Syllabus

NOTE: This syllabus is a work-in-progress. It is impossible for me to forecast how quickly or slowly we will cover material. This document, and the course calendar, will be updated as we go on the Telesis system. At the end of the semester, I will post a complete syllabus on my homepage.

L32 5732. Topics in QPM: Bayesian Inference
Tuesday 1:00-3:00
Eliot Hall 314

Instructor Information

Andrew D. Martin, Ph.D.
Associate Professor of Political Science
Office Address: Eliot 326
Telephone Number: 314.935.5863
Fax Number: 314.935.5856
Homepage: http://adm.wustl.edu
E-mail: admartin@wustl.edu
Office Hours: By appointment only (email to schedule)

Course Description

This course is an introduction to the use of Bayesian statistics in the social sciences. We will cover foundational issues, Markov chain Monte Carlo methods, and models for cross-sectional, over-time, and clustered data. Much of the specific content will be dictated by student demand. We will also spend a good deal of time on practical issues related to statistical computation.

My expectation is that by the end of the course students will be in a position to *read* relevant methodological literatures, and *implement* models using modern software tools. Students who excel in the course will be in a position to develop and fit innovative models of their own, thus *contributing* to the field of political methodology.

Most of the learning in this course will take place outside of the classroom. One cannot learn to be a good applied data analyst, or methodologist, but listening to lectures and reading textbooks. Rather, learning takes place through experience and practice. I expect that this course will take between ten and twenty hours a week of time outside of class for reading, working on problem sets (both analytical and practical), writing code, and analyzing data (both simulated and real). If a student cannot devote this amount of effort to the class, I would encourage them to find another course this semester. Ultimately students will only get from this course what they put into it.

Why should a student work so hard for a course like this? A number of reasons come to mind:

- Applied Bayesian statistics is, from my perspective, the most exciting and promising field of research in political methodology.
- Students who master these skill will be encouraged to attend the political annual methodology meeting, and will be attractive on the political methodology job market.
- All students who complete the course will have a paper in hand ready for submission to a peer-reviewed journal.
- Much of the work, especially computer programming, is satisfying.

My responsibility as the instructor is to keep things on track, guiding us through the most important material. Your responsibility as students is to faithfully meet the course requirements, exceed them when necessary, and work very hard to master the material.

Primary Textbook

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 2003. *Bayesian Data Analysis, Second Edition*. Boca Raton, FL: Chapman Hall/CRC. ISBN: 158488388X.

Requirements and Evaluation

The requirements for this course are simple---do the readings ahead of time, attend class, and complete all assignments with the utmost professionalism. I will lecture during most of the class sessions, but the content of my lectures will be determined by students, some based on issues raised in the online discussion.

To facilitate this interaction, we will be using the discussion board on the Telesis system. All questions about the readings, problem lets, etc. should be posted on the discussion board. I will contribute to the board, and I encourage students to respond to questions when they have the answer. *I will not answer private email about substantive course material, nor about problems with software.* All questions of this sort should be posted on the discussion board. I will respond to some questions, others I will not. Part of the learning process is figuring out how to do things without assistance.

There are two formal requirements for the course. First, each week I will assign a handful of problems, most of them from the Gelman text. Your answers should be professionally prepared (ideally using LaTeX), and submitted to me in class each week. Note that solutions to nearly all problems are available on Professor Gelman's website. You can use them in whatever manner you wish, but I strongly suggest that you do not use them at all. If you lean on this crutch throughout the semester, you will take very little from this course. Some of the problems will be drawn from material you are reading for a particular class; others will come from the material discussed in class.

Second, each student will complete an article-length research paper that uses Bayesian methods to answer a substantive research question. I will work with each student on their paper. Final course papers are due December 15, and I hope that all of them are shortly thereafter submitted to journals. Throughout the course of the semester, various parts of the project will be due. These should be new projects (i.e., projects not done previously in another class) unless prior arrangements are made. If you do not have a project in mind, I encourage you to find another class.

Problem sets account for 40% of the course grade, participation in the discussion forum accounts for 20% of the course grade, and the final paper counts for 40% of the course grade. *I have changed my policy regarding incompletes. I will not assign in this course (or, for that matter, any other), an incomplete.* All problem sets are due in class, and the final paper is due December 15.

Supplementary and Reference Texts

In addition to the required text, we will read a number of journal articles which will be made available in advance (and added to the syllabus). Some of the following texts might also be useful.

