimation of spatial averages both with and without measurement error in the data.

Chapter 8, *Case Study*, reviews an inventory of the Zürichberg forest (City and Canton of Zurich) comprising 217.9 ha with a "small area" of 17.1 ha. The latter served as a full census, and the tree coordinates were recorded. The stand map for three qualitative variables—developmental stage (four categories), degree of mixture (two categories), and crown closure (two categories)—was established using aerial photographs and "expert judgement." Thus there were 16 factors. Systematic cluster sampling was used for the inventory, and only two sampling schemes were used. Stem density and basal area were estimated using eight different methods, including three forms of kriging. Numerical results are given for both the entire domain and the small area.

Chapter 9, *Optimal Sampling Schemes for Forest Inventory*, points out that meaningful optimality criteria must rely on the superpopulation model, that is, the actual population is one realization of many similar ones. The design-based variance is fixed for a given realization. The average of this variance under the superpopulation model is called the anticipated variance. Optimal sampling schemes are those that minimize the anticipated variance for given costs or, conversely, minimize the costs for a given anticipated variance. A local Poisson model is used with Poisson strata to avoid the pairwise inclusion probabilities in the Horvitz–Thompson estimator. Assuming negligible boundary effects, optimal one-phase one-stage, one-phase two-stage, and optimal two-phase sampling schemes are obtained.

Chapter 10, *The Swiss National Forest Inventory*, is devoted to using the data from the first and second Swiss National Forest Inventories, *SNFI*1 (1983–1985) and *SNFI*2 (1993–1995), to illustrate the optimization technique presented in the previous chapter. SNFI1 used a one-phase, two-stage simple sampling scheme comprising plots with two concentric circles.

Chapter 11, *Estimating Change and Growth*, is a moderately short chapter pertaining to changes in a forest between two time points, specifically the change in the spatial mean. To simplify some of the analysis, the real forest is replaced by a statistical model; that is, there are some physical characteristics of the forest that are not present in the model. The area of the forest is assumed to be the same at the two time points. A threshold height, H_0, is chosen, and a tree is said to *exist* at time t_1 (resp. at time t_2) if its height at that time. The "forest" now consists of all of the trees that exist at time t_1 or time t_2. Each tree is given a unique label. A number of variables are defined:

$E_i(t_k) = 1$ if the ith tree exists at time t_k otherwise it is zero,

$A_i(t_k) = 1$ if the ith tree is alive at time t_k,

$Y_i(t_k) = 1$ is the variable of interest (e.g., the volume) provided $E_i(t_k) = 1$,

$T_i(t_k) = 1$ if the DBH of the ith tree is at least the inventory threshold, provided $E_i(t_k) = 1$.

The product of these four variables, summed over all of the trees, is the total inventory of the forest. The difference between those sums is the net change over the time interval. One of the principal problems discussed in the chapter is how to estimate the inventories as well as the variance of the estimators.

Chapter 12 points out that transect sampling, or line-intersect sampling, is relatively easy to use and is widely used in many disciplines. Two general cases are considered, one in which the line intersects the entire forest but the length of the intersection is a random variable and another in which the length of the line is fixed. The latter references the work of Kaiser (1983). Two different methods are given for choosing/positioning the random line. This sampling method is closely related to the famous "Buffon needle problem."

Appendix A illustrates various sampling schemes using the small area data from the case study. Histograms are given showing the distributions for various statistics for each sampling scheme. Appendix B reviews basic results pertaining to conditional expectation and conditional variance. Appendix C provides the solutions for selected exercises (mostly numerical) in the various chapters. SAS/IML is used for one of the solutions.

Donald E. MYERS
University of Arizona

REFERENCES

Basu, D. (1991), "An Essay on the Logical Foundations of Survey Sampling," in *Foundations of Statistical Inference*, eds. V. P. Godambe and D. A. Sprott, Toronto: Holt, Rinehart and Winston, pp. 203–242.

Casel, C., Särndal, C., and Wretman, J. (1977), *Foundations of Inference in Survey Sampling*, New York: Wiley.

Godambe, V. (1955), "A Unified Theory of Sampling From a Finite Population," *Journal of the Royal Statistical Society*, Ser. B, 17, 269–278.

Kaiser, L. (1983), "Unbiased Estimation in Line Intercept Sampling," *Biometrics*, 39, 965–976.

