Journal of Mathematical Psychology 56 (2012) 64–66

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Book review

Doing Bayesian data analysis in the classroom: An experience based review of John K. Kruschke's (2011) "Doing Bayesian Data Analysis: A Tutorial with R and BUGS".

Psychology students in the research track at the University of Leuven (Belgium) follow six courses in statistics, conveniently named statistics I to VI. Statistics I up till V introduce the students to descriptive statistics and to inferential statistics from the frequentist perspective (t-tests, ANOVA, regression, multilevel modeling, and some factor analysis thrown in along the way for good measure). As Belgian psychology students are not different from their colleagues in the rest of the world, these courses are a pain for most. So it was with some devilish pleasure that we decided to have the final course make the students call into question everything they painstakingly had mastered. Statistics VI would introduce the students to a new perspective: Bayes.

We had quite stringent desiderata for a textbook introducing Bayesian statistics. First, the book should have the right depth (i.e., at an advanced master level in psychology) and scope (i.e., covering most of the techniques the students encountered in their previous statistics courses, but now from a Bayesian perspective). Second, it should require little or no mathematical and programming background. Third, it should have exercises, preferably with a solution manual. Finally, it should have a practical orientation, implying that user-friendly code should be available, preferably in R (which, unlike Matlab, is free and open source).

That's a lot to ask for. Much to our own surprise, we found a book that seemed to match all of these criteria: "Doing Bayesian Data Analysis: A Tutorial with R and BUGS" by John K. Kruschke. After having used Kruschke's book in our course last semester, it is fair to conclude that it lived up to our quite demanding expectations. We therefore fully agree with Smithson (2011), who in his review praised the book as filling a major gap in introductory textbooks on Bayesian statistics. The aim of our review is to share our experiences of having used the book in the classroom to teach the students how to do Bayesian data analysis.

Doing Bayesian Data Analysis is aimed towards the audience we have in the classroom, and starts from scratch. The first part (chapters 2–4) reviews some basic but necessary aspects of probability theory (e.g., what is a probability density function?) and devotes a whole chapter to Bayes' theorem. These chapters provide a very useful brush-up of the students' undergraduate statistics courses. The second part (chapters 5–13) introduces the main concepts on which Bayesian statistics is built, by means of the articulate and detailed Bayesian analysis of a coin tossing problem. It is a pedagogically wise choice and an impressive accomplishment that all concepts and techniques, including MCMC, are presented in the context of such a simple setting. The third and final part of the book (chapters 14–22) applies this freshly gained knowledge to the Generalized Linear Model framework, including ANOVA and regression. This unifying framework is a clear asset of the book because it allows the students to connect Bayesian statistics to what was covered in their previous statistics courses. Chapters 2 and 7 contain gentle introductions to R and BUGS, respectively, targeted at first time users. From chapter 2 onwards, most figures and analyses can be replicated using the extensive and carefully commented code. Further, all chapters end with a set of enlightening exercises, for which a complete solution manual is available, including relevant R and BUGS code.

In several respects, Doing Bayesian Data Analysis even exceeded our already high expectations. Overall, the book is artfully structured and very well written. All chapters, except maybe the one on hierarchical models, excel in clarity. Most topics are presented patiently and elaborately, with a special emphasis on developing the right intuitions. Further, Kruschke's writing style has a contagious enthusiasm that is seldom seen in statistics textbooks. Add the plenty of detailed and relevant graphs and you end up with a book that is highly accessible. Perhaps most importantly, the students, when asked for their opinion at the end of the course, expressed their appreciation for the book's clarity and accessibility. On a 7 point scale ranging from 1 for pathetic to 7 for excellent, the average score given by the students ($n = 10$) was 6.

An additional strength of the book is its compelling demonstration of one of the big advantages the Bayesian approach has compared to the frequentist one. Unlike the latter approach, which critically relies on exact or asymptotic distributions for test statistics, the Bayesian framework grants considerable flexibility in model building. This refreshing and much needed freedom is neatly illustrated in the book by the use of t-distributions instead of the ubiquitous normal ones, at various levels in the model. For instance, for continuous data, one could use the t-distribution as a model for the data or, in a hierarchical model, one could use the t as a shrinkage distribution.

