Applied Bayesian Modeling for Social Scientists
From theory to estimation and inference

Johannes Karreth
Assistant Professor, Department of Politics & International Relations
Ursinus College
Email: jkarreth@ursinus.edu
Course website: http://www.jkarreth.net/bayes-icpsr.html (with links to course materials)
Course meetings: June 15–19, 2020 / 9:00am–5:00pm / University of Michigan, Ann Arbor

This syllabus will see minor updates prior to the course. Please visit www.jkarreth.net/bayes-icpsr.html for the most recent version including additional course readings.

Course description and goals
This course provides an applied introduction to Bayesian data analysis and inference, geared toward participants from the social sciences. Bayesian methods have rapidly grown in the social sciences in recent years and have become a central tool for a wide variety of analytical methods, such as multilevel and measurement models, quantitative text analysis, and network analysis. The goal of this course is to enable participants to immediately use Bayesian tools in their own research and to effectively communicate their Bayesian results to other social science scholars.

Covering both Bayesian theory and applications, the course explores the following topics:

- Why use Bayesian inference?
- Philosophical and theoretical foundations for Bayesian inference
- The mechanics of MCMC tools and sampling
- Building and estimating Bayesian linear and generalized linear models
- Using MCMC output for postestimation, including marginal effects and predicted probabilities
- Bayesian multilevel/hierarchical models
- Bayesian approaches to measurement
- Bayesian tools for model comparison
- Model presentation and communication
- Optimal solutions for workflow and reproducibility

Upon completion of this course, participants will be able to:

- Understand the origins and logic behind Bayesian inference
- Use Bayesian methods for analyzing continuous and categorical outcomes in a regression framework
- Use Bayesian methods for measurement models
- Communicate Bayesian estimation results to practitioners and social science audiences

To allow participants to take full advantage of Bayesian data analysis in their own work, the course also teaches participants how to use the free and open-source software packages R and Stan. Practical examples and applied exercises form an integral part of the course.

**Prerequisites**

The course presumes a working knowledge of the linear regression model. Familiarity with probability theory would also be helpful, but is not formally required. Participants without any prior knowledge of statistics should consider a more basic quantitative methods course.

**Literature**

Participants should have access to:


The following books are recommended as background companions; some of their content will appear throughout this course:


As a general primer for R, I recommend:


As a background guide for mathematical concepts discussed in this short course, I recommend:


Additional readings will be made available to participants during the course.
Software and Preparation

Before the start of the course, participants should try to install the following programs on their laptops:

1. *R* is an open-source software package and available for download at [http://www.r-project.org](http://www.r-project.org).
2. RStudio is a convenient integrated development environment for R and available for free at [http://www.rstudio.com](http://www.rstudio.com).

We will go over how to use these programs on the first day of the course, using a detailed tutorial with step-by-step instructions. We will also have time to catch up on installation problems on the first day and install a number of packages required for Bayesian analysis in R, including rstan, during a course meeting.
Course outline

The following time slots and topics may be modified as the course proceeds. The most current version of this document can be found at [http://www.jkarreth.net/bayes-nus.html](http://www.jkarreth.net/bayes-nus.html) and in the course folder.

- **Lectures** are self-contained mini-units mixing lecture and discussion, with slides provided in the course folder.
- **Labs** are guided tutorials with documented scripts available in the course folder.
- **Assignments** are problem sets that participants may complete to reinforce the material learned in the course on that respective day.

<table>
<thead>
<tr>
<th>Day</th>
<th>Unit</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday</td>
<td>Lecture 1</td>
<td>Why use Bayesian Inference?</td>
</tr>
<tr>
<td>Monday</td>
<td>Lecture 2</td>
<td>Philosophical and theoretical foundations</td>
</tr>
<tr>
<td>Monday</td>
<td>Lecture 3</td>
<td>Bayesian versus frequentist inference</td>
</tr>
<tr>
<td>Monday</td>
<td>Lecture 4</td>
<td>Review: Probability and distributions</td>
</tr>
<tr>
<td>Monday</td>
<td>Lab 1</td>
<td>Introduction to R</td>
</tr>
<tr>
<td>Monday</td>
<td>Assignment 1</td>
<td>R exercises</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lecture 5</td>
<td>Priors</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lecture 6</td>
<td>The mechanics of MCMC tools and sampling</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lecture 7</td>
<td>Building and estimating Bayesian (linear) models</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lab 2</td>
<td>Introduction to Stan and rstanarm</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Assignment 2</td>
<td>Linear regression</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lecture 8</td>
<td>Convergence diagnostics</td>
</tr>
<tr>
<td>Tuesday</td>
<td>Lab 3</td>
<td>Assessing convergence</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lecture 9</td>
<td>Bayesian estimation for binary outcomes</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Assignment 3</td>
<td>Binary logit/probit regression</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lecture 10</td>
<td>Using MCMC output for postestimation</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lab 4</td>
<td>Working with MCMC output</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lecture 11</td>
<td>Bayesian estimation for ordered outcomes</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lecture 12</td>
<td>Bayesian estimation for categorical outcomes</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Lecture 13</td>
<td>Bayesian estimation for count outcomes</td>
</tr>
<tr>
<td>Wednesday</td>
<td>Assignment 4</td>
<td>Postestimation for binary logit/probit regression</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lab 5</td>
<td>Writing customized models in Stan</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lecture 14</td>
<td>Bayesian multilevel models for linear outcomes</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lab 6</td>
<td>Estimating multilevel models</td>
</tr>
<tr>
<td>Thursday</td>
<td>Lecture 15</td>
<td>Bayesian multilevel models for non-continuous outcomes</td>
</tr>
<tr>
<td>Friday</td>
<td>Lecture 16</td>
<td>Bayesian approaches to measuring latent variables</td>
</tr>
<tr>
<td>Friday</td>
<td>Lecture 17</td>
<td>Bayesian tools for model comparison and model checking</td>
</tr>
<tr>
<td>Friday</td>
<td>Lab 7</td>
<td>Communicating results from Bayesian analysis</td>
</tr>
<tr>
<td>Friday</td>
<td>Lecture 18</td>
<td>Optimizing workflow for reproducibility</td>
</tr>
</tbody>
</table>

