



## Book reviews, Fall 2011

Christian Robert

### ► To cite this version:

| Christian Robert. Book reviews, Fall 2011. CHANCE, 2011, 24 (4), pp.58-61. hal-00636655

**HAL Id: hal-00636655**

**<https://hal.science/hal-00636655>**

Submitted on 28 Oct 2011

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Book reviews, Fall 2011

Christian P. Robert  
Université Paris-Dauphine, CEREMADE, IUF, and CREST  
xian@ceremade.dauphine.fr  
<http://xianblog.wordpress.com>

October 28, 2011

## Abstract

This note is made of three book reviews of Lange (2010), Vasishth and Broe (2011), and Stephenson (2008), respectively. They are scheduled to appear in the December 2011 issue of *CHANCE*.

## 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 iii)*

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, eigen-analysis, 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, see vignette) chapters! I simply enjoyed *Numerical Analysis for Statisticians* from beginning till end.

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

*“Many fine textbooks (...) are hardly substitutes for a theoretical treatment emphasising 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 xx)*

From the above, it may sound as if *Numerical Analysis for Statisticians* does not fulfil 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 (see vignette). This is brilliantly executed in two pages and the code is just one keyboard away. The EM algorithm (see vignette) 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.

### 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

$$L^o(\theta|x) = \int L^c(\theta; x, z) dz$$

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  $\theta$ , 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.)

### 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 Gerschgorin discs  $D(a_{ii}, R_i)$ , where  $a_{ii}$  is the  $i$ -th diagonal element of the matrix  $A$  and

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

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

*“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 modelling share the limelight in a broad yet rigorous overview of those parts of numerical analysis most relevant to statisticians.” (page xx)*

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 (see vignette), 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.

### 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  $\varphi(t) = \mathbb{I}_{0 \leq t < 1}$ . 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.

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.

### 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.

## The foundations of Statistics: a simulation-base approach, by Shravan Vasishth and Michael Broe

- **Hardcover:** 194 pages
- **Publisher:** Springer-Verlag, Berlin Heidelberg (2011)
- **Language:** English
- **ISBN-10:** 3642163122

First, the title of this book is a misnomer, in that *The foundations of Statistics* is a light introduction to statistics for mathematically challenged students, using simulation, rather than any reflection on the foundations of our field. It is sadly plagued with errors that show the incomplete grasp of the authors have on their subject.

*“We have seen that a perfect correlation is perfectly linear, so an imperfect correlation will be ‘imperfectly linear’.” (page 128)*

Those authors are Shravan Vasishth (Chair of Psycholinguistics and Neurolinguistics, Postdam, Germany) and Michael Broe (Department of Evolution, Ecology, and Organismal Biology, Ohio State University). Their

purpose there is to teach statistics “in areas that are traditionally not mathematically demanding” at a deeper level than traditional textbooks “without using too much mathematics”, towards building “the confidence necessary for carrying more sophisticated analyses” through R simulation. This is a praiseworthy goal, bound to produce a great book. However, and most sadly, I find the book does not live up to those expectations.

*“Let us convince ourselves of the observation that the sum of the deviations from the mean always equals zero.” (page 5)*

Besides the factual errors and foundational mistakes found therein, a puzzling feature of this book is the space dedicated to expository developments that aim at bypassing mathematical formulae, only to find this mathematical formula provide at the very end of the argument (as, e.g., the binomial pdf). Another difficulty is the permanent confusion between the sampling distribution and the empirical distribution, the true parameters and their estimates. If a reader has had some earlier exposition to statistics, the style and pace are likely to unsettle her. If not, she will be left with gaping holes in her statistical bases: for instance, the book contains no proper definition of unbiasedness (hence a murky justification of the degrees of freedom whenever they appear), of the Central Limit theorem, of the  $t$  distribution, no mention being made of the Law of Large Numbers (although a connection is found in the summary, page 63). This is a strong gap, given the reliance on simulation methods throughout the book. The material therein thus does not seem deep enough to engage in reading Gelman and Hill (2006), as suggested at the end of the book. Having the normal density defined as the “somewhat intimidating-looking function” (page 39)

$$f(x) = \frac{1}{(\sigma\sqrt{2\pi})} E^{-((x-\mu)^2/2\sigma^2)}$$

and with a very unfortunate capital  $E$  certainly does not help. (Nor does the call to `integrate` rather than `pnorm` suggested to compute normal tail probabilities (pages 69-70), as it paradoxically requires more mathematical maturity. A minor point, admittedly.)

