Biometrics 64, 1299–1310
December 2008

BOOK REVIEWS

EDITOR:
THOMAS M. LOUGHIN

VAN GROESEN, E. and MOLENAAR, J. **Continuum Modeling in the Physical Sciences**. SIAM, Philadelphia, 2007. x + 228 pp. US$65.00. ISBN 978-0-898716-25-2.

Successful mathematical modeling requires the ability to formulate a mathematical description of the phenomenon of interest as well as to apply appropriate tools to glean results from the model. An effective modeling text must give attention to developing both of these abilities. The intriguing, well-written book by van Groesen and Molenaar is successful in both respects.

Authors of modeling texts must decide whether to present mathematical concepts through a series of case studies, or whether to present mathematical principles in and of themselves and then apply them in examples. Van Groesen and Molenaar make the latter choice. Their philosophy (described in the preface) is that the case study approach "could disappoint the student, since having digested many particular models does not guarantee that one knows how to proceed when confronted with a new situation." Thus, they have written a text that highlights mathematical concepts relevant to the formulation and analysis of models. In contrast to modeling texts such as Fowler (1997) and Howison (2005), case studies do not play a central role. The exposition of general principles is peppered with examples; only at the conclusion of the book do the authors present full modeling case studies.

The book is based on material originally developed for advanced undergraduates at the University of Twente, and is published by SIAM. Excluding a few minor editing issues (for instance, I noticed that the notation in some figures and figure captions was not consistent with that used in the body of the text) the book lives up to the high production standard typical of SIAM texts. It is divided into six chapters, some of which end with a set of well-thought-out "challenging problems" that interweave the exposition of a particular problem with exercises for the reader to complete.

The first two chapters present concepts useful for the formulation of models. Chapter 1 covers dimensional analysis and scaling and applies those tools to problems from classical mechanics, including pendulums and springs. Also showcased is the well known but still fascinating work of G. I. Taylor, who used dimensional analysis to estimate the strength of the first atomic bomb explosion by using film data showing the expansion of the resulting mushroom cloud. Chapter 2 focuses on two key model components, namely conservation laws and constitutive relations, and provides examples from heat flow, fluid flow, and traffic flow. One of my favorite features of this chapter is a pedagogically useful discussion of continuous versus discrete models, including a derivation of the former from the latter in the case of one spatial dimension.

The next two chapters present the core mathematical ideas that the authors believe to be key tools for dynamical modelers. The authors' unified mathematical approach is to treat ordinary and partial differential equations both as evolution equations describing movement through an appropriately defined state space, and to develop all mathematical tools in that context. This is a sensible approach, and one that I have not seen taken before in a modeling text. Chapter 3 defines the basic notions of state space, presents some fundamental tools including linearization and expansion in basis vectors, and introduces some properties of waves. Chapter 4 discusses stability and robustness. The emphasis on stability is standard. However, students first foraying into modeling often miss the importance of robustness, essentially "stability" with regards to perturbations of the form of the model. They will benefit from the clear treatment given to this topic.

Chapter 5 departs from the dynamical modeling emphasized in earlier chapters to take a look at variational modeling, including a presentation of the classic hanging chain (catenary) example as well as forays into Lagrangian and Hamiltonian mechanics. The authors also use the variational perspective to motivate the development of numerical methods (such as finite element) by an appropriate discretization of the state variables. For information on variational methods, modelers typically have to look beyond modeling texts, and so the inclusion of the variational chapter here is a nice addition to the literature.

Chapter 6 brings the previous chapters together in four case studies. The applications examined are polymer dynamics, fiber spinning, surface water waves, and optics. Each case study begins with a convenient list of the specific concepts from earlier in the book (and their section number) called upon in the subsequent exposition. This makes for a tight integration with the rest of the text. Each case study also includes a useful set of exercises.

As mentioned above, the authors intend the book for an undergraduate audience. I think the material will be best appreciated by students with some background not only in multivariable calculus, ordinary and partial differential equations, and linear algebra, but also elementary physics, fluid dynamics and/or continuum mechanics, and some basic notions from functional analysis. I believe the text may be accessible to the most motivated undergraduates, and will definitely be a useful resource for beginning graduate students. The examples and applications in the book are not from the biological realm (as indicated by the title, the authors' interest is in physics-based models) but because of the generality of the modeling approach, a student or scientist with an interest in biological modeling should find the book to be useful. Supplemented with biological modeling texts such as Edelstein-Keshet (2005), Murray (2002), or de Vries et al. (2006), this book will be a valuable resource for students. It is an equally nice addition to the bookshelves of teachers and researchers—indeed, I plan to use it in both endeavors.

REFERENCES

Edelstein-Keshet, L. (2005). *Mathematical Models in Biology.* Philadelphia: SIAM.

Fowler, A. C. (1997). *Mathematical Models in the Applied Sciences.* Cambridge, U.K.: Cambridge University Press.

Howison, S. (2005). *Practical Applied Mathematics: Modeling, Analysis and Approximation.* Cambridge, U.K.: Cambridge University Press.

Murray, J. D. (2002). *Mathematical Biology.* Berlin: Springer-Verlag.

de Vries, G., Hillen, T., Lewis, M., Schonfisch, B., and Muller, M. (2006). *A Course in Mathematical Biology: Quantitative Modeling with Mathematical and Computational Methods.* Philadelphia: SIAM.

CHAD TOPAZ
Department of Mathematics and
Computer Science
Macalester College
St. Paul, Minnesota 55105, U.S.A.

BERTOIN, J. **Random Fragmentation and Coagulation Processes**. Cambridge University Press, New York, 2006. vii + 280 pp. $65.00/£40.00. 9780521867283.

The present monograph deals with the mathematical theory of coagulation and fragmentation processes. These phenomena occur in various contexts, such as biology, chemistry, and astronomy, and have been studied extensively from the beginning of the last century.

Coagulation processes study the evolution in time of a population that is split into groups, where groups can merge and form larger groups. Similarly, fragmentation processes study the time evolution of a population that is split into groups, where these can break into smaller groups. The monograph discusses "binary" coagulation and fragmentation processes, where in each event, only two groups interact. It also describes Kingman's theory, which discusses more complex processes where several groups can interact in an event.

Chapter 1 of the book gives a solid introduction to the mathematical background of the topic, and makes the monograph more or less self-sufficient. Section 1.6 (Comments) provides a short review of part of the relevant literature. Unfortunately, the literature review is not sufficiently comprehensive, and the motivation behind studying coagulation and fragmentation processes is not sufficiently detailed. In my opinion, the motivation and literature study deserve a separate chapter.

Chapter 4 gives a very good discussion of processes with multiple interactions per event. Chapter 5 describes the procedure for extending the coagulation processes to an infinite population, and the derivation and analysis of Smoluchowski's equations. However, the treatment of coagulation, fragmentation, and combined coagulation and fragmentation processes via integro-differential equations is completely missing. This path has been a relevant part of the research, from its beginning, and the coverage of the topic is not really a complete overview without an appropriate chapter dedicated to the integro-differential equations models.

