STRUCTURED, INTERACTIVE RESOURCES FOR TEACHING BAYESIAN INFERENCE

Gian Carlo Diluvi, Bruce Dunham, Nancy Heckman, Melissa Lee, and Rodolfo Lourenzutti
Department of Statistics, University of British Columbia, Canada
gian.diluvi@stat.ubc.ca

Most statistics undergraduate curricula provide only a brief introduction to Bayesian inference. Furthermore, there is evidence that learners can fail to appreciate core concepts of the Bayesian framework because of the focus on the mathematical formalism that is common in traditional instruction of Bayesian inference. Guided exercises and interactive simulations are promising alternatives for introducing Bayesian inference. However, these two types of resources have thus far been developed separately, which may make them less effective. In this work, we bridge this gap by developing interactive, structured resources in the form of web apps and accompanying activity sheets with guiding exercises. We also incorporate student feedback via think-aloud sessions. This allows us to pinpoint sources of confusion in students, e.g., problems constructing their priors.

INTRODUCTION

The Bayesian statistical framework has become a necessary element of a statistician's toolbox due to its growing use in modern statistical problems (Albert, 2002; Gelman et al., 1995). Many undergraduate curricula, however, have not kept up with the changes and typically only briefly incorporate Bayesian inference as a small part of a course. For example, Dogucu and Hu (2022) surveyed the statistics programs of the 152 highest ranked universities and colleges in the United States and found that only 51 offered a Bayesian course. Of them, only four required students to take a Bayesian course before graduating. Furthermore, instruction of Bayesian inference usually emphasizes its mathematical underpinnings as opposed to focusing on core Bayesian concepts, something that has been referred to as the legacy of *mathematical thinking* in statistics (Brown & Kass, 2009; Hoegh, 2020). These core concepts include, for example, the important role of the prior distribution in incorporating expert knowledge into the statistical analysis and how the prior is combined with the likelihood to update the practitioner's beliefs—encoded in the posterior distribution—about the parameters (see the discussion by Johnson et al., 2020). In the mathematical thinking approach to Bayesian inference, these key ideas remain hidden behind layers of algebra.

Previous work has focused on providing guidelines to teach introductory Bayesian courses (Albert & Hu, 2020; Allenby & Rossi, 2008; Berry, 1997; Hoegh, 2020; Hu, 2020; Hu & Dogucu, 2022; Johnson et al., 2020; Witmer, 2017). Most of these guidelines advocate for the use of both Bayesian-specific introductory textbooks and interactive simulations with which students can engage. Bayesian introductory textbooks (such as Albert & Hu, 2019; Berry, 1995; Johnson et al., 2022; and Kruschke, 2014) provide guided exercises tailored to help students achieve specific learning outcomes (e.g., understand that the posterior density is proportional to the prior density times the likelihood). On the other hand, interactive simulations and web apps (e.g., Albert, 2020; Bárcena et al., 2019; O'Hagan, 1995) can completely hide the mathematical details in the backend, thereby allowing students to devote more attention to core Bayesian concepts. Unlike textbooks, interactive resources are naturally suited to showcase the dynamics of Bayesian inference, such as how the choice of prior distribution impacts the posterior distribution.

So far, these two avenues have been developed separately: interactive simulations seldom contain structure in the form of guided exercises, and textbooks are rarely accompanied by any simulation. However, Lane and Peres (2006) found that interactive resources without structure are less effective at aiding learning. Furthermore, these interactive simulations are usually developed without any student input despite students being the intended end users. In this work, we bridge this gap by developing open-source, online Shiny web apps (v1.7.1; Cheng et al., 2021) and accompanying activity sheets with exercises that guide students towards achieving specific learning outcomes. Following the methodology from Dunham et al. (2018), we also carried out think-aloud sessions where students interacted with a web app (following the prompts of the activity sheet) and answered pre- and post-interaction questionnaires. This allowed us to incorporate student feedback into the development of the web apps and to pinpoint sources of confusion amongst students.

