Abstract

We propose a semester-long Bayesian statistics course for undergraduate students with calculus and probability background. We cultivate students’ Bayesian thinking with Bayesian methods applied to real data problems. We leverage modern Bayesian computing techniques not only for implementing Bayesian methods, but also to deepen students’ understanding of the methods. Collaborative case studies further enrich students’ learning and practice experience to solve open-ended applied problems. Our proposed Bayesian course has an emphasis on undergraduate research, where accessible academic journal articles are read, discussed, and critiqued in the class. With increased confidence and familiarity, students take the challenge of reading, implementing, and sometimes extending methods in journal articles for their course projects. Moreover, students become fluent in R Markdown and latex after completing the course, which are beneficial to their statistics careers beyond this course.

Keywords: Bayesian education, Bayesian thinking, JAGS, statistical computing, statistics education, undergraduate research

*The author gratefully acknowledges the Mathematics and Statistics Department at Vassar College for their support and encouragement to our experiment of Bayesian education at the undergraduate level. The author thanks the Liberal Arts Collaborative for Digital Innovation (LACOL), for their generous funding support to share this course across 10 campuses, resulting in wider student audience and richer pedagogy experience for the author. The author also thanks Jim Alert, a co-author of the forthcoming undergraduate-level textbook of Bayesian modeling, for his experience of Bayesian education and generosity to mentor junior Bayesian educators.
1 Introduction

Statistics educators have been introducing Bayesian topics into the undergraduate and graduate statistics curriculum for the past few decades. At the Joint Statistical Meetings in 1996, the Section on Statistical Education organized an invited session, followed by a series of papers and discussion on the advantages, disadvantages, rationale, and methods for teaching an introductory statistics from a Bayesian perspective, appeared in *The American Statistician*. Berry (1997) suggested how introductory courses can be taught using a Bayesian approach and argued why students in these courses are well served by a Bayesian course. Albert (1997) demonstrated how to make an introductory course more data-oriented and introduced two devices, the Bayes’ box and the Bayes’ scatterplot, from a Bayesian perspective. Moore (1997), on the other hand, argued that it was premature to teach the ideas and methods of Bayesian inference in an introductory statistics course.

The obstacles presented by the author include: 1) Bayesian techniques were little used, 2) Bayesians had not yet agreed on standard approaches to standard problem settings, 3) the requirement of conditional probability can be confusing to beginners, and 4) the teaching and learning of Bayesian inference might impede the trend toward experience with real data and a better balance among data analysis, data production, and inference.

Indeed, prior to the invention and development of the Gibbs sampler and other Markov chain Monte Carlo (MCMC) algorithms in the late 1980s and early 1990s, not only the teaching in the classroom, but also the practice of Bayesian methods, had been very limited. Nevertheless, statistics educators made great effort to innovate, especially to connect to real data problems and apply Bayesian methods to solve these problems. Franck et al. (1988) designed a Bayesian analysis suitable for classroom presentation for a post-calculus probability and statistics course for advanced undergraduates and first-year graduate students. The authors introduced the study of the effects of Lactinex for diarrhea by analyzing data from a treatment group versus a control group, emphasized the intuitive way of specifying the prior distributions for this problem, and worked out the detailed posterior derivation. However, the corresponding computation had to rely on numerical integration.

Things quickly started to change, thanks to the revolutionary computational development, as well as the rapid spread of Bayesian techniques being used in applied problems.
While in 1997, Moore (1997) showed concerned about introducing Bayesian methods at the introductory undergraduate level, the teaching of Bayesian topics at the more advanced undergraduate level, and especially at the graduate level, had taken off. The first edition of Bayesian Data Analysis came out in 1995 (Gelman et al., 1995). Now in its third edition, Gelman et al. (2013) has become such a comprehensive book, that Bayesian practitioners and researchers use it as a reference book, and graduate-level instructors might use it as the textbook in their statistics graduate programs. Other popular graduate-level texts for students in the statistics programs, especially at the PhD-level, include Hoff (2009). For students in statistics programs at the Masters-level, Marin and Robert (2014); McElreath (2016); Reich and Ghosh (2019) are more recent texts.

If we look beyond statistics programs, there exists a large amount of work, including textbooks and articles, about teaching Bayesian methods for non-statisticians, mostly at the graduate level. For example, Rossi et al. (2005) is a textbook for PhD students in marketing and business, Kruschke (2014) and Lee and Wagenmakers (2014) are popular textbooks and/or reference books for students and/or researchers of cognitive science and experimental psychology. Moreover, Gelman (2008) talked about teaching Bayesian methods to graduate students in the social sciences, including political science, sociology, public health, education, and economics. Utts and Johnson (2008) described how the authors developed a course for graduate students in non-statistics departments and later turned it into an international workshop.

How about Bayesian education at the undergraduate level? Most of the educational innovation is taking place in the introductory statistics courses, where Bayesian inference is one of the many topics. Most recently, Eadie et al. (2019) designed an active-learning exercise with m&m’s, and Barcena et al. (2019) designed a web simulator to teach Bayes theorem, with application to the search for the nuclear submarine, USS Scorpion, in 1968. For more advanced-level undergraduate statistics courses, where Bayesian inference is either covered as a topic in a statistical inference/mathematical statistics course or as a Bayesian-analysis course, Kuindersma and Blais (2007) designed teaching Bayesian model comparison with the three-sided coin, focused on physics applications. Rouder and Morey (2019) proposed teaching Bayes’ theorem by looking at strength of evidence as predictive accuracy.
In this article, we propose a semester-long Bayesian statistics course for the undergraduates with a background of multivariable calculus and probability. Students in the course might never have taken a statistics course before, therefore the Bayesian statistics course is their first introduction to statistical thinking and analysis. We cultivate students’ Bayesian thinking and reinforce it by Bayesian computing, including the use of Monte Carlo simulation and MCMC. To allow students to explore non-conjugate and advanced Bayesian models, we introduce widely-used MCMC software, such as Just Another Gibbs Sampler (JAGS), as part of the computing in this course. R Markdown ([Allaire et al., 2019](#)), ggplot2 ([Wickham, 2009](#)) and tidyverse ([Wickham and RStudio, 2017](#)) R packages are introduced and used throughout this course.