In general, the book would be very useful as an up-to-date source for an advanced graduate course on this interesting topic, despite the few shortcomings in the coverage and the literature survey mentioned above. I appreciated the self-sufficient mathematical background, which is rigorously presented. This makes the book appropriate as a graduate textbook. The style of the book, however, is quite formal, and it therefore appeals to the more advanced-level courses.

Shay Gueron
Department of Mathematics
University of Haifa
Haifa, Israel

HARLAMOV, B. **Continuous Semi-Markov Process**. John Wiley and Sons, Hoboken, New Jersey, 2008. 375 pp. $150.00 (not available in £ or €). ISBN 9781848210059.

Multistate models are very useful for describing complicated event history data. Commonly, the Markov model is used to estimate and model the transition intensities/probabilities. However, sometimes fitting a Markov model may be unsatisfactory because the transition times may not be constant or exponentially distributed. The semi-Markov models provide a flexible alternative for modeling the transition intensities. Recently, the semi-Markov process has been considered and adapted by Voelkel and Crowley (1984); Lagakos, Sommer, and Zelen (1978); and Shu, Klein, and Zhang (2007), among others. Thus, a well-written semi-Markov book could serve as a reference book for the statistical research community.

Continuous Semi-Markov Process by Harlamov is an excellent textbook and reference. The main goal of this book is presenting the theoretical aspects of continuous semi-Markov models. However, the readers need to have a solid background in measure-theoretic-based probability plus an adequate knowledge of weak convergence to fully understand and appreciate this book.

The book is well written and nicely laid out. The author presents each result with a clear proof or a short discussion. In the introduction section, the author gives a clear overview of the book by each chapter. Chapters 1, 3, and 7 are particularly valuable to biostatisticians. In Chapter 1, the author provides a background review of random-sequence Markov chains and focuses on constructing a measure of the stepped semi-Markov process, which is used to generalize semi-Markov processes in the remainder of the book. In Chapter 3, the author starts with the definition of general semi-Markov processes and presents their basic properties. In Chapter 7 (Limit Theorems for Semi-Markov Processes), the author presents the sufficient and necessary conditions for weak convergence, the conditions for a stepped semi-Markov process to converge weakly to a general semi-Markov process, a proposition of convergence of semi-Markov walks, and an example of convergence of a sequence of semi-Markov walks to a Wiener process.

Extensive references are given in the book, which provide the relevant literature to the readers. However, many references are published in Russian. Only a few examples are presented in the book and no real data has been presented and analyzed. This reviewer wishes there had been more examples presented along with the text. Without data examples, the book might be difficult for those statisticians who do not have sufficient background in probability theory.

References

Lagakos, S. W., Sommer, C. J., and Zelen, M. (1978). Semi-Markov models for partially censored data. *Biometrika* **65,** 311–317.

Shu, Y. I., Klein, J. P., and Zhang, M. J. (2007). The Cox semi-Markov illness-death model: A large sample study asymptotic theory for the Cox semi-Markov illness-death model. *Lifetime Data Analysis* **13,** 91–117.

Voelkel, J. G. and Crowley, J. J. (1984). Nonparametric inference for a class of semi-Markov processes with censored observations. *Annals of Statistics* **12,** 142–160.

Mei-Jie Zhang
Division of Biostatistics
Medical College of Wisconsin
Milwaukee, Wisconsin 53226, U.S.A.

COOK, D. and SWAYNE, D. **Interactive and Dynamic Graphics for Data Analysis: With `R` and `GGobi`**. Springer, New York, 2007. 180 pp. $59.95/€50.24 (paperback). ISBN 9780387717616.

Dianne Cook and Deborah Swayne's new book, part of Springer's *Use* `R`*!* series, helps fill an important niche in the literature for the `R` community and readers of *Biometrics.* The title of the book indicates the substantial challenge: the book is neither purely methodological nor simply about software. It draws upon the expertise of a wide range of contributors, but Cook (Professor of Statistics, Iowa State) and Swayne (Statistics Research Department, AT&T Labs, Florham Park, NJ) specifically recognize Andreas Buja, Duncan Temple Lang, Heike Hofmann, Hadley Wickham, and Michael Lawrence. I commend the authors and their contributors for the considerable effort spent both in producing this book and in developing the software. `GGobi` (Swayne et al., 2007; Lang et al., 2008) is an open-source visualization program for exploring high-dimensional data. Like `R` (R Development Core Team, 2008), it is freely available on Windows, Mac, and Linux platforms. The software availability and ease of installation should be appreciated by readers.

A first test in `R` (perhaps the equivalent of printing `hello world`) also worked as expected: `library(rggobi)`; `ggobi(mtcars)` produced two windows, the `GGobi` control panel and a graphics device displaying a scatterplot of miles per gallon (`mpg`) versus the number of cylinders (`cyl`). Following instructions outlined in the technical notes, I used the `GGobi` menu (`Tools ≫ Save Display Description`) to create a description file of the plot. This was then loaded easily into R using the `dd_load()` function of *DescribeDisplay* (Wickham, Cook, and Buja, 2007) and the plot was produced using either `plot`() or `ggplot`() (Wickham, 2007). As seen in the book's many well-conceived figures, `GGobi` and the associated packages thus provide useful extensions to `R`'s professional graphical capabilities.

My second attempt was less successful than the first. Without going into details, the problem brings me back to the review of the book: although the preface and technical notes provide minimal sufficient help for producing a first plot, I felt the book (and readers new to `GGobi`) would have benefited from a more extended, user-friendly introduction to using `R` with `GGobi`. Even if this duplicates the substantial complementary material available online, I would have preferred more extensive help for beginners in the opening pages.

The book (or as the authors put it, the language in the book) is aimed at advanced undergraduates and graduate students in disciplines involving the study of multivariate data. The authors also recommend it for researchers and industry statisticians. The book has exercises with each chapter, and could be used as a text or complementary text in a wide range of courses. I think the book succeeds in its appeal to these groups, and I applaud the authors for noting in the introduction that "learning how to perform data analysis is a process that continues long after the student's formal training is complete."

After the introduction, Chapter 2 describes parts of the "toolbox" of `GGobi`, noting (correctly) that excellent coverage of `R`'s graphical capabilities is available elsewhere and need not be duplicated. Again, given the subtitle of the book, I would have preferred a more thorough coverage of `GGobi's` methods, rather than simply being encouraged to "explore the methods in more depth than is covered in this book" (p. 45). Duplication of some of the content of the online `GGobi` manual would have been convenient, in my opinion. Subsequent chapters offer concise yet readable coverage of issues and graphical exploration surrounding missing values, supervised classification, and cluster analysis along with other miscellaneous topics. Each chapter explores interesting datasets, and Chapter 7 describes each of them with exceptional clarity. Though many authors attempt this to some extent, few have been as successful as Cook and Swayne.