In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina.* International Association for Statistical Education. iase-web.org

STRUCTURED, INTERACTIVE RESOURCES

One of our main motivations to develop interactive resources—Shiny web apps in our case—with structure (i.e., with accompanying guided exercises) is that each component can inform the other to help students achieve specific learning goals. For example, consider goals related to how the prior distribution captures practitioners' prior knowledge. Instructors can create a set of exercises that guide students towards the learning outcomes and then develop a web app that complements the exercises (or vice versa). The exercises can be self-reflections about how confident the students are about the "true value" of the parameter. The web app can provide a density plot of the prior with hyperparameters that can be modified by users to reflect different degrees of uncertainty. We refer to the combination of a web app and an accompanying activity sheet as a *resource*. We have so far developed two resources for teaching Bayesian inference in the University of British Columbia (UBC), in western Canada. These are available at StatSpace (Diluvi, 2021; Diluvi & Lourenzutti, 2021), a repository of materials for instruction of introductory statistics developed at UBC.

The activity sheets that we created aim at introducing students to the Bayesian framework by guiding them through a case study. This requires students to think about every step of the Bayesian learning process: specifying the prior, defining the likelihood, and computing the posterior. The Shiny web apps (v1.7.1; Chang et al., 2021) that we developed were inspired by *First Bayes* (O'Hagan, 1995), a teaching package for Bayesian inference. Although we developed the web apps with the learning goals of our activity sheets in mind, we also designed them to be usable for other learning outcomes. Our hope is that instructors can adapt our activity sheets or even create their own while using the same web app. Furthermore, the web apps are open source and instructors can also adapt them if needed. Both resources share similar features, such as learning goals and overall structure, and differ only in data likelihood and selection of the prior distribution. Hence, in this work we only focus on one resource, which we refer to as the *beta-binomial* resource.

The Beta-Binomial Resource

The beta-binomial resource considers the setting in which independent and identically distributed binary data are observed and one wants to infer the probability of "success" for each datum. A standard Bayesian analysis of this setting uses a prior beta distribution, which in the case of Bernoulli data is conjugate: the posterior is also a beta distribution and closed-form parameter updates are available. This is one of the most popular examples to teach introductory Bayesian inference (Dogucu & Hu, 2022).

We chose five learning outcomes for the activity, each related to different overarching objectives. Two learning goals are related to students gaining a deeper understanding of the Bayesian learning process and core concepts. Another learning outcome encourages students to compare Bayesian with frequentist inference, which can help students identify the benefits and drawbacks of each inferential approach. The last two goals ensure that students can manually carry out the statistical analysis and thus operationalize their conceptual knowledge. A preamble to the activity contains information about the Bayesian analysis of the beta-binomial example, including closed-form expressions for the posterior hyperparameters. The preamble ensures that all users have the necessary knowledge to do the activity, which we hope will make it more broadly accessible.

The activity itself consists of five parts. The first part introduces the case study: a data set with information about a space rocket that has been launched multiple times; we are interested in the probability of a launch being successful. The next three parts of the activity cover each step of the Bayesian learning process. First, students are asked to specify a prior distribution by thinking about the case study and their own knowledge about space rocket launches. Then, students define the likelihood and compute a (frequentist) confidence interval. Afterwards, students calculate the posterior distribution via a formula provided for the posterior mean and variance, as well as a (Bayesian) credible interval. They are also asked to compare the credible interval with the confidence interval. In the fifth part, students reproduce the space rocket analysis for a different data set, this time related to the probability of a delivery being on time. Finally, students are asked to consider a prior distribution that reflects too much confidence about the parameter value and to compare the resulting posterior with the one they obtained previously.

The web app consists of multiple tabs, each with different controls and displays; modifying the former affects the latter. Specifically, there are four tabs in the web app: one for specifying a prior distribution, another for selecting a data set, and two for carrying out posterior analyses. The first tab allows users to specify the hyperparameters of the beta prior distribution and prints out the density plot and the prior mean and variance. The activity prompts students to try different values of the hyperparameters and to use the density plots to inform their final choice of hyperparameters. The data and likelihood tab allows users to select one of multiple pre-loaded data sets or to define their own (or randomly generate a) data set. The maximum-likelihood estimate is displayed to allow users to carry out frequentist inference. The third tab is focused only on the posterior distribution and includes a plot of the density as well as different posterior quantiles. Users can also compute credible intervals. The fourth tab provides a summary of the analysis by displaying the prior density, the (scaled) likelihood, and the posterior density in a single plot. The original and updated hyperparameters are also displayed. We leverage these two tabs to encourage students to reflect on the results of the case study and formulate a conclusion to the analysis.

