Bayesian Statistics for the Social Sciences
G4065, Spring 2016

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

March 10, 2016

Course website: https://courseworks.columbia.edu/portal/site/QMSSG4065_001_2016_1/
Course Time: Tuesdays and Thursdays 04:10PM – 05:25PM Mathematics 417
Teaching Assistant: Imad Ali (ima2119@columbia.edu)

Office Hours: Wednesday afternoons, 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. A guide to R and the R packages we will use in this course will be provided on CourseWorks.

Prerequisites

For QMSS students, whatever satisfies to 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.

Piazza

Piazza is a relatively new tool that is available 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 TA, 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 satisfactorally 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 TA 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

- *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.
- *An Introduction to Modern Bayesian Econometrics* by Tony Lancaster, published by Blackwell in 2004. Chapter 1 is available online. Can be purchased used for less than $40.
- *Statistical Rethinking: A Bayesian Course with R Examples* by Richard McElreath, published by Chapman & Hall in 2015. Available here. Chapter 1 (and 12) is available online. McElreath also has YouTube videos that go along with the book here.

Supplementary But Columbia-Licensed Online Textbooks

- *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan, Second Edition* by John Kruschke, published by Elsevier in 2015. Link
- *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 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
- *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. E.T. Jaynes previously completed most of a book of lectures, which has less mathematics, that is freely available here.
- Many more are available via this search on CLIO

Course Outline by Week (TBD)

1. Introduction and Discrete Probability
    - (Tuesday) McElreath, 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.
    - (Thursday) McElreath, Chapter 2
3. 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.
4. Bayesian Principles
    - Lancaster chapter 1 and appendix 1
    - McElreath chapter 3
5. Markov Chain Monte Carlo and Introduction to Stan
    - Tuesday
        - McElreath Sections 8.0, 8.1, and 8.2
        - Kruschke, Chapter 7 "Markov Chain Monte Carlo" (link above in online textbooks section)
    - Thursday
        - Kruschke, Chapter 14 "Stan" (link above in online textbooks section)
6. Bayesian Integration of Quantitative and Qualitative Data
    - Tuesday
        - 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
    - Friday (3PM, 501 Schermerhorn, must RSVP to Betty Howe esh2134@columbia.edu)
        - Macartan Humphreys and Alan Jacobs present their paper "Mixing Methods: A Bayesian Approach" with special guests John Huber and Tim Frye
7. (Generalized) Linear Models
    - Tuesday
        - McElreath chapter 4

  - Lancaster chapter 3
- Thursday
  - McElreath chapter 9
  - Lancaster chapter 5

8. Model Checking and Comparison
   - (Tuesday) Lancaster chapter 2
   - (Thursday) McElreath chapter 6
9. Hierarchical Models
   - McElreath chapters 12 and sections 13.0 – 13.3
   - John Fox and Stanford Weisberg, 2015, "Mixed-Effects Models in R", Link
10. The Stan Language
    - (Tuesday) Bob Carpeter et al. 2015, "Stan: A Probabilistic Programming Language", forthcoming in the *Journal of Statistical Software*. Link 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. Also note that it refers to Stan as of 5 releases ago and some things have a lot changed since.
    - (Thursday) McElreath chapter 11
11. Models for Panel and Time-Series Data
    - (Tuesday) Lancaster Chapter 7
    - (Thursday) Lancaster Chapter 9
12. Causal Inference in Microeconometrics
    - (Tuesday) Lancaster chapter 6
    - (Thursday) Lancaster chapter 8
13. 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) McElreath chapter 14
14. Gaussian Processes
    - (Tuesday) McElreath section 13.4
    - (Tuesday) Carl Edward Rasmussen and Christopher K. I. Williams, 2006, *Gaussian Processes for Machine Learning*, MIT Press. Read chapter 1. Link
    - (Thursday) Carl Edward Rasmussen and Christopher K. I. Williams, 2006, *Gaussian Processes for Machine Learning*, MIT Press. Read chapters 2 – 4. Link