$\widehat{\text{Gov 2003}}$: An Informal Bayesian Data Analysis Class

Spring 2010

Course Details

"Instructor": Patrick Lam (`plam@fas.harvard.edu`)
Class Location: CGIS K354 (most weeks)
Class Time: Tuesdays 4-6pm
Office Hours: Whenever you see me or email me

Overview

The typical methods sequence in the Gov Department in most years has been Gov 2000 and Gov 2001 for the first year, followed by Gov 2002 (a special topics course) and Gov 2003 (a Bayesian course) in the second year. However, due to faculty departures and other circumstances, neither Gov 2002 nor Gov 2003 will be offered this year, which leaves current methodologically inclined second years in a bind. The lack of a Bayesian course in the department is especially troublesome given the increasingly widespread use and popularity of Bayesian methods in the discipline.[1]

The recent trend toward the use of Bayesian methods in political science is not a fad. Until about 10-15 years ago, doing Bayesian analysis in a sophisticated way was not possible because of computing limitations. Most political scientists were also not trained in nor familiar with Bayesian methods. However, with the ever increasing advances in computing and more sophisticated methods training, Bayesian methods will surely become much more popular in the field as a way to build and estimate more complex statistical models. As future scholars, we will put ourselves in a better position by becoming familiar with some of the basics of Bayesian statistics.[2]

To fill the gap between the need to learn Bayesian statistics and the lack of a course offered in the department, we have come up with a solution in the form of $\widehat{\text{Gov 2003}}$ (Gov 2003 hat), an informal class on Bayesian data analysis that attempts to approximate the traditional Gov 2003 class in the department. This class carries with it no formal class credit, so participants will be taking it solely for the purpose of learning. However, participants will be expected to put in as much effort as a normal methods course.

Because of the nature of the course, we will be going at a relatively slower pace to accommodate all participants. The material in this course will likely be significantly harder than anything you have encountered in previous methods courses. However, I think this is a unique learning opportunity, as participants can learn complicated Bayesian methods at a reasonable pace without the pressure of grades or impressing faculty. We will also cover some more advanced methodological topics that are not necessarily Bayesian in nature. This class is an opportunity to ask all the dumb methods questions that you would want to ask, but are afraid to. It will also be much more hands-on and student-oriented than other methods classes.

Prerequisites

Although there will be no strictly enforced prerequisites, knowledge of basic probability, maximum likelihood, and other material from Gov 2001 will be very helpful. We will do some quick review in the beginning.

---

[1] A quick and messy search using the keywords "bayesian" and "data" returned 66 articles from the *APSR*, 34 articles from *IO*, and 111 articles from *Political Analysis* since 2000.

[2] There may be some of you who think that you don't need to learn Bayesian to be a good scholar using quantitative methods in the future. While this may be true, it may also be a risky decision. What if 40 years ago, somebody said they didn't need to learn logit because they knew OLS.

Course Requirements

Every week, I will be assigning short problem sets based on the week's material. Because this is an informal class that carries no credit, I will not be grading these problem sets per se, but I would still like everybody to turn in something. The problems do not have to be done in LaTeX or even correct or complete. I would just like to see that everybody put in valiant effort and I would also like to monitor progress.

In any methods course, you will not learn the material unless you do practice problems. This is especially true in this course. However, because the course carries no official credit, there is every incentive for participants to shirk and not work on the problems. Because this course will not succeed without every participant working hard, we will have the following accountability mechanism for the problem sets:

- At the beginning of every class, participants will be randomly selected to present the problems on the board. The selected few will walk through how they approached the problem and how they tried to solve the problem. You are not expected to necessarily get them all correct. In fact, I think we can all learn more from being wrong than being right. But you will be expected to have thought through the problem and solved it to the best of your abilities.

- Participants will be selected at the beginning of each class through random sampling in R. If for any reason, you will be late or absent from a specific class meeting, please email me beforehand to let me know. If you miss the random selection part in the beginning and do not let me know ahead of time, you will be automatically volunteered to present a problem at a randomly chosen later date.

- If you are randomly chosen to present a problem but you did not work on the problem beforehand, that is perfectly fine. You will have an opportunity to practice your ability to improvise and do math on the board in front of your peers.

The point of this is not to embarrass anybody. This not only serves as an enforcement mechanism, but it also serves as a great opportunity to develop teaching and communication skills (something we all need). Again, you are not expected to get the problems all correct. I expect many of you to struggle, and we will walk through all the problems together, but you are expected to have put in effort to work on the problems during the week. Working with and consulting each other is highly encouraged. The course will only be successful if everybody puts in the effort. But I also think this is a unique opportunity to learn more methods than you ever will in one class.

Because of the nature of the course, I will not accept "auditors". Anybody who sits in the room will be expected to fully participate.

Office Hours

I won't be holding office hours, but if you need me, feel free to find me either via email or search party through CGIS. It would be better to consult your fellow peers first and come find me as a group if you are working on the problems, but I will be happy to meet with you individually as well. If I'm around and relatively free, I can meet you or we can set up an appointment.

Books

While I won't require you to purchase a textbook, it would be a good idea for you to have one if you intend to do any Bayesian statistics outside the class. The material we will cover in class is insufficient and should be supplemented with a textbook on Bayesian analysis. For political scientists, I recommend one or more of the following.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald Rubin. 2004. *Bayesian Data Analysis.* 2nd edition. