Journal of Statistical Software

April 2005, Volume 12, Book Review 5. http://www.jstatsoft.org/

Reviewer: Frauke Kreuter
University of Maryland, College Park

Numerical Issues in Statistical Computing for the Social Scientist

Micah Altman, Jeff Gill, Michael P. McDonald
John Wiley & Sons, Hoboken, NJ, 2004.
ISBN 0-471-23633-0. xv + 323 pp. $94.95.
http://www.hmdc.harvard.edu/numerical_issues/

This is a very interesting book in an area that hasn't gotten much attention in the social sciences, but can expect to have more than just a niche audience with the increase of maximum-likelihood-based applications, a heightened interest in simulations, and a general appreciation for computational statistics. *Numerical Issues in Statistical Computing for the Social Scientist* is the right book for any social scientist who has stumbled upon error messages relating to convergence problems, non-invertible, or ill-conditioned matrices, and who is not just interested in some rough guidance on what to watch out for, but rather wants to understand the source of these problems down to effects of errors in floating point arithmetic.

Micah Altman, Jeff Gill, and Michael P. McDonald state in their preface that this book is intended to serve multiple purposes: Introducing new principles, algorithms and solutions while at the same time serving as a guide to statistical computing. There is a benefit to including new research results in a guidebook, but with it comes the challenge to find the right level of difficulty. As a result, the book seems bimodal, sometimes missing the intermediate applied researcher, who writes modest programs within a given statistical package.

The book begins with a very accessible and engaging motivational chapter that raises the awareness of the limits of statistical software, introducing precision and algorithmic problems together with a social science example on predicting the number of suicides by children. Chapter 1 also serves as an outline for the topics covered in later chapters, and emphasizes nicely the importance of replication. Chapters 2 and 3 provide background in statistical computing. Chapter 2 focuses on problems that can occur with the implementation of algorithms and points to limitations of algorithms used for statistical computations. Chapter 3 on the other hand discusses overall strategies to evaluate the performance of existing software using a benchmark tests. While there are certainly limits to how much substantive researchers are interested in performing benchmark tests on existing software, the authors state an important point related to why one should care about numerical accuracy: "free market dictates that supply will meet demand". Chapter 4 extends the general discussion of computational threats to nonlinear algorithms and their sensitivity to implementation. The second half of the book is a selection of special topics, beginning with Chapter 5 on Markov Chain Monte Carlo estimates

(MCMC). Chapter 6 covers in detail problems with inverting the Hessian matrix; explaining reasons why it may not be invertible, the substantive meaning attached to this problem, and what fruitful solution strategies there are. This chapter exemplifies the difficulty in balancing the reading level for various audiences. Less advanced readers should bring some patience, skimming over the introduction and being pleased with a clear and pedagogical explanation of singularity and saddelpoints in section 6.5. In Chapter 7 numerical sensitivity analysis is applied to King's solution to the problem of ecological inference. Chapter 8, 9, and 10, each written by a "guest" author (McCullough, LeSage, and Allison) are somewhat different from the previous chapters. They review the numerical issues discussed in earlier chapters in the light of a specific application (Nonlinear Estimation, Spatial and Logistic Regression). Each of these three chapters is clearly structured, shorter than the previous ones, and full of hands-on recommendations for substantive researchers who apply these methods. Chapter 11 provides a summary for the entire book, as well as short and pointed recommendations for practical use.

As seen in the brief overview, the chapters move from general problem discussions to specific applications. Reading the chapters sequentially is, however, not necessary. Social scientists familiar with logistic regression might prefer starting with Allison's gentle introduction to convergence problems to make themselves familiar with terms used throughout the book. The final chapter, which gives general recommendations for replication and accurate analysis might also be a good start before going into the detailed middle chapters. The book has an extensive ten-page subject index that can in addition serve as a guide for selected reading; albeit one should not expect to find all error message key words in the index, for different terminology is used in different disciplines. Finally, the accompanying web page should be mentioned, where code and data are provided, as well as a growing number of very useful links to related topics. Readers should be sure to check the extensions and corrigenda to avoid being puzzled while working through the equations.

In sum, this book is a good reference for social scientists that are involved in computational statistics. The necessary prerequisites mentioned in the preface –basic introductory statistics course, some knowledge in matrix algebra, elementary knowledge of calculus, and rudimentary programming experience - are somewhat misleading, as students and researchers in the social sciences without some advanced statistical training and familiarity with Bayesian terminology, or those whose matrix algebra classes date back, might at some points get lost in terminology and the in-depth details that seem more appropriately geared towards programmers of statistical algorithms. It is very clear, however, that certain chapters can, with a knowledgeable instructor, be a nice supplement to applied statistics classes in the social sciences.

Reviewer:

Frauke Kreuter
Joint Program in Survey Methodology
University of Maryland, College Park, MD 20742, United States of America
E-mail: fkreuter@survey.umd.edu

Journal of Statistical Software *Published:* 2005-04-01
April 2005, Volume 12, Book Review 5.
http://www.jstatsoft.org/