HAL
open science

Book Reviews (Spring 2012)

Alessandra Iacobucci, Christian Robert

To cite this version:

Alessandra Iacobucci, Christian Robert. Book Reviews (Spring 2012). CHANCE, 2012, 25 (2), pp.57-61. 10.1080/09332480.2012.685375 . hal-00664781

HAL Id: hal-00664781
https://hal.science/hal-00664781v1

Submitted on 31 Jan 2012

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, Spring 2012

Alessandra Iacobucci[1,2] and Christian P. Robert[1,3,4]
[1]Université Paris-Dauphine, CEREMADE, [2]CNRS, [3]IUF, and [4]CREST
iacob,xian@ceremade.dauphine.fr

January 31, 2012

Abstract

This note is made of a review of the R book by Matloff (2011) and of three Bayesian book reviews of Bolstad (2009), Christensen et al. (2010), and Nzoufras (2009) respectively. They are scheduled to appear in the next issue of *CHANCE*.

The first of those reviews is written by Alessandra Iacobucci, while the three last reviews are written by Christian P. Robert.

The Art of R Programming: A Tour of Statistical Software Design, by Norman Matloff

- **Paperback:** xxiv + 373 pages
- **Publisher:** No Starch Press, San Francisco
- **Language:** English
- **ISBN-10:** 1593273843

As Rob J. Hyndman enthusiastically declares on his blog[1], "this is a gem of a book". I would go even further and argue that *The Art of R programming* is a whole mine of gems.

The book is well constructed, and has a very coherent structure. After an introductory chapter, where the reader gets a quick overview on `R` basics that allows her to work through the examples in the following chapters, the rest of the book can be divided in three main parts. In the first part

[1] `http://robjhyndman.com/researchtips/matloff/`, Nov. 2011.

(Chapters 2 to 6) the reader is introduced to main `R` objects and to the functions built to handle and operate on each of them. The second part (Chapters 7 to 13) is focussed on general programming issues: `R` structures and object-oriented nature, I/O, string handling and manipulating issues, and graphics. Chapter 13 is all devoted to the topic of debugging. The third part deals with more advanced topics, such as speed of execution and performance issues (Chapter 14), mix-matching functions written in `R` and `C` (or `Python`), and parallel processing with `R`.

Even though this last part is intended for more experienced programmers, the overall programming skills of the intended reader "may range anywhere from those of a professional software developer to 'I took a programming course in college'." (p. xxii).

With a fluent style, Matloff is able to deal with a large number of topics in a relatively limited number of pages, resulting in an astonishingly complete yet handy guide. At almost every page we discover a new command, most likely *the* command we had always looked for and done without by means of more or less cumbersome roundabouts.

As a matter of fact, it is possible that there exists a ready-made and perfectly suited `R` function for nearly anything that comes up to one's mind. Users coming from compiled programming languages may find it difficult to get used to this wealth of functions, just as they may feel uncomfortable not declaring variable types, not initializing vectors and arrays, or getting rid of loops. Nevertheless, through numerous examples and a precise knowledge of its strengths and limitations, Matloff masterly introduces the reader to the flexibility of `R`. He repeatedly underlines the functional nature of `R` in every part of the book and stresses from the outset how this feature has to be exploited for an effective programming.

> *"One of the most effective ways to achieve speed in* `R` *code is to use operations that are* vectorized*, meaning that a function applied to a vector is actually applied individually to each element." (p. 40).*

The result is so convincing that it pushes even the strictest code purist to free herself from prejudices and surrender to the pleasures of an interpreted language. This probably was the hardest challenge in writing *The Art of R Programming*, and the author brilliantly met it.

The climax is unquestionably attained in the final chapters, where Matloff introduces some advanced and unusual topics with remarkable clarity and briskness. Within a few pages, he manages to tackle the object-oriented side of `R`, to advise and instruct the reader on debugging and performance issues, to show how to deal with `R` and `C` (or `Python`) mixed codes, and

finally to open new perspectives by presenting the different approaches to parallel R. There is even a mention of GPU programming, a short paragraph certainly inexhaustive, but still instructive. To my knowledge, this is the only `R` handbook in which parallel programming with R is tackled with some degree of detail.[2] Also, the importance and prominence given to debugging are commendable, since this topic is often and mistakenly disregarded in most programming handbooks—except those explicitly written on the subject.