To address the concerns raised by Moore ([1997](#)), we emphasize the use of real data in lecture examples, homework exercises, computing labs and exams. When possible, we invite guest lecturers as preview for new topics; for example, inviting a Bayesian cognitive scientist at the same institute to give a guest lecture on how their lab develops Bayesian hierarchical models to understand and analyze people’s learning patterns. To further enrich students’ learning experience, we introduce case studies, mostly in the second half of the semester, where students are encouraged to explore extension of learned methods and/or create new approaches, to solve open-ended applied problems. Students work in pairs to collaborate, and in-class discussion is facilitated by before-class online discussion on Moodle.

We highly resonate with Cobb ([2015](#))’s five imperatives to “rethink our undergraduate curriculum from the ground up,” and especially the last and the most important imperative: teach through research. We achieve teaching through research by first exposing undergraduate students to accessible academic journal articles. A successful experiment is to have students read, discuss, and critique Casella and George ([1992](#)), a concise and nicely written article about the Gibbs sampler. With a reading guide composed of several questions, students respond to these discussion questions before class on Moodle, facilitating and reinforcing small-group and entire-class discussions during lecture. To further enrich students’ learning and research experience, we design computing labs to allow students to replicate simulation studies and graphical results presented in the paper, deepening their understanding of the methods and practicing their computing skills.
The exposure to accessible academic journal articles has gained students' confidence and critical thinking skills, both of which are further challenged and cultivated in their course projects, an important capstone experience. Students are encouraged to brainstorm project ideas from day one, and the project topics come from a wide range: economics and finance, cognitive science, computer science, sports analytics, sociology, and epidemiology, among others. Majority of the students are prepared to read and learn from academic journal articles, some theoretical and others applied, as a first step of their course projects. Such endeavors have been hugely gratifying for the students and the instructor, and a good number of course projects evolve into independent studies for subsequent semesters. We include a list of “tested out” and accessible journal articles in Appendix C, and sample project schedules/progresses are presented in Section 3.5 where more details about the course projects are discussed.

The remainder of this article is organized as follows: as Bayesian computing is an important and interwoven aspect of this course, Section 2 describes our choices, approaches, and philosophy of our Bayesian computing resources. We then proceed to provide a detailed description of the 13-week course in Section 3, including the course topics and prerequisites, homework, computing labs and exams, as well as the three key components of the course, case studies, discussing and critiquing journal articles, and course projects. The article ends with a discussion of students' experience, challenges and future ideas with Section 4. Supplementary materials include sample in-class R scripts, computing labs in R Markdown and homework exercises, and are available online.

2 Computing in The Course

Statistical computing is an important and interwoven component of any modern Bayesian statistics course. In our Bayesian statistics course for undergraduates, our choices of Bayesian computing resources influence every aspect of the course, from the prerequisites to the assignments and assessment. We therefore describe our choices, approaches, and philosophy of our Bayesian computing resources in this section, before proceeding to introduce the other aspects of the course.

A quick survey of available Bayesian textbooks, almost all at the graduate-level, reveals
how much weight the authors have put on Bayesian computing. For most introductory textbooks, including *Bayesian Statistics and Marketing* (Rossi et al., 2005), *Bayesian Essentials with R* (Marin and Robert, 2014), and *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (McElreath, 2016), the authors have created companion R packages to facilitate their teaching and students’ learning of Bayesian methods. Other introductory textbooks, such as *Bayesian Statistical Methods* (Reich and Ghosh, 2019), use JAGS or JAGS-like software extensively. Appendix A lists information about the sampled Bayesian textbooks, including their computing resources and target audience.

2.1 Goals and Approaches

At the undergraduate-level, we expect students to use computing techniques for implementing Bayesian methods in applied problems. Moreover, going through the computing aspect of Bayesian inferences enhances students’ understanding of the methods themselves. We achieve these two goals by a two-stage process:

- Stage 1, first 1/3 of the course: In conjugate cases (e.g. Beta-Binomial, Normal-Normal, Gamma-Normal, Gamma-Poisson), implement and compare exact solutions and Monte Carlo approximation solutions to posterior and predictive inferences.

- Stage 2, second 2/3 of the course: Introduce JAGS for implementing Gibbs samplers, compared to self-coded Gibbs samplers for simple cases. From then on, use JAGS for subsequent topics, for example, Bayesian hierarchical modeling and Bayesian linear regression.

2.2 Exact Solutions versus Simulation Solutions

The focus of Stage 1 is to get students familiar with R programming, and Monte Carlo techniques in simulating posterior and predictive distributions. In conjugate cases, analytical posterior and predictive distributions are available, which are great examples for students to compare the exact solutions and simulation solutions. Furthermore, simulating the predictive distributions and performing posterior predictive checks give students ample opportunities to distill the essence of Bayesian computing. It is therefore desirable not to
introduce JAGS or any other MCMC-solving software at this stage, avoiding the tendency to use these software as a “black box”.

