Teaching Bayesian Statistics

Diogo Ferrari, *University of California, Riverside, USA*

The questions discussed in this article are: What challenges emerge when teaching introductory-level Bayesian statistics to non-statistics graduate students who want to become well-informed users of Bayesian methods, and what content-selection strategies can instructors adopt to overcome those challenges? Before addressing these questions, we may ask why this consideration is important and if there are any distinctive challenges to teaching Bayesian statistics that differ from teaching non-Bayesian statistics.

Bayesian statistics regained popularity among statisticians and applied researchers in various disciplines in the 1990s. By the end of the decade, scholars were debating whether the Bayesian approach to inference should be taught at the undergraduate level. The third edition of the 1997 issue of *The American Statistician* dedicated its *Teachers' Corner* section to the topic. It included articles by Moore (1997), Berry (1997), and Albert (1997) and discussions by well-known statisticians: Jeffrey Witmer, Thomas Short, Dennis Lindley, David Freedman, and Richard Scheaffer. Moore (1997) advocated for not teaching Bayesian statistics at the elementary level because Bayesian methods were not widely used yet, conditional probabilities are difficult to understand, there were no suitable textbooks, and—unlike the frequentist approach—there was no standard set of prescriptive procedures to teach students. Berry (1997) and Albert (1997) countered these points, arguing that Bayesian inference is more intuitive and easier to grasp than frequentist approaches.

We should not teach a subject simply because it is easy or avoid teaching it because it is difficult. Usefulness to answer substantive questions should be one criterion to decide what to teach and, in that regard, Bayesian methods have grown significantly in importance since *The American Statistician* first debated the value of teaching Bayesian statistics in elementary courses. Since then, scholarly work using Bayesian statistics, or developing solutions that rely on Bayesian approaches, has only increased. This can be attributed in part to the popularization of papers by, for example, Geman and Geman (1984) and Gelfand and Smith (1990), who used stochastic methods to approximate marginal distributions and values of integrals, particularly the Metropolis algorithm developed in the 1950s (Metropolis et al. 1953) and later generalized in the 1970s by Hastings (1970). This popularization was accompanied by the technological development of computers that allowed implementations of Metropolis–Hastings algorithms to estimate posterior probabilities for virtually any problem, especially those that are difficult to solve using non-Bayesian methods. Moreover, there now are better textbook options to teach Bayesian statistics at different levels (e.g., Ghosh, Delampady, and Samanta 2007; Hoff 2009; Kruschke 2015) and some written by political scientists (e.g., Gelman et al. 2014; Gill 2014; Jackman 2009). There also are good introductory papers (e.g., Jackman 2004) and articles that suggest specific content and class schedules (e.g., Utts and Johnson 2008).

Nevertheless, certain challenges of teaching Bayesian statistics persist. Before discussing these challenges in detail, I distinguish between teaching statistics from a Bayesian perspective (as discussed in the *Teachers' Corner* section mentioned previously) and teaching Bayesian statistics. Teaching statistics from a Bayesian perspective introduces students to statistical reasoning using the Bayesian paradigm. This is a broader debate that involves core questions about statistical education in general and alternative ways to think about inference.

I use the phrase "teaching Bayesian statistics" to refer to the immediate and pressing challenges of teaching Bayesian statistics to non-statisticians under the current organization of university-level statistical education. I focus on this aspect of statistical education—that is, teaching Bayesian statistics rather than teaching statistics from a Bayesian perspective.

The question of teaching Bayesian statistics is pressing because, as discussed previously, academic literature using Bayesian methods has become ubiquitous since the 1990s, which increases the pressure to include Bayesian statistics in the graduate curriculum. The challenges of teaching Bayesian statistics range from pedagogical methods to content-selection issues. There already is a large literature dedicated to investigating and proposing sound pedagogical practices to foster inclusion in the classroom, promote diversity, and improve students' learning experience and retention; some literature provides specific examples and recommendations to teach statistics to non-statisticians (Gelman 2005; Gelman and Nolan 2017). Therefore, I discuss another aspect of this challenge—namely, selecting content that facilitates students' learning experience.

The challenges of teaching introductory Bayesian statistics courses to undergraduate students with no background in statistics are quite different from those of teaching graduate students who possess some background in statistics or Bayesian statistics in particular. Students taking Bayesian statistics often have some previous experience in statistics—enough to understand conditional probabilities, conditional moments, and basic inference. This article focuses on these students.

