POLS 585: Bayesian Statistical Modeling

Emory University
Spring 2011

Meeting room:	Math & Science Center N302
Meeting time:	Tuesday, 2:30-5:30pm
Instructor:	Drew Linzer
Email:	dlinzer@emory.edu
Phone:	404-727-0697
Office:	Tarbutton 102
Office hours:	Monday, 2:00-4:00pm

Overview

The logic of Bayesian statistical inference provides an intuitively appealing and philosophically satisfying approach to the use of quantitative information to describe and predict social phenomena. But the practical benefits of Bayesian statistical modeling extend far beyond the purely theoretical ones. Knowing how to specify and estimate Bayesian models opens up a vast array of new opportunities for creative analyses of quantitative data. Recent advances in the power of desktop computing have made implementing Bayesian statistical models highly accessible to the applied researcher.

This course will teach the theory and practice of model building and inference using Bayesian statistics, especially through the use of Markov chain Monte Carlo algorithms. Students will gain an introduction to the WinBUGS software for Bayesian analysis, which, alongside R, is one of the most flexible and valuable statistical tools currently available to quantitative political scientists. By the end of the semester, students should be able to devise, estimate, interpret, and communicate the results of their own novel Bayesian statistical models, tailored to the unique characteristics of their own particular data sets, following their own theoretical ideas about their substantive area of interest.

The class is divided into three modules. We will begin with an overview of the Bayesian approach to data analysis, and how these ideas are implemented in practice. The second portion of the course is devoted to the specification and interpretation of multilevel models for data that can be usefully organized into groups. We finish with an overview of Bayesian item response theory, and its applications to ideal point estimation and model-based measurement of latent variables.

Student Expectations

Each class meeting will begin with one or more short student presentations on that week's readings. These presentations should describe the motivation for, and methodological approach involved in, any one of the assigned articles, and provide a starting point for further discussion. Demonstrating that you understand the reading is just as important as presenting it in a clear and interesting manner. We will have a sign-up for these presentations on the first day of class.

To practice using the techniques and software in this class, there will be a series of (mostly) weekly problem sets, due at the beginning of class the following week. Collaboration in small groups of two or three students is encouraged.

Throughout the course, students should be continually thinking about how Bayesian methods might be applied to their own research. For the final project, each student will create an original Bayesian statistical model, and apply it to data for a substantive problem of their choosing. Students will present their model and results, and receive feedback from the class, as part of a two-week mini-symposium we will hold at the end of the semester. A final paper will be due after the last week of class that presents the model findings in greater detail, and incorporates responses from the symposium.

Students are expected to complete all of the assigned readings prior to coming to class, even if not giving a presentation. I have deliberately selected textbooks and articles on the basis of their accessibility and applicability to emerging and important issues in political science research. If you would like any more theoretical or technical background material for anything we discuss in class, please contact me independently, and I will be happy to point you towards the appropriate literature. Research in Bayesian statistics is a massive field, and way more has been written than what I can realistically teach (and expect anyone to learn) in just one semester.

Required Texts

These textbooks are excellent, comprehensive guides to the theory and practice of Bayesian statistical modeling, tailored to applications in the social sciences. We are lucky: this is a fast-moving field and it is only very recently that resources such as these exist.

Gill, Jeff. 2008. *Bayesian Methods: A Social and Behavioral Sciences Approach*, Second Edition. Chapman & Hall/CRC.

Gelman, Andrew and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

Both books are available for purchase at the Emory University bookstore.

Other Useful Texts

For a complementary perspective and/or greater detail on the topics we will be covering this semester, you may wish to consult any of these recommended books.

Albert, Jim. 2009. *Bayesian Computation with R*, Second Edition. Springer.

Bolstad, William M. 2004. *Introduction to Bayesian Statistics*. Wiley.

Carlin, Bradley P. and Thomas A. Louis. 2009. *Bayesian Methods for Data Analysis*, Third Edition. Chapman & Hall/CRC.

Congdon, Peter. 2007. *Bayesian Statistical Modelling*, Second Edition. Wiley.

Fox, Jean-Paul. 2010. *Bayesian Item Response Modeling: Theory and Applications*. Springer.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2003. *Bayesian Data Analysis*, Second Edition. Chapman & Hall/CRC.

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences*. Wiley.

Johnson, Valen E. and James H. Albert. 1999. *Ordinal Data Modeling*. Springer.

van der Linden, Wim J. and Ronald K. Hambleton, eds. 1997. *Handbook of Modern Item Response Theory*. Springer.

Lynch, Scott M. 2010. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists*. Springer.

Raudenbush, Stephen W. and Anthony S. Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*, Second Edition. Sage.

Computer Software

Statistical programming and computational simulation are crucial to modern-day Bayesian data analysis. We will be making extensive use of the R statistical computing environment. A working knowledge of R is a non-negotiable prerequisite for this class. Please download the most current version of R from `http://www.r-project.org`, as well as a decent text editor. As in previous semesters, I strongly recommend that Windows users purchase a student licence for WinEdt version 5.5, which you can download at `http://www.winedt.com`. The RWinEdt package provides a seamless interface between the WinEdt text editor and the R computing environment. Mac users have a variety of other options.