The rare typographical error is unavoidable ("analysis of variance," p. 39), but other errors are puzzling. If the book were used for a course, however, students might be challenged (and thus more interested and engaged) to be critical rather than blindly trusting. For example, Figure 4.3 (top) omits the axes, making the conclusion at the top of page 72 puzzling; and such a pattern in the previous chapter indicated missing values, which is not the case here. In Figure 4.3 (bottom), some symbols use shaded characters while other are unshaded, and this difference is not explained. The scatterplot matrix of Figure 2.6 does not, in fact, include "the univariate ASH plot for each variable" in the diagonal displays (as claimed on page 26). Then, the associated correlation matrix is not symmetric; using the data provided online, I calculate the correlation between BD and CW to be 0.9925 (for the blue species), which on page 26 is reported as both 0.99 and 1.00.

Other wording choices may display a sense of humor, or hide additional insight lurking unseen from other sources. For example, with respect to the olive oil data, the authors wonder on page 77 whether there is a "problem with the quality of the samples" and conjecture on page 99 that "the most plausible story is that the Sicilian oils used borrowed olives from neighboring areas. This is interesting! Data analysis is detective work." Perhaps the interested student would seek out the primary source for more information, but I did not see how this conclusion was supported by the analysis. Without knowing more, I might guess that Sicilian olives could share the characteristics of some of the neighboring areas. In their cluster analysis they imply that the lack of a gap between two groups of points is a reason to "keep looking" (page 72); although this is followed by the discovery of a substantial gap, I think it may send the wrong signal to students: in many (if not most) cases, the classification will not be perfect, even in-sample. At the same time, I was encouraged by their use of training and test samples later in the chapter.

Readers of Cook and Swayne's book are advised to consider *Graphics of Large Datasets: Visualizing a Million* by Unwin, Theus, and Hofmann (2006). This book places greater emphasis on methodology without being tied to specific software (yet readers are encouraged to explore `MANET` [Unwin et al., 1996] and `Mondrian` [Theus, 2002], in particular). Other books, such as *R Graphics* by Paul Murrell (2006) and *Lattice: Multivariate Data Visualization with R* by Sarkar (2008) are also excellent recent additions to the literature on `R` graphics, though filling a somewhat different niche. Readers should also be aware of the release of *Interactive Graphics for Data Analysis: Principles and Examples* by Martin Theus and Simon Urbanek in October 2008.

In summary, *Interactive and Dynamic Graphics for Data Analysis: With* `R` *and* `GGobi` and the associated software are useful additions to the field. I will borrow a phrase the authors use on the opening page: these resources help us "orient ourselves in the sea of information."

References

Lang, D. T., Swayne, D., Wickham, H., and Lawrence, M. (2008). *RGGobi: Interface between R and GGobi.* R package version 2.1.9. Available at `http://www.ggobi.org/rggobi`, Last accessed September 29, 2008.

Murrell, P. (2006). *R Graphics.* Boca Raton, Florida: Chapman & Hall/CRC.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Available at `http://www.R-project.org`, Last accessed September 29, 2008.

Sarkar, D. (2008). *Lattice: Multivariate Data Visualization with R.* New York: Springer.

Swayne, D., Lang, D. T., Cook, D., and Buja, A. (2007). *GGobi.* Version 2.1.7. Available at `http://www.ggobi.org`, Last accessed September 29, 2008.

Theus, M. (2002). Interactive data visualization using Mondrian. *Journal of Statistical Software* **7**.

Unwin, A., Hawkins, G., Hoffman, H., and Siegl, B. (1996). Interactive graphics for data sets with missing values—MANET. *Journal of Computational and Graphical Statistics* **5**, 113–122.

Unwin, A., Theus, M., and Hofmann, H. (2006). *Graphics of Large Dataset: Visualizing a Million.* New York: Springer.

Wickham, H. (2007). *GGplot: An Implementation of the Grammar of Graphics in R.* R package version 0.4.2. Available at http://cran.r-project.org/web/packages/ggplot/, accessed September, 29, 2008.

Wickham, H., Cook, D., and Buja, A. (2007). *DescribeDisplay: R Interface to DescribeDisplay (GGobi plugin).* R package version 0.1.3. Available at `http://www.ggobi.org/describe-display/`, accessed September 29, 2008.

John W. Emerson
Department of Statistics
Yale University
New Haven, Connecticut 06520
U.S.A.

BOSC, D. and BLANKE, D. **Inference and Prediction in Large Dimensions**. Dunod and John Wiley & Sons, Hoboken, New Jersey, 2007. x + 316 pp. $100.00/€67.50. ISBN 9780470017616.

Rapid advances of modern computer technology and its wide application to various disciplines enable one to collect, store, and analyze data of huge sizes and large dimensions. Thus, large dimensional data analysis has become very hot in theoretical and applied research. It has been found that traditional statistical approaches developed based on classical limiting theorems (i.e., those obtained under the assumption of fixed dimension of data and/or the parameters) would lead to intolerable errors when the dimension is large. In applications of modern technologies, the dimensionality of data and/or parameters may be very large (even finite), such as data of curves, images, or movies. Therefore, there is an urgent need to develop new statistical procedures for the large dimensional data analysis. In a sense, the publication of this book will provide some alimentation for large dimensional data analysis.

The book consists of five parts split into 11 chapters. Part I of the book introduces a general theory for statistical prediction of large dimensional problems. In Chapter 1, the basic concepts and methodologies of prediction are introduced and some examples of prediction problems are presented. Most of the results presented are associated with squared loss functions. Predictions under general loss functions and those under Bayesian setups are also briefly mentioned. Some optimality criteria are briefly addressed. Chapter 2 provides some asymptotic properties of large dimensional predictions. Results in this chapter are mainly concerned with parametric predictions for stochastic processes, especially time series, such as causal-invertible autoregressive-moving average ($ARMA[p, q]$) process. Some prediction results are introduced for small and large time lag predictions under quadratic risks.

Part II discusses statistical inferences by projections. Chapter 3 considers the estimation by adaptive projection. The parameters φ of the underlying distribution are characterized as a functional of the distribution, which can be expressed as the expansion of φ with respect to an orthonormal base in the parameter space. Then the estimation problem reduces to one of estimating the coefficient sequence. Under quadratic risks, the optimal estimation turns out to be a projection problem. For some setups, under respective conditions, convergence rates of the estimation of the parameters are presented. When the number of parameters (or equivalently, the number of nonzero coefficients) is infinite, one needs to first truncate the sequences before optimal estimation can be obtained. Adaptive approaches are employed in this case. Chapter 4 considers the functional tests of goodness of fit. It begins with a discussion on Pearson's χ^2 test of goodness of fit and then generalizes to a functional test based on estimation by projection developed in Chapter 3. The efficiency of the functional test is then discussed. Finally, the functional test approach is extended to testing regressions and testing for stochastic processes. Chapter 5 extends the projection idea to nonparametric prediction, in particular, the prediction for conditional distribution and that for conditional distribution function are discussed.

