Bayesian Statistics for the Social Sciences
GR5065, Spring 2019

Lecturer: Ben Goodrich (benjamin.goodrich@columbia.edu)
Verify that the date below is recent! Syllabus subject to change!

February 22, 2019

Course website: https://courseworks2.columbia.edu/courses/54170
Course Time: Tuesdays and Thursdays 04:10PM – 05:25PM in Math 417
Teaching Assistants: Terry Zhang and Xiaoliang Zhou

Office Hours: Tuesday afternoons, sign up on Canvas

Ben Goodrich's office is in IAB room 270I (near 270B, basically go to IAB 410 and then down the stairs two floors)

Course Description

An introduction to Bayesian statistical methods with applications to the social sciences. This course will be less technical than similar courses sometimes offered by the Statistics Department. Considerable emphasis will be placed on regression modeling and model checking. The primary software used will be Stan, which students do not need to be familiar with in advance. We will access the Stan library via R, so some experience with R is necessary.

Prerequisites

For QMSS students, whatever satisfies to data analysis requirement, typically GR4015. Any non-QMSS students interested in taking this course should have a comparable background to a second-semester QMSS student, which is basic probability, familiarity with calculus, linear regression, generalized linear modeling (such as logit models), and some computer programming in R.

Grading

Grading will be based $\frac{1}{6}$ on class participation, $\frac{1}{2}$ on the bi-weekly assignments, and $\frac{1}{3}$ on the final exam. Asking one (public) conceptual question on CampusWire per week (separate from any specific questions you have about your homework) or substantially discussing another student's question is considered good class participation.

CampusWire

CampusWire is a beta version of a tool that is available https://campuswire.com/c/GF74E0E61 using code 5748. Rather than emailing questions directly to the professor or TAs, you should post on CampusWire. That way, other students can answer your question, benefit from an answer that the professor or TA provides, ask follow-up questions, etc. There is also Reddit-style upvoting and the statistics collected by CampusWire go into the participation portion of your grade.

If your question pertains to an ongoing homework assignment, your grades, or similar, then you should click on the option to make your post only visible to "Instructors and TAs". Otherwise, you should post to "Everyone in the class" and avoid direct messaging the instructor and TAs. There is an option to post in Stealth Mode, in which case no one will know it was you that asked the question, but doing so obviously cannot count toward the class participation component of your course grade.

Books

It is not required that you purchase any books for this course. However, here are some links to online books and other resources that we will utilize in multiple weeks.

- *A Mathematics Course for Political and Social Research*, by Will H. Moore and David A. Siegel, published by Princeton University Press in 2013. Available for free here, or you can buy it here. Siegel also has a video course based on this book here.
- *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, Second Edition* by John Kruschke, published by Elsevier in 2015. Available for free Link, or you can buy it here.
- *Regression and Other Stories*, by Andrew Gelman, Jennifer Hill, and Aki Vehtari (to be) published by Cambridge University Press in 2019. Some chapters will be made available during the semester.
- *A Student's Guide to Bayesian Statistics* by Ben Lambert, published by SAGE in 2018. We will primarily utilize Lambert's mostly self-contained YouTube videos, but the underlying book can be purchased on Amazon for about $30 (paperback) or $20 (Kindle). You can read the first few chapters online.