The challenges that instructors face in preparing these students to be educated consumers are slightly different than those we encounter in educating and developing thoughtful users of Bayesian analysis. Although the following discussion

doi:10.1017/S1049096521001244
https://doi.org/10.1017/S1049096521001244 Published online by Cambridge University Press

can serve both cases, I focus on the latter goal. With these considerations, I restate the introductory questions: What challenges can emerge when teaching Bayesian statistics, and how can strategic content selection overcome those challenges?

The question of teaching Bayesian statistics is pressing because academic literature using Bayesian methods has become ubiquitous since the 1990s, which increases the pressure to include Bayesian statistics in the graduate curriculum.

Although the challenges can vary among groups of students, at least two challenges are likely to emerge for those with some background in non-Bayesian statistics but no background in Bayesian statistics—which often is the case for graduate students taking their first course in Bayesian methods.

The first challenge is the "mindset shift" toward Bayesian inference reasoning, and it is a consequence of not teaching statistics from a Bayesian perspective. If the standard (implicit or explicit) choice in a university's statistical curriculum is to develop a non-Bayesian intuition when teaching statistical thinking in introductory courses, then the first challenge for instructors teaching Bayesian statistics is to help students with a mindset shift between Bayesian and non-Bayesian ways of thinking about inference.

If the standard (implicit or explicit) choice in a university's statistical curriculum is to develop a non-Bayesian intuition when teaching statistical thinking in introductory courses, then the first challenge for instructors teaching Bayesian statistics is to help students with a mindset shift between Bayesian and non-Bayesian ways of thinking about inference.

Bayesian statistics requires a specific mindset to think about inference, which differs in important ways from the way that non-Bayesian statisticians conduct the inference, draw conclusions, and evaluate the quality of the inferential procedure. Bayesian approaches often are briefly and superficially covered—if covered at all—in standard introductory statistics courses. It is fair to state that most if not all of the elementary and introductory courses at undergraduate and graduate levels focus primarily on frequentist approaches, perhaps blended with likelihood methods. I do not have complete data about the curriculum of all universities; however, I am inclined to think that this is the typical case. Many departments do not offer separate, optional Bayesian methods in their curriculum. Courses on statistics usually are understood as having a non-Bayesian perspective. If readers are skeptical about the bias toward non-Bayesian statistics in standard curriculum, note that we do not see courses titled "Introduction to Frequentist Statistics" as often as we see "Introduction to Bayesian Statistics." Course titles such as "Introduction to Statistical Inference" and "Introduction to Statistics" often refer to non-Bayesian methods. Therefore, it is not unreasonable to assume that if students have some background in inference, it often is grounded in non-Bayesian perspectives.

I am not advocating that, in general, students should learn statistics from a Bayesian perspective at the elementary level. However, there surely are a few consequences for not doing so, including that instructors may need to facilitate a mindset shift for students between inferential paradigms. For better or worse, new content will pass through the filter of students' previous statistical backgrounds in non-Bayesian methods. Instructors can use this to the students' advantage. The next section suggests how instructors can utilize students' previous experiences in non-Bayesian statistics.

A second challenge, for similar reasons, is that students lack familiarity with various statistical distributions commonly used in Bayesian analysis, as well as their basic features: the support of the distributions, the functional form of the density, the kernel, the meaning of the parameters, and the shape of the distribution as a function of the parameters. Different from non-Bayesian parametric approaches, we must select distributions for the data *and* the unknown parameters. Selecting a distribution for the data in Bayesian analysis is not different than selecting for non-Bayesian analysis. However, Bayesian analysis employs a larger set of statistical distributions that typically are not covered in non-Bayesian courses, including Beta, Gamma, Dirichlet, Wishart, and many others. All other topics that are common in Bayesian analysis—for example, Markov chain Monte Carlo (MCMC) and sensitivity analysis—rely on a solid grasp of these elements and a transition to the Bayesian inference mindset.