*“The key idea for inferential statistics is as follows: If we know what a ‘random’ distribution looks like, we can tell random variation from non-random variation.” (page 9)*

The above quote gives a rather obscure and confusing entry to statistical inference. Especially when it appears at the beginning of a chapter (Chapter 2) centred on the binomial distribution. As the authors seem reluctant to

introduce the binomial probability function (pdf) from the start, they resort to an intuitive discourse based on (rather repetitive) graphs (with an additional potential confusion induced by the choice of an illustrative binomial probability equal to  $p = 0.5$ , since  $p^k(1-p)^{n-k}$  is then constant in  $k$ ). In Section 2.3, the distinction between binomial and hypergeometric sampling is not mentioned, i.e. the binomial approximation to the hypergeometric distribution is used without any warning. The fact that the mean of the binomial distribution  $B(n, p)$  as  $np$  is not established and the one that the variance is  $np(1-p)$  is not stated until the appendix, page 168. (Conversely, the book spends pages 36-39 showing through an R experiment that “the sum of squared deviations from the mean are smaller than from any other number”.)

*“The mean of a sample is more likely to be close to the population mean than not.” (page 49)*

This quote is the concluding summary about the Central Limit theorem, following an histogram with 8 bins showing that “the distribution of the means is normal”. It is itself followed by a section on “ $s$  is an Unbiased Estimator of  $\sigma$ ”. This unfortunately fake result (here,  $s$  is the standard estimator of the standard deviation  $\sigma$ , which cannot be unbiasedly estimated) seems to indicate that the authors are unaware that the transform of an unbiased estimator is generally biased. The introduction of the  $t$  distribution is motivated by the “fact that the sampling distribution of the sample mean is no longer be modelled by the normal distribution” (page 55). With such fundamental flaws in the presentation, it is difficult to recommend the book at any level. Especially at the most introductory level where students or/and instructors have no other referential.

*“We know that the value is within 6 of 20, 95% of the time.” (page 27)*

I am also dissatisfied with the way confidence and testing are handled. The above quote, which replicates the usual fallacy about the interpretation of confidence intervals (since 6 is a realisation of a random variable), is found only a few lines away from a (correct) warning about the inversion of confidence statements. This warning is only detailed much later: “it’s a statement about the probability that the hypothetical confidence intervals (that would be computed from the hypothetical repeated samples) will contain the population mean” (page 59). The book spends a large amount of pages on hypothesis testing, presumably because of the own interests of the authors, however it is unclear a neophyte could gain enough expertise from



those pages to conduct her own tests. Worse, statements like (page 75)

$$H_0 : \bar{x} = \mu_0$$

show a deep misunderstanding of the nature of both testing and random variables, in the line of the earlier confusion between samples and distributions. A similar confusion appears in the ANOVA chapter (e.g. formula (5.51) on page 112). And as follows:

*“The research goal is to find out if the treatment is effective or not; if it is not, the difference between the means should be ‘essentially’ equivalent.” (page 92)*

The following chapters cover analysis of variance (5), linear models (6), and linear mixed models (7), all of which face fatal foundational deficiencies, similar to the ones pointed above. I quite understand that the authors wrote the book in a praiseworthy goal to reach to less sophisticated audiences and to the best of their abilities, however I remain amazed that the book did not undergo a statistician’s review before being published. As is, it cannot deliver the expected outcome on its readers, i.e. cannot train them towards more sophisticated statistical analyses. As a non-expert on linguistics, I cannot judge of the requirements of the field and of the complexity of the statistical models it involves. And I acknowledge that linguists, esp. students, do not have a strong mathematical background. Nor do I know of any available and valuable alternative at this level. However, I maintain that even the most standard models and procedures should be treated with the appropriate statistical rigour. In conclusion, it unfortunately seems to me the book cannot endow its intended readers with the proper perspective on statistics.