Among the sharpest passages of the book, I definitely include the ones on scope and environment issues, to which are devoted both a long section in Chapter 17 and a tiny simple yet enlightening example as early as page 9.

> *"Note carefully the role of* `w`*. The* `R` *interpreter found that there was no local variable of that name, so it ascended to the next level [...] where it found a variable* `w` *with value 12. [...]. It is possible (though not desirable) to deliberately allow name conflicts in this hierarchy. [...] In such a situation the innermost environment is used first." (p. 153).*

The message is clear: know exactly what you want to implement, keep track of all your objects, and scoping will not be an issue but another tool.

> *"In* `C`*, we would not have functions defined within functions [...]. Yet, since functions are objects, it is possible—and sometimes desirable from the point of view of the encapsulation goal of object-oriented programming—to define a function within a function; we are simply creating an object, which we can do anywhere." (p. 152-3).*

Another little gem is Section 7.9 on recursion, a concept that Matloff presents in a very clear and intuitive way. This section ends with one the most inspired extended examples proposed in the book, where recursion is used to implement a binary search tree. Other interesting extended examples are those about discrete-event simulation (§ 7.8.3), Markov chains (§ 8.4.2) and polynomial regression (§ 9.1.7), though these applications may be a little too challenging for readers lacking a solid background in statistics.

Although *The Art of R Programming* is a book of many virtues, there are in my opinion some flaws.

The presence of raw `R` code starting from the first few pages encourages the user to test her understanding straight away while reading, making *The*

[2] I only found a hint of it in Alder (2010), yet no programming details are given therein.

Art of R Programming a sort of plug-and-play guide through `R`. Unfortunately, the pleasure of real-time testing is spoiled by two things. First, the reader has to copy those codes line by line. This is unquestionably useful for the many simple examples scattered throughout the book. However, it may represent an inexhaustible source of typos, both pointless and annoying—not to mention time-consuming—when it comes to more complicated programs like those expounded in the many Extended Example sections. Second, the databases are unavailable so some applications are simply unusable.[3] I am referring here to virtually all the extended examples in Chapters 5 and 6 on data frames, factors and tables. In particular, I find the application on the aids for learning Chinese dialect (§ 5.4.3) so over-elaborate to be nearly worthless. I would certainly suggest designing a dedicated `R` package assembling all the necessary material for a fully profitable training with the book, like the package `mcsm` [4] conceived by Robert and Casella for reproducing the results contained in their book (2010) on Monte Carlo methods with `R`.

In addition, `R` can certainly handle huge databases with great ease, and maybe I am giving way to my personal preferences here, but I find that two whole chapters on data frames and factors (adding up to almost 40 pages!) are perhaps too much. On the contrary, I believe that the "traditional" graphic package should have deserved more space and consideration, not only in the devoted chapter (Chap. 12), but generally throughout the book. Indeed, the author suggests some good handbooks on the subject by Murrel (2011) and Wickham (2010), but these are too detailed and advanced to be used for general purposes.

Despite an overall concise style, there are some long-winded passage and repetitions, especially in the applications, where certain lines of code are definitely redundant. I was likewise puzzled by the total absence in the book of the command separator `;`, which would have considerably shortened and lightened some unnecessarily long examples. Also, a separate and more detailed index of `R` commands and functions would be helpful.

Finally, a minor but curious point about the assignment operator. I find the issue of `<-` vs. `=` particularly fascinating—and a bit perturbing, since this leaves in fact an ambiguity in the definition of such a fundamental operator. Still, there seem to be two main streams and no general agreement. Reading on various blogs and discussion forums, I found no decisive nor robust argument in favor of either. Matloff approaches the issue of `<-` vs.

[3] I managed to find the abalone data set for extended examples of § 2.9.2 and § 4.4.3 at `http://archive.ics.uci.edu/ml/datasets/Abalone` thus discovering this interesting repository, but for the rest my research was rather inconclusive.

[4] cran.r-project.org/web/packages/mcsm/

= in assignments as soon as page 4. As stated in the book, "The standard assignment operator in R is <-. You can also use =, but this is discouraged, as it does not work in some special situations.". I was really eager to see these "special situations" shown in concrete examples. Unfortunately, they are nowhere to be listed in the book.

Notwithstanding these minor defaults, *The Art of R programming* is enriching, enjoyable and definitely worthwhile keeping as a reference while working with R. I highly recommend it to programmers, academic researchers and students in computational statistics willing to be quickly operational in writing R software. And it is undoubtedly a really useful reading for any R user.

Further references

Adler, J. (2010), *R in a Nutshell*, O'Reilly, New York.

Robert, C. P. and G. Casella (2010), *Introducing Monte Carlo Methods with R*, Springer-Verlag New York.

Murrel, P. (2011), *R Graphics*, 2nd ed., Chapman & Hall/CRC Press, New York.

Wickham, H. (2010), *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed., Springer-Verlag New York.

Warning! Before I launch into the following reviews of three Bayesian books—that I first read as a member of the DeGroot Prize committee—, let me warn the reader that, as an author and co-author of four books on Bayesian statistics and computational methods, my views on the books cannot but be subjective: I do favour the way we approached Bayesian and computational methods and, after reading the following books, I would still have written our books the way we did.

Understanding computational Bayesian statistics by William Boldstad

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

- **ISBN-10:** 0470046090

Understanding computational Bayesian statistics is covering 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* and not at the book itself, the current book appears to be a continuation of the earlier one, going beyond the Binomial, Poisson, and normal cases, to cover generalised linear models, via MCMC methods. (In this respect, it corresponds to Chapter 4 of our *Bayesian Core*.) The book is associated with Minitab macros and an R package (written by James Curran), Bolstad2, in continuation of Bolstad, 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 (Chap. 8), Poisson regression (Chap. 9), and normal hierarchical models (Chap. 10). Again, the overall feeling is that the book should be understandable to 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] 90's in that it does not incorporate advances from 2000 onwards (as seen from the reference list) like 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)*

To get into more details, I find the book introduction to Bayesian statistics (Chap. 1) somehow unbalanced with statements like the above and like "statisticians have long known that the Bayesian approach offered clear cut advantages over the frequentist approach" (p.1) [which makes one wonder why there is any frequentist left!], or "clearly, the Bayesian approach is more straightforward [than the frequentist p-value]" (p.53). because antagonistic presentations are likely to be lost to the neophyte. (I also disagree with the statement 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, p.10) is alas found in many books and papers, thus cannot truly be blamed on the author. The statement that ancillary statistics "only work in exponential families" (footnote 5, p.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 θ_1 found by marginalizing θ_2 out of the joint posterior density, and the profile likelihood function of θ_1 turn out to have the same shape" (p.15) [under a flat prior] sounds wrong to me.

> *"It is not possible to do any inference about the parameter θ from the unscaled posterior." (page 25)*

The chapter about simulation methods (Chap. 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 (p.34), omitting the bias due to normalising the importance weights.

The chapter on conjugate priors (Chap. 4), although fine, feels as if it does not belong to this book but should rather be in the previous Bolstad's *Introduction to Bayesian Statistics*. Esp. as it is on the long side. The following Chap. 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 5×5 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 one being entitled "Statistical inference from a Markov chain Monte Carlo sample" (Chap. 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 (p.145)

$$P(\theta, A) = \prod_{j=1}^{J} P_j(\theta_j, A_j|\theta_{-j})$$

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" (p.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" (p.169), which sounds like a waste of simulation power, as burn-in and thin-in of a Markov chain are two different features. The author disagrees with the [my] viewpoint that keeping all the draws in the estimates improves on the precision: e.g., "one school considers that you should use all draws (...) However, it is not clear how good this estimate would be" (p.168) and "values that were thinned out wouldn't be adding very much to the precision" (p.169). I did not see any mention made of effective sample size and the burn-in size is graphically determined via autocorrelation graphs, Gelman-Rubin statistics, and a rather fantasist use of coupling from the past (pp.172-174). (In fact, the criterion is a forward coupling device that only works for independent chains. See Møller and Waagepetersen, 2003 and Robert and Casella (2004) for expositions on coupling from the past, as well as my vignette on perfect simulation in the previous issue of *CHANCE*.)