Part III introduces statistical inference by kernel methods. Chapter 6 begins with an introduction to kernel estimation of density and nonparametric regression based on an independent and identically distributed sample. It then extends the kernel estimation to the case where the sample is a dependent sequence, or equivalently, a stochastic process in discrete time. The kernel estimation method is extended to the continuous time case in Chapter 7. The superoptimality of the estimation is discussed for both the density and the nonparametric regression. Also, some limiting properties, such as pointwise and uniform almost sure convergence, are presented. At the end of the chapter, predictions by kernel approach are addressed. Chapter 8 discusses the kernel method from sampled data. In real statistical applications, sampling a whole stochastic process in continuous time is not realistic and only a sample observed at a finite number of time instants is available. The estimation accuracy is discussed. At the end of this chapter, some numerical studies are presented.

Part IV, Local Time, contains only Chapter 9, which deals with the empirical density. It begins with a definition of occupation measure and density. Using the occupation density, one can define an unbiased estimator of the density of a stationary process, which is called the empirical density. Obviously, the occupation density is additive with respect to the time intervals and invariant under strictly increasing and derivable transformations. The consistency and convergence rate of empirical density estimators are introduced in this chapter. Also, the empirical density can be approximated by the commonly used density estimators.

Part V considers the theory for linear processes in high dimension. Chapter 10 is a preparation for the next chapter. It gives a general introduction to stochastic processes taking values in Hilbert spaces. Some more detailed properties are discussed for moving average (MA) and autoregressive (AR) processes in Hilbert spaces and AR processes in Banach spaces. Chapter 11 is devoted to applications of the theory developed in previous chapters to linear processes. This chapter discusses the estimation of the mean of a functional linear process, estimation of autocovariance operators, and predictions of autoregressive Hilbert processes and autoregressive process with continuous time (ARC) processes. At the end of the book, an appendix is given which provides some definitions and results, as an aid to readers who may lack some of the necessary background material assumed by the authors.

The book provides good examples that demonstrate the present-day importance of large dimensional data analysis. Some existing approaches and results are presented. However, most of them remain the same as in the traditional approach, such as the projection method and quadratic risks. I feel it is better to characterize the dimension as an increasing function with a certain order of growth, such as dimension being proportional to the sample size. Some examples can be found in Bai and Saranadasa (1996) and Bai and Silverstein (2004). Consequently, the limiting efficiency will depend on the growth rate, and also enable the possibility of new methods to be proposed to correct the errors caused by the "large dimension effect." When formed in this manner, the meaning of "large dimension" becomes more meaningful. Moreover, because of the aid modern computers offer, the corresponding approaches and results should be extended to nonquadratic risks so that the robustness can be reached.

References

Bai, Z. D. and Saranadasa, H. (1996). Effect of high dimension: By an example of a two-sample problem. *Statistica Sinica* **6,** 311–329.

Bai, Z. D. and Silverstein, J. W. (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Annals of Probability* **32,** 553–605.

Z. D. Bai
KLASS & School of Mathematics
and Statistics
Northeast Normal University
and DSAP
National University of Singapore
Singapore

CASELLA, G. **Statistical Design**. Springer, New York, 2008. xxiii + 307 pp. US$84.95/€69.95. ISBN 9780387759647.

Statistical Design is the fifth book by George Casella and is intended to serve as a text for a one-semester graduate-level course focusing on the theoretical details and philosophical issues related to the more commonly used experimental designs. The reader should be familiar with analysis methods for designed experiments and should have taken a sequence in mathematical statistics. Some experience with linear models theory would be helpful as well.

The book is organized into six chapters with the first chapter devoted to basic design concepts including sums of squares, expected mean squares, orthogonal contrasts, and replication versus subsampling. Chapter 2 covers issues connected to completely randomized designs. Chapters 3 and 4 are devoted to topics related to blocking, and Chapter 5 considers split plot and strip plot designs. Chapter 6 briefly discusses incomplete block designs and fractional factorials as well as reference designs and loop designs used in microarray experiments.

Each chapter contains a robust set of problems that often demonstrate key concepts in design. The author also does a nice job of demonstrating many definitions and concepts through real-world examples complete with data. Most datasets used in the book, as well as the `R` programs used to analyze them, are said to appear on the author's website, but as of this writing, several have not yet been posted. I did download a few of the author's `R` programs and, with slight modification to specify the location of the data, they ran as advertised. Out of curiosity, I also imported a couple of datasets to `SAS` and generated output using `PROC GLM` and `PROC Mixed` that matched the author's.

The title *Statistical Design* was chosen to reflect the author's separation of a design into a treatment design (i.e., the combinations of factor levels used) and an experimental design (i.e., the randomization of experimental units to treatment groups). As my design sensibilities have been formed by Milliken and Johnson (1992), who refer to treatment and design structures, and by Hinkelmann and Kempthorne (1994), who refer to treatment and error-control designs, I was quite pleased. However, I was disappointed that these concepts were not more explicitly emphasized throughout the book, and in the case of the Latin square, they appear to be confused.

At times the book seemed to be somewhat uneven in its depth and I often found myself wanting more details. For example, the author goes into great detail developing the proper error terms in the *F*-tests for various designs, especially the machinery for computing expected mean squares, but methods of multiple comparisons appear in "miscellanea" where details are given for only Bonferroni's and Scheffé's methods. I was happy to see a discussion on false discovery rate but was then disappointed that no example was given. I also wished the author had made more use of the matrix form of the models as I often found myself doing so to simplify my calculations.

The chapter on split plots leaves me rather conflicted. On one hand, the author gives a very good explanation of the split plot design and lightheartedly suggests that once one sees a split plot design, almost all the designs one sees in the future will now look like some form of split plot (which has certainly been my experience). The development of the *F*-tests for the split plot with completely randomized whole-plot treatments was very good, especially with respect to the intuition behind the different error terms for the whole-plot and subplot treatments. However, I was disappointed at the author's inclusion of the block-by-subplot-treatment interaction in the model for a split plot with whole plots run in a randomized complete block design. This has the unfortunate effect of creating different error terms for the subplot treatments and the whole-plot-by-subplot-treatment interaction. It also ignores the author's own observation that subplot treatments are observed within the levels of the whole-plot treatments.

Overall, I found reading this book to be worthwhile. I particularly think the author's discussion of blocking is quite interesting as well as the discussion of loop designs versus balanced incomplete block designs. In fact, I intend to use this book as supplementary reading material for my own design course.

