

Bayes and MCMC for Undergraduates

Jeff Witmer

Department of Mathematics

Oberlin College

Oberlin, OH 44074

jeff.witmer@oberlin.edu

Jeff Witmer is Professor, Department of Mathematics, Oberlin College, Oberlin, OH 44074 (E-mail: jeff.witmer@oberlin.edu). The author thanks Murray Clayton, George Cobb, Bill Notz, the referees, associate editor, and editor for helpful comments.

Abstract

Students of statistics should be taught the ideas and methods that are widely used in practice and that will help them understand the world of statistics. Today, this means teaching them about Bayesian methods. In this article I present ideas on teaching an undergraduate Bayesian course that uses Markov chain Monte Carlo and that can be a second course or, for strong students, a first course in statistics.

KEY WORDS: Bayesian inference, Markov chain Monte Carlo, education

I INTRODUCTION

In his recent paper on the future of statistics education, George Cobb writes of the importance of teaching of Bayesian thinking (Cobb 2015) and I heartily agree with his position that Bayesian methods can and should be taught to undergraduates.

One advantage of teaching Bayes using Markov chain Monte Carlo (MCMC) is that the power and flexibility of “Bayes via MCMC” lets students address the following kind of question: Who is a better free throw shooter, Quentin Snider or Willie Cauley-Stein? It might seem that it is easy to answer this question by comparing sample data from each of these two basketball players. When I posed this question to students in my Bayesian course we had data from the 2014-15 season. Snider had a bad year shooting free throws, making only 21 of 38 (55.3%). Cauley-Stein made 79 of 128 free throws (61.7%), so he did better than Snider, and that would seem to settle the matter. But looking at data from

16 guards, including Snider, and 8 centers including Cauley-Stein, one can see that centers generally aren't good free throw shooters. How can we use the data, plus a belief that guards might be better free throw shooters than centers, to estimate how Snider and Cauley-Stein will compare in the long run?

My students used the data to fit a hierarchical model that has a top level with a parameter for overall free throw shooting performance among all players, a middle level that allows for guards to be systematically different from centers, and a lower level that allows players within a position (e.g., guards) to differ from one another. The fitted hierarchical model found a 79% posterior probability that Snider is more skilled than Cauley-Stein at shooting free throws – this despite the fact that Snider did worse than Cauley-Stein in the 2014-15 season. It then came as no surprise to me that in the following season Snider shot 75% from the free throw line while Cauley-Stein made only 64.8% of his free throws.

This example illustrates that Bayes using MCMC is flexible – well beyond what students see in a traditional course. Moreover, Bayes is natural: People use probability in loose, informal ways every day and in a sense, every student is a subjective Bayesian. Indeed, learning to speak is a Bayesian process (Norris, 2006; Dowman et al. 2008) as Bayes' Theorem comports with a natural and fruitful way of thinking.

The use of Bayesian statistics has grown considerably over the past several years. I looked through three years worth of articles in the Applications and Case Studies section

of the *Journal of the American Statistical Association* (issues 500 – 511, covering December 2012 – September 2015). More than half the articles in those issues (57 out of 110) involved Bayesian analyses¹. A recent review of clinical trials at one cancer center found that one-third of phase I or II drug trials used Bayesian designs and analyses (Biswas et al. 2009). An increasing proportion of empirical applications using macroeconomic time series analysis employ Bayesian techniques, about half in recent years (Tallman 2014). Such work is influenced by people like Christopher Sims, who delivered a talk on “Why Econometrics Should Always and Everywhere be Bayesian” (Sims 2007) shortly before he won the Nobel Prize in economics. The growth of Bayesian applications is largely due to better software, particularly for implementing MCMC.

The teaching of Bayesian methods can be done in a second course in statistics, but a Bayesian methods course that uses MCMC can be taught to students who have never taken a statistics course. Whatever the level, a modern Bayesian course should go far beyond simple applications of Bayes’ Theorem. I argue for a course that teaches methods parallel to traditional topics, such as t-tests, regression, and analysis of variance, and that goes beyond what can be done with traditional (frequentist) methods.

In the sections that follow I present an overview of a course that I have taught twice and that has been taken by students with or without any formal background in statistics. I include details of the kinds of topics and techniques we cover.

¹ This does not include another five articles in which only Bayes’ Theorem was used.

II MCMC

Computing power has made Bayesian methods accessible and has changed statistical practice. My course is heavily dependent on computing; indeed, the title of the course is “Bayesian Computation.” We use the Jags program via R to implement Markov chain Monte Carlo and find posterior distributions in a variety of settings. I’m using a book (Kruschke 2015) that requires no mathematics beyond some exposure to first-semester calculus, and even that is optional².