> *"We should alway use proper priors in the hierarchical model, particularly for scale parameters. When improper priors are used (...) overall the posterior is improper." (page 257)*

The final chapters apply MCMC methods to logistic (Chap. 8) and Poisson regressions (Chap. 9), again using an independent proposal in the Metropolis-Hastings algorithm. (Actually, we also used a proposal based on the MLE solutions for the logistic regression in *Introducing Monte Carlo Methods with R*, however it was in an importance sampling illustration for Chapter 4.) It is a nice introduction to handling generalised linear models with MCMC. The processing of the selection of variables (p.195-198 and pp.224-226) could have been done in a more rigorous manner, had Bayes factors been introduced. It is also a nice idea to conclude with Gibbs sampling applied to hierarchical models (Chap. 10), a feature missing in the first edition of our *Bayesian Core*, however the chapter crucially misses an advanced example, like mixed linear models. This chapter covers the possibly misbehaviour of posteriors associated with improper priors, with a bit too strong of a warning (see above), and it also unnecessarily [in my opinion] goes into a short description of the empirical Bayes approach (pp.245-247).

The style of *Understanding computational Bayesian statistics* is repetitive at times, sentences from early paragraphs of a chapter being reproduced verbatim a few pages later. While the idea of opposing likelihood-based inference to Bayesian inference by an illustration through a dozen graphs

(Chap. 1) is praiseworthy, I fear the impact is weakened by the poor 3-D readability of the graphs. Another praiseworthy idea is the inclusion of a "Main points" section at the end of each chapter; however, they should have been more focused in my opinion. Maybe the itemized presentation did not help.

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 (p.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 p.247, by the way.)

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

Further references

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

Møller, J. and Waagepetersen, R. (2003). *Statistical Inference and Simulation for Spatial Point Processes.* Chapman and Hall/CRC, Boca Raton.

Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods.* 2nd ed. Springer-Verlag, New York.

Bayesian ideas and data analysis by 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 and prejudices, I liked it very much! It has a lot in common (at least in spirit) with our *Bayesian*

Core, which may explain why I feel so benevolent towards *Bayesian ideas and data analysis*. Just like ours, the book by Christensen, Johnson, Branscum, and Hanson is indeed focused on explaining the Bayesian ideas through (real) examples and it covers a lot of regression models, all the way to non-parametrics. It contains a good proportion of WinBugs and R codes. It intermingles methodology and computational chapters in the first part, before moving to the serious business of analysing more and more complex regression models. Exercises appear throughout the text rather than at the end of the chapters. As the volume of their book is more important (over 500 pages), the authors spend more time on analysing various datasets for each chapter and, more importantly, provide a rather unique entry on prior assessment and construction. Especially in the regression chapters. The author index is rather original in that it links the authors with more than one entry to the topics they are connected with (Ron Christensen winning the game with the highest number of entries).

> *"Although the prior may work well in the sense that it is easily overwhelmed by the data, on should never forget that in itself it is saying very stupid things, namely that θ is likely to be either huge or essentially 0." (page 71)*

The book is pleasant to read, with humorous comments here and there. (I appreciated Ron's dedication to the South Island and to Kaikoura, New Zealand. But I missed the line on Monte Crisco and had to check on the Web. Even though I grasped the one on 'my niece found a nice niche in Nice'.) The presentation is dense but uses enough codes and graphs to make the going smooth. The sections on testing are presenting a wide range of options, which is not the way I would do it, but fine nonetheless. The authors even expose Neyman-Pearson testing to highlight the distinction with Bayesian approaches. The going gets a bit rough (in terms of measure theory, see page 56) for point null hypotheses, but the authors manage to get the idea clarified through examples. Model checking is proposed via Bayesian p-values