Jose M. Bernardo and Adrian F.M. Smith. 1994. *Bayesian Theory.* New York: Wiley.

Morris H. DeGroot and Mark J. Schervish. 2002. *Probability and Statistics, Third Edition.* Reading, MA: Addison-Wesley.

Merran Evans, Nicholas Hastings, and Brian Peacock. 2000. *Statistical Distributions.* New York: Wiley.

Jeff Gill. 2002. *Bayesian Methods: A Social and Behavioral Sciences Approach.* Boca Raton, FL: Chapman Hall/CRC.

Anthony O'Hagan. 1994. *Kendall's Advanced Theory of Statistics, Volume 2B, Bayesian Inference.* London: Edward Arnold.

Christian P. Robert. 2001. *The Bayesian Choice, Second Edition.* New York: Springer.

Christian P. Robert and George Casella. 1999. *Monte Carlo Statistical Methods.* New York: Springer.

Martin A. Tanner. 1996. *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions, Third Edition.* New York: Springer.

Software

We will primarily use R (http://www.r-project.org) for this course which is available for download from the web. Computation is an essential part of applied Bayesian statistics (or, for that matter, applied statistics). The best way to learn programming is by doing. I will provide some guidance in class, and will distribute some code on Telesis, but software problems are up to students collectively. *Post your questions on the discussion board, and work collaboratively to solve them.* We will be using the packages MCMCpack (http://mcmcpack.wustl.edu) to fit some standard models, along with coda (http://www-fis.iarc.fr/coda/) for summarization. Students will also implement non-standard models themselves in R.

Some students might want to learn how to use WinBUGS (http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml), or an open implementation of the BUGS language called JAGS (http://calvin.iarc.fr/~martyn/software/jags/) which is still very-much in development. These languages are relatively easy-to-use for quickly coding non-standard models. I do not use WinBUGS or JAGS, so I cannot serve as a resource, but you might want to work collaboratively on this.

If you have limited experience with R, the first place to go is the excellent manuals available on the Comprehensive R Archive Network (http://cran.r-project.org/). Begin with *An Introduction to R* and *The R Language Definition*. There are other contributed manuals at various levels of sophistication. There are also three published books I would recommend:

Peter Dalgaard. 2002. *Introductory Statistics with R*. New York: Springer.

John Fox and Georges Monette. 2002. *An R and S Plus Companion to Applied Regression.* Thousand Oaks, CA: Sage.

William N. Venables and Brian D. Ripley. 2000. *S Programming*. New York: Springer.

Calendar

Note: <Red Text> assignments ☑ Expand Topics ☐ Expand Assignments ☐ Include History

Date	Time	Description
09/01/04 Wed	11:00A-12:00P	Introduction
09/07/04 Tue	1:00P-3:00P	NO CLASS
09/14/04 Tue	1:00P-3:00P	Background Gelman, Preface, Chapter 1, and Appendix A. Efron. 1986. "Why Isn't Everyone a Bayesian?". *The American Statistician*. 40:1-5. Box. 1976. "Science and Statistics." *JASA*. 71: 791-799. Western and Jackman. 1994. "Bayesian Inference for Comparative Research." *APSR*. 88: 412-423.
09/21/04 Tue	1:00P-3:00P	One Parameter Models Gelman, Chapter 2.
09/28/04 Tue	1:00P-3:00P	Session 4
10/05/04 Tue	1:00P-3:00P	Session 5

10/12/04 Tue	1:00P-3:00P	Session 6
10/19/04 Tue	1:00P-3:00P	Session 7
10/26/04 Tue	1:00P-3:00P	Session 8
11/02/04 Tue	1:00P-3:00P	Session 9
11/09/04 Tue	1:00P-3:00P	Session 10
11/16/04 Tue	1:00P-3:00P	Session 11
11/23/04 Tue	1:00P-3:00P	NO CLASS
11/30/04 Tue	1:00P-3:00P	Session 13
12/07/04 Tue	1:00P-3:00P	Session 14
12/15/04 Wed	4:00P-5:00P	FINAL PAPERS DUE

Assignments

Assignment One (due: Tue. 09/14/04 01:00P)

Gelman, Chapter 1, Problems 1, 4, 6. Also begin work on 9, which will be assigned for the following week and discussed in class.

Also, write a research design for the paper you will write this semester. It should be about five pages in length. Make sure to identify the key substantive issue, how it fits into a literature (or literatures), and the dataset you plan to analyze. If possible, obtain the data and provide summary statistics.