**Understanding Computational Bayesian Statistics**

William M. Boldstad

Hardcover: 336 pages  
Publisher: John Wiley, first edition (December 2009)  
Language: English  
ISBN-10: 0470046090

Understanding Computational Bayesian Statistics covers the basics of Monte Carlo and (fixed dimension) Markov Chain Monte Carlo methods, with a fair chunk dedicated to prerequisites in Bayesian statistics and Markov chain theory. Even though I have only glanced at the table of contents of Bolstad’s Introduction to Bayesian Statistics, the current book appears to be a continuation, going beyond the binomial, Poisson, and normal cases to cover generalized linear models via MCMC methods. (In this respect, it corresponds to Chapter 4 of Jean-Michel Marin’s and my Bayesian Core: A Practical Approach to Computational Bayesian Statistics.)

The book is associated with Minitab macros and an R package (written by James Curran), Bolstad2, in continuation of Bolstad, which was written for Introduction to Bayesian Statistics. Overall, the level of the book is such that it should be accessible to undergraduate students—MCMC methods being reduced to Gibbs, random walk, and independent Metropolis-Hastings algorithms and convergence assessments being done via autocorrelation graphs, the Gelman and Rubin (1992) intra-/inter-variance criterion, and a forward coupling device. The illustrative chapters cover logistic regression (Chapter 8), Poisson regression (Chapter 9), and normal hierarchical models (Chapter 10). Again, the overall feeling is that the book should be understood by undergraduate students, even though it may make MCMC seem easier than it is by sticking to fairly regular models. In a sense, it is more a book of the (roaring MCMC) 90s in that it does not incorporate advances from 2000 onward (as seen from the reference list) such as adaptive MCMC and the resurgence of importance sampling via particle systems and sequential Monte Carlo.

“Since we are uncertain about the true values of the parameters, in Bayesian statistics we will consider them to be random variables. This contrasts with the frequentist idea that the parameters are fixed but unknown constants.” (Page 3)

I find the book’s introduction to Bayesian statistics (Chapter 1) somehow unbalanced with statements such as the above and “Statisticians have long known that the Bayesian approach offered clear cut advantages over the frequentist approach,” (Page 1) which makes one wonder why there is any frequentist left, or “Clearly, the Bayesian approach is more straightforward [than the frequentist p-value],” (Page 53) because antagonistic presentations are likely to be lost to the neophyte. (I also disagree with the declaration that, for a Bayesian, there is no fixed value for the parameter.) The statement that the MAP estimator is associated with the 0–1 loss function (footnote 4, Page 10) is alas found in many books and papers, thus cannot truly be blamed on the author. That ancillary statistics “only work in exponential families” (footnote 5, Page 13) is either unclear or wrong. The discussion about Bayesian inference in the presence of nuisance parameters (pp. 15–16) is also confusing: “The Bayesian posterior density of $\theta$, found by marginalizing $\theta$, out of the joint posterior density, and the profile likelihood function of $\theta$ turn out to have the same shape” (Page 15) [under a flat prior] sounds wrong to me.

“It is not possible to do any inference about the parameter $\theta$ from the unscaled posterior.” (Page 25)

The chapter about simulation methods (Chapter 2) contains a mistake that one might deem of little importance. However, I do not and here it is: Sampling-important-resampling is presented as an exact simulation method (Page 34), omitting the bias due to normalizing the importance weights.

The chapter on conjugate priors (Chapter 4), although fine, feels as if it does not belong to this book, but should rather be in Bolstad’s Introduction to Bayesian Statistics, especially as it is on the long side. Chapter 5 gives an introduction to Markov chain theory in the finite state case, with a nice illustration on the differences in convergence time through two $5 \times 5$ matrices. (But why do we need six decimals?!)  

“MCMC methods are more efficient than the direct [simulation] procedures for drawing samples from the posterior when we have a large number of parameters.” (Page 127)

MCMC methods are presented through two chapters, the second titled “Statistical Inference from a Markov Chain Monte Carlo Sample” (Chapter 7), which is a neat idea to cover the analysis of an MCMC output. The presentation is mainly one-dimensional, which makes the recommendation to use independent Metropolis-Hastings algorithms found throughout the book [using a $t$ proposal based on curvature at the mode] more understandable, if misguided. The
presentation of the blockwise Metropolis-Hastings algorithm of Hastings through the formula (Page 145)
\[ P(\theta, A) = \prod_{j=1}^{J} P_j(\theta_j, A|\theta_{-j}) \]
is a bit confusing, as the update of the conditioners in the conditional kernels is not indicated. (The following algorithm is correct, though.)

I also disliked the notion that “the sequence of draws from the chain (…) is not a random sample” (Page 161) because of the correlation: The draws are random, if not independent. This relates to the recommendation of using heavy thin-in with a gap that “should be the same as the burn-in time” (Page 169), which sounds like a waste of simulation power, as burn-in and thin-in of a Markov chain are different features. The author disagrees with the [my] viewpoint that keeping all the draws in the estimates improves on the precision: “One school considers that you should use all draws (…) However, it is not clear how good this estimate would be” (Page 168) and “values that were thinned out wouldn’t be adding very much to the precision” (Page 169).

Inevitably (trust me!), there are typing mistakes in the book and they will most likely be corrected in a future printing/edition. I am, however, puzzled by the high number of “the the” or the misspelling (Page 261) of Jeffreys’ prior into Jeffrey’s prior (maybe a mistake from the copy editor). (A few normal densities are missing a 1/2 on Page 247, by the way.)

As a final note, let me point out that Bolstad replied to this review on my blog on October 24, 2011, in a fairly detailed way.

Further Reading

Gelman, A., and D. Rubin. 1992. Inference from iterative simulation using multiple sequences (with discussion). Statist. Science 7:457–511.

Møller, J., and R. Waagepetersen. 2003. Statistical inference and simulation for spatial point processes. Chapman and Hall/CRC: Boca Raton.

Robert, C., and G. Casella. 2004. Monte Carlo statistical methods, 2nd ed. Springer-Verlag: New York.

Here is another Bayesian textbook that appeared recently. I read it within a few days and, despite my obvious biases, I liked it very much! It has a

Bayesian Ideas and Data Analysis

Ronald Christensen, Wesley Johnson, Adam Branscum, and Timothy Hanson

Hardcover: 516 pages
Publisher: CRC Press, first edition (June 2010)
Language: English
ISBN-10: 1439803544

Here is another Bayesian textbook that appeared recently. I read it within a few days and, despite my obvious biases, I liked it very much! It has a