Supplementary But Columbia-Licensed Online Books

- *Introduction to Bayesian Statistics* by William Bolsted and James Curran, published by Wiley in 2017. Link
- *Introduction to Probability* by Joseph Blitzstein and Jessica Hwang, published by CRC Press in 2015. Link Blitzstein also has videos based on this book here.
- *Bayesian Inference in the Social Sciences* edited by Ivan Jeliazkov and Xin-She Yang, published by John Wiley & Sons in 2014. Link
- *Bayesian Statistics for the Social Sciences* by David Kaplan, published by CRC Press in 2014. Link
- *Applied Bayesian Modeling* by Peter Congdon, published by John Wiley & Sons in 2014. Link
- *Bayesian and Frequentist Regression Methods* by Jon Wakefield, published by Springer in 2013. Link
- *Handbook of Markov Chain Monte Carlo*, edited by Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, published by Chapman and Hall / CRC Press in 2011. Link
- *The Theory that Would not Die: How Bayes' Rule Cracked the Enigma Code, Hunted down Russian Submarines, and Emerged Triumphant from two Centuries of Controversy*, by Sharon McGrayne, published by Yale University Press in 2011. Link
- *Bayesian Analysis for the Social Sciences* by Simon Jackman, published by John Wiley & Sons in 2009. Link
- *A First Course in Bayesian Statistical Methods* by Peter Hoff, published by Springer in 2009. Link
- *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists* by Scott Lynch, published by Springer in 2007. Link
- *Probability Theory: The Logic of Science* by E.T. Jaynes, published by Cambridge University Press in 2003. Link. The first three chapters are the best and are ungated here. Many of the remaining chapters are ungated here.
- Many more are available via this search on CLIO

Course Outline by Week

Before the semester starts, watch Sharon McGrayne's talk at Microsoft Research. Her book is linked in the previous section if you are interested in learning more about the history of Bayes' Theorem.

1. Introduction and Discrete Probability
    - (Tuesday) Lambert: overview
    - (Tuesday) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, by Richard McElreath, published by CRC Press in 2016. Chapter 1
    - (Thursday) Lambert: Introductions to probability distributions, discrete random variables, two-dimensional discretes, conditional discrete
    - (Thursday) Moore and Siegel, Chapter 9 and Sections 10.0 – 10.3
2. More Discrete Probability
    - (Tuesday) Moore and Siegel, Sections 10.4 – 10.8
    - (Thursday) Lambert: Introduction to Bernoulli / Binomial, Poisson I, Poisson II
3. Continuous Probability
    - (Tuesday) Lambert: Introduction to continuous distributions, marginal continuous, conditional continuous
    - (Tuesday) Moore and Siegel, Chapter 11. If you have forgotten basic calculus, see also Part II of Moore and Siegel or Calculus Made Easy or Calculus at a Fifth Grade Level.
    - (Thursday) Kruschke, Chapter 6
4. Bayesian Principles
    - (Tuesday) Lambert: intuition, beta distribution
    - (Tuesday) Kruschke, Chapter 2
    - (Thursday) Lambert: duality of likelihoods and probability distributions, why likelihoods are not probability distributions, illusion of uninformative priors
    - (Thursday) *An Introduction to Modern Bayesian Econometrics* by Tony Lancaster, published by Blackwell in 2004. chapter 1
5. (Hamiltonian) Markov Chain Monte Carlo
    - (Tuesday) Lambert: What does it mean to sample?, inverse transform
    - (Tuesday) Kruschke, Chapters 7 and 14
    - (Thursday) Lambert: Hamiltonian MCMC, effective sample size
    - (Thursday) "Everything You *Should* Have Learned About Markov Chain Monte Carlo" by Michael Betancourt Link
    - (Thursday) "Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo" by Cole Monnahan, James Thorson, and Trevor Branch, *Methods in Ecology and Evolution*, forthcoming. Link
6. (Generalized) Linear Models
    - (Tuesday) Lambert: PPC, overfitting I, overfitting II
    - (Tuesday) Gelman, Hill, and Vehtari chapters 8 – 10 (will be distributed the previous Thursday)
    - (Thursday) Kruschke chapter 15
    - (Thursday) Gelman, Hill, and Vehtari chapters 13 – 14 (will be distributed the previous Thursday)
7. Model Checking and Comparison
    - (Tuesday) Lambert: measuring fit, information criteria
    - (Tuesday) Aki Vehtari, Andrew Gelman, and Jonah Gabry (2017) "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC" Link
    - (Thursday) Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman (2018) "Vizualization in Bayesian Workflow" Link. See also the code.

