Reviews of Books and Teaching Materials

Bayesian Methods: A Social and Behavioral Sciences Approach (2nd ed.).

Jeff GILL. Boca Raton, FL: Chapman & Hall/CRC, 2008, xxxvii + 711 pp., $69.95 (H), ISBN: 1-58488-562-9.

The second edition of *Bayesian Methods: A Social and Behavioral Sciences Approach* is a major update from the original version. New material has been added, the order of topics has changed, and the concentration on social and behavioral sciences has been reduced. The result is a general-audience text suitable for a first course in Bayesian statistics at the upper undergraduate level for highly quantitative students, or at the graduate level for students in a wider variety of fields. I just finished using this book as the text for my class which includes graduate students in fields from statistics and computer science to geology, political science, and environmental studies, as well as advanced undergraduates in statistics and bioinformatics. Of the texts I have tried so far in this class, Gill's book has definitely worked the best for me.

In the first edition, the choice of topics appeared targeted toward those most useful in the social sciences, such as a large emphasis on generalized linear models. In the second edition, most of the data examples are drawn from the social sciences, but the topics and methodology are all of the standard items one would want in a general interest course: Bayesian philosophy, likelihoods, normal models, linear regression, priors, robustness and sensitivity, hypothesis testing and Bayes factors, Markov chain Monte Carlo (MCMC), and hierarchical modeling. The material on frequentist approaches to generalized linear models has been moved to an appendix. Several new chapters on MCMC methods have been added to this version, giving a more extensive coverage of both standard implementational issues as well as more advanced techniques such as reversible jump and perfect sampling. The topics are generally presented in a reasonable order, although I found that MCMC seems to be introduced quite early in the implementation of various models despite not formally being covered until much later in the book. Also, exchangeability does not appear until late in the book, in the context of hierarchical models, which is admittedly a reasonable place for such a discussion, but traditionally it usually appears in the first chapter in most Bayesian texts, as it provides the fundamental theoretical underpinning via de Finetti's results. In some sense, the placement of exchangeability is representative of the overall approach of this book—topics are covered from a practical aspect, with the theory motivated by the modeling, rather than the other way around.

From the beginning, the book takes a practical perspective. The whole Bayesian paradigm is introduced through data-analytic concerns. Many students will find this approach more convincing than starting from a definition based on the theory of mathematical coherence and exchangeability. Examples of proper handling of uncertainty and of the use of prior information provide ample motivation, along with numerous other issues that arise naturally in data analysis. Only after the conceptual introduction do the equations start flowing.

Computing is a core component of modern Bayesian statistics, and there are numerous examples of R code provided. There is also an R package, BaM, available that includes data and code for examples in the book.

I have found this book good as a text, as topics are introduced at a more conceptual level, with the mathematical details following. In most cases, sufficient details are provided, and in many cases calculations are done in full, making it easier for students to follow along. One possible complaint is that more advanced students may not find as much technical detail as desired. That is part of the trade-off in keeping the book suitable for a wider audience, and keeping it from being overwhelming for those who are not quite as strong mathematically. That is not to say that the book is "lite," as it explicitly assumes knowledge of calculus and linear algebra, and makes use of those tools throughout. The point is that the book tries to keep a balance between the concepts and the details. For those who do want more details, extensive references are provided.

Where this edition does add quite a lot of detail is in the implementation of MCMC. About a third of the book now relates to MCMC. This content starts with a brief introduction in the first chapter and reappears much later as a full introductory chapter. Only after the conceptual and basic implementational introduction does a full chapter on the theoretical aspects appear. That is followed by a full chapter on implementational details. The book concludes with a chapter on "advanced" MCMC techniques, including simulated annealing, Metropolis coupling, simulated tempering, reversible jump, and perfect sampling. Compared to the rest of the book, this last chapter seems a bit out of place, being rather more advanced. Perhaps the idea was that, for many students, this would be the only book on Bayesian statistics that they own, so they might in the future want more details on these techniques. But I'm not convinced that this is the best use of space in the book, as a second chapter on hierarchical modeling or a second chapter on more advanced linear modeling might have been more helpful.

To summarize, this book fills an important market segment for classes where the canonical Bayesian texts (such as Gelman et al. 1995) are a bit too advanced. The emphasis is on using Bayesian methods in practice, with topics introduced via higher-level discussions followed by implementation and theory. As such it is not the book you would use for a class composed exclusively of Ph.D. students in statistics, but it is a good text for wider audiences.

Herbert K. H. LEE
University of California, Santa Cruz

REFERENCES

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian Data Analysis*, Boca Raton: Chapman & Hall.

Data Preparation for Analytics Using SAS.

Gerhard SVOLBA. Cary, NC: SAS Institute, Inc., 2006, xxii+408 pp., $67.95 (P), ISBN 1-59994-047-7.

Although usage of the term "data preparation" has become more common recently, the activities required to prepare data for statistical analyses have been an integral part of the research process. A greater focus has been placed on data cleaning and data preparation because of the increased availability of massive datasets. Those cleaning data must be skilled at implementing automated data checks because manually checking each inconsistency in a large dataset would be overwhelming. Increases in the availability of powerful computers allows many to work with large datasets, but tasks such as merging and transforming can consume large amounts of memory and time. Therefore, there is a large benefit to planning and determining the appropriate structure (or data mart) and programming efficiently.

Unfortunately, it is not readily apparent who possesses these skills. In statistics-related degree programs, data preparation issues are generally not part of the curriculum, as the focus is on statistical theory and applications. Textbook datasets are generally designed to read into a software package quickly and effortlessly. On the other hand, computer science and information technology programs are unlikely to include instruction on statistical programs.

To address this issue, *Data Preparation for Analytics Using SAS* is one of a handful of books focusing on the topic of data preparation. As the name implies, the book concentrates on the SAS software system and is mostly geared toward statisticians in industry, especially those who perform data mining. The book contains 28 chapters and three appendices split into six parts. The book starts with "Data Preparation: Business Point of View," a part describing the types of questions asked, data sources, and the points of view of the business person, quantitative expert, and data manager. All three of these groups are the audience for the first part; the latter two are the audience for the remaining five parts. Part 2 ("Data Structures and Data Modeling") describes the various forms of datasets (or data marts) involved in data mining analyses. Creating these data marts is explained in the third part, "Data Mart Coding and Content." Next,

  DOI: 10.1198/000313008X370915