Numerical analysis for statisticians, by Kenneth Lange

- **Hardcover:** 620 pages
- **Publisher:** Springer-Verlag, New York, New York; 2nd edition (June 15, 2010)
- **Language:** English
- **ISBN-10:** 1441959440

> *"In the end, it really is just a matter of choosing the relevant parts of mathematics and ignoring the rest. Of course, the hard part is deciding what is irrelevant." (page ???)*

I had missed the first edition of this book and thus I started reading it with a newcomer's eyes (I will thus not comment on the differences with the first edition, sketched by the author in the Preface). Past the initial surprise of discovering it was a mathematics book rather than an algorithmic book, I became engrossed into my reading and could not let it go! *Numerical Analysis for Statisticians* is a wonderful book. It provides most of the necessary background in calculus and enough algebra to conduct rigorous numerical analyses of statistical problems. This includes expansions, eigenanalysis, optimisation, integration, approximation theory, and simulation, in less than 600 pages. It may be due to the fact that I was reading the book in my garden, with the background noise of the wind in tree leaves, but I cannot find any solid fact to grumble about! Not even about the MCMC (Markov Chain Monte Carlo) chapters! I simply enjoyed *Numerical Analysis for Statisticians* from beginning till end.

> *"Many fine textbooks (...) are hardly substitutes for a theoretical treatment emphasizing mathematical motivations and derivations. However, students do need exposure to real computing and thoughtful numerical exercises. Mastery of theory is enhanced by the nitty gritty of coding." (page ???)*

From the above, it may sound as if *Numerical Analysis for Statisticians* does not fulfill its purpose and is too much of a mathematical book. Be assured this is not the case: the contents are firmly grounded in calculus (analysis) but the (numerical) algorithms are only one code away. An illustration (among many) is found in Section 8.4: Finding a Single Eigenvalue, where Kenneth Lange shows how the Raleigh quotient algorithm of the previous section can be exploited to this aim, when supplemented with a good initial guess based on Gerschgorin's circle theorem. This is brilliantly executed in two pages and the code is just one keyboard away. The EM algorithm is immersed into a larger MM perspective. Problems are numerous and mostly of high standards, meaning one has to sit and think about them. References are kept to a minimum, they are mostly (highly recommendable) books, which is a principle I highly approve of for textbooks, plus a few research papers primarily exploited in the problem sections.

> *"Every advance in computer architecture and software tempts statisticians to tackle numerically harder problems. To do so intelligently requires a good working knowledge of numerical analysis. This book equips students to craft their own software and to understand the advantages and disadvantages of different numerical methods. Issues of numerical stability, accurate approximation, computational complexity, and mathematical modeling share the limelight in a broad yet rigorous overview of those parts of numerical analysis most relevant to statisticians." (page ???)*

While I am reacting so enthusiastically to the book (imagine, there is even a full chapter on continued fractions!), it may be feared that graduate students over the World would find the book too hard. However, I do not think so: the style of *Numerical Analysis for Statisticians* is very fluid and the rigorous mathematics are mostly at the level of undergraduate calculus. The more advanced topics like wavelets, Fourier transforms, or Hilbert spaces (for self-reproducing kernels) are very well-introduced and do not require prerequisites in complex calculus or functional analysis. Even measure theory does not appear to be a prerequisite! On the other hand, there is a prerequisite for a good background in statistics. This book will clearly involve a lot of work from the reader, but the respect shown by Kenneth Lange to those readers will sufficiently motivate them to keep them going till assimilation of those essential notions. *Numerical Analysis for Statisticians* is also recommended for more senior researchers and not only for building one or two courses on the bases of statistical computing. It contains most of the math bases that we need, even if we do not know we need them! Truly an essential book to hand to graduate students as soon as they enter a Statistics program..