The beta-binomial activity sheet guides students towards the five learning goals by prompting them to interact with the web app in carefully planned exercises. In fact, we designed some of the elements of the web app for the exercises we had in mind. For example, we purposefully print the maximum-likelihood estimate to allow students to calculate confidence intervals and then ask them to compare them to Bayesian credible intervals—the third learning outcome. We believe that the synergy between interactive simulation and guided exercises makes the resource as a whole more effective than either of its components alone.

Tailoring the web app and activity sheet for each other does not limit the former; activities with different learning goals can be easily planned using the same web app. For example, one can envisage an activity to teach students how to update the posterior distribution when a subsequent data set is observed. (Succinctly, the posterior is used as the new prior. The "second" posterior is the same as if we had observed both data sets to begin with.) An activity sheet based on the beta-binomial web app can prompt students to compute the posterior by using Bayes' rule once (with both data sets combined) and then by using it twice (once with each data set).

THINK-ALOUD SESSIONS

Structured, interactive resources such as those detailed in the previous section naturally bridge the interactive nature of simulations with a textbook approach of having guided exercises for instructor-specified learning outcomes. As discussed so far, these resources were entirely developed by a team of graduate students and faculty members with expertise in Bayesian inference. However, the intended end users of the resources are undergraduate students. In practice, we have often found a mismatch between the perceptions of learners and what subject experts anticipated, and when prompted, students frequently mention areas of confusion in the resources. As an example, previous work by our research group found that students prefer activity questions that ask them to think what would happen if they modified a setting in the web app *before* asking them to modify the setting, something we did not originally anticipate.

One way to address this limitation is to involve students in the development stage of the resources. Dunham et al. (2018) (based on a methodology by Ooms & Garfield, 2008) propose carrying out focus groups and interview sessions whereby students are allowed to interact with the resource while being observed and possibly prompted by the researchers. We implemented a similar approach both to incorporate student feedback into our resources and to get some preliminary information about which Bayesian concepts students struggle with.

Think-aloud Sessions Organization

We decided to carry out one-hour think-aloud sessions (Reinhart et al., 2019) instead of focus groups or interviews. Think-aloud sessions differ from focus groups and interviews in that the researchers become passive observers instead of actively interacting with the students. This allowed us to observe how students interacted with our resource in a setting similar to what we would expect them to face after the resource was deployed.

Our interactions with students were limited to a brief introduction where we explained the purposes of the study and gave them instructions and to a ten-minute discussion at the end where we encouraged them to share their thoughts on the resource. To obtain quantitative information about the effectiveness of the resource, we designed an 18-question survey that students answered before and after interacting with the resource. We gave eight (pre-interaction) and five (post-interaction) minutes to students to answer the questionnaire, which was sufficient for all students. During the remaining thirty minutes, students worked their way through the activity sheet using the web app, without any prompts from the researchers.

Sixteen of the survey's questions were divided into three groups: questions to gauge the students' overall knowledge of Bayesian inference (five questions); questions about the prior distribution (five questions); and questions about the posterior distribution (six questions). We designed the questions to have True/False answers but gave students the option to answer "I don't know" to reduce guessing. We also asked students to self-rate their understanding of Bayesian inference and to provide their unique study subject identifier. The post-activity survey also asked students how their understanding of Bayesian inference changed after the interaction.

Our target population consisted of students taking an intermediate statistics course at UBC. The course introduces mathematical inference, and it is the only undergraduate course at UBC Statistics where Bayesian inference is part of the curriculum. We asked the instructor to share an email inviting the 133 students enrolled in the Fall 2021 offering of the course to participate in our think-aloud sessions. We received responses from 17 students and scheduled three sessions with eight students in total (due to scheduling conflicts with the remaining nine students). The sessions were carried out between August and November 2021 over Zoom, the preferred video conferencing software at UBC. We obtained approval from the university's ethics board prior to the recruitment process and gave all participants a $20CAD (around $16USD) bookstore voucher after the session. We also required students to fill out a consent form before each session.

Findings

We focus our discussion on the prior specification and the likelihood definition because most students only completed these parts due to the short time they had to interact with the resource. To support our arguments, we provide brief anonymized quotes from the think-aloud sessions.