$$\mathbb{P}(m(X) \leq m(x_{\text{obs}}))$$

[which has the drawback of not being invariant by reparameterisation], predictive p-values, Bayes factors, BIC [not a Bayesian criterion!], DIC, and the authors' favourite, the pseudo-marginal likelihood (page 81)

$$\hat{m}(x) = \prod_{i=1}^{n} f_i(x_i|x_{-i}) \tag{1}$$

where the components of the product are the cross-validation predictive densities. This pseudo-marginal likelihood allows for improper priors and, like Aitkin's (2010) integrated likelihood, it is not a Bayesian procedure in that the data is used several times to construct the procedure. In addition, the authors recommend the worst version of Gelfand and Dey's (1994) estimate to approximate these cross-validation predictive densities, which indeed amount to using the dreadful harmonic mean estimate!!! (Chapter 4 is actually the central chapter of the book in my opinion and I could make many more comments on how I would have presented things. Like exchangeability, sufficiency [missing the point on Jeffreys as if they were true priors], a very artificial example of inconsistent Bayes estimators [already discussed there], identifiability [imposed rather than ignored: "Our problem with Bayesian analysis is that it is easy to overlook identifiability issues", p.96].)

Do not try harmonic mean integration

When trying to approximate a marginal density

$$m(x) = \int_\Theta \pi(\theta) f(x|\theta) \mathrm{d}\theta$$

in Bayesian model assessment, the convenient identity

$$\int_\Theta \frac{1}{f(x|\theta)} \pi(\theta|x) \mathrm{d}\theta = \int_\Theta \frac{\pi(\theta)}{m(x)} \mathrm{d}\theta = \frac{1}{m(x)}$$

implies that a sample from the posterior distribution $\pi(\theta|x)$ can be put to direct use to approximate the marginal density in the form of an *harmonic mean*:

$$\hat{m}(x) = 1 \Big/ \frac{1}{N} \sum_{i=1}^{N} 1 \Big/ f(x|\theta_i)\,.$$

This is a simple representation and it does not require further simulations. Unfortunately, it too often suffers from the ultimate defect that the estimator has no variance, hence cannot be relied upon as an approximation. An illuminating example in the normal/normal case is provided on Radford Neal's blog, under the title "The Harmonic Mean of the Likelihood: Worst Monte Carlo Method Ever". Extensions to the above identity can be found that avoid the infinite variance pitfall (Marin and Robert, 2010), however they require additional simulations.

> *"Even though the likelihood function has a place of honor in both frequentist and Bayesian statistics it is a rather artificial construction. If you accept that parameters are artificial constructs, then likelihoods must also be artificial constructs." (page 93)*

Once again, I like very much the second part on regression models. Even though I missed Zellner's g-prior. (Some of the graphs are plain ugly: Fig. 5.4 and Fig. 15.7, for instance.) I do prefer their coverage of MCMC (Chap. 6) to Bill Bolstad's, esp. when the authors argue that they "don't believe that thinning is worthwhile" (p.146). (Gibbs sampling is however missing the positivity constraint of Hammersley and Clifford [and Besag's].) And, again, having whole sections on the prior construction is a very neat thing.

Last but not least, the picture on the backcover of the book is one of Pierre Simon Laplace himself! First, Laplace did much more [than Bayes] for the birth of Bayesian statistics. Second, this avoids us one more replication of the likely apocryphal picture of Thomas Bayes. Great!

Further references

Aitkin, M. (2010). *Statistical Inference: A Bayesian/Likelihood approach.* CRC Press, Chapman & Hall, New York.

Gelfand, A. and Dey, D. (1994). Bayesian model choice: asymptotics and exact calculations. *J. Royal Statist. Society Series B*, **56** 501–514.

Marin, J. and Robert, C. (2010). Importance sampling methods for Bayesian discrimination between embedded models. In *Frontiers of Statistical Decision Making and Bayesian Analysis* (M.-H. Chen, D. Dey, P. Müller, D. Sun and K. Ye, eds.). Springer-Verlag, New York. Chap. 14, 513-553.

