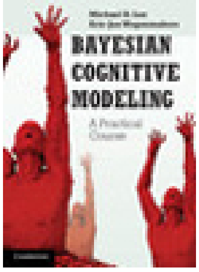

Bayesian Cognitive Modeling

A Practical Course

Michael D. Lee, Eric-Jan Wagenmakers

Book DOI: http://dx.doi.org/10.1017/CBO9781139087759

Online ISBN: 9781139087759

Hardback ISBN: 9781107018457

Paperback ISBN: 9781107603578

Chapter

1 - The basics of Bayesian analysis pp. 3-15

Chapter DOI: http://dx.doi.org/10.1017/CBO9781139087759.003

Cambridge University Press

1 The basics of Bayesian analysis

1.1 General principles

The general principles of Bayesian analysis are easy to understand. First, uncertainty or "degree of belief" is quantified by probability. Second, the observed data are used to update the *prior* information or beliefs to become *posterior* information or beliefs. That's it!

To see how this works in practice, consider the following example. Assume you are given a test that consists of 10 factual questions of equal difficulty. What we want to estimate is your ability, which we define as the rate θ with which you answer questions correctly. We cannot directly observe your ability θ. All that we can observe is your score on the test.

Before we do anything else (for example, before we start to look at your data) we need to specify our prior uncertainty with respect to your ability θ. This uncertainty needs to be expressed as a probability distribution, called the *prior distribution*. In this case, keep in mind that θ can range from 0 to 1, and that we do not know anything about your familiarity with the topic or about the difficulty level of the questions. Then, a reasonable "prior distribution," denoted by $p(\theta)$, is one that assigns equal probability to every value of θ. This uniform distribution is shown by the dotted horizontal line in Figure 1.1.

Now we consider your performance, and find that you answered 9 out of 10 questions correctly. After having seen these data, the updated knowledge about θ is described by the *posterior distribution*, denoted $p(\theta \mid D)$, where D indicates the observed data. This distribution expresses the uncertainty about the value of θ, quantifying the relative probability that each possible value is the true value. Bayes' rule specifies how we can combine the information from the data—that is, the likelihood $p(D \mid \theta)$—with the information from the prior distribution $p(\theta)$, to arrive at the posterior distribution $p(\theta \mid D)$:

$$p(\theta \mid D) = \frac{p(D \mid \theta)\, p(\theta)}{p(D)}. \tag{1.1}$$

This equation is often verbalized as

$$\text{posterior} = \frac{\text{likelihood} \times \text{prior}}{\text{marginal likelihood}}. \tag{1.2}$$

Note that the marginal likelihood (i.e., the probability of the observed data) does not involve the parameter θ, and is given by a single number that ensures that

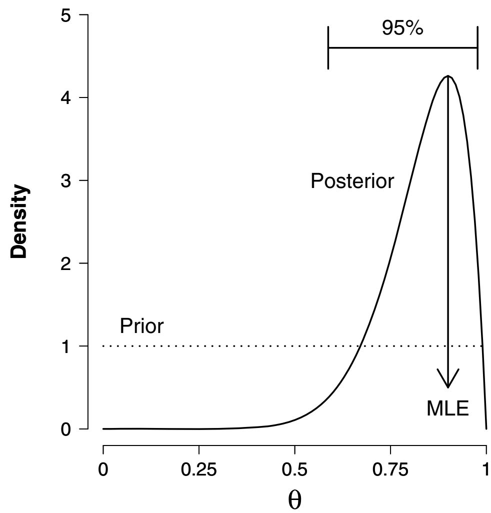

Fig. 1.1 Bayesian parameter estimation for rate parameter θ, after observing 9 correct responses and 1 incorrect response. The mode of the posterior distribution for θ is 0.9, equal to the maximum likelihood estimate (MLE), and the 95% credible interval extends from 0.59 to 0.98.

the area under the posterior distribution equals 1. Therefore, Equation 1.1 is often written as

$$p(\theta \mid D) \propto p(D \mid \theta)\, p(\theta)\,, \tag{1.3}$$

which says that the posterior is proportional to the likelihood times the prior. Note that the posterior distribution is a combination of what we knew before we saw the data (i.e., the information in the prior distribution), and what we have learned from the data. In particular, note that the new information provided by the data has reduced our uncertainty about the value of θ, as shown by the posterior distribution being narrower than the prior distribution.