Students in my course do not become programmers; they just execute code that someone else has written. Moreover, the professor doesn’t need to be adept at creating code. Some familiarity with R is certainly helpful, but it is quite easy to use the scripts that come with the textbook³. Indeed, I have had a few students complain that I don’t lead them down the path of writing code to handle arbitrary situations; instead, I (mostly) have them make simple edits to existing code. To such students I plead “guilty as charged,” as there is only so much that I want to take on in a first course on Bayesian methods. We cover the Bayesian equivalents to t-tests, comparisons of proportions, regression, and ANOVA; in short, we cover the topics that students see from a frequentist perspective if they take my traditional STAT 101 course.

² Other books (Bolstad, 2007, Gill 2008, Albert 2009, Hoff 2009, et al.) are available that also do not presume much sophistication on the part of the audience. There are a number of books that are written at a higher level (Christensen et al. 2011, Gelman et al. 2014, McElreath 2016) that would challenge most undergraduates at the introductory level, but that might be quite suitable for students with strong backgrounds.

³ Moreover, installing the BayesianFirstAid package in R and then changing `t.test()` to `bayes.t.test()`, e.g., is easy for an R user who wants to use MCMC but doesn’t want to work with `rjags` scripts.

Students in my Bayesian course don't actually need to know calculus, although it helps to have some idea of what integration is. Likewise, they don't need to have taken a previous statistics course, although it may help (or hurt?) to have seen frequentist ideas of P-values and confidence intervals as points of comparison when we discuss Bayes' factors and highest density intervals. Many of my students have previously taken both calculus and statistics, but some have taken only one of those two (but at least one of the two).

III A SINGLE PROPORTION

Early in my course I introduce model building with a simple construction of the degree of uncertainty about a Bernoulli parameter, θ . We look at this in three ways. (1) We use a $\text{Beta}(a,b)$ as the prior distribution on θ and get a posterior that is $\text{Beta}(a + z, b + n - z)$, where z is the number of successes in n Bernoulli trials. I don't prove this result about the posterior, but I do present the simple mathematics that shows the posterior mean as a weighted average of the prior mean and the data mean. (2) We use a discrete prior for θ that lives on grid points between 0 and 1. (3) We use MCMC. I show these three approaches in parallel, noting that a discrete prior with many grid points (2) gives a good approximation to the theoretical result (1) and MCMC (3) gives a good approximation to both of these.

As an example, I ask the question "What percentage of students on our campus are vegetarians?" and then present data from a sample of 136 students, 19 of whom are

vegetarians. Appendix 1 shows output from a program that takes a Beta prior plus data and produces a Beta posterior, along with MCMC output from a second program.

Bayesian methods, in contrast to their frequentist counterparts, incorporate information that was available prior to the collection of the data in a clear and consistent way⁴.

To explore the role of the prior, I present pre-posterior predictive analysis with the vegetarian data and ask the question “Is this prior sensible?” To choose a prior for the vegetarian parameter I specified a prior mode – I chose 0.20 as a guess of the campus-side vegetarian proportion – and a weight – I chose 20 in answer to the question “How many observations is my prior worth?” I then entered 0.20 and 20 into a program and saw that a Beta(4.6,15.4) matches those values. From here I generated a random binomial of size 25 for each of 2000 points drawn from that Beta(4.6,15.4) prior. A histogram of the binomial counts was consistent with my expectations, so I was comfortable with the Beta(4.6,15.4) prior. That is, the histogram of possible binomial counts changed the question “Is this a sensible prior?” into the question “Do I expect data that follow this kind of a distribution?”

Although I discuss and use informative priors with my students, for much of the semester we avoid the objectivity versus subjectivity issue by using noninformative or mildly informative priors⁵.

⁴ Regarding current statistical practice, about half of the JASA papers that use Bayesian methods use informative priors and half do not.

IV TWO PROPORTIONS AND HIERARCHICAL STRUCTURE

MCMC makes it easy to use a prior that has a hierarchical structure. This is often a natural choice and one that tends to soften the link between subjective prior belief and the posterior distribution. Appendix 2 presents in more detail the opening example on free throw shooting among guards and centers, using a fairly simple hierarchical prior that has a parameter for each player, but that also allows for guards to be systematically different from centers. The flexibility of this model is a feature that is easily incorporated into Bayesian analyses.

V MEANS

The Bayesian paradigm fosters flexibility aside from the use of hierarchical models. In particular, when using MCMC it is not necessary to stipulate that the error term is Gaussian. A traditional t-test assumes a Gaussian likelihood and uses a t-distribution for the test statistic. In contrast, the command `bayes.t.test(y, mu=30)` in the R package `BayesianFirstAid` accepts a vector of data `y` and runs MCMC for a single mean *using a t likelihood*, with `mu=30` as a null hypothesis comparison point. Allowing the data to come from a long-tailed distribution can render immaterial the often vexing question “What should I do if I think that one of the observations is an outlier?”

⁵ The R package `BayesianFirstAid` incorporates noninformative priors for Bayesian analyses that mimic standard methods; for example, the `prop.test()` command becomes `bayes.prop.test()` and uniform priors are used on parameters.

