Matthew Schofield
Department of Statistics
University of Kentucky
matthew.schofield@uky.edu

Patrick Breheny
Department of Biostatistics
University of Iowa
patrick-breheny@uiowa.edu

The BUGS Book: A Practical Introduction to Bayesian Analysis

David Lunn, Christopher Jackson, Nicky Best, Andrew Thomas, and David Spiegelhalter. Boca Raton, FL: CRC Press, 2013. ISBN 978-1-58488-849-9. xvii+381 pp.

This long-awaited text by the developers of BUGS, the most widely used software for Bayesian data analysis, provides a thorough description of BUGS and how to use it for Bayesian modeling. The first four chapters provide a introduction to Bayesian inference, the BUGS language, and the ideas behind Markov chain Monte Carlo (MCMC) methods. After this introduction, prior distributions are discussed in detail (both default/reference and informative priors for a variety of types of parameters) before moving on to an impressive catalog of examples. These range from familiar (regression, generalized linear models, categorical data) to advanced (spatial models, mixture models, time-to-event models, . . .). Along the way are chapters on modeling issues, model checking and comparison, and hierarchical modeling. The final section is a convenient reference chapter describing the different manifestations of the BUGS engine such as JAGS as well as R interfaces and their implementation. The supporting website, `www.mrc-bsu.cam.ac.uk/bugs/thebugsbook`, contains the code and data to reproduce the examples in the book, as well as solutions to selected exercises.

As the title indicates, the authors aim to provide not merely a BUGS tutorial, but a practical introduction to the art of modeling and analyzing data in the Bayesian paradigm. In particular, the authors state that they "do not assume familiarity with Bayesian methods" (p. xv) on the part of the reader. We believe they do an admirable job of justifying that statement. A feature of the introductory chapters is the emphasis on using simulation to calculate properties of probability distributions, starting as early as the first chapter. In chapter two, they begin using the BUGS language for simulation based probability calculations, that the readers can either verify by hand, or by using their favourite programming language. This helps to familiarize the reader with the underlying purpose of BUGS as being a user-friendly computing environment for simulating data from a wide range of probability distributions. Chapter three then introduces Bayesian inference. The initial focus is on one parameter conjugate examples, where as well as explicitly deriving the posterior distribution, the simulation examples from chapter two are modestly extended to include the data model. These examples help to show that BUGS is not a programming language per se, at least not in sense of R, Python or other such languages. Instead, the BUGS approach may be thought of as a whiteboard on which to write your probability model. We need only specify the conditional distributions that combine to form our probability model and BUGS takes care of the rest. How it does that is introduced in chapter four, which deals with MCMC. Their approach in this chapter is somewhat distinctive in that aside from a relatively in-depth discussion on Gibbs sampling, no other algorithms are discussed in detail. It would be impossible for someone unfamiliar with Bayesian methods and MCMC to use this book to implement even the simplest Metropolis algorithm. Instead, the authors use the chapter to explain the general idea behind MCMC and focus on practical aspects required for applied modeling. These include the role of initial values, assessing convergence and a description of Monte Carlo standard errors.

Many readers of this book will be interested in the discussion about DIC, in particular the calculation of p_D, the effective number of parameters. This comes in chapter eight, the longest of the book, which deals with model checking and comparison. We found the description of various methods to be generally well balanced. In particular, we were happy to see several examples that showed the difficulty in applying DIC as a default procedure, alongside the positive aspects of the approach. In particular, example 8.6.2 shows how two parameterizations $g_1(\theta)$ and $g_2(\theta)$ lead to very different DIC and p_D values despite having an identical posterior distribution for θ. Moreover, the authors discuss how there are numerous estimation approaches involved in calculating both p_D and DIC, and show how the resulting values can vary for a given model. The description of these methods, together with the examples are very helpful in understanding where these methods are appropriate, as well as what their limitation are.

We would have liked to see more discussion involving the relationship between DIC and the choice of model specification. This issue was raised twice, once in chapter 8 and then again in chapter 10. An example is given using the Student's t-distribution. Depending on whether this is specified directly, or using a scale mixture of normal distributions, the corresponding DIC changes. The authors state that this is because "the focus of the prediction problem" has changed. Their advice is that we must take care in identifying the focus of the prediction problem and choose the model selection method accordingly. While we agree, this is easier said than done for many complex statistical models. One of the appealing aspects of Bayesian modeling using MCMC is the straightforward inclusion of data augmentation. This often allows us to take a complex likelihood and specify it in terms of a series of simpler conditional distributions, using the appropriate latent variables. How does one proceed with model selection in such a case? We believe that this is a critical issue regarding p_D and DIC and we would have liked to have seen it receive more attention. In this regard we found the contrast between DIC and Bayes factors in Section 8.9 ("Discussion on model comparison") to be somewhat misleading. It reads as if DIC is the clear winner, while we believe the real "scorecard" is not quite so one-sided. For example, the authors state that computation for DIC is decidedly easier for DIC than Bayes factors. This will not be the case if we have used latent variables and need to integrate over them to obtain the appropriate density for DIC. Moreover, the summary made no mention of the various issues related to estimation of p_D discussed earlier in the chapter.

Chapter ten introduces the notion of hierarchical models. The authors begin by explaining the concept of hierarchical modeling, and show how it can be implemented in BUGS. They then discuss issues related to hierarchical modeling, including placing constraints on the hierarchical distributions, the need for care when specifying prior distributions for variance parameters, and improving convergence through parameter expansion. Examples are used to show the practical difference between models with (i) complete pooling (i.e. assuming parameters are identical), (ii) no pooling (i.e. assuming parameters are unrelated), and (iii) partial pooling (i.e. assuming parameters are similar and specified according to a hierarchical model). There are also examples that help to show various approaches for checking the assumptions of hierarchical models. The material introduced in chapter ten is relied upon in the remaining chapters of the book, with numerous examples including hierarchical models. When teaching out of this book (see below), we found that chapter ten presents a nice introduction to hierarchical modeling and lends itself nicely to be used alongside other books that consider the topic in more detail, such as Gelman and Hill (2006), as suggested by the authors themselves at the end of chapter ten.

Chapter eleven, which presents advanced models across a range of fields, brings everything together and displays the power and generalizability of BUGS. These include time-to-event models, time series models, spatial models, pharmacokinetic models, latent class modeling, splines and Bayesian nonparametrics. Even experienced users should learn some new things from the rich collection of examples provided in this chapter.

We believe *The BUGS Book* will likely to appeal to graduate and undergraduate students in statistics, as well as applied researchers in any quantitative field. While it seems particularly well-suited for use as a self-teaching text, many instructors would likely find it a welcome addition to their courses. One of us (Breheny) used *The BUGS Book* as a textbook for a course titled "Bayesian Modeling in Biostatistics" in the Spring of 2013, and found it to be both easy to teach from and popular with the students. This was particularly true at the beginning of the course, where the plentiful examples and clear, concise explanations eased the transition into Bayesian modeling for students used to the maximum likelihood paradigm.

In summary, *The BUGS Book* should appeal to a wide audience, both as a practical reference for implementing a wide variety of models in BUGS and as a well-written, interesting introduction to Bayesian data analysis.

Matthew Schofield
University of Kentucky
Patrick Breheny
University of Iowa

REFERENCES

Gelman, A. and Hill, J. (2006), *Data analysis using regression and multilevel/hierarchical models*, New York, NY: Cambridge University Press.