References

Hinkelmann, K. and Kempthorne, O. (1994). *Design and Analysis of Experiments, Vol. 1: Introduction to Experimental Design.* New York: John Wiley & Sons.

Milliken, G. A. and Johnson, D. E. (1992). *Analysis of Messy Data, Vol. 1: Designed Experiments*. New York: Chapman & Hall.

Christopher I. Vahl
Department of Statistics
North Dakota State University
Fargo, North Dakota 58105
U.S.A.

GALWEY, N. W. **Introduction to Mixed Modelling: Beyond Regression and Analysis of Variance**. Wiley, West Sussex, England, 2006. x + 336 pp. US$100.00/€69.90. ISBN 9780470014967.

WEST, B. T., WELCH, K. B., and GALECKI, A. T. **Linear Mixed Models: A Practical Guide Using Statistical Software**. Chapman & Hall/CRC, Boca Raton, Florida, 2007. xix + 353 pp. US$79.95/£44.99. ISBN 9781584884804.

Both of these books are aimed at providing an introduction to mixed models for researchers in fields other than statistics. As such they place a heavy emphasis on the more practical aspects of performing a mixed model analysis and focus on the software used to perform the analysis. As a result of this emphasis, neither book would be appropriate for someone interested in the methodological foundations of mixed models.

Both books share a number of common features in which a typical chapter focuses on one aspect of a mixed model. The books each walk through the analysis in sufficient detail that a typical reader would be able to both reproduce the analysis and apply the lessons learned to similar data sets. The books cover the usual mixed models starting with a single random effect and proceeding to random regression models and repeated measures designs.

The basic structure of West et al. is a series of case studies. A case study starts with a description of the study followed by an analysis plan. The book then looks at the details associated with running the analysis in `SAS`, `SPSS`, `R`, `Stata`, and `HLM`. An examination of the results along with a comparison of the results from the different software packages follows. The examination of the same analysis plan with different packages will be particularly useful for a reader familiar with one package who has a need to use a different package. One drawback of this approach can be found in tables that contain the results from each analysis with several columns containing the same numbers. The approach and terminology used in the book is to frame a linear mixed model in terms of a multilevel hierarchical model that should be familiar to researchers in the social sciences.

Galwey takes a more traditional approach to presenting mixed models. In part, this can be seen in the development and presentation of models typically encountered in a course on experimental design. Whereas West et al. use the case study as the focus to bring out statistical concepts, Galwey focuses on a statistical concept and uses the data set to illustrate the concept. Consistent with its emphasis on `GenStat` and `R`, hypothesis testing is based on the use of full and reduced models. Exercises are included at the end of each chapter, which typically revolve around applying the lessons learned to new data sets. The book does include short introductions to topics such as generalized linear mixed models and analysis of spatial data. While the coverage of these topics is necessarily limited, it does provide the reader with a starting point for working with data that do not fit into one of the usual linear mixed models.

Each book provides a reader with the basic tools needed to conduct a mixed model analysis. Because the focus is on the practical aspects, the coverage of statistical concepts behind BLUP and REML is minimal. The approach taken by Galwey is to give the reader a feeling for BLUP and REML. For example, the discussion of the differences between BLUE and BLUP (Chapter 5) focuses on the sampling properties while avoiding defining what it means to be a BLUP. West et al. do a better job of introducing the statistical concepts at the beginning of the book. However, each book tends to treat the sampling properties of an empirical BLUP calculated using estimated variances and of a BLUP calculated using the true variances as the same.

I found the discussion of algorithms for estimating variance components in Galwey (Chapter 10, Section 10.9) problematic. My reading of Section 10.9 would indicate that there are two algorithms available for obtaining REML estimates. The two algorithms are the average information (AI) algorithm and Fisher scoring. Further, with the AI algorithm you can use sparse matrix methods and with Fisher scoring you can use absorption. The erroneous impressions that the reader may be left with are that there are only two algorithms, that sparse matrix methods cannot be used with Fisher scoring, and that absorption cannot be used with the AI algorithm. While West et al. do a better job describing the algorithms in general terms, the recommendation against Fisher scoring (p. 31) because of problems of determining the expected Hessian matrix is troublesome given that one of Fisher scoring's advantages is that the calculations are simplified.

I would see either of these books being useful to someone who wanted to understand the process of setting up, conducting, and evaluating a mixed model analysis. West et al. would be more appropriate for a researcher with problems where model selection is a major component of the analysis. West et al. would also be useful to a practicing statistician who is familiar with mixed models and wishes to use a new statistical package. Galwey, on the other hand, takes a more designed-experiment approach to mixed models. Coupled with its focus on statistical concepts along with the exercises it would be easy use Galwey as a text for a service course in mixed models.

Stephen D. Kachman
Department of Statistics
University of Nebraska
Lincoln, Nebraska 68583, U.S.A.

BISWAS, A., DATTA, S., FINE, J. P., and SEGAL, M. R. (eds). **Statistical Advances in the Biomedical Sciences: Clinical Trials, Epidemiology, Survival Analysis, and Bioinformatics**. John Wiley & Sons, Inc., Hoboken, New Jersey, 2008. xxix + 582 pp. $127.50/€92.90. ISBN 9780471947530.

This book is divided into five parts. Part I includes five chapters on clinical trials. Part II consists of four chapters on epidemiology. Part III includes eight chapters on survival analysis, while part IV is focused on bioinformatics and contains ten chapters. Finally, there are three additional chapters in a final section of the book labeled as Miscellaneous Topics.

As the authors began this project, they had three primary goals. These were:

1. Develop a well organized and multifaceted presentation of cutting-edge research in biomedical applications of statistics, putting them all under one umbrella;
2. Identify new directions of research or outstanding problems in each area for future researchers and provide a detailed list of references in order to facilitate self-study; and
3. Accomplish these two forgoing objectives in a way accessible to people outside of academia as well as those within.

The authors pursued these objectives by identifying a strong international team of leading researchers as the contributing authors. For the most part, this group of contributors adhered quite closely to a uniform presentation style throughout the book. Each chapter begins with a clear introduction that explains a biomedical/statistical problem. This is followed by a literature review and an expository example or case study; together this organization makes the technical material accessible to a broader audience than might otherwise be attained.

The comprehensiveness of the topic coverage varies within the major sections. It is more comprehensive within the Survival Analysis and Bioinformatics sections than it is within the Clinical Trials and Epidemiology sections of the book. For example, the five chapters in the Clinical Trials portion of the book include a chapter on phase I clinical trials; another on phase II clinical trials; a third on a highly specialized topic in phase III clinical trials, namely response adaptive designs; a chapter devoted to inverse sampling for clinical trials; and one on the cluster randomized trial. Topics such as group sequential interim analysis, adaptive designs, incomplete data and imputation, and noncompliance are areas that are receiving considerable attention in the current research literature, but receive no mention here. A similar comment can be made about the section on epidemiology. There is one chapter dealing with dynamic modeling in AIDS research, another on spatial epidemiology, another on modeling disease dynamics with cholera as the case study, and finally, a chapter on measurement error and misclassification. Each chapter in these two portions of the book is well written, but collectively the breadth of topic coverage is somewhat narrow.

