Applied Statistical Modeling and Inference: Bayesian APSTA-GE 2123, Spring 2019 (last seven weeks)

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

November 25, 2018

Course website: TBD
Course Time: Mondays 08:55AM – 12:00PM from April 1 to May 13 in room TBD. This technically includes a lecture period followed by a computer lab, but we will combine them so just bring your laptops.
 Grader: TBD

 Office Hours: Either right after class or we'll have to schedule an online chat because Ben works fulltime at Columbia

Course Description

An introduction to Bayesian statistical methods with applications to the social sciences. 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

Steinhardt students presumably have taken an introduction to probability in the fall and a course on frequentist inference in the first seven weeks of the spring semester. Any non-Steinhardt student interested in taking this course should have taken similar courses, which would cover basic probability (with calculus), linear regression, generalized linear modeling (such as logit models), and some computer programming in R.

In addition, all students should go through the first four weeks of material (including lecture videos, homework problems, etc.) in the corresponding Columbia course, which are available at https://courseworks2.columbia.edu/courses/72393 .

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 project. 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/ using code ####. 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.

- *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, Second Edition* by John Kruschke, published by Elsevier in 2015. Available for free here, 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 2018. 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).

Course Outline by Week

1. (Hamiltonian) Markov Chain Monte Carlo for Bayesian Inference

   - Lambert: What does it mean to sample?, inverse transform, Hamiltonian MCMC, effective sample size
   - Kruschke, Chapters 7 and 14
   - "Everything You *Should* Have Learned About Markov Chain Monte Carlo" by Michael Betancourt Link
   - "Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo" by Cole Monnahan, James Thorson, and Trevor Branch (2017), *Methods in Ecology and Evolution*, 8 (3): 339–48. https://onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12681

2. (Generalized) Linear Models

   - Lambert: PPC, overfitting I, overfitting II
   - Gelman, Hill, and Vehtari chapters 8, 12, and 13 (will be distributed the previous week)
   - Kruschke chapter 15

3. Model Checking and Comparison

- Lambert: measuring fit, information criteria
- Aki Vehtari, Andrew Gelman, and Jonah Gabry (2017) "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC" *Statistics and Computing*, Volume 27, Issue 5, pp 1413-1432 Link
- Jonah Gabry, Daniel Simpson, Aki Vehtari, Michael Betancourt, and Andrew Gelman (2019) "Vizualization in Bayesian Workflow" to be published in the *Journal of the Royal Statistical Society, Series A* Video, Paper, Appendix, Code
- 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
- Juho Piironen, Markus Paasiniemi, and Aki Vehtari (2018) "Projective Inference in High-dimensional Problems: Prediction and Feature Selection", arXiv:1810.02406, link

4. Bayesian Regression Models using Stan

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

5. Hierarchical Models

- *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
- *Statistical Rethinking with brms, ggplot2, and the tidyverse* by A. Solomon Kurz (2018) Link (especially chapter 12)
- "Advanced Bayesian Multilevel Modeling with the R Package brms" by Paul-Christian Bürkner (2018), *The R Journal* (2018) 10:1, pages 395-411 Link

6. The Stan Language

- Lambert: coding bespoke distributions and writing a Stan program
- 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.
- Work through Rasmus Bååth's Beginners Exercise: Bayesian computation with Stan and Farmer Jöns

7. Stan Applications

- Macartan Humphreys and Alan M. Jacobs, 2015, "Mixing Methods: A Bayesian Approach", *American Political Science Review*, 109(4), 653 – 673 Video, paper, and code available here.