The solid line in Figure 1.1 shows the posterior distribution for θ, obtained when the uniform prior is updated with the data. The central tendency of a posterior distribution is often summarized by its mean, median, or mode. Note that with a uniform prior, the mode of a posterior distribution coincides with the classical maximum likelihood estimate or MLE , $\hat{\theta} = k/n = 0.9$ (Myung, 2003). The spread of a posterior distribution is most easily captured by a Bayesian $x\%$ credible interval that extends from the $(100-x)/2^{\text{th}}$ to the $(100+x)/2^{\text{th}}$ percentile of the posterior distribution. For the posterior distribution in Figure 1.1, a 95% Bayesian credible interval for θ extends from 0.59 to 0.98. In contrast to the orthodox confidence interval, this means that one can be 95% confident that the true value of θ lies in between 0.59 and 0.98.

Exercises

Exercise 1.1.1 The famous Bayesian statistician Bruno de Finetti published two big volumes entitled *Theory of Probability* (de Finetti, 1974). Perhaps surprisingly, the first volume starts with the words "probability does not exist." To understand why de Finetti wrote this, consider the following situation: someone tosses a fair coin, and the outcome will be either heads or tails. What do you think the probability is that the coin lands heads up? Now suppose you are a physicist with advanced measurement tools, and you can establish relatively precisely both the position of the coin and the tension in the muscles immediately before the coin is tossed in the air—does this change your probability? Now suppose you can briefly look into the future (Bem, 2011), albeit hazily. Is your probability still the same?

Exercise 1.1.2 On his blog, prominent Bayesian Andrew Gelman wrote (March 18, 2010): "Some probabilities are more objective than others. The probability that the die sitting in front of me now will come up '6' if I roll it ... that's about 1/6. But not exactly, because it's not a perfectly symmetric die. The probability that I'll be stopped by exactly three traffic lights on the way to school tomorrow morning: that's well, I don't know exactly, but it is what it is." Was de Finetti wrong, and is there only one clearly defined probability of Andrew Gelman encountering three traffic lights on the way to school tomorrow morning?

Exercise 1.1.3 Figure 1.1 shows that the 95% Bayesian credible interval for θ extends from 0.59 to 0.98. This means that one can be 95% confident that the true value of θ lies between 0.59 and 0.98. Suppose you did an orthodox analysis and found the same confidence interval. What is the orthodox interpretation of this interval?

Exercise 1.1.4 Suppose you learn that the questions are all true or false questions. Does this knowledge affect your prior distribution? And, if so, how would this prior in turn affect your posterior distribution?

1.2 Prediction

The posterior distribution θ contains all that we know about the rate with which you answer questions correctly. One way to use the knowledge is *prediction*.

For example, suppose you are confronted with a new set of 5 questions, all of the same difficulty as before. How can we formalize our expectations about your performance on this new set? In other words, how can we use the posterior distribution $p\,(\theta \mid n = 10, k = 9)$—which, after all, represents everything that we know about θ from the old set—to *predict* the number of correct responses out of the new set of $n^{\text{rep}} = 5$ questions? The mathematical solution is to integrate over the posterior,

$\int p\,(k^{\mathrm{rep}} \mid \theta, n^{\mathrm{rep}} = 5)\, p\,(\theta \mid n = 10, k = 9)\ \mathrm{d}\theta$, where k^{rep} is the predicted number of correct responses out of the additional set of 5 questions.

