management, the three pillars of data mining. The result is a book with many nice features that has elements of interest for each of the group cited above. The book’s wide-ranging character means that almost every subset of the intended audience will have some sections that are largely review and some sections that are perhaps too difficult.

The format chosen for this book probably makes the most sense for students who will someday be consumers, rather than producers, of data mining analyses. The book features an introductory chapter that is an overview of data mining and the data mining process. This chapter would be an excellent primer for managers who will be on the receiving end of data mining operations and want to have good questions in hand to get maximum benefit from the analysis. The five chapters that follow on methodology could serve a similar role, although the level of mathematical sophistication used in the writing might be problematic for many readers in the business and information technology sectors. On page 156, for instance, in the midst of a chapter titled “Statistical Data Mining” the author spends a paragraph showing that the normal distribution is a type of probabilist concerned with entropy in the context of stochastic processes, and small print does not draw in the reader. Nearly 20 years later, those petition. As Spurrier noted in his review, its sparse use of graphics, absence of color, and small print does not draw in the reader. Nearly 20 years later, those weaknesses are magnified, because modern books on the subject are more inviting. However, the book is a good source of problems, many of which were taken from British university examinations. The best aspect of the text is the provided solutions (with commentary), which are as detailed as one could wish.

J. Wade Davis
University of Missouri–Columbia

REFERENCES
Beaumont, G. P. (1972). *Elementary Mathematical Statistics*, London: McGraw-Hill.
Spurrier, J. D. (1987). Review of Probability and Random Variables, by G. P. Beaumont, *Journal of the American Statistical Association*, 82, 947.

Entropy.

Andreas Greven, Gerhard Keller, and Gerald Warneck (eds.). Princeton, NJ: Princeton University Press, 2003. ISBN 0-691-11338-6. xiv + 358 pp. $69.50.

This book is a collection of expanded and “textified” versions of presentations at a 2000 symposium on entropy held at the Max Planck Institute of Complex Systems in Dresden. Quoting from the Preface, it is meant to address a “general consensus” at the symposium that “it would be useful for the community of mathematicians and physicists to have the essence of the talks and discussions available in a book that conveys the thought processes and discussions at the conference.”

That community includes precious few statistical scientists—mainly a set of probabilists concerned with entropy in the context of stochastic processes, some of whom are authors of papers in this volume. Like most compilations, it is not a good point of entry for nonexperts. For the experts (and near-experts), it seems to have pluses and minuses. The style of most of the papers is more expository than that of many in-your-face-inaccessible compilations. There is a good mixture of points of view, including thermodynamics, quantum physics, stochastic processes, and information theory. On the negative side, there is the inevitable unevenness and the possibility that no one person will be interested in more than a handful of the 17 papers in the volume.

Perhaps it is best to treat the editors’ statement quoted in the first paragraph at face value. A very exciting conference—at least to those who attended—was held, and this book is an attempt to capture what happened, not only for those who attended, but also for those who did not. Today, however, there are many other vehicles besides (presumably unreviewed) published volumes for doing this, many of which do not entail spending $69.50.

Alan F. Karr
National Institute of Statistical Sciences

Survival Analysis: A Self-Learning Text (2nd ed.).

David Kleinbaum and Mitchel Klein. New York, Springer, 2005. ISBN 0-387-23918-9. xv + 590 pp. $84.95.

The most meaningful accolade that I can give to this text is that it admirably lives up to its title. As a faculty member in a biostatistics department who has never completed a class in survival analysis, this book comes as a real blessing. Of course, the authors have not geared their text toward naive statistics and biostatistics professors who lacked the foresight to incorporate survival analysis into their graduate curricula! Rather, the targeted audience consists of researchers in medicine and public health who seek an intuitive, conceptually driven introduction to the topic. This makes the success of their contribution all the more impressive. Presenting an understandable, cogent description of a potentially confusing technical construct, such as a hazard function, is a much more daunting challenge than presenting mathematical formulas and relying on quantitative expertise to supplant intuition.