We rarely expect an effect to be exactly zero. Thus, Bayesian reasoning focuses on parameter estimation, rather than hypothesis testing. Appendix 3 shows a comparison of two means using either a web applet or an easily edited R script. The output gives the estimated difference in the means, along with a 95% credible interval for that difference.

V REGRESSION

The usual presentation of regression in an introductory course imposes the condition that the error term has a normal distribution. Outliers are then considered and points are either included or excluded from the analysis. But just as with means, in a regression model fit using MCMC it is easy to use a t distribution for the error term, with the degrees of freedom being a parameter estimated from the data. Appendix 4 details how this can be done in the context of a well-known regression example.

VI OTHER TOPICS

I spend some time on one-way and two-way Bayesian Analysis of Variance and I show an example or two of logistic regression; other professors would make other choices.

Now that Stan is available I mention Stan and Hamiltonian Monte Carlo, but I don't expect students to learn how to use Stan.

I also spend time “looking under the hood,” as it were, regarding MCMC. I don't expect students to become adept at coding in Jags (or BUGS or Stan), but I do want them to use code and to understand what MCMC is and how it works. To that end, I go beyond what the textbook shows (i.e., what I do is completely optional, and I would skip this if I were

not teaching strong students) and spend about three weeks on a unit in which I introduce Markov chains, with the canonical example of rain/shine and a 2×2 matrix of transition probabilities. I then use R to raise the transition matrix to a large power to show the limiting, stabilized distribution. Next, I show a chain with three states and a simple walk of a few dozen steps among the three states before talking about a setting with many states and a long chain. Then I introduce more Markov chain ideas: reducibility, reversibility, the ergodic property, etc. I follow that with a “proof” that the Metropolis algorithm works. I put proof in quotes here because I do a lot of hand-waiving leading up to the final step, in which I show that Metropolis uses a reversible transition process, which completes the “proof.” Finally, I say that Gibbs sampling is of the same spirit as Metropolis, but can be more efficient, etc. In summary, I try to make the MCMC programs that we use be not a black box, but perhaps a gray box instead. But what I require of students is that they manipulate these programs (R scripts) in weekly homework in which they need to read in data and edit the lines that specify the prior and the likelihood function.

VIII SUMMARY

Students should be exposed to the world of Bayesian statistics. At a minimum they should see an application of Bayes’ Theorem, but they should also see MCMC and fully Bayesian analyses of data. We have had the necessary computing power for quite a while, but we haven’t had easy-to-use software and textbooks that make Bayes accessible to undergraduates. That has recently changed, so that today a Bayesian course that teaches

MCMC is available to a variety of students. A Bayesian course should appear at the entry level of the curriculum⁶.

The world of statistics users has moved into using Bayesian methods widely, and to good result. It is time for statistics educators to join in.

APPENDIX 1

Veggie %. Figure 1 shows the result of running a program called `BernBetaExample.R` after editing a few lines of the code. One needs to input the two Beta parameters plus the data, where “number of flips” is the number of observations in the sample and “number of heads” is, in our example, the number of vegetarians.

```
Prior = c(4.6,15.4)    # Specify Prior as vector with the two shape parameters.  
N = 136                # The total number of flips.  
z = 19                # The number of heads.
```

⁶ I recognize that curricular decisions vary greatly from place to place and I don't mean to minimize the constraints that local politics can impose. Some faculty will find that the only feasible place to add a Bayesian course is as an upper-level elective. My main hope is that the course harnesses the power of MCMC.