Bayesian modeling using WinBUGS by Ioannis Ntzoufras

- **Hardcover:** 520 pages
- **Publisher:** John Wiley, first edition (February 2009)
- **Language:** English
- **ISBN-10:** 047014114X

Yes, yet another Bayesian textbook: Ioannis Ntzoufras' *Bayesian modeling using WinBUGS* was published in 2009 and it got an honourable mention at the 2009 PROSE Award. (Nice acronym for a book award! All the mathematics books awarded that year were actually statistics books.) (Note this book is not to be confused with the very recent *Bayesian population modeling using WinBUGS*, by Marc Kéry and Michael Schaub, which is about Bayesian analysis for ecology.)

> *"As history has proved, the main reason why Bayesian theory was unable to establish a foothold as a well accepted quantitative approach for data analysis was the intractability involved in the calculation of the posterior distribution." (page 1)*

The book launches into a very quick introduction to Bayesian analysis, since, by page 15, we are "done" with linear regression and conjugate priors. This is somehow softened by the inclusion at the end of the chapter of a few examples, including one on the Greek football team in Euro 2004, but nothing comparable with Christensen et al.'s initial chapter of motivating examples. Chapter 2 on MCMC methods follows the same pattern: a quick and dense introduction in about ten pages, followed by 40 pages of illuminating examples, worked out in full detail. CODA is described in an Appendix. Compared with Bayesian ideas and data analysis, Bayesian modeling using WinBUGS spends time introducing WinBUGS and Chapter 3 acts like a 20 page user manual, while Chapter 4 corresponds to the WinBUGS example manual. Chapter 5 gets back to a more statistical aspect, the processing of regression models (including Zellner's g-prior). up to ANOVA. Chapter 6 extends the previous chapter to categorical variables and the ANCOVA model, as well as the 2006-2007 English premier league. Chapter 7 moves to the standard generalised linear models, with an extension in Chapter 8 to count data, zero inflated models, and survival data. Chapter 9 covers hierarchical models, with mixed models, longitudinal data, and the water polo World Cup 2000.

> *"Although this [the harmonic mean] estimator is simple, it is quite unstable and sensitive to small likelihood values and hence is not recommended." (page 393)*

While most chapters rely on DIC for model comparison, the last two chapters of Bayesian modeling using WinBUGS open on other model comparison approaches like the posterior predictive p-value, residual values, cross-validation (with, once again!, the dreaded harmonic mean estimator! and, once again, Geisser and Eddy's conditional predictive ordinates),

keeping the introduction of Bayes factors for Chapter 11, with an immediate criticism through the Jeffreys-Lindley-Bartlett paradox, maybe because "Bayes factors cannot be generally calculated within WinBUGS unless sophisticated approaches are used" (p.390). Surprisingly, and as clearly stated in the above quote, the computational section warns about the poor performances of the harmonic mean estimator without making the connection with the earlier proposal of the very same estimator (p.375). After reviewing the most standard approaches for marginal approximation, Ntzoufras falls back on a Laplace approximation to the likelihood function. This chapter also covers variable selection by Gibbs sampling, stochastic search, the Carlin and Chib (1995) method and reversible jump MCMC, the later being summarily expedited in half a page! It concludes with the (non-Bayesian) information criteria, AIC and BIC.

> *"Bayesian statistics suddenly became fashionable, opening new highways for statistical research." (page 2)*

On the material (!) side, while the presentation is overall very nice, I dislike the fonts (which are imposed by John Wiley, as I seem to remember) and the fact that the text within a page seems to have slided down to the bottom: I mean, each page of text ends up (or down) very close to the physical bottom of the page. Nothing important, obviously, but a slight impression of cramming. (See, e.g., pages 3 or 38, where any additional subscript would have been sticking out of the book!) A further nitpicking remark is that the examples start as indented and then loose their indentation after a paragraph or two, which does not help in identifying examples as a whole within the text. I like the idea of highlighting R/WinBUGS code with grey background (as we did in *Bayesian Core*), however the rendering of the two-column opposition of algorithm and R code is unfortunately difficult to read. Some graphs are given as screen copies, which reduce their readability for no proper reason.