Computationally, you can think of this procedure as repeatedly drawing a random value θ_i from the posterior, and using that value to every time determine a single k^{rep}. The end result is $p\,(k^{\mathrm{rep}})$, the posterior predictive distribution of the possible number of correct responses in the additional set of 5 questions. The important point is that by integrating over the posterior, all predictive uncertainty is taken into account.

Exercise

Exercise 1.2.1 Instead of "integrating over the posterior," orthodox methods often use the "plug-in principle." In this case, the plug-in principle suggests that we predict $p(k^{\mathrm{rep}})$ solely based on $\hat{\theta}$, the maximum likelihood estimate. Why is this generally a bad idea? Can you think of a specific situation in which this may not be so much of a problem?

1.3 Sequential updating

Bayesian analysis is particularly appropriate when you want to combine different sources of information. For example, assume that you are presented with a new set of 5 questions of equal difficulty. You answer 3 out of 5 correctly. How can we combine this new information with the old? Or, in other words, how do we update our knowledge of θ? Consistent with intuition, Bayes' rule entails that the prior that should be updated based on your performance for the new set is the posterior that was obtained based on your performance for the old set. Or, as Lindley put it, "today's posterior is tomorrow's prior" (Lindley, 1972, p. 2).

When all the data have been collected, however, the order in which this was done is irrelevant. The results from the 15 questions could have been analyzed as a single batch; they could have been analyzed sequentially, one-by-one; they could have been analyzed by first considering the set of 10 questions and next the set of 5, or vice versa. For all these cases, the end result, the final posterior distribution for θ, is identical. Given the same available information, Bayesian inference reaches the same conclusion, independent of the order in which the information was obtained. This again contrasts with orthodox inference, in which inference for sequential designs is radically different from that for non-sequential designs (for a discussion, see, for example, Anscombe, 1963).

Thus, a posterior distribution describes our uncertainty with respect to a parameter of interest, and the posterior is useful—or, as a Bayesian would have it, necessary—for probabilistic prediction and for sequential updating. To illustrate, in the case of our binomial example the uniform prior is a beta distribution with parameters $\alpha = 1$ and $\beta = 1$, and when combined with the binomial likelihood this yields a posterior that is also a beta distribution, with parameters $\alpha + k$ and $\beta + n - k$. In simple *conjugate* cases such as these, where the prior and the posterior belong to the same distributional family, it is possible to obtain analytical solutions for the posterior distribution, but in many interesting cases it is not.

1.4 Markov chain Monte Carlo

In general, the posterior distribution, or any of its summary measures, can only be obtained analytically for a restricted set of relatively simple models. Thus, for a long time, researchers could only proceed easily with Bayesian inference when the posterior was available in closed-form or as a (possibly approximate) analytic expression. As a result, practitioners interested in models of realistic complexity did not much use Bayesian inference. This situation changed dramatically with the advent of computer-driven sampling methodology, generally known as Markov chain Monte Carlo (MCMC: e.g., Gamerman & Lopes, 2006; Gilks, Richardson, & Spiegelhalter, 1996). Using MCMC techniques such as Gibbs sampling or the Metropolis–Hastings algorithm, researchers can directly sample sequences of values from the posterior distribution of interest, forgoing the need for closed-form analytic solutions. The current adage is that *Bayesian models are limited only by the user's imagination.*

In order to visualize the increased popularity of Bayesian inference, Figure 1.2 plots the proportion of articles that feature the words "Bayes" or "Bayesian," according to Google Scholar (for a similar analysis for specific journals in statistics and economics see Poirier, 2006). The time line in Figure 1.2 also indicates the introduction of WinBUGS, a general-purpose program that greatly facilitates Bayesian analysis for a wide range of statistical models (Lunn, Thomas, Best, & Spiegelhalter, 2000; Lunn, Spiegelhalter, Thomas, & Best, 2009; Sheu & O'Curry, 1998). MCMC methods have transformed Bayesian inference to a vibrant and practical area of modern statistics.