### 2.3 Why and How to Use JAGS

The shift from self-coding to available software such as JAGS takes place in covering the Gibbs sampler and MCMC diagnostics. Simple cases, such as a two-parameter Normal model, are used to introduce the definition and derivation of full conditional posterior distributions, the keys to writing one’s own Gibbs sampler. Given example R scripts of a Gibbs sampler (involving functions and loops), students practice writing Gibbs samplers to explore important aspects of MCMC, for example:

- **Question 1:** Do initial values of the parameters matter?
  
  **Exercise 1:** Write a new Gibbs sampler with different initial values and compare results.

- **Question 2:** Does the sampling order of the parameters matter?
  
  **Exercise 2:** Write a new Gibbs sampler with a different order of the parameters and compare results.

Once students have a solid understanding of the mechanics of Gibbs samplers and have gained the ability to write their own Gibbs samplers, JAGS software is introduced, focusing on its descriptive nature of the specified Bayesian models and its comparison to a self-coded Gibbs sampler. To show JAGS’s descriptive nature, Figure 1 presents the JAGS script, with the expressions of the sampling density and the prior distributions to its right. To compare JAGS output to the output of a self-coded Gibbs sampler, students are prompted to revisit previously covered aspects of MCMC, further distilling the keys to MCMC and its diagnostics.

Using JAGS achieves the aforementioned two goals: JAGS not only “frees up” students to explore non-conjugate priors and advanced Bayesian models, especially in the subsequent units of Bayesian hierarchical modeling and Bayesian linear regression; it also enhances students’ understanding of the Bayesian models being implemented, because even though students might not know the actual algorithms that JAGS performs, they need to be
absolutely clear about how to write the JAGS script to implement the Bayesian models they have intended to use. We believe at the undergraduate-level, JAGS is sufficient and self-directed, that there is no need to provide an R package for the Bayesian statistics course. We recognize the need at the graduate-level, especially for students in non-statistics departments, an R package can be hugely beneficial, as done by Rossi et al. (2005); Marin and Robert (2014); McElreath (2016). We also acknowledge that JAGS is not applicable for every available Bayesian method. For one, JAGS is not suitable for implementing Bayesian variable and model selection, an important topic that will require other computing resources, among which bayess R package is a good choice (Robert and Marin, 2013).

Figure 1: The JAGS script to express the sampling density and the prior distributions.

```r
modelString <- "
model{

# The sampling density
for (i in 1:N) {
    y[i] ~ dnorm(mu, phi)
}

# The prior distributions
mu ~ dnorm(mu_0, phi_0)
phi ~ dgamma(alpha, beta)
}
"
```

- The sampling density:

$$Y_1, \cdots, Y_n \mid \mu, \sigma \overset{i.i.d.}{\sim} \text{Normal}(\mu, \sigma).$$

- The prior distributions:

$$\mu \sim \text{Normal}(\mu_0, \sigma_0),$$

$$1/\sigma^2 = \phi \sim \text{Gamma}(\alpha, \beta).$$

2.4 rmarkdown, ggplot2 and tidyverse

Ever since the advocated use of R Markdown (Allaire et al., 2019) in introductory statistics courses (Baumer et al., 2014), statistics educators for all levels of undergraduate statistics courses have incorporated R Markdown into their curriculum. R Markdown’s ability to produce high-quality and reproducible documents, reports and presentations has made it popular, not only among statistics educators but also among statistics researchers. For similar reasons, the ggplot2 and tidyverse R packages are popular and widely used (Wickham, 2009; Wickham and RStudio, 2017).
We interweave the teaching and learning of R Markdown, `ggplot2` and `tidyverse` in several inter-connected components of the course, which will be described in detail in Section 3. We typically provide sample R scripts in lecture using `ggplot2` and `tidyverse` syntax. Students are provided with an R Markdown file containing all the in-class R scripts, in addition to the lecture slides themselves (as an aside, the lecture slides are prepared using R Markdown). For practice, computing labs are in R Markdown, with sample R scripts using `ggplot2` and `tidyverse` syntax. Completion and grading of computing labs are conducted in R Markdown. We have observed our students using R Markdown to complete R portions of homework, take-home R portions of exams, preparing report for case studies and even creating presentation slides/posters for their projects, while all these do not require a certain format. Students’ growth of their knowledge, skills, familiarity and confidence with R Markdown and R programming has been extraordinary through the course of a semester. Samples of in-class R scripts and computing labs in R Markdown are presented in our supplementary materials. Section 3.1 describes additional R resources used in the course.

3 The Course

This section gives a detailed description of the course. We start with the course topics and prerequisites in Section 3.1. Section 3.2 describes homework, computing labs and exams in the course. Sections 3.3 through 3.5 present and discuss three key components of the course one-by-one: case studies, discussing and critiquing journal articles, and course projects.

3.1 Course Topics and Prerequisites

The course has 5 topics:

1. Bayesian inference for a proportion
2. Bayesian inference for a mean
3. Gibbs sampler and MCMC
4. Bayesian hierarchical modeling
5. Bayesian linear regression

The first two topics cover Bayes theorem, conjugate prior, posterior distribution and predictive distribution, with a focus of comparison of the exact solutions versus the simulation solutions in one-parameter Bayesian models. The third topic introduces Gibbs sampler and other MCMC algorithms, where JAGS is first introduced and practiced. With solid understanding of basic Bayesian methods and adequate R programming skills, the last two topics introduce two main Bayesian methodologies, hierarchical modeling and linear regression, within the contexts of interesting applied problems. Our list of 5 topics might not seem comprehensive. However, it will become evident soon (Sections 3.3 through 3.5), that students are exposed to a much wider range of Bayesian methods in this course: some are innovative extension of methodologies covered in class, as in case studies; others are much more advanced methods students encounter in their course projects. In fact, the short list of 5 topics is intentional: it provides adequate coverage of basic Bayesian concepts, inference methods and computing techniques, while leaving enough time and space for students to dive into research with what they have gained for what they want to do.

