Bayesian Statistics for the Social Sciences
G4065, Spring 2017

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

April 20, 2017

Course website: https://courseworks.columbia.edu/portal/site/QMSSG4065_001_2017_1/
Course Time: Tuesdays and Thursdays 04:10PM – 05:25PM Room TBA
Teaching Assistants: Terry Zhang

Office Hours: Monday evenings, sign up on CourseWorks

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 would be helpful.

Prerequisites

For QMSS students, whatever satisfies data analysis requirement, typically G4015. Any non-QMSS students interested in taking this course should have a comparable background to a second-semester QMSS student, which is basic probability, linear regression, generalized linear modeling (such as logit models), and some computer programming (but not any particular language).

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 Piazza per week (separate from any specific questions you have about your homework) or substantially discussing another student’s question is considered good class participation.

Piazza

Piazza is a relatively new tool that is available here or via CourseWorks (click on the Piazza tab in the bottom left and it may ask you a few questions the first time). Rather than emailing questions directly to the professor or TAs, you should post on Piazza. That way, other students can answer your question, benefit from an answer that the professor or TA provides, ask follow-up questions, etc. If you ask a question or a follow up question, be sure to click the Resolved button when the question is satisfactorily resolved.

If your question pertains to an ongoing homework assignment, your grades, or similar, then you should post in private mode, in which case only the professor and TAs will be able to see your post and respond. Otherwise, you should post in public mode. There is an option to post anonymously, in which case no one will know it was you that asked the question, but only named public posts count toward the class participation component of your course grade.
Required Textbooks


- *Regression and Other Stories*, by Andrew Gelman and Jennifer Hill, (to be) published by Cambridge University Press in 2017. Some chapters will be made available on CourseWorks during the semester.

Supplementary But Columbia-Licensed Online Books


- *Bayesian Inference in the Social Sciences* edited by Ivan Jeliazkov and Xin-She Yang, published by John Wiley & Sons 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


- *A First Course in Bayesian Statistical Methods* by Peter Hoff, published by Springer in 2009. 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’ Rule.

1. Introduction and Discrete Probability

   - (Tuesday) *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*, by Richard McElreath, published by CRC Press in 2016. Chapter 1
   - (Thursday) Moore and Siegel, Chapters 9 – 10.

2. Continuous Probability

   - (Tuesday) Moore and Siegel, Chapter 11. If you have forgotten basic calculus, see also Part II of Moore and Siegel.
   - (Tuesday) Kruschke, Chapter 6

3. Matrix Algebra and Multivariate Probability

   - (Tuesday) Moore and Siegel, Chapter 12

4. **Bayesian Principles**

• (Tuesday) Kruschke, Chapter 2

• (Thursday) *An Introduction to Modern Bayesian Econometrics* by Tony Lancaster, published by Blackwell in 2004. [chapter 1](#).

5. **(Hamiltonian) Markov Chain Monte Carlo**

• (Tuesday) Kruschke, Chapters 7 and 14

• (Thursday) “Everything You Should Have Learned About Markov Chain Monte Carlo” by Michael Betancourt [Link](#)


6. **Bayesian Integration of Quantitative and Qualitative Data**

• (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)

7. **(Generalized) Linear Models**

• (Tuesday) Gelman and Hill chapters 4 – 7 (pick up outside the QMSS office in IAB 807)

• (Thursday) Gelman and Hill chapters 8 – 9 (will be distributed in class Tuesday)

8. **Model Checking and Comparison**


• (Thursday) Vehtari, Gelman, and Gabry (2016) “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC” [Link](#)

9. **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.

• (Thursday) John Fox and Stanford Weisberg, 2015, “Mixed-Effects Models in R”, [Link](#)

10. **Bayesian Regression Models using Stan**

• (Tuesday) “brms: An R Package for Bayesian Multilevel Models using Stan” by Paul-Christian Bürkner, forthcoming in the *Journal of Statistical Software*. [Link](#)

• (Thursday) “brms Reference Manual” by Paul-Christian Bürkner. [Link](#)

11. **The Stan Language**

• 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](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.

12. **Miscellaneous Stan Topics**

• (Tuesday) Work through Rasmus Bååth’s *Beginners Exercise: Bayesian computation with Stan and Farmer Jöns*

13. Missing Data
   • (Tuesday) Jackman, section 5.2.6 (link given above in online books)
   • (Thursday) Gelman and Hill chapter 12 (will be distributed Tuesday)

14. Review