Journal of Statistical Software

July 2004, Volume 11, Book Review 6. http://www.jstatsoft.org/

Reviewer: George Michailidis
University of Michigan

Computational Statistics Handbook with MATLAB

Wendy L. Martinez and Angel R. Martinez
Chapman & Hall/CRC, Boca Raton, Florida, 2002.
ISBN 1-58488-229-8. xvii + 591 pp. $89.95.

The stated goals of this book by its authors are (page xv on the Preface):

- To make computational statistics techniques available to a wide range of users, including engineers and scientists, and
- To promote the use of MATLAB by statisticians and other data analysts.

Before we examine how successful their undertaking is, we provide a brief summary of the contents of the book. Chapter 1 gives an overview of the topics covered in the book together with a brief description of MATLAB code and availability of MATLAB resources on the Web. Chapter 2 provides a basic discussion of probability concepts, whereas Chapter 3 discusses sample statistics (mean, variance, moments, etc.) and parameter estimation methods (maximum likelihood, method of moments). Chapter 4 is about generating discrete and continuous random variables with an emphasis on the inversion and rejection methods. Chapter 5 discusses exploratory data analysis concepts with a strong emphasis on graphical methods. A fairly detailed description of how to take advantage of the advanced graphics capabilities of MATLAB to make nice and informative plots is provided. Chapter 6 introduces the reader to Monte Carlo and boostrap methods and their use for hypothesis testing and construction of confidence intervals, while Chapter 7 covers cross-validation, jackknifing and its combinations with the bootstrap. Linear regression is introduced as a way to motivate the concept of cross-validation. Chapter 8 discusses density estimation methodology (histograms, kernel-based techniques, finite mixtures), where an application of the EM algorithm appears. Selected classification and clustering techniques are presented in Chapter 9, which also contains a nice discussion of ROC curves. Chapter 10 covers non-parametric regression, including local smoothing and kernel methods, as well as regression trees. Chapter 11 introduces the readers to Markov Chain Monte Carlo basics and discusses the Metropolis-Hastings and the Gibbs sampler algorithms, together with convergence monitoring methods. Finally, Chapter 12 focuses on modeling, visualizing and simulating spatial point processes. There are also four appendices that provide a quick introduction to MATLAB (Appendix A), a summary of

the notation used in the book (Appendix B), a discussion of several projection pursuit indices (Appendix C) and the source code of some of the more complicated MATLAB functions discussed in the book (Appendix D).

Almost all the concepts discussed in the book are illustrated with MATLAB code, which although not always particularly efficient, is nonetheless well documented and easy to understand. Every chapter is complemented with several exercises of varying degree of difficulty and with an informative set of 'Further Reading' notes.

The book is fairly successful in achieving the authors' second goal, but definitely falls short with respect to their first goal. MATLAB has become the language of choice for engineers and applied mathematicians and is also quite popular within some statistics areas (e.g. wavelet analysis and time series analysis in general). By providing well documented, although not optimized, code for many widely used statistical techniques, the authors would help expand the MATLAB community within statistics.

On the other hand, there are several important topics not covered at all in the book, such as generalized linear models, mixed and random effects models, time series analysis, and survival analysis, that can be found in a book with similar scope (see Venables and Ripley 2002). There are also some important omissions from the topics covered, such as linear and quadratic discriminant techniques and the k-means clustering algorithm in Chapter 9, spline methods in Chapter 10, krieging in Chapter 12, and a more in depth discussion of linear regression, just to name a few. Furthermore, the presentation of the material is rather superficial and a discussion of the strengths and weaknesses of the various techniques is lacking. On the positive side, the material is well motivated and its exposition is clear and lucid.

Nevertheless, I think that this book is perfectly appropriate as a textbook for an introductory course on computational statistics. It covers many useful topics, which in combination with the well-documented code make the underlying concepts easy to grasp by the students. It also brings to the forefront the powerful graphics capabilities of MATLAB.

Overall, this is a very nice book to be used in an undergraduate or Masters level computational statistics course. It would also prove useful to researchers in other fields that want to learn and implement quickly some advanced statistical techniques. However, it falls short as a computational statistics handbook, both in terms of depth and to some extent of breadth of coverage of topics.

References

Venables WN, Ripley BD (2002). *Modern Applied Statistics with S. Fourth Edition*. Springer. ISBN 0-387-95457-0, URL http://www.stats.ox.ac.uk/pub/MASS4/.

Reviewer:

George Michailidis
University of Michigan
Department of Statistics
Ann Arbor, MI 48109, USA
E-mail: gmichail@umich.edu

Journal of Statistical Software *Published:* 2004-07-04
July 2004, Volume 11, Book Review 6.
http://www.jstatsoft.org/