### Further references

GELMAN, A. and HILL, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, Cambridge, UK.

## Anathem, by Neal Stephenson

- **Paperback:** 981 pages
- **Publisher:** William Morrow (first edition, 2008)

- **Language:** English
- **ISBN-10:** 0061694940

Published in 2008, *Anathem* is a wonderful book, especially for mathematicians, and while it qualifies as a science-fiction book, it blurs the frontiers between the genres of science-fiction, speculative fiction, documentary writings and epistemology. This explains why it was reviewed in *Nature*. In parallel, the book was awarded the 2009 Locus SF Award. And got top-rank best seller position in the New York Times. So *Anathem* has true sci-fi characteristics, including Arthur C. Clarke's bouts of space opera with a *Rama*-like vessel popping out of nowhere. But this is not the main feature that makes Stephenson's book so unique and fascinating, enough to deserve a review in CHANCE.

*"The Adrakhonic theorem, which stated that the square of a right triangle hypotenuse was equal to the sum of the squares of the other two sides..." (page 128)*

While the story is universal enough to appeal to all readers, witness the above-mentioned award, what I find most endearing about the book is the connection with mathematical thinking and the many ways mathematicians share characteristics with monks. The universe imagined by Stephenson segregates scientists and philosophers into convents, under strict rules that prevent any theoretical discovery to be turned into a technological application. This appears to have been imposed by the secular powers after scientific experiments on anti-matter, nuclear fusion or genetic engineering ran out of control. The final twist in the story is how the scientists manage to escape this prohibition while apparently adhering to it (and save the planet on the side).

*"Do you think it is the case that for every proof you and the other Edharians work out on a slate, the aliens have a proof in their own system that corresponds to it? That says the same thing expresses the same truth?" (page 314)*

Beside the attractive order of a monastic closed environment (and the appeal of being shut from the outside world!), one appeal of Stephenson's construct is the almost detective enquiry of the "mathematical monks" who end up solving fundamental quantum theory questions about parallel universes. Toying with all those theories is obviously another appeal of *Anathem*, as is fishing for well-known principles like Occam's razor, the travelling salesman

problem, cord theory, Einsteinian causality cone, Schrödinger’s cat, Platonic realism, and so on. (There are just too many quotes I could have included in this review!) Some may find the deliberate hiding of standard theory behind a novlangue a nuisance but this is rather light and one quickly gets used to it. (Until the aliens start detailing their parallel universes, at least.) Maybe the title should have been *Amathem* rather than *Anathem*, given that the monastic communities are called math, but this is a minor quibble. (It took me a while to realize that anathem was not an English word, as I had extrapolated the French *anathème* to its English-sounding counterpart!)

*“I was playing the Teglon. The objective of the game was to build the pattern outward from one vertex and pave the entire Decagon in such a way that the groove formed a continuous, unbroken curve from the first vertex to the last.” (page 552)*

Compared with earlier works of Stephenson like the wonderful cyberpunk novel *Snow Crash*, *Anathem* is closer to *The Baroque Cycle*, a novelised account of 17th and 18th centuries science, involving characters like Newton and Leibniz, and to the excellent *Cryptonomicon*. Both *Anathem* and *The Baroque Cycle* reflects on how deep is pondering on the nature of Science and the philosophy of scientific discovery. That Stephenson manages to turn those reflections into a lively and fascinating story—much more so here than in *The Baroque Cycle*—demonstrates how impressive an author he is. Even some readers may complain of lengthy passages, I think the book should greatly appeal to scientific minds for its wealth of considerations on mathematics, physics, and philosophy.

## Further references

ROHN, J. (2008). Imprisoned by intelligence. *Nature* **456**, 446–447.

STEPHENSON, N. (2000). *Snow Crash*. Bantam Spectra Book.

STEPHENSON, N. (2002). *Cryptonomicon*. Avon.

STEPHENSON, N. (2006). *Quicksilver (The Baroque Cycle No. 1)* Harper-Torch.