In addition, a large part of the instruction in this course will involve teaching you how to use the WinBUGS software package to specify and estimate Bayesian statistical models. WinBUGS is a Windows implementation of the BUGS (*B*ayesian inference *U*sing Gibbs *S*ampling) project, and is free to download at `http://www.mrc-bsu.cam.ac.uk/bugs` (see their health warning at the bottom of the page). The R package R2WinBUGS enables complete integration of WinBUGS into R. Non-Windows users should instead download and install JAGS (*J*ust *A*nother *G*ibbs *S*ampler) from `http://www-fis.iarc.fr/~martyn/software/jags`. JAGS syntax is nearly identical to that of WinBUGS and can be called from R using commands contained in the rjags package.

Class schedule

January 18: Introduction. Probability, evidence, and learning from data.

Cohen, Jacob. 1994. The Earth Is Round ($p < .05$). *American Psychologist*. 49(12): 997-1003.

Carey, Benedict. Journal's Paper on ESP Expected to Prompt Outrage. *New York Times*, January 5, 2011.

Carey, Benedict. You Might Already Know This... *New York Times*, January 10, 2011.

Wagenmakers, Eric-Jan, Ruud Wetzels, Denny Borsboom, and Han van der Maas. 2010. Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi. Working Paper, University of Amsterdam.

January 25: The basics of Bayesian inference.

Gill, Chapters 1, 2, and 5, and Appendix A.

Western, Bruce, and Simon Jackman. 1994. Bayesian Inference for Comparative Research. *American Political Science Review*. 88(2): 412-423.

Jackman, Simon. 2004. Bayesian Analysis for Political Research. *Annual Review of Political Science*. 7: 483-505.

February 1: Sampling algorithms and Markov chain Monte Carlo.

Gill, Chapters 8 and 9.

MacKay, David J.C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press. Chapters 28-29.

Jackman, Simon. 2000. Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo. *American Journal of Political Science*. 44(2): 369-398.

February 8: Implementing MCMC methods.

Gill, Chapters 11-12 and Appendix C.

Gelman & Hill, Appendix C.

Spiegelhalter, David, Andrew Thomas, Nicky Best, and Dave Lunn. 2003. WinBUGS User Manual.

Sturtz, Sibylle, Uwe Ligges, and Andrew Gelman. 2005. R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*. 12(3).

February 15: Practical issues in model selection and comparison.

Gill, Chapters 6 and 7.

Gelman & Hill, Chapter 24.

Kass, Robert E. and Adrian E. Raftery. 1995. Bayes Factors. *Journal of the American Statistical Association*. 90(430): 773-795.

Quinn, Kevin M., Andrew D. Martin, and Andrew B. Whitford. 1999. Voter Choice in Multi-Party Democracies: A test of competing theories and models. *American Journal of Political Science*. 43(4): 1231-1247.

February 22: Multilevel/hierarchical models.

Gill, Chapter 10.

Gelman & Hill, Chapters 1, 11-13, 16-17, and Section 14.2.

Gelman, Andrew, Boris Shor, Joseph Bafumi, and David Park. 2007. Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut? *Quarterly Journal of Political Science*. 2(4): 345-367.

March 1: Multilevel model checking, interpretation, and presentation.

Gelman & Hill, Chapters 18-19 and 21.

Shor, Boris, Joseph Bafumi, Luke Keele, and David Park. 2007. A Bayesian Multilevel Modeling Approach to Time-Series Cross-Sectional Data. *Political Analysis*. 15(2): 165-181.

March 8: No class; Spring break.

March 15: Multilevel regression and poststratification, AKA Mister P.

Gelman & Hill, Sections 14.1 and 14.2.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls. *Political Analysis*. 12(4): 375-385.

Lax, Jeffrey R. and Justin H. Phillips. 2009. How Should We Estimate Public Opinion in The States? *American Journal of Political Science*. 53(1): 107-121.

March 22: Item response theory and ideal point estimation.

Clinton, Joshua, Simon Jackman and Douglas Rivers. 2004. The Statistical Analysis of Roll Call Data. *American Political Science Review*. 98(2): 355-370.

Bafumi, Joseph, Andrew Gelman, David Park, and Noah Kaplan. 2005. Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation. *Political Analysis* 13(2): 171-187.

Curtis, S. McKay. 2010. BUGS Code for Item Response Theory. *Journal of Statistical Software*. 36: Code snippet 1.

March 29: Hierarchical and bridging approaches to IRT.

Martin Andrew D., and Kevin M. Quinn. 2002. Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999. *Political Analysis*. 10(2): 134-153.

Bailey, Michael A. 2007. Comparable Preference Estimates across Time and Institutions for the Court, Congress, and Presidency. *American Journal of Political Science*. 51(3): 433-448.

Bafumi, Joseph and Michael C. Herron. 2010. Leapfrog Representation and Extremism: A Study of American Voters and Their Members in Congress. *American Political Science Review*. 104(3): 519-542

April 5: Adapting IRT model assumptions to suit different (theoretical) DGPs.

Jackman, Simon. 2004. What Do We Learn from Graduate Admissions Committees? A Multiple-Rater, Latent Variable Model with Incomplete Discrete and Continuous Indicators. *Political Analysis*. 12(4): 400-424.

Lauderdale, Benjamin E. 2010. Unpredictable Voters in Ideal Point Estimation. *Political Analysis*. 18(2): 151-171.

Pemstein, Daniel and Stephen A. Meserve and James Melton. 2010. Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type. *Political Analysis*. 18(4): 426-449.

April 12: Mini-symposium: Day 1

April 19: Mini-symposium: Day 2