Introduction to Bayesian Data Analyis

NYU Applied Statistics for Social Science Research

Instructor: eric.novik@nyu.edu

05 Nov 2022

1 Course description

This course will introduce students to Bayesian data analysis and Bayesian workflow, which is a process that we follow to develop statistical models, test their efficacy, and make model improvements. This is a deep subject with a rich history and we can only offer an introductory treatment during a seven-week course. Nonetheless, motiviated students should get enough experience from this course to learn how to think like a Baysian, be able to fit standard models usint Bayesian inference, evalute the quality of the model, and make predictions.

During the course we will make extensive use of simulations, the R language, Stan, and the `rstanarm` package.

2 Times, Location, and Office Hours

In-person weekly lectures will be held every Monday from 5 – 8 PM in Room XYZ beginning on March 21 and ending on May 15. Office hours will be held over Zoom on Tuesdays from 5 – 6 PM.

3 Course prerequisites

Students should be familiar with the R language, as we will not have time to provide the introduction to R. Ideally, students will be comfortable with probability and basic differential and integral calculus and have taken a basic regression course. Students don't have to be familiar with Stan, or `rstanarm`, and no prior knowledge of Bayes is expected.

4 Course materials

The core text for the course will be "Bayed Rules!" Johnson, Ott, and Dogucu (2022) which is available for free online and as a paper copy from CRC Press.

Students are expected to have access to R and RStudio or another R programming environment of their choice. Students should follow the setup instructions from the Getting set up chapter.

For those who are interested in a more advanced treatment of the subject, the following are three popular choices.

- Statistical Rethinking, McElreath (2020)
- A Student's Guide to Bayesian Statistics, Lambert (2018)
- Bayesian Data Analysis, Gelman et al. (2014)
- Stan User's Guide, Stan Development Team (2022)

Aki Vehtari from Aalto Univsity compiled an excellent list of mostly Bayesian texts here.

5 Course Outline

The following tentative schedule assumes seven 3-hour sessions during which we plan to cover the following material.

- **(1) Bayesian Workflow**

  We will briefly discuss the history of Bayesian inference and introduce the key components of the Bayesian workflow — from model development, to model testing and decision-making. We will dig into the key components of the so called Bayes' rule.

  - Bayes Rules!, Chapters 1 and 2
  - Bayesian Workflow Gelman et al. (2020), Chapter 1: Introduction
  - Optional meta-track: "Not Only Defended But Also Applied": The Perceived Absurdity of Bayesian Inference, Gelman and Robert (2013)

- **(2) Conjugate models: Beta-Binomial**

  In otder to demonstrate the Bayesian machinery, we will go over a few simple models for which analytical solutions are available. In practice, we seldom use conjugate models, but they have pedagogical value, are very fast, and allow us to validate numerical methods.

  - Bayes Rules!, Chapters 3 and 4
  - Bayesian Workflow Gelman et al. (2020), Chapter 2: Before fitting a model
  - Optional meta-track: Statistical Modeling: The Two Cultures, Breiman (2001)

- **(3) Other conjugate models and introduction to posterior sampling**

  - Bayes Rules!, Chapters 5 and 6
  - Bayesian Workflow Gelman et al. (2020), Chapter 3: Fitting a model
  - Optional meta-track: Reflections on Breiman's Two Cultures of Statistical Modeling, Gelman (2021)

- **(4) MCMC and posterior inference**

  - Bayes Rules!, Chapters 7 and 8
  - Bayesian Workflow Gelman et al. (2020), Chapter 4: Using constructed data to find and understand problems
  - Optional meta-track: TBD

- **(5) Baysian linear regression and model evaluation**

  - Bayes Rules!, Chapters 9 and 10
  - Bayesian Workflow Gelman et al. (2020), Chapter 6: Evaluating and using a fitted model
  - Optional meta-track: TBD

- **(6) Expanding the linear model and modeling counts**

  - Bayes Rules!, Chapters 11 and 12
  - Bayesian Workflow Gelman et al. (2020), Chapter 7: Modifying a model
  - Optional meta-track: TBD

- **(7) Logistic regression and introduction to hierarchical models**

  Hierchical and multi-level models are workhorses of Bayesian analysis.

  - Bayes Rules!, Chapters 13 and 15
  - Bayesian Workflow Gelman et al. (2020), Chapter 5: Addressing computational problems
  - Optional meta-track: TBD

6 Grading

There will be six homework assignments, each worth 10% and an analysis project which can be completed in groups of up to three people. All assignments must be completed in R Markdown, Quarto, or similar systems suitable for scientific publication.

- Homework: 60%
- Analysis project: 30%
- Class participation: 10%

7 Other Resources

Following resources may be helpful to those who need a refresher in the prerequisites:

- R for Data Science, Wickham and Grolemund (2016)
- Calculus Made Easy, Thompson (1980)
- Calculus, Herman, Strang, and OpenStax (2016)
- YouTube: Essence of Calculus, Sanderson (2018)
- Intoduction to Probability, Blitzstein and Hwang (2019)
- Introduction to Probability Cheatsheet v2, Chen (2015)

References

Blitzstein, Joseph K., and Jessica Hwang. 2019. *Introduction to Probability*. Second edition. Boca Raton: crc Press/Taylor & Francis Group.

Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)." *Statistical Science* 16 (3): 199–231. https://doi.org/10.1214/ss/1009213726.

Gelman, Andrew. 2021. "Reflections on Breiman's Two Cultures of Statistical Modeling." *Observational Studies* 7 (1): 95–98. https://doi.org/10.1353/obs.2021.0025.

Gelman, Andrew, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. 2014. *Bayesian Data Analysis, Third Edition*. https://doi.org/10.1007/s13398-014-0173-7.2.

Gelman, Andrew, and Christian P. Robert. 2013. ""Not Only Defended But Also Applied": The Perceived Absurdity of Bayesian Inference." *The American Statistician* 67 (1): 1–5. https://doi.org/10.1080/00031305.2013.760987.

Gelman, Andrew, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. 2020. "Bayesian Workflow." *arXiv:2011.01808 [Stat]*, November. http://arxiv.org/abs/2011.01808.

Herman, Edwin, Gilbert Strang, and OpenStax. 2016. *Calculus Volume 1*. https://d3bxy9euw4e147.cloudfront.net/oscms-prodcms/media/documents/CalculusVolume1-OP.pdf.

Johnson, Alicia A., Miles Q. Ott, and Mine Dogucu. 2022. "Bayes' Rule." In, 17–48. Chapman; Hall/CRC. https://doi.org/10.1201/9780429288340-2.

Lambert, Ben. 2018. *A Student's Guide to Bayesian Statistics*. 1st edition. Los Angeles: SAGE Publications Ltd.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in r and STAN*. 2nd edition. Boca Raton: Chapman; Hall/CRC.

Sanderson, Grant. 2018. "Essence of Calculus," November. https://www.youtube.com/playlist?list=PLZHQObOWTQDMsr9K-rj53DwVRMYO3t5Yr.

Stan Development Team. 2022. *Stan Modeling Language Users Guide and Reference Manual*. 2.30 ed. https://mc-stan.org/docs/stan-users-guide/index.html.

Thompson, Silvanus P. 1980. *Calculus Made Easy: Being a Very-Simplest Introduction to Those Beautiful Methods of Reckoning Which Are Generally Called by the Terrifying Names of the Differential Calculus and the Integral Calculus*. 3d ed. New York: St. Martin's Press.

Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. First edition. Sebastopol, CA: O'Reilly. https://r4ds.had.co.nz/.