Bayesian Modeling for the Social Sciences I - Introduction and Application

ICPSR Summer Program 2016

Ryan Bakker (Weeks 1–2)
Associate Professor
Department of Political Science
University of Georgia
rbakker@uga.edu

Johannes Karreth (Weeks 3–4)
Assistant Professor
Department of Political Science, Rockefeller College of Public Affairs & Policy
University at Albany, SUNY
jkarreth@albany.edu

This syllabus may be updated prior and during the workshop. Please visit www.jkarreth.net/bayes-icpsr.html for the most recent information on topics, labs, rooms, and assigned readings.

Course: 9am–11am / Room: See the Summer Program handbook.
Office: Helen Newberry building, rooms TBA
Office hours: 12-2pm, M-F. Feel free to stop by the office any time and come in if the door is open. We’re also happy to schedule meetings at most other times during the day.

Teaching assistants

Carolin Maney Purser, University of Georgia
Email: carolin@uga.edu
Office: Helen Newberry building, room TBA
Office hours: TBA

Chase Meyer, University of Georgia
Email: chasebm@uga.edu
Office: Helen Newberry building, room TBA
Office hours: TBA
Course description

This course introduces the basic theoretical and applied principles of Bayesian statistical analysis in a manner geared toward students and researchers in the social sciences. The Bayesian paradigm is particularly useful for the type of data that social scientists encounter given its recognition of the mobility of population parameters, its ability to incorporate information from prior research, and its ability to update estimates as new data are observed. The course begins with a discussion of the strengths of the Bayesian approach for social science data and the philosophical differences between Bayesian and frequentist analyses. Next, the course covers the theoretical underpinnings of Bayesian modeling and provides a brief introduction to the primary estimation algorithms. The bulk of the course focuses on estimating and interpreting Bayesian models from an applied perspective. Participants are introduced to the Bayesian forms of the standard statistical models taught in regression and MLE courses (i.e., linear, logit/probit, Poisson, etc.). Additional topics include measurement models, model comparison, and an in-depth treatment of multilevel modeling. Participants should have a solid understanding of the linear model and matrix algebra and some exposure to models with limited dependent variables. The course relies mostly on R and WinBUGS/JAGS for estimation, with a short segment on Stan, a new, but rapidly growing tool for Bayesian inference. Prior experience with R is preferred but not assumed or necessary. We offer lab sessions to familiarize participants with R, WinBUGS, JAGS, and Stan.

Goals. Upon conclusion of this course, we aim for participants to be able to:

- appreciate the fundamental differences and similarities between frequentist and Bayesian approaches to inference
- apply Bayes’ rule to the regression context
- formulate linear and generalized linear models in the Bayesian framework
- estimate linear and generalized linear models in the Bayesian framework using flexible code
- exploit the advantages of Bayesian estimation with regard to
  - incorporating prior information
  - incorporating uncertainty in parameter estimates
  - dealing with missing data
  - measuring latent concepts
  - incorporating variance at multiple levels of observation
- present and communicate results from Bayesian (and frequentist) estimation in an effective manner
- have fun learning new methods and better understanding familiar ones!

A note on computing. This course mostly uses JAGS and WinBUGS/OpenBUGS as the primary software options to fit Bayesian models, with one unit toward the end dedicated to Stan. We access JAGS and WinBUGS through R. Most lectures build on JAGS and WinBUGS/OpenBUGS. The languages of these two programs are nearly identical. WinBUGS and its sibling OpenBUGS run on Macs only with the appropriate Windows emulation software, but can be a bit buggy. JAGS runs on all platforms, including Macs. We offer special Mac-friendly lab sessions and support both JAGS and WinBUGS/OpenBUGS. JAGS code for all models encountered in this course and other JAGS-specific code and examples are provided.

Course resources

Z-Drive. All slides, code used in course sessions, and problem sets will be posted on the Z-Drive. Participants can access the Z-Drive from any computer in the three computer labs in the Helen Newberry building.

Course website with additional materials: Additional code, a JAGS tutorial, and other materials for weeks 3–4 are posted on Johannes’ website: http://www.jkarreth.net.
Reading materials

Books
The main texts used in this course are:


You may also find the following titles useful for many of the topics discussed in this course. They are available in the ICPSR Summer Program Library for borrowing:


As a general primer for R, we recommend:


Articles
All articles listed in the syllabus are available on the Z-Drive and through the University of Michigan library website from the campus network.

Software
This course relies mostly on R and JAGS/WinBUGS/OpenBUGS/Stan, but may also briefly discuss Stata as an alternative for some applications. We provide assistance installing R and JAGS/WinBUGS/OpenBUGS on your computers in the first week of the course. The labs at the Helen Newberry building have all necessary software as well. R, WinBUGS and JAGS are available at no cost from:

- [http://www.cran.r-project.org](http://www.cran.r-project.org)
- [http://www.mrc-bsu.ca.ac.uk/bugs](http://www.mrc-bsu.ca.ac.uk/bugs)
- [http://www.openbugs.net/w/FrontPage](http://www.openbugs.net/w/FrontPage)
- [http://mcmc-jags.sourceforge.net](http://mcmc-jags.sourceforge.net)

Each website links to relevant documentation and user manuals. There is a learning curve for these programs, but you need not have any computer programming background to learn them rather easily—just patience and desire. Our goal is to make you as comfortable as possible with these programs by the end of this course so that you will be able to use them with ease at your home institutions and in your own work.