The EM algorithm

The EM algorithm was introduced in 1977 by Dempster, Laird and Rubin in a paper that remains one of the most quoted statistics papers (to wit, currently 22,485 links on Google scholar!). EM stands for expectation-maximisation and it is an algorithm that aims at maximising likelihoods with a latent structure, like mixtures of distributions. Since the (observed) likelihood writes like an integrated completed likelihood

$$L^o(\theta|x) = \int L^c(\theta; x, z)\mathrm{d}z$$

the EM algorithm proceeds iteratively by computing an expected log-likelihood (E-step)

$$Q(\theta|x, \theta^{(t)}) = \mathbb{E})\theta^{(t)}[\log L^c(\theta; x, Z)|X],$$

where the expectation integrates out Z conditional on X and for a parameter value $\theta^{(t)}$. And then by maximising (M-step) $Q(\theta|x, \theta^{(t)})$ in θ, thus obtaining the new value $\theta^{(t+1)}$. By a convexity argument, each EM step increases the observed likelihood. The EM algorithm thus ends up in a local if not necessarily the global mode of the observed likelihood. Numerous extensions to the original scheme are found in the literature. (See also the Wikipedia article on Expectation-maximization algorithm, which contains an illustration for the mixture problem.)

Wavelets

Wavelets form a special type of function basis used to decompose functions into scale components. Because of the simultaneous use of two types of basis (the mother and the father wavelets), accounting for a multiresolution analysis, it is more efficient than the older Fourier transforms, which use sinusoids as a basis. For instance, a mother wavelet is the sinc function

$$\text{sinc}(t) = \frac{\sin(2\pi t) - \sin(\pi t)}{\pi t}$$

while an example of a father wavelet is the Haar wavelet $\varphi(t) = \mathbb{I}_{0\le t<1}$. This representation of functions is quite useful in data compression, being for instance at the basis of the JPEG 2000 standard. It is also a branch of non-parametric statistics since the 1990's.

MCMC methods

Markov chain Monte Carlo methods are a special branch of simulation (or Monte Carlo) methods where the distribution to be simulated f is the limiting and stationary distribution of the Markov chain. Two major groups of MCMC algorithms are Gibbs samplers on the one hand and Metropolis–Hastings algorithms on the other hand. The former uses a substitute Markov kernel $y \sim Q(x^t, y)$ and the Metropolis–Hastings acceptance probability

$$\rho(x^t, y) = 1 \wedge \frac{f(y)}{f(x^t)} \frac{Q(y, x^t)}{Q(x^t, y)}\,.$$

The later relies on full conditional distributions to simulate f one component at a time. While the theoretical convergence of MCMC methods is almost always garanteed, the practical implementation may face difficulties. However, MCMC methods have greatly contributed to the dissemination of Bayesian techniques in applied fileds since the 1990's.

The Gerschgorin's theorem

This theorem is used in the resolution of linear systems involving matrices A with a large condition number (i.e. a large ratio between the largest and the smallest absolute eigenvalues of A). It states that every eigenvalue of a matrix A lies within at least one of the Gershgorin discs $D(a_{ii}, R_i)$, where a_{ii} is the i-th diagonal element of the matrix A and

$$R_i = \sim_{j \neq i} |a_{ij}|\,.$$

Therefore, the information provided by the theorem about the magnitude of the eigenvalues allows for a preconditioning step that greatly reduces the condition number.

Further references

Daubéchies, I. (1992). *Ten Lectures on Wavelets.* SIAM.

Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). *J. Royal Statist. Society Series B*, **39** 1–38.

MacLachlan, G. and Krishnan, T. (1997). *The EM Algorithm and Extensions.* 2nd ed. John Wiley, New York.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. and Teller, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Phys.*, **21** 1087–1092.

Meyer, Y. (1990). *Ondelettes et opérateurs.* Hermann, Paris.

Nasson, G. (2008). *Wavelet Methods in Statistics with R.* Springer-Verlag, New York.

Robert, C. and Casella, G. (2011). A history of Markov chain Monte Carlo-subjective recollections from incomplete data. *Statist. Science*, **26** 102–115.

Silverman, B. and Vassicoulos, J. (2000). *Wavelets: the Key to Intermittent Information?* Oxford University Press.