8. More Things You Can Do with Draws from the Posterior Distribution

    - (Tuesday) Yuling Yao, Aki Vehtari, Daniel Simpson, and Andrew Gelman (2018) "Regularized stacking of Bayesian predictive distributions using leave-one-out cross-validation". Forthcoming in *Bayesian Analysis*. Link
    - (Thursday) Lambert: Bayes factors and their sensitivity to priors
    - (Thursday) Quentin Gronau et al. (2017) "A Tutorial on Bridge Sampling", *Journal of Mathematical Psychology*, Volume 81 80 – 97 Link
    - (Thursday) Juho Piironen, Markus Paasiniemi, and Aki Vehtari (2018) "Projective Inference in High-dimensional Problems: Prediction and Feature Selection", arXiv:1810.02406, link

9. Bayesian Regression Models using Stan

    - (Tuesday) "brms: An R Package for Bayesian Multilevel Models using Stan" by Paul-Christian Bürkner (2017), *Journal of Statistical Software*, 80:1 Link
    - (Tuesday) "brms Reference Manual" by Paul-Christian Bürkner. Link
    - (Thursday) "Monotonic Effects: A Principled Approach for Including Ordinal Predictors in Regression Models" by Paul-Christian Bürkner and Emmanuel Charpentier (2018) Link

10. Matrix Algebra and Multivariate Probability

    - (Tuesday) Moore and Siegel, Chapter 12
    - (Thursday) Wolfgang Karl Härdle and Léopold Simar, 2012, "Theory of the Multinormal", in *Applied Multivariate Statistical Analysis, Third Edition* edited by Wolfgang Karl Härdle and Léopold Simar, Springer. Link.

11. Hierarchical Models

    - (Tuesday) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, by Richard McElreath, published by CRC Press in 2016. Chapter 12 Do not worry too much about the R code that is specific to McElreath's rethinking R package; instead see Kurz's book below
    - (Tuesday) *Statistical Rethinking with brms, ggplot2, and the tidyverse* by A. Solomon Kurz (2018) Link (especially chapter 12)
    - (Thursday) "Advanced Bayesian Multilevel Modeling with the R Package brms" by Paul-Christian Bürkner (2018), *The R Journal* (2018) 10:1, pages 395-411 Link

12. The Stan Language

    - (Tuesday) Lambert: coding bespoke distributions and writing a Stan program
    - (Tuesday) Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). "Stan: A Probabilistic Programming Language". *Journal of Statistical Software*, 76(1), 1 - 32. doi:http://dx.doi.org/10.18637/jss.v076.i01. Do not worry too much about references to the "command line"; we will be using the rstan R package to interface with Stan and can obtain all the same information.
    - (Thursday) Work through Rasmus Bååth's Beginners Exercise: Bayesian computation with Stan and Farmer Jöns

13. Stan Applications

    - (Tuesday) Macartan Humphreys and Alan Jacobs presentation of their paper "Mixing Methods: A Bayesian Approach" with very special guests John Huber and Tim Frye. Link
    - (Thursday) Macartan Humphreys and Alan M. Jacobs, 2015, "Mixing Methods: A Bayesian Approach", *American Political Science Review*, 109(4), 653 – 673 Link
    - (Thursday) "Supplementary Materials" to Macartan Humphreys and Alan M. Jacobs, 2015, "Mixing Methods: A Bayesian Approach", *American Political Science Review*, 109(4), 653 – 673 Link

14. Missing Data

    - (Tuesday) Jackman, section 5.2.6 (link given above in online books)
    - (Tuesday) Stef van Buuren, 2012, *Flexible Imputation of Missing Data*, Chapman and Hall / CRC Press. Chapter 1 link
    - (Thursday) Gelman and Hill chapter ??? (will be distributed Tuesday)

15. Review