The first edition of this book appeared in 1996, written by Kleinbaum (1996) alone. The presentation was based on a “lecture–book” format; each page is split into two columns, with expository material appearing on the right and illustrations, formulas, and summarizations appearing on the left. The left-column material mirrors what might comprise a set of slides for a classroom lecture; hence the name of the format is apt.

The first edition contained chapters on Kaplan–Meier survival curves and the log-rank test, the Cox proportional hazards (PH) model, the PH assumption, the stratified Cox procedure, and extension of the Cox PH model for time-dependent covariates. The second edition adds chapters on parametric survival models, recurrent-event survival analysis, and competing-risk survival analysis. The computer Appendix provides step-by-step instructions for using STAT (version 7.0), SAS (version 8.2), and SPSS (version 11.5). The now-defunct package SPIDA, featured prominently in the first edition, has been dropped. Each chapter concludes with a detailed outline, practice exercises, a test, and
answers to the practice exercises. Although not designed as a textbook for an introductory course in applied survival analysis, the book could be easily used for this purpose.

Joseph Cavanaugh  
The University of Iowa

**REFERENCE**

Kleinbaum, D. G. (1996), *Survival Analysis: A Self-Learning Text*, New York: Springer-Verlag.

**Handbook of Beta Distribution and Its Applications.**

Arjun K. Gupta and Saralees Nadarajah. New York: Dekker, 2004, ISBN 0-8247-5396-8. viii + 571 pp. $179.95.

The beta distribution is a versatile tool for the statistician, and this handbook can be considered a detailed user’s manual. The 22 contributed pieces range from nearly prose-free collections of relevant formulas associated with the beta to some “how-to” essays on applications. The set of applications ranges from statistical standards, like stochastic processes and hierarchical models, to specific applications in soil science and graph theory. Many of the individual articles include very extensive reference lists, a particularly valuable feature in a handbook of this type. This book would be a good addition to any statistics group’s reference library, where it would be of use to researchers looking to incorporate a beta distribution into a model. It also may be of great value to graduate students in a statistics theory class who find that many classic homework problems from a first course are solved in these pages.

Richard J. Cleary  
Bentley College

**Univariate Discrete Distributions (3rd ed.).**

Norman L. Johnson, Adrienne W. Kemp, and Samuel Kotz. Hoboken, NJ: Wiley, 2005. ISBN 0-471-27246-9. xix + 646 pp. $140.00.

This classic reference is thoroughly updated from the second edition published in 1992 (Johnson, Kotz, and Kemp 1992). The Preface does an excellent job noting changes in the latest edition. These include improved coverage of order-k and q-series distributions, Lagrangian distributions, and mixture distributions. The authors provide new material on chain binomial models; the intervened Poisson distribution; minimum, maximum, and condensed negative binomial distributions; and computer generation of various random variables. The most substantial revision is in Chapter 11, which now includes distributions associated with fully parameterized regression models for discrete data, such as the Tweedie–Poisson family, Poisson lognormal, Poisson polynomial, and Ifrén’s double-Poisson and double-binomial models. Other enhancements include cleaner typesetting and finer partitioning of chapter sections and subsections. The resulting effect is increased readability and easier navigation. The book is worth buying for its extensive bibliography alone—nearly 100 pages. The authors note that most of the 400 new citations refer to work published since 1992. The breadth of information presented is encyclopedic, but its presentation is more contextual and guided. The authors continue to do a praiseworthy job of making the material accessible in the third edition. This book should be on every library’s shelf.

J. Wade Davis  
University of Missouri–Columbia

**REFERENCE**

Johnson, N. L., Kotz, S., and Kemp, A. W. (1992), *Univariate Discrete Distributions* (2nd ed.), New York: Wiley.

**State-Space and Unobserved Component Models: Theory and Applications.**

Andrew Harvey, Siem Jan Koopman, and Neil Shephard. New York: Cambridge University Press, 2004. ISBN 0-521-83595-X. xiv + 380 pp. $50.00.

