Journal of Statistical Software

March 2007, Volume 19, Book Review 1. *http://www.jstatsoft.org/*

Reviewer: Jan de Leeuw
University of California at Los Angeles

Computational Statistics

Geof H. Givens and Jennifer A. Hoeting
Wiley-Interscience, Hoboken, NJ, 2005.
ISBN 0-471-46124-5. 418 pp. USD 99.95 (P).
http://www.stat.colostate.edu/computationalstatistics/

There seems to be an agreement between textbook writers that *computational statistics* is statistics that uses computation, often a lot of computation, while *statistical computing* is computational mathematics for statisticians. Both fields are hybrids, but it seems that in computational statistics the emphasis is on the statistics and in statistical computation the emphasis is on the tools. Of course, not everybody rigorously uses the distinction, and often it is quite difficult to draw the line.

Initially, most textbooks were in statistical computing (Chambers 1977; Kennedy and Gentle 1980), but starting with the influential work of Thisted (1988) (which was published after the desktop revolution) we see a shift towards computational statistics. Of the more recent textbooks, Gentle (2002) is computational statistics, while Gentle (1998a,b) are statistical computation. Gentle is working on another book covering optimization, clearly also in statistical computing. Lange (1999) on numerical analysis and Lange (2004) on optimization are solidly in statistical computation. The recent Marin and Robert (2007) is definitely computational statistics.

Statistical computation includes general numerical analysis (floating point, rounding error), numerical linear algebra, optimization, approximation, integration, and Monte Carlo methods. There is usually very little, if anything, on differential equations, partial or otherwise. This makes the field quite different from, say, *applied mathematics*.

The book we are reviewing here is partly in statistical computation and partly in computational statistics. For statistical computation it has chapters on continuous (Chapter 2) and discrete (Chapter 3) optimization, on numerical integration (Chapter 5), on simulation and Monte Carlo integration (Chapter 6), and on Markov chain Monte Carlo (Chapters 7 and 8). In addition, there are computational statistics chapters on EM optimization methods (Chapter 4), on bootstrapping (Chapter 9), on nonparametric density estimation (Chapter 10) and on bivariate and multivariate smoothing (Chapters 11 and 12).

Is it a problem that the two fields are freely mixed? Only in rare cases, I think. One possible problem is that numerical linear algebra and error analysis for floating point computation

are missing. That is an unfortunate omission, I think. The authors try to justify this in the preface by saying they want to concentrate on the big picture, but I think textbooks cannot afford to sweep too much of the nitty-gritty under the rug. You can leave it out of the lectures, but it should be in the textbook.

Another possible consequence of not distinguishing the statistical and computational aspects of the problem is the chapter on the EM algorithm. It takes the likelihood as its starting point and immediately embeds the algorithm in a maximum likelihood framework. That may be historically accurate, but we see from Lange (2004) that it is much clearer to separate the computational core of the algorithm from its statistical applications. The same thing is true, I think, for the Bootstrap, which can be quite naturally presented as a numerical or Monte Carlo method to evaluate derivatives. The authors do a pretty good job separating MCMC from Bayesian inference, but in other chapters the computational techniques and the statistical applications are pretty tightly intertwined.

Sine this is a *Journal of Statistical Software* review, we should say something about software. It is missing. In the preface, the authors have some obligatory references to S-PLUS, R, and MATLAB, but they take the point of view that "good students can learn as they go". This is undoubtedly true, but rigorous application of this point of view makes it unnecessary to write textbooks as well. It seems that the authors have realized this as well, since the website of the book (which is in its fourth printing) now has code and examples in R. Not in S-PLUS and MATLAB, to be sure.

The discussion of the various techniques, both of the computational and statistical aspects, is competent and clear. The accepted wisdom is presented, there are very few idiosyncrasies. So we can characterize this as a solid book, perhaps even as a good book, even though some things are missing and even though some topics would benefit from a clearer separation of the computation and the statistics. I would have no hesitation recommending it to working statisticians and quantitative empirical scientists. I would have just a little bit more hesitation using it in graduate courses. It's pretty expensive.

References

Chambers JM (1977). *Computational Methods for Data Analysis*. Wiley.

Gentle JE (1998a). *Numerical Linear Algebra for Applications in Statistics*. Springer-Verlag.

Gentle JE (1998b). *Random Number Gneeration and Monte Carlo Methods*. Springer-Verlag.

Gentle JE (2002). *Eelements of Computational Statistics*. Springer-Verlag.

Kennedy WJ, Gentle JE (1980). *Statistical Computing*. Marcel Dekker.

Lange K (1999). *Numerical Analysis for Statisticians*. Springer-Verlag.

Lange K (2004). *Optimization*. Springer-Verlag.

Marin JM, Robert CP (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. Springer-Verlag.

Thisted RA (1988). *Elements of Statistical Computing*. Chapman & Hall.

Reviewer:

Jan de Leeuw
Department of Statistics
University of California at Los Angeles
Los Angeles, CA 90095-1554, United States of America
E-mail: deleeuw@stat.ucla.edu
URL: http://www.cuddyvalley.org/

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 19, Book Review 1 *Published:* 2007-03-23
March 2007