



Review of *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Second Edition, by Richard McElreath, Chapman and Hall, 2020

Jim Albert

To cite this article: Jim Albert (2020): Review of *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Second Edition, by Richard McElreath, Chapman and Hall, 2020, Journal of Statistics Education, DOI: [10.1080/10691898.2020.1806761](https://doi.org/10.1080/10691898.2020.1806761)

To link to this article: <https://doi.org/10.1080/10691898.2020.1806761>



© The Author(s). Published with license by Taylor and Francis Group, LLC.



Published online: 10 Sep 2020.



Submit your article to this journal [↗](#)



Article views: 873



View related articles [↗](#)



View Crossmark data [↗](#)

BOOK REVIEW



Review of *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, Second Edition, by Richard McElreath, Chapman and Hall, 2020

Jim Albert

Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH

1. Bayesian Texts

Due to the developments in Bayesian applications and associated computational resources in the past 30 years, there are many Bayesian texts currently available. Some of the texts, such as Gelman et al. (2013), Carlin and Louis (2008), and Hoff (2009), are directed toward graduate students in statistics. Other texts, such as Kruschke (2014), Gill (2014), and Jackman (2009) are directed toward graduate students in specific applied disciplines such as psychology, social, and behavioral sciences. Given the ready availability of Bayesian software, some texts, such as Lunn et al. (2012) and Link and Barker (2009), focus on the use of software such as WinBugs or JAGS in Bayesian computation.

2. Plan of *Statistical Rethinking*

Statistical Rethinking has a similar focus to other applied Bayesian texts. The text is directed to graduate-level researchers in the natural and social sciences who have had a basic course in regression. The plan is to raise the reader's knowledge of the basic tenets of Bayesian thinking, Bayesian computation, and Bayesian statistical regression modeling, so they can make reasonable choices and interpretations for statistical modeling in their own studies.

McElreath's approach is to use few mathematical formulas in presenting probability distributions and models. For example, the formal definition of a normal density is only presented as part of the "overthinking" material. Instead, the modeling and associated methods are presented by use of R code. This means that the text requires some introductory knowledge of R to understand much of the code that is integrated into the text. To facilitate the Bayesian computations, the author has written the *rethinking* R package that facilitates the computational methods described in the book.

3. Structure of the Text

As mentioned above, the text is written for applied graduate students or professionals who have some previous exposure to regression but limited background in calculus. The book is intended for a course that follows the chapters sequentially from the beginning. Chapter 1, with the interesting title "The

Golem of Prague," discusses popular statistical methods and associated philosophy, arguing that it is desirable to move away from "black-box" tools for testing null hypotheses. Chapter 2 presents the basic elements of Bayesian data analysis in the context of learning about a proportion. Since simulation will be the main tool in implementing posterior inference and predictive checking, Chapter 3 illustrates the use of simulation in summarizing a posterior of a proportion and for implementing posterior predictive checks.

Chapters 4–6 provide an introduction to the broadly applicable Bayesian general linear model. Chapter 4 describes inference and prediction for a single predictor variable and Chapters 5 and 6 describe many of the issues, such as masked relationships, conditional independence, and confounding, that occur when one has several predictor variables.

Chapter 7 presents an overview of measures for Bayesian model selection. It explains the problems of overfitting and underfitting when there are many potential explanatory variables. This chapter takes a side road to describe the concepts of entropy, information, and divergence, and then gives an overview of measures of predictive accuracy such as Akaike information criterion, deviance information criterion, and widely applicable information criterion. Chapter 8 introduces interactions in multiple regression where the association between the response and predictor depends on the value of another predictor.

Chapter 9 provides a gentle overview of Markov chain Monte Carlo (MCMC) algorithms including the Metropolis algorithm, Gibbs sampling, and Hamiltonian Monte Carlo that is the basis of the Stan software (Carpenter et al. 2017). This chapter describes MCMC diagnostics and gives insights as to when these MCMC algorithms can fail.

The Hamiltonian Monte Carlo algorithm is the primary computational tool for the regression chapters to follow. Chapter 10 motivates the class of generalized linear models as the distributions that have the largest entropies, that is, contain the least information given particular constraints on the outcome variable. Chapter 11 provides illustrations of logistic and Poisson modeling for count response data and Chapter 12 deals with situations where the response variable is overdispersed, contains more zero observations than predicted, or is categorical with ordered outcomes.

Chapters 13 and 14 provide introductions to multilevel models from a Bayesian perspective. Chapter 13 focuses on the varying intercept situation where one wishes to simultaneously estimate means from different groups. This varying intercept situation is illustrated in the situation where one wishes to learn about the average waiting times at different cafés. Chapter 14 generalizes to the scenario where intercepts and slopes vary in the regressions in the different groups. This situation can be applied to the café example where one wishes to learn about both the average wait times and the differences between morning and afternoon average wait times between different cafés.

The book is concluded in Chapters 15 and 16 with some advanced modeling topics. Chapter 15 deals with regression models with measurement errors in the predictors and missing data, and Chapter 16 considers Bayesian fitting of several non-linear models motivated by scientific theory.

4. Story Telling

One of the most appealing aspects of *Statistical Rethinking* is the exposition of the Bayesian material. McElreath does more than provide general guidance in the construction and interpretation of the regression models. He has a colorful manner of providing intuition into the Bayesian concepts that underlie these methods.