The course prerequisites include multivariable calculus and probability. We emphasize a solid review of the following probability material: events and partitions, axioms of probability, discrete and continuous random variables, joint distributions, conditional distributions and independent random variables. In addition, we believe the importance of high-level familiarity with transformation of random variables, evident in the derivation-focused (before MCMC) and application-driven (analyzing the effects of lactinex for diarrhea) Bayesian analysis proposed by Franck et al. (1988). Furthermore, in reviewing probability material, we give students ample opportunities to work with joint distributions, especially joint densities of conditionally independently and identically (i.i.d.) distributed random variables, often needed in expressing joint likelihood functions and joint prior distributions later in the course.

Due to the statistics curriculum structure at Vassar College, that students could come into this Bayesian course with only the two aforementioned prerequisites, we do not assume students to have any statistics background. This requires very little mention of the classical/Frequentist methods in this Bayesian course - mostly in the introduction lecture where
these two paradigms are briefly compared and discussed. Although we do not cover specific comparisons of the classical/Frequentist and the Bayesian methods, students with prior exposure classical/Frequentist methods often venture to include comparisons and discussions in their course projects, which is highly encouraged.

We also do not assume prior R experience of students, though most likely they have had some exposure. To ensure students are ready in terms of statistical computing, we assign three DataCamp courses: Introduction to R, Intermediate R, and Introduction to the Tidyverse, all of which are available through the DataCamp For The Classroom. Our experience has shown that completing all three within the first couple of weeks of the semester, and/or prior R experience equivalent to the three courses, could prepare students well for the statistical computing in this Bayesian course.

One more note about linear algebra. Without requiring it, we deviate from the typical matrix presentations of Bayesian linear regression. Instead, we focus on simple and independent univariate prior distributions for the regression coefficients. Time permitting, Multivariate Normal models (which require linear algebra) are covered in case studies in specific context, such as missing data imputation with Bivariate Normal models. This arrangement unfortunately limits our ability (not that we will have the time) to talk about the details of variable selection and model selection, two topics heavily rely on matrix algebra. Time permitting, we supplement using functions in publicly available R packages, such as the ModChoBayesReg function in the bayess R package (Robert and Marin [2013]), to implement accessible methods of variable selection and model selection.

3.2 Homework, Computing Labs and Exams

Homework is assigned about every two weeks, mainly in the first half of the semester. Each homework is composed of two portions: a written portion, focused on exercises on topics such as deriving the posterior distributions; and an R portion, focused on implementing

---

Footnotes:

1For more information about DataCamp, visit [datacamp.com](https://datacamp.com).
2The course link: [https://www.datacamp.com/courses/free-introduction-to-r](https://www.datacamp.com/courses/free-introduction-to-r).
3The course link: [https://www.datacamp.com/courses/intermediate-r](https://www.datacamp.com/courses/intermediate-r).
4The course link: [https://www.datacamp.com/courses/introduction-to-the-tidyverse](https://www.datacamp.com/courses/introduction-to-the-tidyverse).
5For more information about the DataCamp For The Classroom, visit [https://www.datacamp.com/groups/education](https://www.datacamp.com/groups/education).
Bayesian inference methods with real data. The written portion enhances students’ understanding of key Bayesian concepts: Bayes theorem, conjugate prior, posterior distribution and predictive distribution. The R portion, on the other hand, gives students ample opportunities to practice using R for Bayesian inferences, such as Bayesian hypothesis testing, Bayesian credible intervals, Bayesian prediction and posterior predictive checks. The gentle and slow pace of the homework R portion allows students with little or no prior R experience to gain familiarity and confidence and hones everyone’s R programming skills. Samples of homework are presented in our supplementary materials.

Computing labs are assigned about every two weeks, throughout the semester. As mentioned in Section 2.4, the lab are prepared and expected to be finished using R Markdown, which greatly help students gain familiarity and skills with this important tool of modern statistics and data science. We usually spend some time in class to get everyone started on the lab - this practice turns out to be very useful since students are exposed to the task in class and clarification questions can be easily answered by the instructor and communicated to the whole class.

The topic of each lab is closely connected to the lecture material. For example, as mentioned before, in the Gibbs sampler and MCMC topic, students read, discuss, and critique a journal article, which contains description of the authors’ designed simulation studies and graphical presentation of the simulation results. A computing lab is designed to walk through some key components of their simulation designs, and ask students to replicate their results. Completing the lab in turn deepens students understanding of the journal article, and their knowledge of the Gibbs sampler.

Two exams are given, one in Week 6, and the other in Week 11 (the course is 13-week long). Similar to the format of homework, exams have an in-class portion focused on Bayesian thinking and theoretical derivations, and a take-home portion focused on Bayesian computing to implement appropriate methods to solve applied problems.

3.3 Case Studies

Like Allenby and Rossi (2008), we believe all aspects of a Bayesian analysis are communicated best through interesting case studies. Case studies are great ways for applied
Bayesian analysis, where one constructs appropriate prior distributions, develops the likelihood, computes the posterior distributions, and finally communicates their results to address the questions of interest.