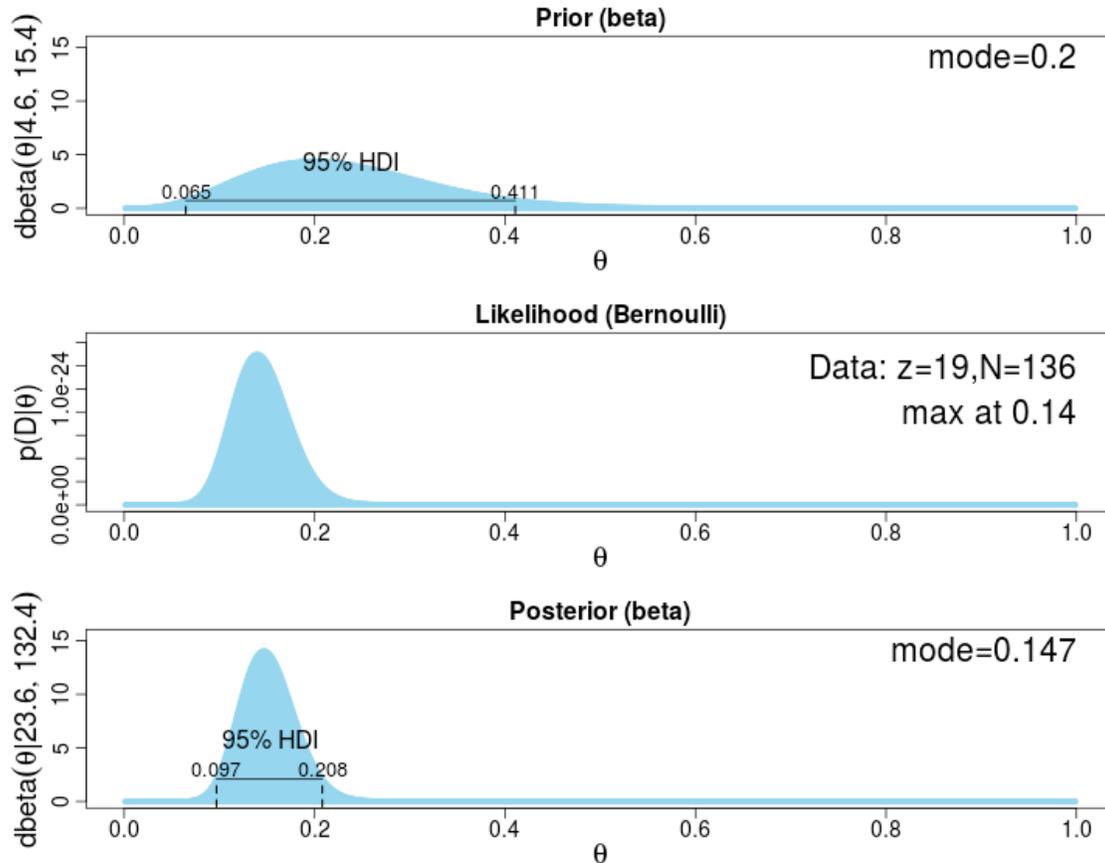


Figure 1. Prior, likelihood, and posterior distribution of the percentage of vegetarians on campus.

Figure 2 shows the result of using MCMC, running a program called Jags-Ydich-Xnom1subj-MbernBeta-Veggie.R. The rather long program name says “I am going to use Jags as my MCMC engine, the response variable Y is dichotomous, the predictor X is a single subject nominal variable, the model is a Bernoulli likelihood with a Beta prior, and I’m analyzing the vegetarian data so I’m adding ‘Veggie’ at the end.” The program displays the posterior mode of 0.146 rather than the posterior mean of 0.145, but the user

has the option of asking for the mean. The only thing the user needs to specify is the data⁷, by inputting the numbers 19 and 136 in the following line:

```
myData = data.frame(y = c(rep(0, 136 - 19), rep(1, 19)))
```

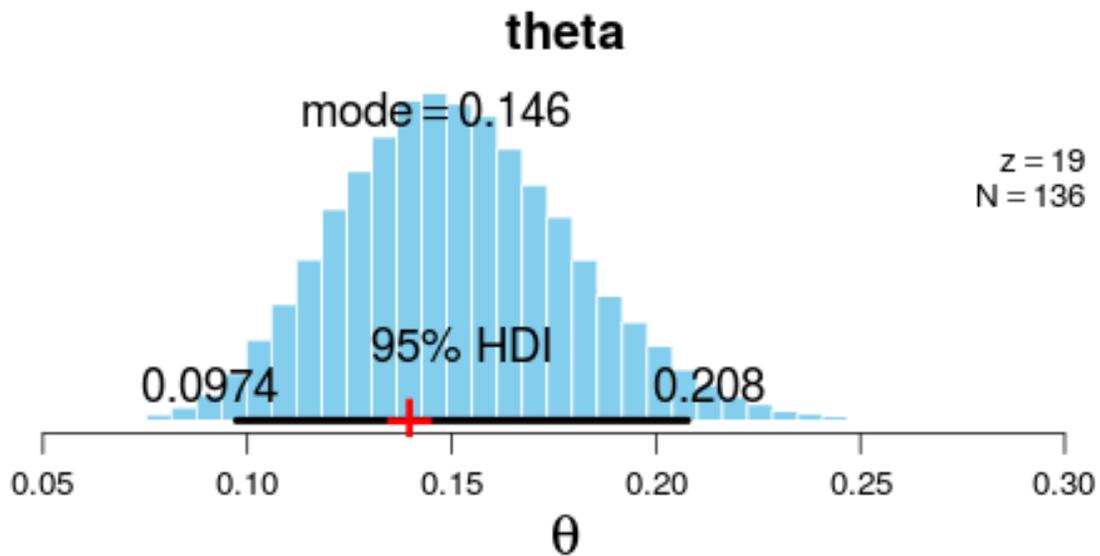


Figure 2. Output from running MCMC with a Beta(4.6,15.4) prior and data of 19 out of 136 students being vegetarians. The 95% highest density interval from MCMC agrees with the theoretical 95% HDI shown in Figure 1.