Matérn, B. (1986), *Spatial Variation*. Lecture Notes in Statistics, Vol. 36, Berlin: Springer-Verlag.

Elementary Statistics Using JMP®, by Sandra D. SCHLOTZHAUER, Cary, NC: SAS Institute Inc., 2007, ISBN 978-1-59994-375-6, 480 pp., $79.95.

The topics covered in this book are very similar to those that would be covered in an introductory statistics class. When compared to the introductory guide to JMP manual available with the software, I find that this book covers many of the same topics. But this book includes more details on the topics, along with more step-by-step instructions for basic analyses. Included with these step-by-step instructions are numerous screen captures, allowing the user to follow the instructions more easily.

This book makes a very minimal set of assumptions regarding the reader's previous experience with JMP statistical software, which makes it very appealing for the new user. The book has very little statistical theory and formulas coverage; thus, as the author clearly warns, the book "is not a replacement for statistical texts" (p. 9). The author does include a reference list in an appendix at the end of the book for statistical texts that can be consulted for more information.

The author uses a wide variety of data sets to illustrate the analysis steps and interpretation of the output. All of the data sets are available from the JMP help menu or from the CD included with the book, which makes them very accessible. At the end of each main chapter are a number of exercises, which are helpful for the reader desiring more practice in using the software. In addition to these exercises are a summary section gathering the key points from the chapter and a "JMP steps" section providing a convenient reference to all of the tasks covered in the chapter. These "JMP steps" sections also are included on the CD–ROM that comes with the book.

After two short chapters on getting started in JMP, Chapter 3 covers data tables, including data input and manipulation. Chapter 4 covers the *Distribution* platform, including how to obtain histograms and basic summary statistics, such as the mean and standard deviation, for a set of data. Chapter 5 discusses Chart and Treemap graphs and describes how to copy output to another software package, such as Microsoft Word. Chapters 6 and 7 cover some of the statistical theory, including the normal distribution, the central limit theorem, confidence intervals, and hypothesis testing. These two chapters would be a nice refresher for someone with previous exposure to basic statistical methods.

Chapter 8 covers comparisons of two groups, including paired and nonparametric tests in addition to the traditional two-sample t-test. Chapter 9 covers comparisons for multiple groups and multiple comparison procedures. These two chapters use the *Fit Y by X* platform for all the analyses. Chapters 10 and 11 cover simple linear and multiple regression in both the *Fit Y by X* and *Fit Model* platforms, including checking assumptions and regression diagnostics. Finally, Chapter 12 covers methods for contingency tables, including chi-squared and Fisher's exact tests.

This book is not meant to be a comprehensive look at all of the features of JMP. Rather, it should be considered a carefully selected subset of features pertinent to an introductory statistics course. When discussing the output from a particular analysis, the focus is on the key portions of the output, such as p values, although the author generally includes a brief description of each portion of the output. I found the numerous handy tips and tricks interspersed throughout the book valuable.

In summary, *Elementary Statistics Using JMP®* is a good introduction to JMP statistical software that would be a good supplementary text for an introductory statistics class. It also is very accessible for a new user of the software who has some previous exposure to basic statistical concepts.

Willis A. JENSEN
W. L. Gore & Associates, Inc.

Observed Confidence Levels, by A. M. POLANSKY, Boca Raton, FL: Chapman & Hall/CRC, 2008, ISBN 978-1584888024, xvi + 271 pp., $79.95.

This book covers theory and applications of observed confidence levels and the closely related problem of regions, a topic introduced by Efron and Tibshirani (1998). An illustration of the simplest method is informative. Suppose $\boldsymbol{\theta}$ is a vector of multivariate location or regression parameters. Draw b bootstrap samples and compute the bootstrap estimate $\hat{\boldsymbol{\theta}}_i^*$ of $\boldsymbol{\theta}$. Then the estimated percentile observed confidence level for a region of interest Ψ is the proportion of the $\hat{\boldsymbol{\theta}}_i^*$ that lie in Ψ: $\hat{\alpha}_{perc}(\Psi) \approx \frac{1}{b}\sum_{i=1}^{b} I(\hat{\boldsymbol{\theta}}_i^* \in \Psi)$ where the indicator $I(\hat{\boldsymbol{\theta}}_i^* \in \Psi)$ is one if $\hat{\boldsymbol{\theta}}_i^*$ is in Ψ and zero, otherwise.