On the "sin" of using the data twice

Several aspects of the books covered in this review relate to the problem of "using the data twice". What does that mean? Nothing really precise, actually. The accusation of "using the data twice" found in the Bayesian literature can be thrown at most procedures exploiting the Bayesian machinery without actually being Bayesian, i.e. which cannot be derived from the posterior distribution. For instance, the integrated likelihood approach of Aitkin (2010) avoids the

difficulties related with improper priors π_i by *first* using the data x to construct (proper) posteriors $\pi_i(\theta_i|x)$ and then *secondly* using the data in a Bayes factor

$$\int_{\Theta_1} f_1(x|\theta_1)\pi_1(\theta_1|x)\,\mathrm{d}\theta_1 \Big/ \int_{\Theta_2} f_2(x|\theta_2)\pi_2(\theta_2|x)\,\mathrm{d}\theta_2$$

as if the posteriors were priors. This obviously solves the improperty difficulty (Robert, 2001), but it creates a statistical procedure outside the Bayesian domain, hence requiring a separate validation since the usual properties of Bayesian procedures do not apply. Similarly, the whole *empirical Bayes* approach falls under this category, even though some empirical Bayes procedures are asymptotically convergent. The pseudo-marginal likelihood of Geisser and Eddy (1979) is defined by (1). While it also allows for improper priors, it does use the same data in each term of the product and, again, it is not a Bayesian procedure.

Once again, from first principles, a Bayesian approach should use the data only once, namely when constructing the posterior distribution on every unknown component of the model(s). Based on this all-encompassing posterior, all inferential aspects should be the consequences of a sequence of decision-theoretic steps in order to select optimal procedures. This is the ideal setting while, in practice, relying on a sequence of posterior distributions is often necessary, each posterior being a consequence of earlier decisions, which makes it the result of a multiple use of the data? For instance, the process of Bayesian variable selection is on principle clean from the sin of ?using the data twice?: one simply computes the posterior probability of each of the variable subsets and this is over. However, in a case involving many (many) variables, there are two difficulties: one is about building the prior distributions for all possible models, a task that needs to be automatised to some extent; another is about exploring the set of potential models. Resorting to projection priors as in the intrinsic solution of Pérez and Berger (2002, Biometrika, a much valuable article!), while unavoidable and a "least worst" solution, means switching priors/posteriors based on earlier acceptances/rejections, i.e. on the data. Second, the path of models truly explored by a computational algorithm—a minuscule subset of the set of all models—will depend on the models rejected so far, either when relying on a stepwise exploration or when using a random walk MCMC algorithm. Although this is not crystal clear (there is actually plenty of threads for supporting the opposite view!), it could be argued that the data is thus used several times in this process.

"All predictive diagnostics presented above have the disadvantage of

double usage of the data." (page 375)

In conclusion, and in reflection with their respective titles, *Bayesian modeling using WinBUGS* feels more technical than *Bayesian ideas and data analysis*, even though their coverage is *in fine* very similar. Not only do they both insist on methodology much more than theory, but they also similarly emphasize the application aspect through numerous examples based on real data. The later is slightly more philosophical and for this reason (as well as typographical comfort) more to my own (personal and subjective) taste. I figure the choice of one versus the other as a textbook will very much depend on the intended audience. More mature statistical students may favour *Bayesian ideas and data analysis*, while more applied students could benefit more from *Bayesian modeling using WinBUGS*.

Further references

Aitkin, M. (2010). *Statistical Inference: A Bayesian/Likelihood approach.* CRC Press, Chapman & Hall, New York.

Carlin, B. and Chib, S. (1995). Bayesian model choice through Markov chain Monte Carlo. *J. Royal Statist. Society Series B*, **57** 473–484.

Geisser, S. and Eddy, W. (1979). A predictive approach to mdoel selection. *J. American Statist. Assoc.* 74: 153–160.

Robert, C. (2001). *The Bayesian Choice* (second edition). Springer-Verlag, New York.