MINDSET SHIFT USING SIDE-BY-SIDE EXAMPLES AND A FOUR-STEP MODELING APPROACH

Students who have some statistical inference background but are learning Bayesian statistics for the first time often struggle to understand the difference between estimating an unknown parameter in a non-Bayesian manner and updating the posterior distribution of the "parameter" in a Bayesian framework.

https://doi.org/10.1017/S1049096521001244 Published online by Cambridge University Press

The confusion is not difficult to clarify, but for that reason it can be overlooked by instructors. It is common to read phrases like "the estimator of the parameter ($\boldsymbol{\theta}$)" and "the true value of $\boldsymbol{\theta}$." Many researchers use a half-Bayesian approach in which they adopt the Bayesian computation apparatus but assume that a single fixed value of the parameters generated the data. If students are not familiar with the procedures, this half-Bayesian approach can contribute to their confusion. Using this approach, it is correct to state that the posterior distribution approaches that fixed value of $\boldsymbol{\theta}$ when the size of the data increases, but it is conceptually incorrect using the Bayesian framework to state the inferential question as learning the "true" value of $\boldsymbol{\theta}$ because $\boldsymbol{\theta}$ is a random variable—even when we assume that there is an underlying true fixed value of the parameter behind the data-generating process. More precisely, what we learn in Bayesian approaches is not the "true" value of $\boldsymbol{\theta}$ but rather its distribution —or the distribution of our beliefs about possible values of $\boldsymbol{\theta}$— after we take into account the data. Arguably, there is nothing wrong with a half-Bayesian approach; however, it can be confusing for a student learning the subject for the first time.

To facilitate a mindset shift from a non-Bayesian to a Bayesian paradigm, instructors can combine a four-step modeling approach when developing examples and periodically provide side-by-side solutions using the frequentist—or perhaps the "likelihoodist"—approach and the Bayesian paradigm.

To facilitate a mindset shift from a non-Bayesian to a Bayesian paradigm, instructors can combine a four-step modeling approach when developing examples and periodically provide side-by-side solutions using the frequentist—or perhaps the "likelihoodist"—approach and the Bayesian paradigm.

The four-step approach follows four modeling steps that the instructor can illustrate explicitly—and with students' participation—in every example of application of Bayesian models. The four steps are as follows:

1. Select the data model.
2. Select the prior model.
3. Derive the posterior.
4. Compute the posterior quantities of interest.

There are many benefits to explicitly following these four steps in every example and emphasizing each one as the examples progress. One benefit is that students do not see the analysis as a "bag of models" that we merely mechanically plug into the data but rather as a logical process in which the analyst makes a series of modeling decisions in each step. Instructors can use this to emphasize that the quality and adequacy of those modeling decisions must be assessed and to demonstrate that such an assessment is an essential part of the modeling process. Students will better understand the importance of grasping the logic and steps behind Bayesian modeling and not that they simply must learn an assortment of seemingly unrelated models. Another benefit is that the four steps empower students by providing a structure for modeling substantive problems and they can follow these steps themselves.

I use this four-step procedure as a guideline and motivation for all other topics I cover in introductory courses, and I explicitly restate and follow the steps in all examples. In terms of motivation to introduce other topics, step 2 leads naturally to issues of prior selection (e.g., Jeffrey's prior, conjugate priors, objective Bayes, and sensitivity) and how to handle them. Some challenges emerge for students when selecting the prior in step 2, which is discussed in the next section. Steps 3 and 4 lead naturally to issues of computing integrals when there is no analytical solution, which helps to motivate Monte Carlo methods, MCMC, and the issues that emerge with those procedures, including MCMC implementation, convergence diagnostics, and stopping rules.

I explicitly construct every example following these four steps and I often present Bayesian and non-Bayesian solutions side by side to help students recognize differences between the paradigms. Here is an example of a side-by-side comparison that instructors can adapt and use in their own course. Suppose we installed an alarm system in our house and we go away on a trip. We set up the alarm to ring on our cell phone, and we can remotely activate an additional anti-burglary tech system if we believe a burglary is in progress. The anti-burglary system is expensive to recharge after it has been activated, so we should use it only if we are highly confident that there is a burglar. These are our indicators:

$$X = \begin{cases} 1, & \text{alarm is active} \\ 0, & \text{alarm is not active} \end{cases}$$

$$\boldsymbol{\theta} = \begin{cases} 1, & \text{there is a burglary happening} \\ 0, & \text{there is no burglary happening} \end{cases}$$

The alarm manual contains information summarized in table 1.