Observed confidence levels have many applications and can be used to complement hypothesis testing and confidence regions. There are several competing methods for observed confidence levels, including the percentile method, and not all of the methods need the bootstrap. Most of the chapters derive several methods of observed confidence levels, asymptotic accuracy based on Edgeworth expansions, simulations, examples, discussion on how to estimate the observed confidence levels using the statistical software R and exercises.

The text is at a Ph.D. level because of the asymptotic theory, but many of the ideas are simple and may be of great use. The text is useful for researchers who want to learn about observed confidence levels, and the topic of observed confidence levels would be a useful addition to a course on resampling methods such as the bootstrap.

The wide variety of regions Ψ that can be examined lead to many interesting applications. Chapter 1 introduces observed confidence levels, the method of regions and example applications. Chapter 2 covers single parameter problems. For example, if a confidence interval for θ needs to lie in the range $[-2, 2]$, then regions $\Psi_0 = (-\infty, -2)$, $\Psi_1 = [-2, 2]$ and $\Psi_2 = (2, \infty)$ may be of interest. Chapter 3 considers multiple parameter problems where a region such as $\Psi = \{\boldsymbol{\theta} : \theta_1 \le \theta_2 \le \cdots \le \theta_7\}$ may be of interest. Chapter 4 considers applications for linear models and regression. Chapter 5 gives applications for nonparametric regression and for nonparametric density estimation where a region might be "all densities with two or more modes." Chapter 6 gives further applications including principal component analysis where a region might be "four principal components are needed to account for over 90% of the variability in the data." Chapter 7 considers connections with hypothesis testing including multiple comparisons while Appendix A reviews asymptotic theory.

The website (*www.math.niu.edu/~polansky/oclbook/*) contains R functions and data sets.

David J. OLIVE
Southern Illinois University

REFERENCE

Efron, B., and Tibshirani, R. J. (1998), "The Problem of Regions," *The Annals of Statistics*, 26, 1687–1718.

Editor Reports on New Editions, Proceedings, Collections, and Other Books

Statistical Models and Methods for Biomedical and Technical Systems, edited by Filia VONTA, Mikhail NIKULIN, Nikolaos LIMNIOS, and Catherine HUBER-CAROL, Boston: Birkhäuser, 2008, ISBN 987-0-81764-464-2, 556 pp., $139.00.

This edited volume, the outcome of the International Conference on Statistical Models for Biological and Technical Systems, showcases the current research interests in survival analysis and reliability. It comprises 37 papers grouped into 9 parts:

- Cox Models, Analysis and Extensions
- Reliability Theory-Degradation Models
- Inferential Analysis
- Analysis of Censored Data
- Quality of Life
- Inference for Processes
- Designs
- Measures of Divergence, Model Selection, and Survival Models
- New Statistical Challenges.

The book's style and format are fine. Its organization makes it useful for its target audience. As the editors note, it will be useful to broad interdisciplinary readerships of researchers and practitioners in applied probability, statistics, and related areas. The volume includes 37 tables and 55 figures, a nice feature. However, like many other edited volumes, it suffers from an overly broad range and wide selection of topics, some related and some seemingly unrelated.

Recent Advances in Linear Models and Related Areas, edited by SHALABH and Christian HEUMANN, Heidelberg, Germany: Springer Physica-Verlag, 2008, ISBN 987-3-79082-063-8, 446 pp., $189.00.

This edited volume attempts to present some recent developments in linear models and related areas. I find the term "related areas" difficult to comprehend; generally speaking, the term is used generously, which is also the case in this volume. This collection comprises 24 papers to honor the contributions of Helge Toutenburg, an accomplished and established statistician, on his 65th birthday. Topics covered include boosting, Cox regression models, time series models, cluster analysis, information theory, and matrix theory, to mention a few.

The volume covers an array of topics related to linear models, some applied and some of a theoretical nature. To my surprise, I only found one paper on the Lasso (least absolute selection and shrinkage operator); research in this area is of great interest and is ongoing. Two references to partial regression models are given at the end. This volume will be useful and of interest to the intended audience—experienced researchers as well as practitioners in the field of linear models.