For each day, the core reading usually provides substantial details for the units discussed on that day. Background readings typically address questions you may have during and after course. Sample applications demonstrate the techniques encountered on the respective day.
Day 1 Monday, June 15

Lecture 1 Why use Bayesian Inference?
Lecture 2 Philosophical and theoretical foundations
Lecture 3 Bayesian versus frequentist inference
Lecture 4 Review: Probability and distributions
Lab 1 Introduction to R
Assignment 1 R exercises at the end of Lab 1

Core reading:
- Jackman, chapter 1 & appendix B.

Background:

Day 2 Tuesday, June 16

Lecture 5 Priors
Lecture 6 The mechanics of MCMC tools and sampling
Lecture 7 Building and estimating Bayesian (linear) models
Lab 2 Introduction to Stan and rstanarm
Assignment 2 Linear regression
Lecture 8 Convergence diagnostics
Lab 3 Assessing convergence

Core reading:
- Jackman, chapters 2–5.

Background:
• Gelman, Andrew and Shirley, Kenneth. 2011. “Inference from Simulations and Monitoring
Convergence.” Chap. 6 in Handbook of Markov Chain Monte Carlo, ed. by Brooks, Steve et al.,

Sample applications:

• Blais, André, Guntermann, Eric, and Bodet, Marc A. 2017. “Linking Party Preferences and the
Composition of Government: A New Standard for Evaluating the Performance of Electoral

Day 3 Wednesday, June 17

Lecture 9 Bayesian estimation for binary outcomes
Assignment 3 Binary logit/probit regression
Lecture 10 Using MCMC output for postestimation
Lab 4 Working with MCMC output
Lecture 11 Bayesian estimation for ordered outcomes
Lecture 12 Bayesian estimation for categorical outcomes
Lecture 13 Bayesian estimation for count outcomes
Assignment 4 Postestimation for binary logit/probit regression

Core reading:

• Jackman, chapters 6 & 8

Background:

• King, Gary, Tomz, Michael, and Wittenberg, Jason. 2000. “Making the Most of Statistical
Analyses: Improving Interpretation and Presentation.” American Journal of Political Science
• Hanmer, Michael J. and Kalkan, Kerem Ozan. 2013. “Behind the Curve: Clarifying the Best
Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent
Versus Maximum Likelihood Estimation of Treatment Effects in Bivariate Probit Instrumental
Variable Models.” Political Science Research and Methods 7 (3): 651–659.
• Gelman, Andrew et al. 2008. “A weakly informative default prior distribution for logistic and

Sample application:

• Karreth, Johannes. 2018. “The Economic Leverage of International Organizations in Inter-
• Stegmueller, Daniel. 2013b. “Modeling Dynamic Preferences: A Bayesian Robust Dynamic
Day 4 Thursday, June 18

Lab 5 Writing customized models in Stan
Lecture 14 Bayesian multilevel models for linear outcomes
Lab 6 Estimating multilevel models
Lecture 15 Bayesian multilevel models for non-continuous outcomes

Core reading:
- Gelman & Hill, chapter 17, 18, 19, 24, 25

Background:

Sample applications:
Day 5 Friday, June 19

Lecture 16 Bayesian approaches to measurement
Lecture 17 Bayesian tools for model comparison and model checking
Lab 7 Communicating results from Bayesian analysis
Lecture 18 Optimal solutions for workflow and reproducibility

Core reading:

Background:

- On factor models:

- On IRT models:

- On model comparison:

Sample applications for measurement models:

· Williams, Rob et al. Forthcoming. “A latent variable approach to measuring and explaining peace agreement strength.” Political Science Research and Methods.

Sample applications for model comparison:

- Gelman, Andrew and Rubin, Donald B. 1995. “Avoiding Model Selection in Bayesian Social Research.” *Sociological Methodology* 25:165–173 (Background on BMA, read if you’re interested)

Last updated: February 10, 2020  
http://www.jkarreth.net/bayes-icspr.html