This volume is a collection of papers presented at the Academy Colloquium of the Royal Netherlands Academy of Arts and Sciences, held in Amsterdam on August 29–September 3, 2002. The volume, like the colloquium, honors James Durbin of the London School of Economics, a luminary in the field.

Such volumes are sometimes less than the sum of the parts, but that is not so in this case. This volume opens with a wonderfully readable “Introduction to State-Space Time Series Analysis” by Durbin himself. (And ends, fittingly, with a paper by his son Richard Durbin on “Finding Genes in the Human Genome With Hidden Markov Models.”) The other two papers in the first section, by Peter Whittle on decision aspects and Simon Maskell on particle filtering, combined with Durbin’s paper, represent an excellent introduction to the topic that alone makes the book worth buying.

The other sections focus on testing, Bayesian inference, and applications, which, in addition to genomics, include financial variability, stochastic volatility, and macroeconomics. Perhaps inevitably, the papers in these sections are uneven and sometimes isolated from one another. The coherence and strength of the first section mitigates these effects substantially, however, by providing an entry point to the remainder of the volume.

Alan F. Karr  
National Institute of Statistical Sciences

**Estimation in Conditionally Heteroscedastic Time Series Models.**

Daniel Straumann. New York: Springer, 2005. ISBN 3-540-21135-7. xi + 228 pp. $79.95 (P).

This book, volume 181 of Springer’s *Lecture Notes in Statistics Series*, has evolved from the authors’ dissertation. The book studies estimation methods in generalized autoregressive moving average time series models (the so-called “ARCH/GARCH paradigms”). ARCH/GARCH models have proven useful in finance and other areas where raw observations have a white noise structure, but correlations lurk in functions of the series (such as the absolute value or square). The stochastic structure of ARCH/GARCH models is relatively complex, and, accordingly, estimation in such models is not necessarily straightforward.

The author begins with several chapters narrating the mathematical tools and limit theory that are needed in the ensuing analyses. A review of estimation for autoregressive moving average series is first presented. From there, the author embarks on quantifying the stochastic traits of ARCH/GARCH models and the large-sample properties of ARCH/GARCH estimators.

Although this text is not for the mathematically weak, it is largely self-contained. For those with the required mathematical competence and breadth, the book should be very attractive. Indeed, the discourse is masterfully presented, and the author does a superb job linking many relevant areas of stochastic (limit theory, quasi-likelihood methods, ergodic theory, Markov chain stability, and heavy-tailed models, to name a few). Many historical tidbits are provided. Overall, I highly recommend this text to those in the area.

Robert Ludd  
Clemson University

**Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg.**

Donald W. K. Andrews and James H. Stock (eds.). New York: Cambridge University Press, 2005. ISBN 0-521-84441-X. xiii + 573 pp. $75.00.

This book is a Festschrift honoring Tom Rothenberg on his retirement from the Economics Department at University of California Berkeley. It contains 23 research articles partitioned into 4 parts: identification and efficient estimation, asymptotic approximations, inference involving potentially nonstationary time series, and nonparametric and semiparametric inference. The authors are a distinguished cast that includes Ole E. Barndorff-Nielsen, Peter Bickel, David A. Freedman, and Andrew C. Harvey.

There is something here for both the econometrician and the technically oriented statistician. Taken as a whole, the articles touch on many classical and cutting-edge topics, including structural equation modeling, causation, financial volatility, parametric bootstrap methods, empirical likelihood, unit root time series, robust estimation techniques, structural breaks and changepoints, nonlinear models, and weighted least squares techniques. I found most of the articles to be well prepared and insightful. Typographical errors are sparse, and the presentation is reasonably uniform across articles.
This homogeneity noted, some of the articles are presented at the research level and demand advanced technical skills to read, whereas others are more expository in nature. Matching my interests in time series, the part on inference for nonstationary time series, where the general theme was unit root behavior, was my favorite. Others’ mileage will vary, but I encourage those in this general area to troll the table of contents for something interesting.