The sections on survival analysis and bioinformatics each provide coverage of broad research activities within each of these fields. The level of writing in each of these chapters again facilitates the reader's introduction to each of the topics. Several of the chapters nicely work at the interface of several of the topic areas, namely survival analysis and bioinformatics. These chapters are particularly rich in both theoretical content and ideas for future research. Graduate students will find these chapters valuable.

The methods described in the text overall are applicable to many areas of medical research. While the topics addressed in bioinformatics cover important areas such as biomarker discovery, prediction of RNA signals, and assessment of differentially expressed genes, it is noteworthy that several chapters present techniques that can be applied to complex oncology data, repeated event data, time-varying covariate data, as well as data with spatial features. This enriches the book's appeal to readers focused on biostatistical discovery as well as statistical applications.

The authors have done an excellent job of meeting the objectives they put forward in the preface. They have produced an authoritative volume of readable chapters. While the topic coverage may not be as comprehensive in some sections of the book as it is in others, the chapters are well written and will be understandable to graduate students in biostatistics and statistics. The book will have an important place as a reference book on the shelf of many professional biostatisticians working in a biomedical research environment. Additionally, it should be useful as a special topics text for graduate students in biostatistics and statistics graduate programs.

ROBERT F. WOOLSON
Department of Biostatistics
Bioinformatics & Epidemiology
Medical University of South Carolina
Charleston, South Carolina 29425, U.S.A.

COOK, T. D. and DeMETS, D. L. (eds). **Introduction to Statistical Methods for Clinical Trials**. Chapman and Hall/CRC Press, Boca Raton, Florida, 2008. xxiii + 439 pp. US$69.95/£37.99. ISBN 9781584880271.

Another book on clinical trials—why? What is wrong with the existing ones? Well, to answer the second question, perhaps nothing is wrong with the existing ones; to answer the first question, perhaps a different approach, differences of emphasis, different styles, and so on, all add to the richness of the literature. This book does take a different approach and different emphasis. It aims at quite a high technical level so offers a really solid methodological background to the statistical issues in clinical trials.

This is an edited text, and these are always prone to problems of inconsistencies of content as well as style. While I think there is good consistency of content, styles across the chapters do differ and there seems to be some unnecessary overlap. For example, after an introductory/background chapter, Chapter 2 is entitled "Defining the Question." This is an important (and often missed) issue for a statistically orientated book about clinical trials—but why such a chapter should contain formulae for how to calculate the chi-squared statistic for a 2-by-2 table, or how to calculate the Mann-Whitney U-statistic, is not clear. (The t-statistic gets a mention but we do not get to see how to calculate it.) A further point on different contributors' styles is that most chapters conclude with a short set of exercises/problems—but not all do; and there seems little reason for those that do rather than those that do not.

On a different issue, there is an extreme insistence on the intention-to-treat principle and great criticism of anything that deviates from this; clinical trials being described as "fundamentally hypothesis testing instruments." To a large part I agree but I also think this type of thinking has put unnecessary constraints on the design and analysis of early development (e.g., phase II) studies and, hence, on our ability to learn sufficiently from them. Exploratory analyses should be interpreted for what they are—exploratory. But that is not to say that all clinical trials should be devoid of such analyses.

Related to the authors' (or is it the editors'?) view that clinical trials are "fundamentally hypothesis testing instruments," they also challenge the value of estimates of effect size. Significance tests (at least in well-run studies) are "always valid" (according to the text), but estimates of effect size are not (or may not be) because of selected populations, idealistic trial conditions, and so on. This is something I have long believed and I am pleased to see it written and justified, but it does leave me a bit worried. I thought 20 or so years ago, statisticians had much success in convincing clinicians that P-values were not all that was important—*how much* benefit (or harm!) was at least as important. I would be left a little worried if ever some of my clinical colleagues were to pick up on this challenge to the perceived wisdom of estimation. And besides, while absolute effects are almost certainly unbelievable, relative effects (or differences from comparator) might be more believable.

In addition to chapters broadly applicable across most trials, there are two quite specific chapters: one on survival analysis and one on longitudinal data. Both are very good (and useful), but others could have been considered: binary data or nonparametric methods, as examples. There is a general chapter on "Selected Issues in the Analysis" which covers (among other things) missing data, subgroup analyses, and multiplicity. Overall, I was very impressed by all the content but if I had to pick out the best of the bunch, I would cite a chapter on data monitoring and interim analyses contributed by Kyungmann Kim, Thomas Cook, and Dave DeMets. The (technical) statistical content is the main focus of the book and this is what helps it to stand apart from most others on clinical trials (even the more obviously statistically orientated ones). It takes the reader to quite a technical background that would serve him or her well if moving on to research problems in the various areas covered, yet does not lose sight of practical issues.

Production and finishing is generally to a high standard although there are a few typographical/style "errors" that should have been picked up in the copy editing. There are some differences in style of listing references, probably symptomatic of its being an edited book with many contributors. But these are all minor issues and I did not notice any gross and important errors. I was, however, amused by an entry in the references to Sir Austin Bradford-Hill, listed as "Hill, S. A. B." Ironically, the adjacent reference is to "Hill, A."

So despite a few niggling points (which have perhaps been overemphasized in this review), my overall summing up has to be very positive. There is much material in this book that is not for the casual reader and it is very much aimed at statisticians. Nonstatisticians will be comfortable with some of the content but will occasionally turn a page and remember why they gave up maths at school! For those of us with the interest (and need) to grapple with these more statistical issues I wholeheartedly recommend it.

Simon Day
Roche Products Ltd.
Welwyn Garden City
United Kingdom

CHOW, S.-C., SHAO, J., and WANG, H. **Sample Size Calculations in Clinical Research**, 2nd edition. Chapman and Hall/CRC, New York, 2008. x + 465 pp. $88.95/£44.09. ISBN 9781584889823.

This is the second edition of a widely used book on sample size selection. The first edition of this book has gotten quite favorable reviews. Shanmugam (2004) noted that it is "a good reference book for researchers in clinical trials." In a second review, Shanmugam (2005) described the first edition as "a fascinating book." He added that "applied statisticians, health and medical researchers will like it a lot. Statistical consultants will be fond of the book as a reference guide."

The second edition has been carefully updated from the first edition. For example, the discussion of power and sample size for the false discovery rate contains a careful review of the work of Jung, Chow, and Young (2005) that came out after the first edition.