Figure 3 is a histogram of 2000 binomial counts from a pre-posterior predictive analysis in the vegetarian setting.

⁷ A uniform prior is used by default. I used a Beta(4.6, 15.4) prior when creating Figure 2, which had no material effect on the results when compared to using a Beta(1,1) prior.

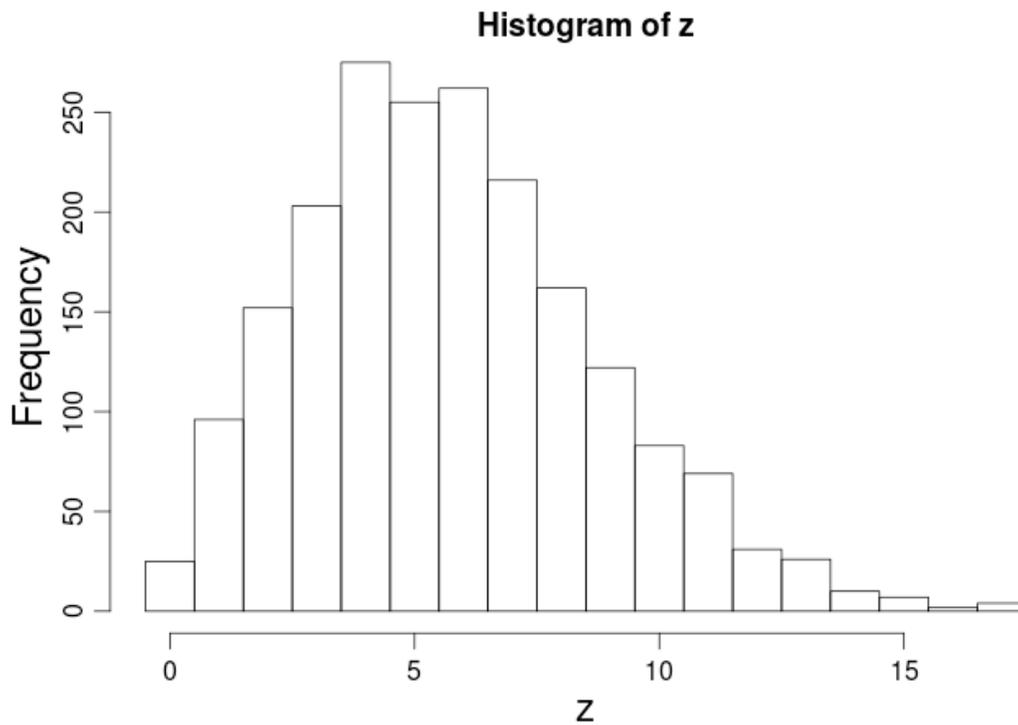


Figure 3. Pre-posterior predictive analysis of vegetarian question when the prior is Beta(4.6,15.4).

APPENDIX 2

Free Throw Shooting and Hierarchical structure. I collected data on the centers and guards for the teams that made it to the “Elite Eight” in the 2015 NCAA basketball tournament. The free throw shooting performance of guards ranged from a low of 55.3% to a high of 88.8% success, compared to centers who ranged from a low of 47.9% to a high of 77.7%. A naïve analysis might pool together all hits and misses for each position and compare the aggregate guard percentage (78.6%) to the aggregate center percentage (61.3%), as if all guards were interchangeable and all centers were interchangeable. A

conservative analysis might treat all players as separate, ignoring the fact that guards are similar to one another and centers are similar to one another.

It is easy to take a middle path of fitting a hierarchical model and conducting a Bayesian analysis, which my students see during week six of the semester. Denote each player's ability with a parameter θ_{Player} , let the θ s for the centers come from a $\text{Beta}(a_1, b_1)$ distribution and the θ s for the guards come from a $\text{Beta}(a_2, b_2)$ distribution, and let the parameters of the two Betas come from hyperpriors that describe prior belief about typical free throw shooting success. E.g., using a $\text{Beta}(30, 10)$ for the mode, ω , (where $\omega = (a-1)/(a+b-2)$) and a diffuse gamma distribution on the concentration, κ , (where $\kappa = a+b$) says that we expect basketball players to make about 75% of their free throws, without specifying that guards are better than centers. The posterior distributions tell us that there is a 92% chance that in general guards are better free throw shooters than centers. Beyond that, a hierarchical model allows us to pool information among centers and among guards, which leads to the following comparison. Willie Cauley-Stein, a center, made 79 of 128 free throws, for a 61.7% success rate. Quentin Snider, a guard, made 21 of 38 free throws, for a 55.3% success rate. The small number of attempts by Snider combined with the fact that he is a guard suggests that in the long run he will do quite a bit better than 55.3% and we find that there is a 79% posterior probability that Snider is more skilled than Cauley-Stein at shooting free throws⁸.

⁸ During the 2015-16 season, Snider made 45 of 60 free throws (75%) while Cauley-Stein made 81 of 125 free throws (64.8%). Given the Bayesian analysis above, this Snider advantage comes as no surprise.