We therefore focus on case studies for the second half of the semester, in place of traditional homework assessment. By the time we complete the Gibbs sampler and MCMC topic, students have experience with one-parameter and multi-parameter models, and the basic computing skills to implement Gibbs samplers if needed. These have prepared them for Bayesian analysis in new applications; for example, missing data imputation with Bivariate Normal models. Case studies like this one could introduce important applied problem (i.e. missing data imputation) and new Bayesian methods (i.e. Bivariate Normal models) simultaneously, further strengthening students’ learning experience.

As the semester progresses to cover Bayesian hierarchical modeling and Bayesian linear regression, more and more interesting case studies are ready to be introduced; for example, hierarchical Gamma-Poisson models for analyzing marriage rates in Italy during World War II. Time permitting, more advanced and accessible Bayesian methods, such as latent class models and text mining techniques, are great case study topics with interesting context.

Most case studies are open-ended, meaning that students are encouraged to try out ideas, existing and new, to solve the analysis problems at hand. In the aforementioned hierarchical Gamma-Poisson case study, although this specific modeling technique is not covered in lecture, students have experience with other types of hierarchical models and the Gamma-Poisson conjugacy; the Italy marriage rates context is a great example to explore the Gamma-Poisson model with a hierarchical flavor, and students overall learn and implement new models very well in applied settings.

Logistic-wise, case studies are assigned in pairs, encouraging collaborative work. Resources permitting, students work with a different partner in every case study. Pairs upload their case study writeup (not required in R Markdown format, though most students choose to do so) on Moodle before class discussion. During lectures, students first discuss their approaches and findings in a small group, and we all then discuss and critique different approaches as an entire class.
3.4 Journal Articles

We believe undergraduate students could and should be reading academic journal articles as part of their education, provided that the articles are with the right content and at the right level. Journals such as *The American Statistician* are great sources of high-quality and accessible journal articles for undergraduates. *Casella and George* (1992) has been our favorite choice, as a great resource of learning the Gibbs sampler. Concise and nicely written, *Casella and George* (1992) gives a simple explanation of how and why the Gibbs sampler works, illustrates its properties with a two-by-two simple case, and designs simulation studies and analyzes their results. Furthermore, as an early paper, some of the introduced aspects of the Gibbs sampler in *Casella and George* (1992), such as how to obtain independent parameter draws, could be different from current practice. All these provide wonderful opportunities for students to read and learn about the development of the Gibbs sampler, and to discuss and critique different practices. We provide a reading guide containing a series of questions, available in Appendix B. Similar to case studies, students will post their responses on Moodle Discussion Board before class discussion, and we have small-group followed by entire-class discussions during lectures.

A somewhat unexpected but certainly welcome outcome of exposing undergraduates to academic journal articles is that our students have gained confidence and are willing to take the challenge to read, understand, implement, and sometimes extend methods from journal articles. In many cases, when working on the course project, students encounter journal articles. Typically, reading the journal article(s) is the first step of students’ course project. Depending on the goal and scope of the project and the level of the article(s), sometimes such reading leads to replication and a new application of the proposed methods, as in cases of a Bayes-by-Backprop project with the article by *Blundell et al.* (2015) and an option pricing project with the article by *Ho et al.* (2011); sometimes such reading leads to introducing the proposed methods in innovative ways, as in cases of a Dirichlet Process project with the article by *Teh* (2010). It has been a great pleasure to witness students’ growth from a focus on completing assignment to a focus on conducting research, and discussing and critiquing *Casella and George* (1992) have been a key to strengthening students’ confidence and critical thinking. Details of course projects will be described next.
and a list of “tested out” and accessible journal articles are available in Appendix C.

3.5 Course Projects

Cobb (2015) proposed five imperatives to “rethink our undergraduate curriculum from the ground up.” The 5th imperative, teach through research, is considered most important among the five and the one served by the other four: flatten prerequisites, seek depth, embrace computation, and exploit context. Cobb’s five imperatives can be implemented through our Bayesian statistics course, and we cannot resonate more with the teach through research, in the form of a course project.

Students in the Bayesian statistics course are encouraged to brainstorm project ideas from day one. In addition to examples of Bayesian methods solving interesting applied problems prepared by the instructor, the introduction lecture includes a few video clips of students’ projects from previous semesters (a 2-minute introduction video of the course project is required as part of students’ project submission). This immediately shifts the focus of interesting and exciting projects considered by us, the instructors, to those by them, the fellow students.

The variety of project topics and interests not only showcases what students could achieve in their projects, but also motivates them to choose what they want to explore. As evident in the list of journal articles from students’ course projects in Appendix C, common interests and topics can be observed. In the case of Vassar College: we have a good number of double majors of Mathematics/Statistics and Economics, leading to groups of students working on projects related to economics and finance; we have a Bayesian cognitive scientist faculty in the Cognitive Science Program, therefore relevant courses and research teams, leading to groups of students analyzing experimental data to explore learning theories. Hot topics such as neural networks inevitably attracts students’ attention, reflected in their project interests and choices.

Within the first week of the semester, students are encouraged to indicate their project interests through a self introduction post on Moodle, from which a list of potential project topics are extracted. The list of the topics and the students interested in each topic are shared with a Google Doc, which students can freely browse and add more thoughts. It is
at this stage that students start to find shared interests with each other, and slowly project teams start to form. By Week 6, students settle down on their project topics, and submit a one-page project proposal. Each project team (up to 3 students) needs to meet with the instructor before submitting the project proposal, and detailed feedback of feasibility and advice is given by Week 7, midway of a 13-week semester.