Table 1
Anti-burglary Example

	$\boldsymbol{\theta}=0$ (no burglary)	$\boldsymbol{\theta}=1$ (burglary)
$x=0$	$p(x=0\mid\boldsymbol{\theta}=0)=0.95$ (alarm off, no burglary)	$p(x=0\mid\boldsymbol{\theta}=1)=0.01$ (alarm off, burglary happening)
$x=1$	$p(x=1\mid\boldsymbol{\theta}=0)=0.05$ (false alarm)	$p(x=1\mid\boldsymbol{\theta}=1)=0.99$ (alarm on, burglary happening)

https://doi.org/10.1017/S1049096521001244 Published online by Cambridge University Press

If we hear the alarm ($X=1$), is there a burglary happening ($\boldsymbol{\theta}=1$)? A non-Bayesian way to answer this question is to use the likelihood function to estimate $\boldsymbol{\theta}$. We choose the value of $\boldsymbol{\theta}$ that maximizes $p(X=1|\boldsymbol{\theta})$. Comparing $p(X=1|\boldsymbol{\theta}=0)=0.05$ to $p(x=1|\boldsymbol{\theta}=1)=0.99$, we choose $\boldsymbol{\theta}=1$. That is, we choose the $\boldsymbol{\theta}$ that makes the data more likely. This is the estimation step. Regardless of anything else that may be occurring, if we hear the alarm, we should adopt $\boldsymbol{\theta}=1$. There is no way to solve this problem using a frequentist approach, but we can decide to reject (or not) that estimated value using a likelihood method to estimate $\boldsymbol{\theta}$. I use these points to briefly discuss inference and compare the paradigms.

If we select $\boldsymbol{\theta}=1$, we are still not completely certain if there is a burglary happening because there is a small chance that it is a false alarm. What is the chance that there is a burglary happening if we hear the alarm? Formally restated, what is $p(\boldsymbol{\theta}|x=1)$? Only the Bayesian paradigm can answer that question, which is as follows:

$$p(\boldsymbol{\theta}|X=1)=\frac{p(X=1|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\sum_{\boldsymbol{\theta}=}^{1}p(X=1|\boldsymbol{\theta})p(\boldsymbol{\theta})}$$

The only unknown is $p(\boldsymbol{\theta})$, the prior. I use two possible choices for that quantity. In one case, we have no clue about the overall probability of burglary ($p(\boldsymbol{\theta}=1)=p(\boldsymbol{\theta}=0)=0.5$). In the other scenario, we know that the neighborhood is extremely safe ($p(\boldsymbol{\theta}=0)=0.9$). In this case, we reason about the probability of $\boldsymbol{\theta}$—that is, the fixed observed data $X=1$ changed our belief about $\boldsymbol{\theta}$.

We can work with other examples to emphasize the differences between confidence intervals and posterior probabilities. Here is one possible example: Assume that there is a probability $\boldsymbol{\theta}$ that a random person will vote for a party A. We take a random sample and 40 of 100 people state that they will vote for that party.

I discuss with students the selection of the data model that results in the choice of a binomial distribution (i.e., step 1 of the four-step procedure) and a uniform distribution for the prior (i.e., step 2) and show (i.e., step 3) that:

$$p(\boldsymbol{\theta}|x)\propto p(x|\boldsymbol{\theta})p(\boldsymbol{\theta})=\boldsymbol{\theta}^{x}(1-\boldsymbol{\theta})^{n-x}=\boldsymbol{\theta}^{40}(1-\boldsymbol{\theta})^{60}$$

We can use a simplified version of non-Bayesian inference side by side in this example. We can compute the point estimate and the confidence interval for the problem. Thus, we can show that in frequentist approaches, we compute the point estimate; however, we must ask what would happen with that estimator if we had observed a different sample. It helps to show that we must rely on the sampling distribution of the estimator—usually based on asymptotic results—think about repeated sampling, and construct confidence intervals based on the variance of the estimator. For this example, it is as follows:

$$\hat{\boldsymbol{\theta}}=\frac{40}{100}\quad;\quad CI\left(\hat{\boldsymbol{\theta}}\right)=\left[0.4-t_{\alpha}\sqrt{\mathbb{V}\mathrm{ar}\left[\hat{\boldsymbol{\theta}}\right]},0.4+t_{\alpha}\sqrt{\mathbb{V}\mathrm{ar}\left[\hat{\boldsymbol{\theta}}\right]}\right]$$
$$=[0.303,0.496]$$

I use simplified versions of non-Bayesian procedures because teaching advanced, non-Bayesian procedures is not