Figure 4 is a screenshot of output comparing Snider and Cauley-Stein. The red plus sign shows the difference in sample proportions, but the posterior is not centered there. Instead, the hierarchical model produces a posterior with a mode of about 0.07, with 79% of the distribution above zero.

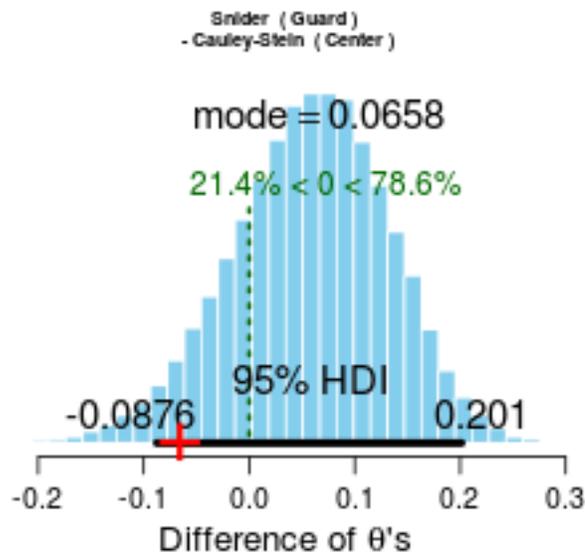


Figure 4 Posterior with 95% highest density interval for the difference in free throw shooting abilities of Snider and Cauley-Stein.

APPENDIX 3

Comparing two populations. Myocardial blood flow (MBF) was measured for two groups of subjects after five minutes of bicycle exercise (Namdar et al. 2006). One group was given normal air to breath (“normoxia”) while the other group was given a gas mixture to breathe with reduced oxygen (“hypoxia”) to simulate high altitude. The data (ml/min/g) are

Normoxia: 3.45, 3.09, 3.09, 2.65, 2.49, 2.33, 2.28, 2.24, 2.17, 1.34

Hypoxia: 6.37, 5.69, 5.58, 5.27, 5.11, 4.88, 4.68, 3.50

We want to know what these data indicate about the difference between training at normal altitude (normoxia) and training at high altitude (hypoxia).

One can go to the website http://sumsar.net/best_online/ and enter the data and hit the “Click to start!” button. Within a few seconds results appear that are based on a model of independent t distributions for the data. Diffuse (non-informative) priors are used for the population means, the degrees of freedom, and the population standard deviations. The posterior distribution of $\mu_1 - \mu_2$ is graphed and the mean is seen to be -2.65. The 95% Highest Density Interval is (-3.53, -1.73) and the probability that $\mu_1 < \mu_2$ is nearly 100%. Figure 5 is a screenshot.