Students struggled much more than we anticipated with the prior specification. For example, some students thought they were choosing the true value of the parameter to be used in the activity. Other students tried to minimize the prior variance, which is probably due to the habit of avoiding high variance quantities in traditional statistics instruction. In their words, "it [the variance] kind of has an extremum, a kind of optimal value at certain a and b [prior hyperparameters]." These misunderstandings were not reflected in the results of the questionnaire: seven out of eight students correctly answered (post-interaction) that the practitioner's prior beliefs are incorporated through the prior, and also that the prior is a distribution over the values of the parameters, not the data. This suggests that although conceptually the students know what the prior is, their understanding might be too abstract or superficial to allow them to operationalize their knowledge. As a final example supporting this idea, a group of students mentioned that "we are not familiar with this stuff [space rockets]." We would have expected these students to specify a vague or somehow uninformative prior. However, they did not manage to operationalize their (correct) understanding of prior distributions and instead stuck with the relatively informative default prior: "we should specify a large variance, maybe?" said one student. "We're not experts," argued the other student in the group. "Yeah, just move to [the next] question," concluded the first student.

The sessions also provided us with feedback about the activity. Originally, the case study description was at the end of the preamble. During the first think-aloud session, we realized that students skimmed that part and thus struggled with the rest of the activity. For the rest of the sessions, we included the case study description in its own section of the activity. Speaking of future users, one student said it would be ideal if "they don't spend time reading lots of text. [It] might hinder their understanding." Too much text in an activity sheet is also feedback that we have received from other similar ongoing projects. A potential solution is to include the activity in the web app itself, or alternatively to split the activity so that students do not see all the text at once. We also noticed that students are easily led by words related to familiar concepts. For example, the prior specification section asks students how *confident* they are about their knowledge of the parameter. Two students thought they were meant to specify a confidence interval: "I think for this the probability might be calculated by the confidence interval." "Yeah," immediately agreed the other student in the group.

Of the eight students, three considered that their understanding of Bayesian inference improved slightly after interacting with the resource, whereas four said their knowledge remained about the same. One student said they felt more confused, although we believe this might be due to the Dunning-Kruger effect (Kruger & Dunning, 1999). Specifically, the student—perhaps overconfidently—ranked their understanding of Bayesian inference relatively high in the pre-interaction questionnaire, but through the activity probably realized they did not understand the concepts as much as they thought. We note that we did not expect the resource to have a tremendous impact on students' understanding of Bayesian inference due to the limited interaction time. The fact that some students reported positive experiences is, in our view, encouraging.

CONCLUSION

Typically, interactive simulations and guided exercises to teach introductory Bayesian inference have been developed separately. In this work, we developed structured, interactive resources for the instruction of Bayesian statistics. These allow students to get the best of both worlds by interacting with the dynamics of Bayesian inference while working towards learning outcomes designed by the instructor. Furthermore, we involved students in the development of these resources via think-aloud sessions. These allowed us to pinpoint sources of confusion, such as specifying the prior distribution. We are currently planning to carry out think-aloud sessions to explore how to best aid students in understanding and specifying their prior distribution. We are also working on developing more resources, potentially with more advanced settings than the beta-binomial example. Finally, future work can be used to create more interactive activities as well, for example by embedding them in the web app.

REFERENCES

Albert, J. (2002). Teaching introductory statistics from a Bayesian perspective. In B. Phillips (Ed.), *Developing a statistically literate society. Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6).* ISI/IASE. https://iase-web.org/documents/papers/icots6/3f1_albe.pdf?1402524960

Albert, J. (2020). *Apps for teaching Bayes*. Retrieved April 7, 2022, from https://bayesball.github.io/ProbBayes/Bayes_Shiny_Apps.html

Albert, J., & Hu, J. (2019). *Probability and Bayesian modeling*. Chapman & Hall.