Table 2

Comparing Bayesian and Non-Bayesian Approaches

Non-Bayesian (Typical) Approaches	Bayesian (Typical) Approach
$\boldsymbol{\theta}$ is a **fixed** and **unknown** quantity	$\boldsymbol{\theta}$ is **random** and **unknown**
X is "**random**": Some procedures care about the samples that we could have observed	X is "**fixed**": We care only about the sample that we did observe
Learning about $\boldsymbol{\theta}$:	
Estimate $\boldsymbol{\theta}$ with $\hat{\boldsymbol{\theta}}$: maximum likelihood, method of moments, least square, etc.	We compute the posterior: $p(\boldsymbol{\theta}\|x)\propto p(x\|\boldsymbol{\theta})p(\boldsymbol{\theta})$
Estimation summaries:	
Point estimates; standard errors and confidence intervals of the estimator	Posterior mean, median, quantiles, and regions of high (posterior) probability
Properties of the estimation:	
a. Bias (small-sample properties)	a. Coverage (in some cases)
b. Root mean squared error (RMSE)	b. Sensitivity to the choice of the prior
c. Consistency, efficiency (asymptotic properties)	
Inference:	
Reject hypothetical values of $\boldsymbol{\theta}$ that make the observed data highly unlikely to happen	Consider values of $\boldsymbol{\theta}$ with high posterior probability
Measures of quality of the inference:	
Bias, efficiency, consistency, power of the test, sensitivity to model assumptions, etc.	a. Posterior predictive checks
	b. Bayes Factor, Bayesian Information Criterion (BIC), Deviance Information Criterion (DIC)
	c. If MCMC was used: Convergence checks
	d. Sensitivity to prior distribution

https://doi.org/10.1017/S1049096521001244 Published online by Cambridge University Press

Table 3

Example of a Table for Selection of Prior Distributions

Candidate	Support	Density	Kernel
Uniform	[0,1]	$p(\boldsymbol{\theta})=1$	1
Beta	[0,1]	$p(\boldsymbol{\theta}\|\alpha,\beta)=\frac{1}{B(\alpha,\beta)}\boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}$	$\boldsymbol{\theta}^{\alpha-1}(1-\boldsymbol{\theta})^{\beta-1}$
Truncated Normal	[0,1]	$p(\boldsymbol{\theta}\|\mu,\sigma,0,1)=\frac{1}{\sigma(\Phi(1)-\Phi(0))}\phi(\boldsymbol{\theta})$	$\exp\left\{-\left(\frac{\boldsymbol{\theta}-\mu}{\sigma}\right)^2\right\}$

the goal of the course. Readers can find good discussions and additional examples to compare different inferential paradigms in Christensen (2005). The number of details to include depends on the students' background. The idea is to leverage the understanding of Bayesian inferential reasoning using students' previous backgrounds and to avoid creating confusion between the paradigms.

During the course, I incrementally construct table 2 until it is fully filled, which naturally occurs as we approach the final lectures. The table is stylized to emphasize key differences, and it works well to help with the mindset shift between Bayesian and non-Bayesian paradigms. After constructing intentionally simple examples of Bayesian inference—such as those presented previously—to compare against non-Bayesian cases, I present the table again and derive the non-Bayesian and Bayesian inference, side by side. Although I use the four-step procedure to develop each example, I compare Bayesian and non-Bayesian approaches in only some of them.

In summary, in addition to helping with the mindset shift and to internalizing and understanding the Bayesian framework, the side-by-side examples and the four-step approach serve three other purposes: (1) to ground the comparison between Bayesian and other approaches and make their differences more apparent; (2) to describe a simple step-by-step procedure to derive Bayesian models; and (3) to motivate the discussion of other topics, such as MCMC and sensitivity analysis.

A BAG OF PRIOR DISTRIBUTIONS

Another challenge that can emerge when teaching Bayesian statistics is related to the process of selecting prior distributions (i.e., step 2 in the four-step procedure). Students often lack familiarity with both the process of selecting a distribution to model the data and the properties of various distributions often used in Bayesian analysis to model the prior. These properties include the support, the functional form of the density, the kernel, the meaning of the parameters, and the shape of the distribution as a function of the parameters. It is likely that many students have never learned about Beta, Gamma, and other common distributions. This issue is easy to address and, for that reason, it also often is overlooked by instructors—which can lead to confusion later when introducing more complex models.

Arguably, the support of a distribution and how its shape depends on the parameters are the more important aspects for practical applications. Therefore, it is worth covering those topics using examples, which can be constructed using the four-step procedure described in the previous section.