For a concrete and simple illustration of Bayesian inference using MCMC, consider again the binomial example of 9 correct responses out of 10 questions, and the associated inference problem for θ, the rate of answering questions correctly. Throughout this book, we use WinBUGS to do Bayesian inference, saving us the effort of coding the MCMC algorithms ourselves.[1] Although WinBUGS does not work for every research problem application, it will work for many in cognitive sci-

[1] At this point, some readers want to know how exactly MCMC algorithms work. Other readers feel the urge to implement MCMC algorithms themselves. The details of MCMC sampling are covered in many other sources and we do not repeat that material here. We recommend the relevant chapters from the following books, listed in order of increasing complexity: Kruschke (2010a), MacKay (2003), Gilks et al. (1996), Ntzoufras (2009), and Gamerman and Lopes (2006). An introductory overview is given in Andrieu, De Freitas, Doucet, and Jordan (2003). You can also browse the internet, and find resources such as `http://www.youtube.com/watch?v=4gNpgSPal_8` and `http://www.learnbayes.org/`.

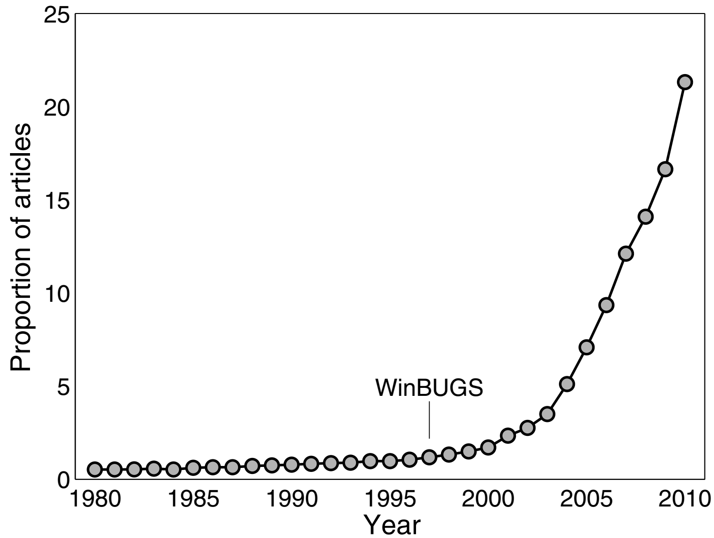

Fig. 1.2 A Google Scholar perspective on the increasing popularity of Bayesian inference, showing the proportion of articles matching the search "bayes OR bayesian -author: bayes" for the years 1980 to 2010.

ence. WinBUGS is easy to learn and is supported by a large community of active researchers.

The WinBUGS program requires you to construct a file that contains the model specification, a file that contains initial values for the model parameters, and a file that contains the data. The model specification file is most important. For our binomial example, we set out to obtain samples from the posterior of θ. The associated WinBUGS model specification code is two lines long:

```
model{
  theta ~ dunif(0,1) # the uniform prior for updating by the data
  k ~ dbin(theta,n)  # the data; in our example, k = 9 and n = 10
}
```

In this code, the "$\sim$" or twiddle symbol denotes "is distributed as", `dunif(a,b)` indicates the uniform distribution with parameters a and b, and `dbin(theta,n)` indicates the binomial distribution with rate θ and n observations. These and many other distributions are built in to the WinBUGS program. The "#" or hash sign is used for comments. As WinBUGS is a declarative language, the order of the two lines is inconsequential. Finally, note that the values for k and n are not provided in the model specification file. These values constitute the data and they are stored in a separate file.

When this code is executed, you obtain a sequence of MCMC samples from the posterior $p(\theta \mid D)$. Each individual sample depends only on the one that immediately preceded it, and this is why the entire sequence of samples is called a *chain*.