The book also addresses recent approaches, such as transdimensional MCMC to perform model comparison (chapter 10). This approach requires users to think about model indicators as binary parameters (with a distribution) and pseudo-priors, which is a radical departure from how students are used to think about model comparison (e.g., making use of F-tests). Despite their novelty for the students, Kruschke admirably succeeds in explaining these concepts.

The book closes with a section on how to report a Bayesian analysis (chapter 23). Given the lack of established standards in psychological research on reporting a Bayesian analysis, these guidelines are extremely useful—not necessarily as a canon, but definitely as a stimulating starting point.

Despite these strengths, we felt that two issues deserve more attention. First, default or reference priors, such as the Jeffreys prior, the Berger–Bernardo prior or the g prior are largely neglected. An in depth treatment of this complicated topic is probably not needed for an introductory textbook, but providing a brief discussion of this important research area and directing interested readers to the appropriate references seems essential for any textbook on Bayesian statistics.

0022-2496/$ – see front matter
doi:10.1016/j.jmp.2011.12.001

Another underexposed topic relates to Kruschke's stance in the nowadays schism that seems to be present in current Bayesian statistics (in particular in psychology, but also other fields). On the one hand, there are researchers who attempt to solve their inferential problem by estimating model parameters; and on the other hand there are those who aim to choose the best model from a finite set of possible models. Kruschke, belonging to the first group, somewhat disregards model checking and model comparison. For example, a posterior predictive check (PPC) is only recommended as being optional in a Bayesian analysis. Of course, one can have methodological worries about the PPC, but our view is that if a researcher, or his audience, does not know that the model fits the data reasonably well the parameter estimates are not meaningful at all. Further, the ability of Bayesian model comparison to balance goodness-of-fit with complexity is only briefly mentioned. Few readers will get this feature based on the book alone, which is unfortunate, since the automatic implementation of an Ockham's razor is one of the key strengths of the Bayesian method.

Just like Kruschke, we wanted the students to learn how to actually *do* a Bayesian analysis. After their graduation, most students will need to be able to apply statistics rather than to develop foundational theory, rendering practical skills more valuable than deep theoretical insights. Hence, the focus in our course was less working out gritty theoretical details (e.g., de Finetti's theorem), and more rolling up sleeves and analyzing data in a Bayesian way. Thus, we set up our course to have as few theoretical lectures as possible, leaving room for plenty hands on experience. In particular, the first two parts of the book were covered in a few theoretical lectures. These concepts were further explored in three computer lab sessions, in which the students worked on the exercises from the covered chapters.

After these sessions, the students were asked to conduct a Bayesian data analysis themselves. Unlike the data sets used in most other statistics courses, the data sets for this analysis were not streamlined, trimmed and preprocessed for analysis. Rather, the students were expected to analyze data they gathered themselves for their master thesis. They could use any of the methods from the third part of the book, or a combination thereof. These chapters were not covered in class. Instead, each student was required to study at least one of these chapters on their own in order to successfully apply them to the data analysis. Whenever they encountered technical or conceptual issues during their project, the students could seek advice from us or from the teaching assistant.

Doing a Bayesian analysis seemed daunting to most of the students. Additionally, getting up to speed with two new languages – `R` and `BUGS`, which are very similar to each other but each have some peculiarities of their own – turned out to be quite difficult for most. Nevertheless, the projects went relatively well. By the end of the one-semester course, all students were presenting quite successfully a Bayesian analysis of their own, unique data. Given that only a few months earlier they all were Bayesian illiterates, this is no less than impressive. Most of the credit for this almost spectacular result should go to Doing Bayesian Data Analysis and to the well-documented code that is available in the book and on the accompanying website.