McElreath has a unique, story-telling style of writing that gets the reader interested in the subject matter. For example, the Metropolis, Gibbs sampling, and Hamiltonian Monte Carlo algorithms are often described in texts in mathematical terms and algorithms that can escape the reader who does not have the background in mathematics or computer science. Instead, McElreath describes these algorithms in Chapter 9 by use of parables. Metropolis sampling is introduced by a story about Good King Markov who visits the different islands of his kingdom by use of a random mechanism. In another story introducing Hamiltonian sampling, King Monty plans on visiting the citizens of his kingdom according to the population density and his advisor Hamilton devises a method of moving the king's vehicle governed by the vehicle's momentum. These parables provide the reader with the intuition implicit in these MCMC algorithms.

I also found Chapter 7 to be a very appealing introduction and description of Bayesian model selection criteria. This chapter, titled "Ulysses' Compass," motivates the model selection problem by contrasting the Ptolemaic and Copernican models for describing the movements of bodies in the solar system. The two models produce the same predictions, but the Copernican model required fewer circles and should be preferred by the principle of Ockham's razor. Section 7.1 provides a simple regression example illustrating overfitting and underfitting and Section 7.2 gives some background material on entropy and divergence. By the time the information criteria are introduced in Section 7.4, one understands that these criteria are really just estimates of the out-of-sample Kullback–Leibler divergence of a regression model.

Story telling is also used to motivate multilevel modeling. Chapter 13 describes an individual Clive Wearing, an accomplished musician who suffered from anterograde amnesia. This person was unable to form long-term memories, forgetting

music he played only a few minutes ago. Statistical models sometimes have anterograde amnesia in that they estimate parameters from some clusters, forgetting information from different clusters. This type of amnesia is not helpful for learning about the world. It is desirable to share information among clusters to obtain more accurate estimators, and this motivates the use of multilevel models.

5. Computation

Currently, many popular Bayesian regression models can be fit through simple functions, and the user can be oblivious to the MCMC algorithms that are used in the model. McElreath provides an appealing introduction to Bayesian computation in stages. First, in Chapters 3 and 4, he illustrates the use of a grid approximation where the posterior is computed on a fine grid of parameter values. At a more sophisticated level, McElreath uses a normal approximation to the posterior for the regression models in Chapters 4–6. Last, after he introduces MCMC, Hamiltonian Monte Carlo is the main computational tool in the remaining chapters.

To provide a bridge between the model description and the fitting, McElreath uses special functions `quap()` and `ulam()` where one uses a script defining the Bayesian model including the sampling component and the prior. I found this approach better than trying to explain all of the modeling implicit in the use a special MCMC Bayesian function like `brm()` in the `brm` package.

Pedagogically, I think this approach to computation is attractive. Simulation is introduced early for posterior and predictive calculations for the grid approximation and normal approximation methods. McElreath's approach prepares the reader to use other software tools such as `RStan` for the MCMC fitting of the Bayesian models for their applications.

6. What's New in the Second Edition?

The second edition is a considerable update to the original text. There is a greater emphasis on the use of prior predictive simulation for understanding the implications of a particular choice of prior. The first chapter on multiple regression has been split into two chapters and directed acyclic graphs are used to discuss issues of causal inference. The `rethinking` package has been revised and the Bayesian fitting functions allow for more general model structures. Also there are new types of models illustrated, such as smoothing splines in Chapter 4, robust regression in Chapter 7, and the models in a new Chapter 16 that are not within the generalized linear mixed model framework.

7. Overall Assessment

Overall, I think *Statistical Rethinking* is an excellent introduction to modern applied Bayesian modeling very suitable for masters students in statistics and doctoral students from other disciplines. As the title suggests, the book is directed to students who are trying to make sense of the foundations implicit in the statistical methods learned in a previous course. The text is very readable with interesting stories to motivate and illustrate statistical concepts. Even if one has some background in

Bayesian thinking, I would think that there would be something to learn in this book with the variety of regression models and insights into Bayesian computational algorithms and predictive measures. Although this text is not specifically directed to undergraduates, I believe that this text would be a great resource for statistics instructors who wish to develop Bayesian modules at the undergraduate level. I actually recently used part of Chapter 9 to provide some intuition on Hamiltonian sampling for my graduate course in statistical computing.

I only had a few concerns with the text. For the doctoral student in statistics who is developing new Bayesian methods, I think there is a need for some mathematical rigor that is not contained in this text. Also, the success of the computational material rests on the usefulness of the author's *rethinking* package and given the constant updates to the Stan software, I wonder about the stability of the author's fitting functions described in the text in future years. These concerns aside, I believe *Statistical Rethinking* will be a valuable resource for practitioners in the applied sciences who are interested in learning about the fundamental concepts of Bayesian modeling and applying these models in their own research.

References

- Carlin, B. P., and Louis, T. A. (2008), *Bayesian Methods for Data Analysis*, Boca Raton, FL: CRC Press.
- Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017), "Stan: A Probabilistic Programming Language," *Journal of Statistical Software*, 76, 1–43.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013), *Bayesian Data Analysis*, Texts in Statistical Science, Boca Raton, FL: CRC Press.
- Gill, J. (2014), *Bayesian Methods: A social and Behavioral Sciences Approach* (Vol. 20), Boca Raton, FL: CRC Press.
- Hoff, P. D. (2009), *A First Course in Bayesian Statistical Methods* (Vol. 580), New York: Springer.
- Jackman, S. (2009), *Bayesian Analysis for the Social Sciences* (Vol. 846), New York: Wiley.
- Kruschke, J. (2014), *Doing Bayesian Data Analysis: A Tutorial With R, JAGS, and Stan*, Boston, MA: Academic Press.
- Link, W. A., and Barker, R. J. (2009), *Bayesian Inference: With Ecological Applications*, Boston, MA: Academic Press.
- Lunn, D., Jackson, C., Best, N., Thomas, A., and Spiegelhalter, D. (2012), *The BUGS Book: A Practical Introduction to Bayesian Analysis*, Boca Raton, FL: CRC Press.