This edition of the book, like the last one, is aimed at all players in medical and pharmaceutical research, including biostatisticians, doctors, and scientists. Somewhat drolly, the authors note that the aim of this book is to avoid the "(i) wrong test for the right hypotheses, (ii) right test for the wrong hypotheses, (iii) wrong test for wrong hypotheses, or (iv) right test for the right hypotheses with insufficient power."

The book begins with a comprehensive discussion of the American Food and Drug Administration regulatory requirements for sample size calculations. The brief review of confounding, interaction, crossover and parallel designs, subgroup and interim analysis, randomization, and rare events is a helpful summary of trial design aspects that can affect sample size.

The remaining chapters include formulae to calculate sample size for the comparison of means or variances, large sample tests for proportions, goodness of fit and contingency tables, time-to-event data, group sequential methods, bioequivalence testing, dose–response studies, microarray studies, Bayesian sample size calculation, and nonparametrics. The discussion of vaccine clinical trials is well done, and difficult to find in most other discussions of power and sample size.

One of the strengths of the book is the organizational structure. Each chapter contains comprehensive references, examples, and practical considerations. The book is clearly laid out and easy to read. The table of contents and the index are comprehensive, which makes the book quite useful as a reference. Despite the utility of the formulae, one deficit of the book is the lack of any software, or reference to software. Users of this book have to locate or program the formulae and algorithms on their own.

In summary, this is a useful, comprehensive compendium of almost every possible sample size formula. The strong organization and carefully defined formulae will aid any researcher designing a study.

References

Jung, S. H., Chow, S. C., and Young, S. S. (2005). Sample size calculation for multiple testing in microarray data analysis. *Biostatistics* **6,** 157–169.

Shanmugam, R. (2004). Book review. *Journal of Statistical Computation & Simulation* **74,** 387–390.

Shanmugam, R. (2005). A review of advanced biostatistics and statistical genetics books. *Journal of Statistical Computation & Simulation* **75,** 761–769.

D. H. Glueck
Colorado School of Public Health
University of Colorado Denver
Denver, Colorado 80262, U.S.A.

RUBINSTEIN, R. Y. and KROESE, D. P. **Simulation and the Monte Carlo Method, 2nd edition**. John Wiley and Sons, Hoboken, New Jersey, 2008. xvii + 345 pp. $99.95/€77.99. ISBN 9780470177945.

Except for the title and the first author, I would not recognize this as the second edition of a book I first read over 25 years ago. I had to check carefully, but I now am sure that "second edition" is in reference to the book by the first author in 1981. In the preface, the authors use the standard phrase from promotional blurbs, "long-awaited second edition." The first edition was reviewed in *Biometrics* in 1983 (vol. 39, p. 302), but this edition is substantially different.

The book is organized into nine chapters and an appendix. The first chapter, "preliminaries," has twelve sections on basic topics in probability, two sections on statistics, and one section on optimization and duality. The organization and contents of this chapter illustrate my main criticism of the book: the units, either chapters or sections, are too diverse to allow sufficient depth of coverage, and further, the units do not have a strong logical continuity. After the first two chapters, on preliminaries and random number generation, the subsequent chapters seem to be a random selection of topics. The topics themselves, for example, "counting via Monte Carlo" or "the cross-entropy method," are all interesting and include useful examples, but I feel that a more unified organization would have resulted in a much better book. I would not choose this book as the textbook for a course, but I would readily refer to it as a source of supplemental examples.

Many items appear essentially in laundry lists. For example, when the Kullback–Leibler divergence measure is introduced, the general φ-divergence measure is defined (with a typo; the "p" should be an "h") and five versions of this measure are given. There is no further comment; the only one used subsequently is the Kullback–Leibler divergence measure, which is used in the cross-entropy method of choosing a reference vector in importance sampling. Would other measures work equally well, or are there instances where one should be preferred over the others? Otherwise, why develop the general formula and list the five instances?

Each chapter has 10 to 20 exercises. They generally appear to be interesting, but, with some notable exceptions, are rather easy. Some require programming. No particular software is required, although the authors give sample `Matlab` programs in various places in the text. There is a "detailed" (in the authors' words) solutions manual, but I did not see it.

The book abounds with acronyms. Fortunately the authors provide a list, but many terms, such as "VM" (variance minimization) or "SF" (score function), do not seem to justify the overhead required for the reader's recognition, especially because a given acronym is used only in one chapter. (The problem is when the reader returns to that chapter for a quick fact check, but does not want to proceed sequentially through the chapter.) My dislike of the multitude of acronyms is related to the diversity of topics.

If an acronym is very standard in a field, or if in a single document, a great deal of discussion is devoted to a concept, use of an acronym is justified.

The references are listed at the ends of the chapters, which given the diversity of topics makes sense. The references also illustrate my criticism of lack of unity; in one chapter one book is given as the reference for a particular topic, and in a different chapter, another book given for the same topic (and, in my opinion, at least in some cases, if I were to give the two references for the given topic, I would have reversed the selection because of the context). Most of the references are from the last century. (For example, in chapters 6 and 7, of the 46 references, 31 are from the 1900s).

Despite my criticisms about the organization, I enjoyed reading the book, and found the individual examples quite interesting.

James E. Gentle
Department of Computational
and Data Sciences
George Mason University
Fairfax, Virginia 22030, U.S.A.

SHMULEVICH, I. and DOUGHERTY, E. R. **Genomic Signal Processing**. Princeton University Press, Princeton, New Jersey, 2007. xiii + 298 pp. US$60.00/£35.00, ISBN 0-69111-762-4.

This book is written from an engineering perspective on topics related to genomic signal processing (GSP), all of which are essential building blocks in the emerging field of biological network modelling and systems biology. According to the authors, "GSP is the engineering discipline that studies the processing of genomic signals . . . (the goal of GSP) is to integrate the theory and methods of signal processing with the global understanding of functional genomics." The goal of the book is to provide a rigorous mathematical foundation in the area while also connecting readers to real-world problems. The authors are eminently qualified to write about the topic based on their rich experiences and outstanding publications in the field.

The authors first present a terse, self-contained introduction to molecular biology, genomics, and proteomics. As can be imagined, 21 pages of coverage is neither elementary enough for a novice outside the field nor advanced enough for a researcher already inside the field. The main body of the book consists of two parts. The first part is dedicated to genetic networks with deterministic models in Chapter 2 and stochastic models in Chapter 3. The basic properties of graphic models and insightful discussions of popular Boolean networks and Bayesian networks are covered. Differential equation models are touched upon in a short section. In the second part of the book, the authors discuss topics in supervised machine learning, covering the fundamental concepts of classification in Chapter 4 and an important topic of regularization for feature selection in Chapter 5. The content focuses on the large p (variables) and small n (sample size) setting, which is commonly seen in genomic and proteomic problems. Many important propositions and principles contained are not well known to most microarray practitioners. For example, cross-validation has been viewed as gold standard in almost all the microarray literature. However, the authors assert: "Unfortunately, it is often used with neither justification nor mention of its serious shortcomings in small-sample setting . . . Resubstitution, leave-one-out and even 10-fold cross-validation are generally outperformed by bootstrap and bolstered estimators." Finally the book ends with a short chapter about clustering with only minimal material, due to the difficult and subjective nature of this topic with few mathematical foundations established in the literature.