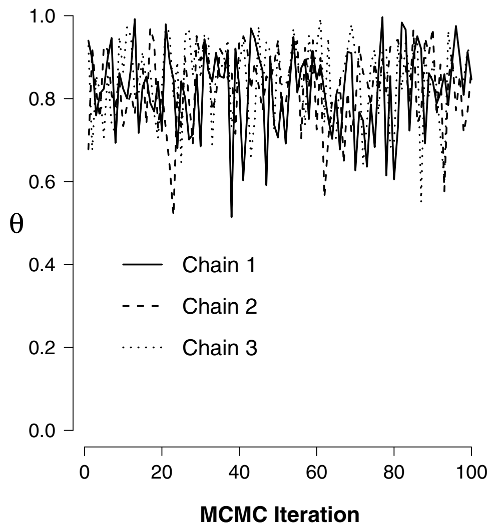

Fig. 1.3 Three MCMC chains for rate parameter θ, after observing 9 correct responses and 1 incorrect response.

In more complex models, it may take some time before a chain converges from its starting value to what is called its stationary distribution. To make sure that we only use those samples that come from the stationary distribution, and hence are unaffected by the starting values, it is good practice to diagnose convergence. This is an active area of research, and there is an extensive set of practical recommendations regarding achieving and measuring convergence (e.g., Gelman, 1996; Gelman & Hill, 2007).

A number of worked examples in this book deal with convergence issues in detail, but we mention three important concepts now. One approach is to run multiple chains, checking that their different initial starting values do not affect the distributions they sample from. Another is to discard the first samples from each chain, when those early samples are sensitive to the initial values. These discarded samples are called *burn-in* samples. Finally, it can also be helpful not to record every sample taken in a chain, but every second, or third, or tenth, or some other subset of samples. This is known as *thinning*, a procedure that is helpful when the chain moves slowly through the parameter space and, consequently, the current sample in the MCMC chain depends highly on the previous one. In such cases, the sampling process is said to be autocorrelated.

For example, Figure 1.3 shows the first 100 iterations for three chains that were set up to draw values from the posterior for θ. It is evident that the three chains are "mixing" well, suggesting early convergence. After assuring ourselves that the chains have converged, we can use the sampled values to plot a histogram, construct a density estimate, and compute values of interest. To illustrate, the three chains from Figure 1.3 were run for 3000 iterations each, for a total of 9000 samples from the posterior of θ. Figure 1.4 plots a histogram for the posterior. To visualize how the

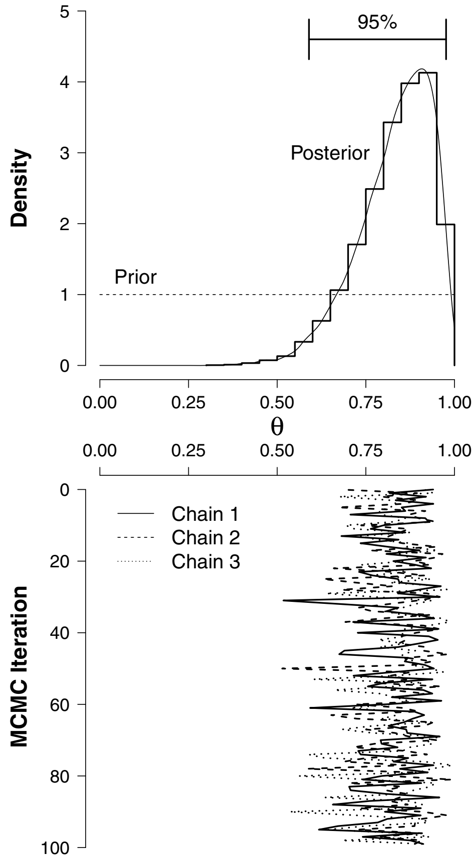

Fig. 1.4 MCMC-based Bayesian parameter estimation for rate parameter θ, after observing 9 correct responses and 1 incorrect response. The thin solid line indicates the fit of a density estimator. Based on this density estimator, the mode of the posterior distribution for θ is approximately 0.89, and the 95% credible interval extends from 0.59 to 0.98, closely matching the analytical results from Figure 1.1.

histogram is constructed from the MCMC chains, the bottom panel of Figure 1.4 plots the MCMC chains sideways; the histograms are created by collapsing the values along the "MCMC iteration" axis and onto the "θ" axis.