REFERENCES

Ahmed, S. E., Doksum, K. A., Hossain, S., and You, J. (2007), "Shrinkage, Pretest and Absolute Penalty Estimators in Partially Linear Models," *Australian & New Zealand Journal of Statistics*, 49, 435–454.

Hossain, S., Doksum, K. A., and Ahmed, S. E. (2009), "Positive Shrinkage, Improved Pretest and Absolute Penalty Estimators in Partially Linear Models," *Linear Algebra and Its Applications*, 430, 2749–2761.

Random Effect and Latent Variable Model Selection, edited by David DUNSON, New York: Springer, 2008, ISBN 987-0-38776-720-8, 174 pp., $69.95.

Random-effects and latent variable models play an essential role in the analysis of multivariate data. This volume consists of 7 papers in the form of separate chapters, divided into two parts for the reader's convenience. The first part covers the frequent likelihood ratio test and score tests for zero variance components, whereas the second part concentrates on Bayesian methods for random effect selection in linear mixed effects, generalized linear, and structural equation models. The titles of the papers are as follows:

Part I: Random Effects Models

1. Likelihood Ratio Testing for Zero Variance Components in Linear Mixed Models
2. Variance Component Testing in Generalized Linear Mixed Motion for Longitudinal/Clustered Data and Other Related Topics
3. Bayesian Model Uncertainty in Mixed Effects Models
4. Bayesian Variable Selection in Generalized Linear Mixed Models.

Part II: Factor Analysis and Structural Equations Models

1. A Unified Approach to Two-Level Structural Equation Models and Linear Mixed Effects Models
2. Bayesian Model Comparison of Structural Equation Models
3. Bayesian Model Selection in Factor Analytic Model.

The topics are appropriate and timely. The chapters are well written and well organized. In summary, this book presents methods incorporating model uncertainty in random-effects and latent variable models. This concise, self-contained (in some sense), application-based volume will be beneficial to graduate students and researchers interested in both theory and application. It also is suitable as a text for a graduate seminar or a supplementary reference for a course on this topic. Finally, I must commend the editor for providing a nice summary and discussion on the topics covered in the book in the Preface; I enjoyed reading it!

Introductory Statistics With R (2nd ed.), by Peter DALGAARD, New York: Springer-Verlag, 2008, ISBN 978-0-38779-053-4, 364 pp., $59.95.

The review of the first edition of this book (Sandry 2003) provides a detailed exploration of the book's contents. The revisions, reorganization, and additions in the second edition certainly enhance the book's value. In this edition, author has updated the text and code to R version 2.6.2. In addition, he has revised the exercises and sketches for the answers in Appendix D. This is a welcome addition to the new edition that will be appreciated by its users. Furthermore, in this edition the introductory chapter is significantly revised, reorganized, and expanded. This chapter is now divided into two chapters ("Basics" and "R Environment"). New chapters on more advanced data handling tasks, rates, and Poisson regression and nonlinear curve fitting have been added. The new edition is well written, and the new materials are well incorporated. Like the first edition, this edition will continue to be useful to the target audience, and I can safely recommend it to them. In passing, I would like to remark here that the author has been a member of the R Core Team since 1997.

REFERENCE

Sandry, T. D. (2003), Review of *Introductory Statistics With R*, by P. Dalgaard, *Technometrics*, 45, 274.

Forthcoming Reviews

Books listed here have been assigned for review in the past quarter. Publication of their reviews or reports generally would occur within the next four issues of the journal. Persons interested in reviewing specific books must notify the editor soon after the publication date for the book. Persons interested in being reviewers or reviewing specific books should contact the editor, Ejaz Ahmed, by electronic mail (*techeditor@uwindsor.ca*).

A First Course in Order Statistics, by B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, SIAM

A First Course in Statistical Programming With R, by W. John Braun and Duncan J. Murdoch, Cambridge University Press

Level Crossing Methods in Stochastic Processes, by Percy H. Brill, Springer

Observed Confidence Levels—Theory and Applications, by Alan M. Polansky, Chapman & Hall/CRC Press

Optimum Experimental Designs, With SAS, by A. C. Atkinson, A. N. Donev, and R. D. Tobias, Oxford University Press

Simulation and Inference for Stochastic Differential Equations With R Examples, by Stefano Iacus, Springer