It also is useful to provide a table such as table 3 that students can reference quickly when trying to find a prior model. Following the previous example with the proportions, we can present a table with prior options and compare their properties. This develops the students' intuition of the process behind prior selection, introduces some distributions, emphasizes the distributions' support, and teaches how to think about their parameters and how they affect the shape of the distribution. Which prior should be selected from the previous example about the likelihood of a random person voting for party A? Some of the options are as follows:

This exercise of prior selection is repeated in each example I present and allows students to choose among the options of priors I provide, which usually is a superset with options the support of which is outside of the values that the parameter $\boldsymbol{\theta}$ can take. The exercise can be supplemented with animations to show the effect of the parameter values on the shape of the prior; subsequently, similar animations can be used to show their effect on the posterior (i.e., sensitivity).

DISCUSSION

Important topics could not be covered in this article due to space limitations, including decisions about the scope of sensitivity analysis and the issues with MCMC estimation and diagnostics. These are certainly essential elements that instructors must address when teaching Bayesian statistics, and they have their own challenging aspects. I also did not discuss pedagogical approaches, including active-learning methods and a flipped classroom.

I focused instead on two issues related to teaching Bayesian statistics for non-statistician students at the graduate level who have some background in statistics: the mindset shift from non-Bayesian to Bayesian reasoning about inference and the familiarity with a bundle of distributions for prior selection, which typically is not taught in other statistics courses.

https://doi.org/10.1017/S1049096521001244 Published online by Cambridge University Press

It is up to instructors to identify whether their students are facing these challenges. Quizzes and checkpoints can be useful for that purpose. The four-step approach to build Bayesian models and inference, frequently using side-by-side examples of Bayesian and non-Bayesian approaches, and the table of priors—all suggested here—can help if students are experiencing the learning difficulties discussed in this article. ■

REFERENCES

Albert, Jim. 1997. "Teaching Bayes' Rule: A Data-Oriented Approach." *The American Statistician* 51 (3): 247–53.

Berry, Donald A. 1997. "Teaching Elementary Bayesian Statistics with Real Applications in Science." *The American Statistician* 51 (3): 241–46.

Christensen, Ronald. 2005. "Testing Fisher, Neyman, Pearson, and Bayes." *The American Statistician* 59 (2): 121–26.

Gelfand, Alan E., and Adrian F. Smith. 1990. "Sampling-Based Approaches to Calculating Marginal Densities." *Journal of the American Statistical Association* 85 (410): 398–409.

Gelman, Andrew. 2005. "A Course on Teaching Statistics at the University Level." *The American Statistician* 59 (1): 4–7.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2014. *Bayesian Data Analysis.* Second edition. Boca Raton, FL: Chapman & Hall/CRC.

Gelman, Andrew and Deborah A. Nolan. 2017. *Teaching Statistics: A Bag of Tricks.* Oxford: Oxford University Press.

Geman, Stuart, and Donald Geman. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6:721–41.

Ghosh, Jayanta K., Mohan Delampady, and Tapas Samanta. 2007. *An Introduction to Bayesian Analysis: Theory and Methods.* Berlin, Germany: Springer Science & Business Media.

Gill, Jeff. 2014. *Bayesian Methods: A Social and Behavioral Sciences Approach.* Boca Raton, FL: CRC Press.

Hastings, Wilfred K. 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Applications." *Biometrika* 57 (1): 97–109.

Hoff, Peter D. 2009. *A First Course in Bayesian Statistical Methods.* New York: Springer.

Jackman, Simon. 2004. "Bayesian Analysis for Political Research." *Annual Review of Political Science* 7:483–505.

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences.* Hoboken, NJ: John Wiley & Sons.

Kruschke, John. 2015. *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan.* Second edition. Cambridge, MA: Academic Press.

Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *Journal of Chemical Physics* 21 (6): 1087–92.

Moore, David S. 1997. "Bayes for Beginners? Some Pedagogical Questions." In *Advances in Statistical Decision Theory and Applications*, ed. Subramanian Panchapakesan and Narayanaswamy Balakrishnan, 3–17. New York: Springer.

Utts, Jessica, and Wesley O. Johnson. 2008. "The Evolution of Teaching Bayesian Statistics to Nonstatisticians: A Partisan View from the Trenches." *The American Statistician* 62 (3): 199–201.

https://doi.org/10.1017/S1049096521001244 Published online by Cambridge University Press