Of course, the projects did not go always as smooth as one would have hoped for. It is enlightening to list the main obstacles that emerged when the ideas and code of the book were put to practice, by real students, on real problems, with real data.

1. *Missing data*. Many statistics textbooks treat the problem of missing data as an advanced one. However, as soon as one focuses on real psychological data, it is not advanced at all but very basic and very real. Almost all of the students brought a data set with missing data. Now there are two possible problems with missing data: A practical and a conceptual one. The conceptual one is the hardest to tackle because it touches the heart of the statistical analysis: Are you allowed to draw the conclusions you drew in the light of the missing data? Even if you sidestep this problem by telling the students that they do not have to worry about it for the moment (due to the limited time available etc.), there is still the practical problem: often the code from the book did not work because of missing (or unbalanced) data. Our – extremely unsatisfying – ad hoc solution to this problem was to rearrange the data in such a way the missingness would not get in the way of making the code run.
2. *Convergence*. Convergence problems are one of the bottlenecks of the widespread use of Bayesian statistics. Indeed, almost all of the students encountered some problems with convergence of the MCMC chains. The book (appropriately) stresses the importance of checking that the chains have converged, but offers little advice on steps to take when they do not. Additional examples or even an additional chapter addressing ways of improving convergence would have been very welcome. We have found, for example, advanced yet practical techniques such as parameter expansion or hierarchical centering to be very helpful.
3. *Adaptability*. The code that comes with the book is clean and works well for the purposes it has been written. However, it is not always straightforward to change the code to be applicable to more complex but still common designs. For instance, one student had a three-factorial design with only a single observation per cell. Although she eventually succeeded, it took quite some time and effort to adapt the code to her relatively modest needs.
4. *Informative priors*. One extremely powerful feature of the Bayesian method is the fact that it is (at least conceptually) easy to incorporate prior knowledge into the analyses by means of an informative prior. Some students got over their initial worries that informative priors corrupt the supposed objectivity of science and wanted to take advantage of this feature. Unfortunately, all non-toy applications in the book use generic or mildly informed priors. Having no examples to fall back on, the students struggled with how to add prior information in their analyses.

The practical problems encountered by the students do not derogate the book's immense value. For one, one could argue that having to go beyond what is available in the textbook is not necessarily bad from an educational perspective. Further, as said before, all students were able to rise above these difficulties and to successfully complete their project. Our experiences do suggest that the usefulness of Doing Bayesian Data Analysis could be further increased if a user community was set up containing user contributed content, such as code adapted to deal with missing data, code adapted to deal with advanced designs, convergence tips and tricks, and worked examples of how to add meaningful prior information. And it should not stop just there. If properly protected, this data base could also, for example, contain user contributed exam questions based on the book. For most of these issues, we have some code and examples available that we are happy to share.

Overall, Kruschke is to be applauded for his incredible efforts at writing such a highly accessible and useful textbook on Bayesian statistics. Doing Bayesian Data Analysis is an impressive piece of work that presents a major step in the dissemination of the Bayesian approach into mainstream psychology and will shape the way future psychologists will deal with their data. We are delighted to use it again in our course and wholeheartedly recommend it to anyone who wants to acquaint students with Bayesian statistics—both its bright concepts as well as its grim realities.

About 20 years ago, the second author of this review enrolled in college and bought the 800 page textbook for Statistics I. It contained a chapter on the history of statistics, which briefly mentioned Bayesian statistics and claimed that the Bayesian approach would become increasingly important in statistical theory and practice in the future. It is thanks to books like Doing Bayesian Data Analysis that this shiny future is – finally – beginning to emerge.

References

Smithson, M. (2011). Book review. *Journal of Mathematical Psychology*, *55*, 397–398.

Wolf Vanpaemel *
Francis Tuerlinckx
Faculty of Psychology and Educational Sciences, University of Leuven, 3000 Leuven, Belgium
E-mail address: wolf.vanpaemel@ppw.kuleuven.be (W. Vanpaemel).

1 December 2011
Available online 4 January 2012

* Corresponding reviewer.