From Week 8, each project team creates a weekly schedule to complete the project. There are 2 credit-bearing check points for every team: a methodology draft by Week 10 and a project draft by Week 12. On the last day of class in Week 13, teams present their projects at a poster session. The poster session lasts for 75 minutes. Typically we break all teams into 3 sessions, and each session is allocated with 15 minutes, with a 5-minute discussion break between sessions. This arrangement allows students to present their own posters and to explore other students’ work. The 5-minute discussion break invites students to share their thoughts after learning about other students’ projects. In addition, each team submits a 2-minute introduction video about their project, and every student is asked to watch the videos before the poster session. These 2-minute videos help presenters give a high-level pitch about their work. It also helps everyone to plan their poster session better, for example, to spend more time on a poster which they are curious about based on the introduction video.

The detailed weekly schedule and specific tasks highly depend on the project topic and scope. Table 1 is a sample weekly schedule for a team of two students, whose project title is “A Bayesian Group-Wise Modality Comparison of P300 ERPs.” The corresponding journal article is Number 19 in Appendix C. Table 2 is a sample weekly schedule for a team of two students, whose project title is “Predicting NCAA March Madness Upsets Using BART”. The corresponding journal articles are Number 7 and Number 8 in Appendix C. Both projects require a certain amount of literature review, finding and cleaning dataset(s), implementing newly acquired Bayesian methods, and analyzing and presenting the final results. The first team’s work involves frequent communications with a Cognitive Science professor, while the second team’s work involves learning about the bartMachine R package.
Table 1: Sample weekly schedule # 1. Note: Prof. X is a cognitive science professor; Prof. Y is the Bayesian statistics course instructor; Z is one of the team members.

| Week     | Due                                      | Progress                                                                                                                                 |
|----------|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Week 8   | Readings for the Time Series Model in the textbook. | We worked on reading the paper and getting a better understanding of the time series model. An intent is also to visit Prof. X’s office hours on Tuesday to understand hyperpriors and priors chosen in the paper from an EEG and cognitive science standpoint. |
| Week 9   | Slides explaining the variables being used and prior information. | We met with Prof. X to help us understand the model. Also met with Prof. Y to clarify our plan of how to simplify the model shown in the paper for our own application to the data Z collected. |
| Week 10  | Methodology part draft.                   | Finished Methodology Draft.                                                                                                                                                                      |
| Week 11  |                                          | Worked on analysis of the Gibbs Sampler and furthered our progress on the slides for presentation. Also met with Professor X.                                                                   |
| Week 12  | Project draft.                           | Submission of Final Draft and JAGS script were worked on. We also met with Professor Y.                                                                                                           |
| Week 13  | Poster session.                          |                                                                                                                                                                                                |
| Week    | Due                                      | Progress                                                                                                                                 |
|---------|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| Week 8  | Find an appropriate data set.            | Have a data set of 2008-2018 games wrangled and prepared for model fitting                                                            |
| Week 9  | Explore with various models (LRMC), Decide scope of the model. | We have decided to go with the BART model (Chipman et al. 2008).                                                                           |
| Week 10 | Methodology part draft.                  | Previously had about 15 slides of methodology, condensed down into 9. Want to explain the structure of the model and then later on will also assist in R implementation recommendations. |
| Week 11 | [goal] Familiarize with R implementation (Package bartMachine) and run a few models with different predictor vectors. |                                                                                                                                            |
| Week 12 | Project draft.                           |                                                                                                                                            |
| Week 13 | Poster session.                          |                                                                                                                                            |

Course projects naturally grow into independent studies in the following semesters. Past and current independent study topics stemming from this Bayesian statistics course include: Bayesian estimation of future realized volatility, Bayesian nonparametric models, Bayesian time series, and Bayesian variable and model selection. Students in these independent studies are engaged in almost the entire process of applied statistics research: literature review, collect/find datasets, implement methods, analyze the results, and write a journal-style article/report. Moreover, working with students on topics of their interests exposes the instructor to new research ideas, and some projects, with tuning, can be turned into a new topic in the future iterations of this Bayesian statistics course, further enriching the
4 Epilogue and Discussion

Students’ experience - through informal and formal evaluations - has been overall positive. Despite challenging and heavy workload, students recognize their knowledge building and skills building in this course. Many have expressed positive experience with the course project, and the structured weekly schedule and progress report has been highly appreciated.

Since the introduction of the course in Fall 2016, we have been running it once every academic year at Vassar College. We have had two 1st place winners in the intermediate statistics category of the Undergraduate Class Project Competition (USCLAP), organized by the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) and the American Statistical Association (ASA). Interestingly, a course project’s 2-minute introduction video on “Bayes by Backprop: Weight Uncertainty in Neural Networks” on YouTube caught attention beyond our course. Probably by keyword search, the video received more than 100 views and even a like. There are a few ongoing independent studies stemmed from students’ projects, which have the potential to be published in academic journals.

We recognize the challenges of teaching an undergraduate-level Bayesian statistics course for students with limited background. For one, our course contains nontrivial R programming and statistical methods, which can be especially challenging to students with weak R programming backgrounds. We have been supplementing by in-class R examples, designated computing labs, and DataCamp courses, and we see room for improvement on this end. With no requirement of linear algebra, we have little means (and little time) to cover important topics such as Bayesian model selection and variable selection. We see room for improvement, and we believe it is highly likely for instructors at other institutions to have linear algebra as one of the prerequisites, such that topics of model selection and variable selection can be incorporated into their courses.

The same goes for prior exposure to statistics, especially the classical/Frequentist

---

For more information about the USCLAP, visit [https://www.causeweb.org/usproc/usclap](https://www.causeweb.org/usproc/usclap)
paradigm. While Vassar College students might come into our course with no statistics background, we can easily envision instructors at other institutions to include comparisons and discussions of classical/Frequentist and Bayesian methods. These comparisons and discussions can be added into topics such as two-sample comparisons (proportion and mean), hierarchical/multi-level modeling, and linear regressions.