Robert LUND
Clemson University

The Basics of S–PLUS (4th ed.).

Andreas Krause and Melvin Olson. New York: Springer, 2005. ISBN 0-387-26109-5. xxii + 442 pp. $59.95 (P).

Now in its fourth edition, this paperback continues to evolve smoothly, keeping up with changes in S–PLUS and its underlying language, S. The publication of this edition corresponds with the release of S–PLUS version 7. First published in 1997, this is the book that I recommend most often for beginning users of the software. The authors are careful to point out differences between UNIX and Windows versions of S–PLUS. Although the book’s primary focus is on command-line usage (as it should be), the treatment of the graphic user interface (GUI) has improved for both Windows and UNIX/LINUX implementations. The book contains more than four dozen screen shots of the S–PLUS GUI, along with step-by-step instructions on how to perform common tasks through the GUI. The occasional S–PLUS user would most likely benefit from the chapter titled “Tips and Tricks.” The book does a good job incorporating the new functions introduced in version 7, as well as mentioning those features that are included only in the “Enterprise Developer” edition of the software. As primarily an R user, I most appreciate the final chapter, which covers the differences between R and S–PLUS. If you own a first or second edition of this book, I recommend updating to the fourth edition.

J. Wade Davis
University of Missouri–Columbia

Recent Advances in Quantitative Methods in Cancer and Human Health Risk Assessment.

Lutz Eddler and Christos P. Kitsos (eds.). Chichester, U.K.: Wiley, 2005. ISBN 0-470-85756-0. xxviii + 463 pp. $130.00.

This edited volume in Wiley’s Series in Probability and Statistics “aims to present new concepts and methods for cancer and human health risk which account for the wealth of biological data and biological and medical concepts.” The material contained in this text reflects the editors’ strong belief “that risk assessment has to be quantitative and that it cannot develop properly without a correct appraisal of the stochastic nature of risk and its adequate mathematical treatment.” Much of the work contained in this volume was presented at the International Conference on Cancer Risk and Assessment—Mathematical, Statistical and Computational Methods held August 23–25, 2003 in Athens, Greece, at the Department of Mathematics of the Technological Education Institute. Comprising 25 self-contained chapters, this volume clearly shows the international nature of the conference in the diversity of the more than 50 contributing authors. The chapters are grouped into six parts, each with a brief introductory section intended to give some guidance and additional information on the topic. The chapters are grouped as follows: Cancer and Human Health Assessment, Biological Aspects of Carcinogenesis, Modeling for Cancer Risk Assessment, Statistical Approaches for Carcinogenesis Studies, Specific Modeling Approaches for Health Risk Assessment, and Case Studies. Overall, this book provides a convenient collection of quantitative research in cancer and risk assessment that is both rigorous and up-to-date.

J. Wade Davis
University of Missouri–Columbia

Advances in Neural Information Processing Systems 17: Proceedings of the 2004 Conference.

Lawrence K. Saul, Yair Weiss, and Léon Bottou (eds.). Cambridge, MA: MIT Press, 2005. ISBN 0-262-19534-8. 1668 pp. $100.00.

This volume contains papers presented at the 18th annual conference on Neural Information Processing Systems held in British Columbia on December 13–16, 2004. The articles consider a broad range of topics including algorithms, imaging, probability, clustering, speech and signal processing, and learning theory (among others). With 207 total papers, this volume is quite heavy. However, because the articles are limited to 8 pages, the variety makes for nondrowsy reading.

The papers seem well prepared in general. This should be expected, because they were culled from more than 800 submissions. The papers appear in alphabetical order by leading author, so there is little continuity of flow. Most articles have a computer science flavor, and the terms familiar to a statistician may have a different nomenclature here. For statisticians, this volume will resemble a potluck dinner—maybe one will find something interesting, maybe not. Those with interest in computer science should be encouraged to investigate further.

Robert LUND
Clemson University