In the top panel of Figure 1.4, the thin solid line represent a density estimate. The mode of the density estimate for the posterior of θ is 0.89, whereas the 95% credible interval is $(0.59, 0.98)$, matching the analytical result shown in Figure 1.1.

The key point is that the analytical intractabilities that limited the scope of Bayesian parameter estimation have now been overcome. Using MCMC sampling, posterior distributions can be approximated to any desired degree of accuracy. This

Box 1.1 Why isn't every statistician a Bayesian?

"The answer is simply that statisticians do not know what the Bayesian paradigm says. Why should they? There are very few universities in the world with statistics departments that provide a good course in the subject. Only exceptional graduate students leave the field of their advisor and read for themselves. A secondary reason is that the subject is quite hard for someone who has been trained in the sampling-theory approach to understand. . . . The subject is difficult. Some argue that this is a reason for not using it. But it is always harder to adhere to a strict moral code than to indulge in loose living. . . . Every statistician would be a Bayesian if he took the trouble to read the literature thoroughly and was honest enough to admit that he might have been wrong." (Lindley, 1986, pp. 6–7).

book teaches you to use MCMC sampling and Bayesian inference to do research with cognitive science models and data.

Exercises

Exercise 1.4.1 Use Google and list some other scientific disciplines that use Bayesian inference and MCMC sampling.

Exercise 1.4.2 The text reads: "Using MCMC sampling, posterior distributions can be approximated to any desired degree of accuracy." How is this possible?

1.5 Goal of this book

The goal of this book is to show, by working through concrete examples, how Bayesian inference can be applied to modeling problems in cognitive science. Bayesian data analysis has received increasing attention from cognitive scientists, and for good reason.

1. Bayesian inference is *flexible*. This means that Bayesian models can respect the complexity of the data, and of the processes being modeled. For example, data analysis may require the inclusion of a contaminant process, a multi-level structure, or an account of missing data. Using the Bayesian approach, these sorts of additions are relatively straightforward.
2. Bayesian inference is *principled*. This means that all uncertainty is accounted for appropriately, and no useful information is discarded.

Box 1.2 Common sense expressed in numbers

"The Bayesian approach is a common sense approach. It is simply a set of techniques for orderly expression and revision of your opinions with due regard for internal consistency among their various aspects and for the data. Naturally, then, much that Bayesians say about inference from data has been said before by experienced, intuitive, sophisticated empirical scientists and statisticians. In fact, when a Bayesian procedure violates your intuition, reflection is likely to show the procedure to have been incorrectly applied." (Edwards et al., 1963, p. 195).

3. Bayesian inference yields *intuitive conclusions*. This reflects the fact that Bayesian inference is normative, stipulating how rational agents should change their opinion in the light of incoming data. Of course, it can nevertheless happen that you occasionally find a Bayesian conclusion to be surprising or counter-intuitive. You are then left with one of two options—either the analysis was not carried out properly (e.g., errors in coding, errors in model specification) or your intuition is in need of schooling.
4. Bayesian inference is *easy to undertake*. This means that with the software packages used in this book, Bayesian inference is often (but not always!) a trivial exercise. This frees up resources so more time can be spent on the substantive issues of developing theories and models, and interpreting results when they are applied to data.