*Mac and JAGS users*: See Johannes’ website for more information on installing JAGS.
Homework assignments

Homework exercises are assigned in class. Our goal is to make sure participants receive sufficient feedback to complete all assignments successfully. We distribute between 2 and 4 assignments per week. They are mostly computer-based with the exception of the first assignment. Please email your assignments to the TAs as PDF files and include [Bayes2016] in the subject line. Also always include all code you used to complete your assignments. The TAs will aim to return graded assignments to you within 2-3 days with comments via email. We (the instructors and TA) are happy to provide help with assignments during office hours: don’t be afraid to come by and ask.

Labs

We offer several labs with guided hands-on exercises. The lab sessions will be held in the computer labs at the Helen Newberry building. Please see the schedule for dates. Additional labs and exact locations will be announced in class and posted in the updated version of this syllabus. Likely topics:

1. Installing and using R
2. Installing and accessing JAGS/BUGS from R
3. Obtaining convergence diagnostics using R
4. Using R and RMarkdown for an integrated and reproducible workflow for Bayesian (and frequentist) statistics
5. Model presentation
6. Using Stan

Preparing for each workshop meeting

To get the most out of this workshop, we recommend that you read the assigned background & textbook readings for each day in depth and skim one of the applied studies if one is assigned. Particularly in the second half of the workshop, we list a larger number of applied works—pick one that is closest to your area of interest.

You should also feel encouraged to come to TA and instructor office hours on any day of the workshop to follow up on topics discussed during workshop meetings and to discuss how any topic we discussed might relate to your own work.

Course content and schedule

The following dates and topics may be modified as the course proceeds. The most recent version of the syllabus will always be at www.jkarreth.net/bayes-icpsr.html.

Monday, July 20
No course meeting

Recommended: Dave Armstrong’s Introduction to the \LaTeX{} Text Processing System, 5:30pm–7:30pm.

Day 1, Tuesday, June 21
Introduction: Background and Basics of Bayesian Inference

Please read:
· Gill: Chapter 1.

Day 2, Wednesday, June 22
Review of Generalized Linear Models

Refresher:
· Gill: Section 2.2.
· Gelman & Hill: Chapter 6.

Day 3, Thursday, June 23
Probability and Bayes’ Rule

Please read:
· Gill: Chapter 2.

Lab 1: Installing and using R.

Day 4, Friday, June 24
Priors

Please read:
· Gill: Chapter 4.

HW 1 assigned: Prior and posterior distributions.
Day 5, Monday, June 27
Sampling Methods and Introduction to the BUGS/JAGS Language

Please read:

- Gill: Chapters 9 & 10.

Lab 2: Installing and accessing JAGS/BUGS from R

Day 6, Tuesday, June 28
Convergence Diagnostics

Please read:


HW 2 assigned: Becoming familiar with WinBUGS/JAGS.
Lab 3: Obtaining convergence diagnostics using R.

Day 7, Wednesday, June 29
The Normal Distribution; Priors (ctd.)

Please read:

- Gill: Chapter 3

Day 8, Thursday, June 30
The Bayesian Linear Model

Please read:

- Gill: Chapter 5.

HW 3 assigned: Linear model.
Day 9, Friday, July 1: Missing Data

Please read:


**HW 4 assigned: Debugging BUGS/JAGS code.**

Monday, July 4
Lab (optional)

Lab 4: Using R and RMarkdown for an integrated and reproducible workflow for Bayesian (and frequentist) statistics.

Day 10, Tuesday, July 5

Binary Outcomes

If you'd like a refresher for generalized linear models and their interpretation, please read:

- Gelman & Hill, Chapter 5.

**HW 5 assigned: Logistic regression model.**

Day 11, Wednesday, July 6

Ordered and Categorical Outcomes

If you'd like a refresher on today's models, please read:

- Gelman & Hill, section 6.5.

Also, please read one of:

HW 6 assigned: Ordered or multinomial logit model.

Day 12, Thursday, July 7
Count outcomes
If you'd like a refresher on today's models, please read one of the following:

- Gelman & Hill, section 6.2.
- Ntzoufras, sections 7.4 and 8.3

Also, please read one of:


HW 7 assigned: Poisson model.

Day 13, Friday, July 8
Measurement Models
Please read one of these sample applications:


HW 8 assigned: Factor or IRT model.

Day 14, Monday, July 11

Bayes Factors and Bayesian Model Averaging

Please read:


Day 15, Tuesday, July 12

Model Checking and Model Presentation

Multilevel Models (Intro)

Please read:
- Gill: Chapters 6 & 7.

**HW 9 assigned: Model checking for linear regression.**

---

### Day 16, Wednesday, July 13

**Multilevel Models (Fundamentals)**

Please read:

- Gelman & Hill: Chapter 16 or/and Gill: Chapter 10
- Gelman & Hill: Chapter 11 (for a refresher on multilevel models).

**HW 10 assigned: Multilevel model.**

---

### Day 17, Thursday, July 14

**Multilevel Models (non-continuous outcomes; time-series cross-sectional data as multilevel data)**

**Multilevel regression with poststratification (MRP)**

Please continue to read:

- Gelman & Hill: Chapter 17 (Chapter 15 for a refresher).

as well as any of these empirical articles using MLMs that is/are in your area of interest:

Day 18, Friday, July 15
Bayesian analysis of spatial data
Using Bayesian Modeling in Your Applied Work

Please read:


If interested, please read the following for background and applications of spatial modeling using Bayesian inference:

- Lunn et al.: Section 11.3.