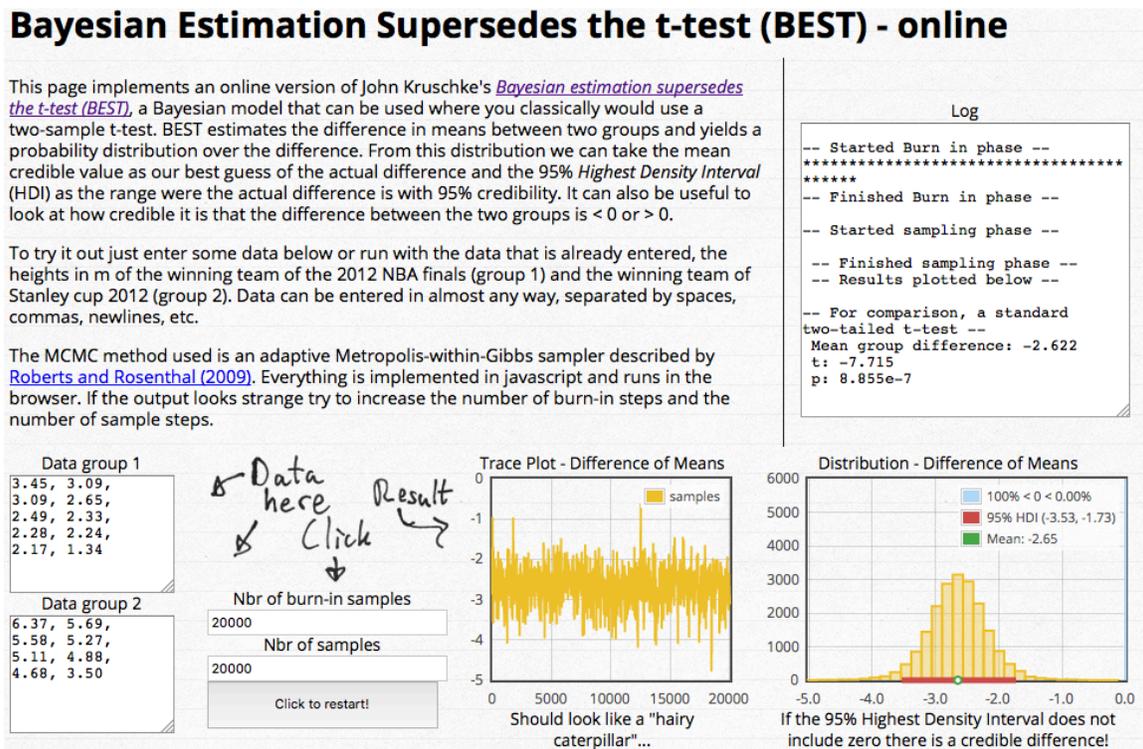


Figure 5 Screenshot of using the BEST website.

My students would analyze these data using an R script that accompanies the textbook. The student can easily choose a t likelihood rather than a normal likelihood, as this amounts to simply changing “ $y[i] \sim \text{dnorm}(\dots)$ ” to “ $y[i] \sim \text{dt}(\dots)$ ” within the code. But it isn’t even necessary to make that change, as the textbook programs include a script that has a t density pre-selected.

Figure 6 is a screenshot of output from running the script with these data.

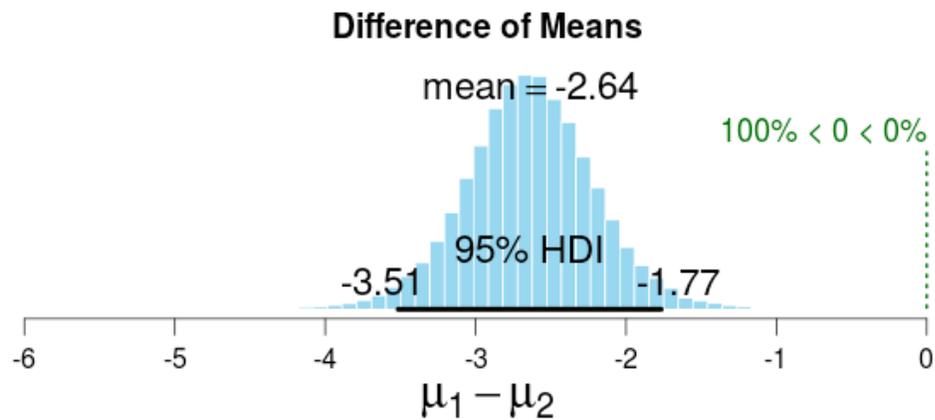


Figure 6 Posterior distribution of the difference in means between hypoxia and normoxia groups.

To get this output, I took an existing R script – downloaded from the textbook website – and edited three lines, stating the name of the data file, the name of the response variable, and the name of the predictor (group) variable:

```
myDataFrame = read.csv( file="ExerciseHypoxia2.csv" )  
yName="MBF"  
xName="Oxygen"
```

All I needed to do after that was to run the program. I might have changed the default comparison point from zero (“Are the two means equal?”) to something else (e.g., “Is the difference in means at least 3?”), and I might have changed the default, vague, prior to an informative prior, but those are options, not requirements.

APPENDIX 4

Regression. A well-known data set contains the winning (i.e., gold medal) men’s long jump distance in the Olympics and the year. Bob Beamon’s phenomenal jump in 1968 of 8.9 meters broke the previous world record by 55 cm and results in the 1968 data point being an outlier. Under a frequentist analysis the 95% confidence interval for the slope of the regression line for predicting jump distance (Y) from year (X) is (1.12, 1.69) if all of the data are used but is (1.14, 1.60) if the 1968 data point is deleted. The residual standard error when all data are used is 0.238 but this changes to 0.191 if the 1968 data point is removed. During week eleven of the semester I cover Bayesian regression. Rather than delete the 1968 point, we conduct a Bayesian analysis that replaces the usual condition of normally distributed errors with the condition that $Y|X$ has a t distribution on v degrees of freedom, with v as a random variable. This leads to fitting a Bayesian regression model with four parameters: the slope and intercept of the line, the standard deviation of the error term, and the degrees of freedom.

To fit this model, I have my students use an R script called `Jags-Ymet-Xmet-Mrobust-Example.R`, the name of which indicates that both the response Y and the predictor X are continuous (metric) and the model is robust (i.e., a t rather than a normal likelihood).

They need to edit the script to specify where the data are to be found and what the names are of the variables. This is done with the following lines, the fourth of which rescales the response variable to make the output easier to read:

```
library(Stat2Data)  
data(LongJumpOlympics)  
myData = LongJumpOlympics  
myData$Year = (myData$Year - 1900)/100  
xName = "Year" ; yName = "Gold"
```

Using the non-informative priors that are built into the script produces a 95% HDI on the slope of (1.13, 1.67). The posterior mode for the degrees of freedom is approximately 18 and the posterior mode for σ is 0.212, which sits between the residual standard errors that the two frequentist analyses yield.