The book provides rigorous mathematical definitions and many classical propositions for the main elements in GSP. Most examples given in the book are concise and artificial only for the purpose of elaborating on the mathematical concepts or theory. It is a good reference book for both researchers working in applications and methodological development. It also provides a complementary view to readers from statistics or biometrics community. These readers may find the treatment of problems very different. On the other hand, they may also find some familiar tools missing in the book; for example, the famous ridge regression, lasso, and elastic net are not even mentioned in the regularization topic in Chapter 5. In summary, the book is well laid out, pleasant to read, and with very few typos. It is not an ideal textbook or tutorial for students entering the field or for readers hoping to see many applications and case studies. For that purpose, one might consider the book *Genomic Signal Processing and Statistics* (Dougherty et al., 2005).

Reference

Dougherty, E. R., Shmulevich, I., Chen, J., and Wang, Z. J. (2005). *Genomic Signal Processing and Statistics*. New York: Hindawi Publishing Corporation.

George C. Tseng
Department of Biostatistics
University of Pittsburgh
Pittsburgh, Pennsylvania 15261, U.S.A.

BRIEF REPORTS BY THE EDITOR

ANDERSON, D. A. **Model Based Inference in the Life Sciences: A Primer on Evidence**. Springer, New York, 2008. xxiv + 184 pp. \$39.95/€35.26. ISBN 9780387740737.

Much like Burnham and Anderson (2002), also (co-)authored by David Anderson, this book's main focus is on modeling using information criteria. The main difference is that the present book is written for an audience that does not have a high degree of mathematical or statistical training. Otherwise, the spirit of this book and its main message are essentially the same as those of Burnham and Anderson (2002), as are some of tracts of text and certain examples. The author advocates model selection based on an a priori reduced subset of feasible models, obtained through knowledge of the subject, rather than on any automated search method. The small-sample-corrected Akaike's information criterion is recommended to rank and weight models, and multimodel inference is discussed. The writing style is pragmatic and appropriate for someone without advanced statistical training. Readers looking to recommend a book on information-criteria-based modeling to colleagues who are not statisticians, or looking to locate such a book for their libraries are likely to be satisfied with this book. However, those who already own Burnham and Anderson (2002) and want to learn alternative approaches to modeling will not find anything new here.

Reference

Burnham, K. P. and Anderson, D. A. (2002). *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*, 2nd edition. New York: Springer.

CAMPBELL, M. J., MACHIN, D., and WALTERS, S. J. **Medical Statistics: A Textbook for the Health Sciences, 4th edition**. John Wiley and Sons, Chichester, England, 2007. ix + 331 pp. \$37.50/€24.90. ISBN 9780470025192.

Previous editions of this book were reviewed in *Biometrics* **47,** p. 348 (1st edition) and *Biometrics* **49,** p. 1286 (2nd edition). The book is intended for a target audience of students (and some professionals) in the health sciences. The authors obviously do not intend their readers to become self-sufficient statisticians just by reading this book: Chapter 1 includes a section entitled, "How a statistician can help." Instead the authors apparently hope for their readers to be able to think statistically in the contexts of their own work. This is a far more sensible goal, with a far greater chance of success.

The authors state that this new edition is a "total revamp" of the textbook, although it is intended to retain the low technical level of its predecessors. Indeed, I found that the book places very little emphasis on calculations. The authors instead stress general statistical understanding, appropriate uses of techniques, and interpretation of results. For example, there is a "Points when reading the literature" section in most chapters, a great idea that recognizes that the most common opportunity for their target audience to use statistical thinking is in the interpretation of someone else's published analysis.

Coverage of individual statistical methods is brief and shallow, but considering the scope of topics introduced within the book, the alternative would be an 800-page tome. The authors focus on the most simplified forms of analysis of categorical and numerical data, including touching on linear and logistic regression, survival analysis, and reliability of diagnostic tests. I suspect that individual instructors will often have a quibble with one or another of these simplified presentations, but again, the authors would have that the readers understand the principles of an analysis rather than the details. Chapters on observational studies, randomized trials, and sample size issues similarly give readers something to think about, without providing a wealth of solutions. Provided that the lesson on "How a statistician can help" is repeated often throughout a course, this book seems to me to be a very useful course text for an audience of students in the health sciences.

RABE-HESKETH, S. and SKRONDAL, A. **Multilevel and Longitudinal Modeling Using `Stata`, 2nd edition**. Stata Press, College Station, Texas, 2008. xxxiii + 562 pp. $59.00. ISBN 9781597180405.

The first edition of this book was reviewed in *Biometrics* **62,** p. 951. The reviewer summarized that "Established users of `Stata` who already possess knowledge of random effects and longitudinal models will be well served by this book," while noting that "the presentation is inextricably tied to the software," rendering it not very useful to non-`Stata` users. This is a considerably expanded version, nearly double the size of the original. Much of the added material serves to delineate more clearly between statistics and software.

The book now begins with a new chapter reviewing linear regression, including applications in `Stata`. New introductory material has been added to the inaugural chapter on basic random-effects models, making the book more accessible to readers who are inexperienced in the use of these models. Another new chapter has been created based upon expanded material from sections on longitudinal, panel, and growth data. This serves also to allow the development of simple random-intercept and random-coefficient models to transpire more concisely. A third new chapter on discrete-time survival data has been added. Finally, throughout the book separate sections and subsections entitled "Estimation with `Stata`" help to separate the discussion of the models from the discussion of the fitting of the models using `Stata`. This improves the readability of the book and opens it up to a potentially broader audience.

GOLLERY, M. **Handbook of Hidden Markov Models in Bioinformatics**. Chapman & Hall/CRC Press, Boca Raton, Florida, 2008. xx + 156 pp. $59.95/£31.99. ISBN 9781584886846.

This is not a book about hidden Markov models (HMMs), per se. Rather, it is more of a book on how to use software packages based on HMMs for the purpose of doing bioinformatics analyses and database searches. Clearly, the book's audience is those who want to use the tools rather than those who want to develop them. One can read the book cover to cover and still not really know what HMMs are or how they work beyond a very superficial level. Indeed, a colleague of mine who does research on HMMs, Rachel Altman, read the book and agreed with this assessment. She further commented, "In my opinion, the researchers in this field are using the term HMM because (i) there are some hidden states and (ii) there is a Markovian component to their models. However, I personally would not classify their models as HMMs" (in the same sense as is used in current methodological research).