Albert, J., & Hu, J. (2020). Bayesian computing in the undergraduate statistics curriculum. *Journal of Statistics Education, 28*(3), 236–247. https://doi.org/10.1080/10691898.2020.1847008

Allenby, G. M., & Rossi, P. E. (2008). Teaching Bayesian statistics to marketing and business students. *The American Statistician, 62*(3), 195–198. https://doi.org/10.1198/000313008X330801

Bárcena, M. J., Garín, M. A., Martín, A., Tusell, F., & Unzueta, A. (2019). A web simulator to assist in the teaching of Bayes' theorem. *Journal of Statistics Education, 27*(2), 68–78. https://doi.org/10.1080/10691898.2019.1608875

Berry, D. A. (1995). *Basic statistics: A Bayesian perspective*. Wadsworth.

Berry, D. A. (1997). Teaching elementary Bayesian statistics with real applications in science. *The American Statistician, 51*(3), 241–246. https://doi.org/10.2307/2684895

Brown, E. N., & Kass, R. E. (2009). What is statistics? *The American Statistician, 63*(2), 105–110. https://doi.org/10.1198/tast.2009.0019

Chang, W., Cheng, J., Allaire, J. J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2021). *Shiny: Web application framework for R* (Version 1.7.1) [Computer software]. RStudio. https://shiny.rstudio.com/

Diluvi, G. C. (2021). *Bayesian analysis of normal data*. StatSpace. https://statspace.elearning.ubc.ca/handle/123456789/404

Diluvi, G. C., & Lourenzutti, R. (2021). *Bayesian analysis of binary data*. StatSpace. https://statspace.elearning.ubc.ca/handle/123456789/403

Dogucu, M., & Hu, J. (2022). The current state of undergraduate Bayesian education and recommendations for the future. *The American Statistician*. Advance online publication. https://doi.org/10.1080/00031305.2022.2089232

Dunham, B., Lee, M., & Yapa, G. (2018). Comparison of testing and evaluation methods for new resources in statistical education. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018) Kyoto, Japan.* ISI/IASE. https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_8D3.pdf?1531364292

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (1995). *Bayesian data analysis* (3rd ed.). Chapman and Hall.

Hoegh, A. (2020). Why Bayesian ideas should be introduced in the statistics curricula and how to do so. *Journal of Statistics Education, 28*(3), 222–228. https://doi.org/10.1080/10691898.2020.1841591

Hu, J. (2020). A Bayesian statistics course for undergraduates: Bayesian thinking, computing, and research. *Journal of Statistics Education, 28*(3), 229–235. https://doi.org/10.1080/10691898.2020.1817815

Hu, J., & Dogucu, M. (2022). Content and computing outline of two undergraduate Bayesian courses: Tools, examples, and recommendations. *Stat, 11*(1), Article e452. https://doi.org/10.1002/sta4.452

Johnson, A. A., Ott, M. Q., & Dogucu, M. (2022). *Bayes rules! An introduction to applied Bayesian modeling.* CRC Press.

Johnson, A. A., Rundel, C., Hu, J., Ross, K., & Rossman, A. (2020). Teaching an undergraduate course in Bayesian statistics: A panel discussion. *Journal of Statistics Education, 28*(3), 251–261. https://doi.org/10.1080/10691898.2020.1845499

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Kruschke, J. (2014). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd ed.). Academic Press.

Lane, D. M., & Peres, S. C. (2006). Interactive simulations in the teaching of statistics: Promise and pitfalls. In A. Rossman & B. Chance (Eds). *Working cooperatively in statistics education. Proceedings of the Seventh International Conference on Teaching Statistics (ICOTS7)*. ISI/IASE. https://iase-web.org/documents/papers/icots7/7D1_LANE.pdf?1402524965

O'Hagan, T. (1995). *First Bayes* [Computer software]. Retrieved April 5, 2022, from https://www.tonyohagan.co.uk/1b/

Ooms, A., & Garfield, J. (2008). A model to evaluate online educational resources in statistics. *Technology Innovations in Statistics Education, 2*(1). https://doi.org/10.5070/T521000033

Reinhart, A., Evans, C., Luby, A., Orellana, J., Meyer, M., Wieczorek, J., Elliot, P., Burckhardt, P., & Nugent, R. (2022). Think-aloud interviews: A tool for exploring student statistical reasoning. *Journal of Statistics and Data Science Education, 30(*2), 100–113. https://doi.org/10.1080/26939169.2022.2063209

Witmer, J. (2017). Bayes and MCMC for undergraduates. *The American Statistician, 71*(3), 259–264. https://doi.org/10.1080/00031305.2017.1305289