At this point you may be champing at the bit, eager to apply the tools of Bayesian analysis to the kinds of cognitive models that interest you. But first we need to cover the basics and this is why Parts I, II, and III prepare you for the more complicated case studies presented in Part IV. This is not to say that the "elementary" material in Parts I, II, and III are devoid of cognitive context. On the contrary, we have tried to highlight how even the binomial model finds meaningful application in cognitive science.

Perhaps the material covered in this first chapter is still relatively abstract for you. Perhaps you are currently in a state of confusion. Perhaps you think that this book is too difficult, or perhaps you do not yet see clearly how Bayesian inference can help you in your own work. These feelings are entirely understandable, and this is why this book contains more than just this one chapter. Our teaching philosophy is that you learn the most by doing, not by reading. So if you still do not know exactly what a posterior distribution is, do not despair. The chapters in this book make you practice core Bayesian inference tasks so often that at the end you will know exactly what a posterior distribution is, whether you like it or not. Of course, we rather hope you like it, and we also hope that you will discover that Bayesian statistics can be exciting, rewarding, and, indeed, fun.

1.6 Further reading

This section provides some references for further reading. We first list Bayesian textbooks and seminal papers, then some texts that specifically deal with WinBUGS. We also note that Smithson (2010) presents a useful comparative review of six introductory textbooks on Bayesian methods.

1.6.1 Bayesian statistics

This section contains an annotated bibliography of Bayesian articles and books that we believe are particularly useful or inspiring.

- Berger, J. O. & Wolpert, R. L. (1988). *The Likelihood Principle* (2nd edn.). Hayward, CA: Institute of Mathematical Statistics. This is a great book if you want to understand the limitations of orthodox statistics. Insightful and fun.
- Bolstad, W. M. (2007). *Introduction to Bayesian Statistics* (2nd edn.). Hoboken, NJ: Wiley. Many books claim to introduce Bayesian statistics, but forget to state on the cover that the introduction is "for statisticians" or "for those comfortable with mathematical statistics." The Bolstad book is an exception, as it does not assume much background knowledge.
- Dienes, Z. (2008). *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*. New York: Palgrave Macmillan. An easy-to-understand introduction to inference that summarizes the differences between the various schools of statistics. No knowledge of mathematical statistics is required.
- Gamerman, D. & Lopes, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Boca Raton, FL: Chapman & Hall/CRC. This book discusses the details of MCMC sampling; a good book, but too advanced for beginners.
- Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. This book is an extensive practical guide on how to apply Bayesian regression models to data. WinBUGS code is provided throughout the book. Andrew Gelman also has an active blog that you might find interesting: `http://andrewgelman.com/`
- Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo in Practice*. Boca Raton, FL: Chapman & Hall/CRC. A citation classic in the MCMC literature, this book features many short chapters on all kinds of sampling-related topics: theory, convergence, model selection, mixture models, and so on.
- Gill, J. (2002). *Bayesian Methods: A Social and Behavioral Sciences Approach*. Boca Raton, FL: CRC Press. A well-written book that covers a lot of ground. Readers need some background in mathematical statistics.