REFERENCES

Albert, J. (1997), "Teaching Bayes' Rule: A Data-Oriented Approach," *The American Statistician*, **51**, 247 – 253.

Albert, J. (2002), "Teaching Introductory Statistics from a Bayesian Perspective," *Proceedings of the Sixth International Conference on Teaching Statistics*. Cape Town, South Africa.

Albert, J. (2009), *Bayesian Computation with R*. New York, NY: Springer.

Albert, J. and Rossman, A. (2009), *Workshop Statistics: Discovery with Data, A Bayesian Approach*. (http://bayes.bgsu.edu/nsf_web/workshop.bayes.pdf)

Bayarri, M. and Berger, J. (2004), “The Interplay of Bayesian and Frequentist Analysis,” *Statistical Science*, 19, 58 – 80.

Berger, J. and Berry, D. (1988), “Statistical Analysis and the Illusion of Objectivity.” *The American Scientist*, 76, 159 – 165.

Berry, D. (1995), *Statistics: A Bayesian Perspective*. Belmont, CA: Wadsworth.

Biswas, S., Liu, D., Lee, J., and Berry, D. (2009), “Bayesian Clinical Trials at the University of Texas M. D. Anderson Cancer Center,” *Clinical Trials*, 6, 205 – 216.

Bolstad, W. (2007), *Introduction to Bayesian Statistics, 2nd ed.* Hoboken, NJ: Wiley.

Christensen, R., Johnson, W., Branscum, A., and Hanson, T. E. (2011), *Bayesian Ideas and Data Analysis*. Boca Raton, FL: CRC Press.

Cobb, G. (2015), “Mere Renovation is Too Little Too Late: We Need to Rethink our Undergraduate Curriculum from the Ground Up,” *The American Statistician*, **69**, 266 – 282.

Dempster, A. P. (1971), “Model Searching and Estimation in the Logic of Inference,” in V. P. Godambe and D. A. Sprott, *Foundations of Statistical Inference*, pp. 56-78. Holt Rinehart and Winston.

Dowman, M., Savova, V., Griffiths, T. L., Kording, K. P., Tenenbaum, J. B., and Purver, M. (2008), “A Probabilistic Model of Meetings That Combines Words and Discourse Features.” *IEEE Transactions on Audio, Speech, and Language Processing*, **16** (7), 1238 - 1248.

Geisser, S. (1993), *Predictive Inference: An Introduction*. Boca Raton, FL: Chapman & Hall.

Gill, J. (2008), *Bayesian Methods: A Social and Behavioral Sciences Approach*. Boca Raton, FL: Chapman & Hall.

Hoff, P. (2009), *A First Course in Bayesian Statistical Methods*. New York, NY: Springer.

Horton, N. et al. (2014), “Curriculum Guidelines for Undergraduate Programs in Statistical Science,” [online], Available at <http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf>

Huelsenbeck, J. P., Ronquist, F., Nielsen, R., and Bollback, J. P. (2001), “Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology.” *Science*, **294**, 2310 – 2314.

Kruschke, J. (2015), *Doing Bayesian Data Analysis*. Waltham, MA: Elsevier.

Lin, X., Genest, C., Banks, D.L., Molenberghs, G., Scott, D.W., and Wang, J.-L. (2014), *Past, Present, and Future of Statistical Science*, CRC Press.

Madigan, D., Bartlett, P., Buhlmann, P., Carroll, R., Murphy, S., Roberts, G., Scott, M., Tavaré, S., Triggs, C., Wang, J.-L., Wasserstein, R.L., and Zuma, K. (2014), *Statistics and Science: A Report of the London Workshop on the Future of the Statistical Sciences* (<http://bit.ly/londonreport>).

McElreath, R. (2016), *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. Boca Raton, FL: CRC Press.

Moore, D. (1997), “Bayes for Beginners? Some Reasons to Hesitate” (with rejoinder), *The American Statistician*, **51**, 254 – 261, 272 – 274.

Namdar, M., Koepfli, P., Grathwohl, R., Siegrist, P. T., Klainguti, M., Schepis, T., Deleloye, R., Wyss, C. A., Gaemperli, O., and Kaufmann, P. A. (2006). "Caffeine decreases exercise-induced myocardial flow reserve," *Journal of the American College of Cardiology*, **47**, 405–410. Raw data read from Figure 1.

Norris, D. (2006), "The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process," *Psychological Review*, **113** (2), 327–357.

Press, S. J. and Tanur, J. M. (2001), *The Subjectivity of Scientists and the Bayesian Approach*. New York, NY: Wiley.

Sims, C. (2007), "Bayesian Methods in Applied Econometrics, or Why Econometrics Should Always and Everywhere be Bayesian," [online], Available at <http://sims.princeton.edu/yftp/EmetSoc607/AppliedBayes.pdf>

Tallman, E. (2014), Personal communication.