Some of our proposed pedagogy approaches are applicable to many undergraduate-level statistics courses. We believe statistical computing, not only as a means to implement statistical methods, but also to strengthen students’ understanding of the statistical methods, is a crucial and interwoven component of any modern statistics course. Exposing students to accessible academic journal articles enriches students’ learning experience, and boosts students’ confidence and critical thinking. Furthermore, we support “teach through research” by course projects, and we believe in instructors’ mentoring of students’ course projects.

For prospective instructors, we believe the experience of a graduate-level Bayesian statistics course is sufficient, though Bayesian research experience is desirable to teach an undergraduate-level Bayesian course. We have teaching and learning materials publicly available at [https://github.com/monika76five/BayesianStatistics](https://github.com/monika76five/BayesianStatistics). We have a forthcoming textbook for undergraduate Bayesian education in the CRC Texts in Statistical Science series. Details are available at [https://www.crcpress.com/Probability-and-Bayesian-Modeling/Albert-Hu/p/book/9781138492561](https://www.crcpress.com/Probability-and-Bayesian-Modeling/Albert-Hu/p/book/9781138492561).

**Supplementary Materials**

Please see our supplementary materials for sample in-class R scripts and computing labs in R Markdown mentioned in Section 2.4, sample homework mentioned in Section 3.2.

**References**

Albert, J. (1997) Teaching Bayes’ rule: a data-oriented approach. *The American Statistician*, 51, 247–253.
Allenby, G. M. and Rossi, P. E. (2008) Teaching Bayesian statistics to marketing and business students. *The American Statistician, 62*, 195–198.

Barcena, M. J., Garin, M. A., Martin, A., Tusell, F. and Unzueta, A. (2019) A web simulator to assist in the teaching of Bayes’ theorem. *Journal of Statistics Education, 27*, 68–78.

Baumer, B., Cetinkaya-Rundel, M., Bray, A., Loi, L. and Horton, N. J. (2014) R Markdown: integrating a reproducible analysis tool into introductory statistics. *Technology Innovations in Statistics Education, 8*.

Berry, D. A. (1997) Teaching elementary Bayesian statistics with real applications in science. *The American Statistician, 51*, 241–246.

Blundell, C., Cornebise, J., Kavukcuoglu, K. and Wierstra, D. (2015) Weight uncertainty in neural networks. In *ICML’15 Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, 1613–1622.

Casella, G. and George, E. I. (1992) Explaining the Gibbs sampler. *The American Statistician, 46*, 167–174.

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.

Eadie, G., Huppenkothen, D., Springford, A. and McCormick, T. (2019) Introducing
Bayesian analysis with m&m’s: an active-learning exercise for undergraduates. *Journal of Statistics Education*, **27**, 60–67.

Franck, W. E., Hewett, J. E., Islam, M. Z., Wang, S. J. and Cooperstock, M. (1988) A Bayesian analysis suitable for classroom presentation. *The American Statistician*, **42**, 75–77.

Gelfand, A. E. and Smith, A. F. M. (1990) Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

Gelman, A. (2008) Teaching Bayes to graduate students in political science, sociology, public health, education, economics, ... *The American Statistician*, **62**, 202–205.

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, CRC Press.

Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (1995) *Bayesian Data Analysis*. Texts in Statistical Science, CRC Press.

Ho, S. W., Lee, A. and Marsden, A. (2011) Use of Bayesian estimates to determine the volatility parameter input in the Black-Scholes and binomial option pricing models. *Journal of Risk and Financial Management*, **4**, 1–23.

Hoff, P. D. (2009) *A First Course in Bayesian Statistical Methods*. Springer Texts in Statistics.

Kruschke, J. (2014) *Doing Bayesian Data Analysis: A tutorial with R, JAGS, and Stan*. Academic Press.

Kuindersma, S. R. and Blais, B. S. (2007) Teaching Bayesian model comparison with the three-sided coin. *The American Statistician*, **61**, 239–244.

Lee, M. D. and Wagenmakers, E. (2014) *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press.

Marin, J.-M. and Robert, C. P. (2014) *Bayesian Essentials with R*. Springer Texts in Statistics, Springer.
McElreath, R. (2016) *Statistical Rethinking: a Bayesian course with examples in R and Stan*. Texts in Statistical Science, CRC Press.

Moore, D. S. (1997) Bayes for beginners? some reasons to hesitate. *The American Statistician*, 51, 254–261.

Reich, B. J. and Ghosh, S. K. (2019) *Bayesian Statistical Methods*. Texts in Statistical Science, CRC Press.

Robert, C. P. and Marin, J.-M. (2013) *bayess: Bayesian Essentials with R*. R package version 1.4.

Rossi, P. (2019) *bayesm: Bayesian Inference for Marketing/Micro-Econometrics*. R package version 3.1.

Rossi, P. E., Allenby, G. M. and McCulloch, R. (2005) *Bayesian Statistics and Marketing*. New York: Wiley.

Rouder, J. N. and Morey, R. D. (2019) Teaching Bayes’ theorem: strength of evidence as predictive accuracy. *The American Statistician*, 73, 186–190.

Teh, Y. W. (2010) Dirichlet Process. In *Encyclopedia of Machine Learning*, 280–287. Springer US.

Utts, J. and Johnson, W. (2008) The evolution of teaching Bayesian statistics to nonstatisticians: a partisan view from the trenches. *The American Statistician*, 62, 199–201.

Wickham, H. and RStudio (2017) *tidyverse: Easily Install and Load the ‘Tidyverse*. R package version 1.2.1.

Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer New York.
Appendix A: Information about several Bayesian textbooks

The following Bayesian textbooks are listed by publication year.

1. *Bayesian Statistics and Marketing* (Rossi et al., 2005):
   - Computing: bayesm package (Rossi, 2019).
   - Target audience: PhD students in marketing and business.

2. *Bayesian Essentials with R* (Marin and Robert, 2014):
   - Computing: bayesss package (Robert and Marin, 2013).
   - Target audience: 2nd year master’s program for students aiming at a professional degree in data processing and statistics.

3. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (McElreath, 2016)
   - Computing: package rethinking (available on the author’s GitHub page).
   - Target audience: researchers in the natural and social sciences, whether new PhD students or seasoned professionals, who have had a basic course on regression.

4. *Bayesian Statistical Methods* (Reich and Ghosh, 2019)
   - Computing: JAGS.
   - Target audience: advanced undergraduate statistics majors, non-statistics graduate students from engineering, ecology, psychology etc. and masters students in Statistics Program.
Appendix B: Reading Guide for Casella and George (1992)

1. [Section 2] How does Gelfand and Smith (1990) suggest to obtain an approximate sample from \( f(x) \)? How is it different from or similar to the approach we talked about in class? What are the advantages and disadvantages of each approach?

2. [Section 2] The authors claim “Gibbs sampling can be used to estimate the density itself by averaging the final conditional densities from each Gibbs sequence.” What is the theory behind this claim? How does Figure 3 support this claim?

3. [Section 2] Two simulations in Figure 1 and Figure 3: What are the similarities and differences? Why the conditional carries more information than the marginal?

4. [Section 3] Write down marginal distribution of \( y \), and verify the conditional probabilities \( A_{y|x} \) and \( A_{x|y} \). Also, verify \( A_{x|x} = A_{y|x}A_{x|y} \) and \( f_xA_{x|x} = f_xA_{y|x}A_{x|y} = f_x \).

5. [Section 4] What is a fixed point integral equation in the bivariate case? How does it help illustrate how sampling from conditionals produces a marginal distribution? [Hint: check equations (3.5), (4.1), and (4.2).]

6. [Section 4] The authors claimed “a defining characteristic of the Gibbs sampler is that it always uses the full set of univariate conditionals to define the iteration.” Explain this claim by illustrating how a Gibbs sampler works with \( k \) parameters \((\theta_1, \theta_2, \cdots, \theta_k)\).

7. [Section 5] Summarize different approaches to sampling the Gibbs sequence.
Appendix C: List of Journal Articles from Students’ Course Projects

The following statistics journal articles are listed by alphabetic order of the first author’s last name.

1. Aha, D. W. and Goldstone, R. L. (1992) Concept learning and flexible weighting. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 534-539.

2. Amini, S. M. and Parmeter, C. F. (2011) Bayesian model averaging in R. *Journal of Economic and Social Measurement*, 36(4), 253-287.

3. Barkan, O. (2017) Bayesian neural word embedding. In *Thirty-First AAAI Conference on Artificial Intelligence*.

4. Blei, D. M., Ng, A. Y and Jordan, M. I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.

5. Blundell, C., Cornebise, J., Kavukcuoglu, K. and Wiersra, D. (2015) Weight uncertainty in neural networks. In *ICML’15 Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, 1613-1622.

6. Brazinskas, A., Serhii, H. and Titov, I. (2016) Embedding words as distributions with a Bayesian skip-gram model. In *NIPS Bayesian Deep Learning Workshop*.

7. Chipman, H. A., George, E. I. and McCulloch, R. E. (2010) BART: Bayesian Additive Regression Trees. *The Annals of Applied Statistics*, 4(1), 266-298.

8. Denison, D. G. T., Mallick, B. K. and Smith, A. F. M. (1998) A Bayesian CART algorithm. *Biometrika*, 85, 363-377.

9. Fernandez, C., Ley, E. and Steel, M. F. J. (2001) Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16(5), 563-576.
10. Gelman, A. (2006) Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis, 1*(3), 515-534.

11. Ho, S. W., Lee, A. and Marsden, A. (2011) Use of Bayesian estimates to determine the volatility parameter input in the Black-Scholes and binomial option pricing models. *Journal of Risk and Financial Management, 4*, 1-23.

12. Joseph, L., Gyorkos, T. W. and Coupland, L. (1995) Bayesian estimation of disease prevalence and the parameters of diagnostic tests in the absence of a gold standard. *American Journal of Epidemiology, 141*(3), 263-272.

13. Liu, J. S. (1994) The collapsed Gibbs sampler in Bayesian computations with applications to a gene regulation problem. *Journal of the American Statistical Association, 89*, 958-966.

14. Murray, J. S. (2018) Multiple imputation: a review of practical and theoretical findings. *Statistical Science, 33*(2), 142-159.

15. Patel, A. D., Gibson, E., Ratner, J., Besson, M. and Holcomb, P. J. (1998) Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

16. Shi, L., Griffiths, T. L., Feldman, N. H. and Sanborn, A. N. (2010) Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic Bulletin & Review, 17*(4), 443-464.

17. Teh, Y. W. (2010) Dirichlet Process. In *Encyclopedia of Machine Learning*, 280-287. Spring US.

18. Tesauro, G., Rajan, V. T. and Segal, R. (2010) Bayesian inference in Monte Carlo tree search. In *UAI’10 Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, 580-588.

19. Wu, W., Wu, C., Gao, S., Liu, B., Li, Y. and Gao, X. (2014) Bayesian estimation of ERP components from multicondition and multichannel EEG. *Neuroimage, 88*, 319-339.