- Hoff, P. D. (2009). *A First Course in Bayesian Statistical Methods*. Dordrecht, The Netherlands: Springer. A clear and well-written introduction to Bayesian inference, with accompanying R code, requiring some familiarity with mathematical statistics.
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Cambridge, UK: Cambridge University Press. Jaynes was one of the most ardent supporters of objective Bayesian statistics. The book is full of interesting ideas and compelling arguments, as well as being laced with Jaynes' acerbic wit, but it requires some mathematical background to appreciate all of the content.
- Jeffreys, H. (1939/1961). *Theory of Probability*. Oxford, UK: Oxford University Press. Sir Harold Jeffreys is the first statistician who exclusively used Bayesian methods for inference. Jeffreys also invented the Bayesian hypothesis test, and was generally far ahead of his time. The book is not always an easy read, in part because the notation is somewhat outdated. Strongly recommended, but only for those who already have a solid background in mathematical statistics and a firm grasp of Bayesian thinking. See `www.economics.soton.ac.uk/staff/aldrich/jeffreysweb.htm`
- Lee, P. M. (2012). *Bayesian Statistics: An introduction* (4th edn.). Chichester, UK: John Wiley. This well-written book illustrates the core tenets of Bayesian inference with simple examples, but requires a background in mathematical statistics.
- Lindley, D. V. (2000). The philosophy of statistics. *The Statistician, 49*, 293–337. One the godfathers of Bayesian statistics explains why Bayesian inference is right and everything else is wrong. Peter Armitage commented on the paper: "Lindley's concern is with the very nature of statistics, and his argument unfolds clearly, seamlessly and relentlessly. Those of us who cannot accompany him to the end of his journey must consider very carefully where we need to dismount; otherwise we shall find ourselves unwittingly at the bus terminus, without a return ticket."
- Marin, J.-M. & Robert, C. P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. New York: Springer. This is a good book by two reputable Bayesian statisticians. The book is beautifully typeset, includes an introduction to R, and covers a lot of ground. A firm knowledge of mathematical statistics is required. The exercises are challenging.
- McGrayne, S. B. (2011). *The Theory that Would not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy*. New Haven, CT: Yale University Press. A fascinating and accessible overview of the history of Bayesian inference.
- O'Hagan, A. & Forster, J. (2004). *Kendall's Advanced Theory of Statistics Vol. 2B: Bayesian Inference* (2nd edn.). London: Arnold. If you are willing to read only a single book on Bayesian statistics, this one is it. The book requires a background in mathematical statistics.
- Royall, R. M. (1997). *Statistical Evidence: A Likelihood Paradigm*. London: Chapman & Hall. This book describes the different statistical paradigms, and highlights

the deficiencies of the orthodox schools. The content can be appreciated without much background knowledge in statistics. The main disadvantage of this book is that the author is not a Bayesian. We still recommend the book, which is saying something.

1.6.2 WinBUGS texts

- Kruschke, J. K. (2010). *Doing Bayesian Data Analysis: A Tutorial Introduction with R and BUGS*. Burlington, MA: Academic Press. This is one of the first Bayesian books geared explicitly towards experimental psychologists and cognitive scientists. Kruschke explains core Bayesian concepts with concrete examples and OpenBUGS code. The book focuses on statistical models such as regression and ANOVA, and provides a Bayesian approach to data analysis in psychology, cognitive science, and empirical sciences more generally.
- Lee, S.-Y. (2007). *Structural Equation Modelling: A Bayesian Approach*. Chichester, UK: John Wiley. After reading the first few chapters from this book, you may wonder why not everybody uses WinBUGS for their structural equation modeling.
- Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, FL: Chapman & Hall/CRC Press. Quoted from the publisher: "Bayesian statistical methods have become widely used for data analysis and modelling in recent years, and the BUGS software has become the most popular software for Bayesian analysis worldwide. Authored by the team that originally developed this software, The BUGS Book provides a practical introduction to this program and its use. The text presents complete coverage of all the functionalities of BUGS, including prediction, missing data, model criticism, and prior sensitivity. It also features a large number of worked examples and a wide range of applications from various disciplines."
- Ntzoufras, I. (2009). *Bayesian Modeling using WinBUGS*. Hoboken, NJ: John Wiley. Provides an accessible introduction to WinBUGS. The book also presents a variety of Bayesian modeling examples, with an emphasis on Generalized Linear Models. See `www.ruudwetzels.com` for a detailed review.
- Spiegelhalter, D., Best, N., & Lunn, D. (2003). *WinBUGS User Manual 1.4*. Cambridge, UK: MRC Biostatistic Unit. Provides an introduction to WinBUGS, including a useful tutorial and various tips and tricks for new users. The user manual has effectively been superseded by *The BUGS Book* mentioned above.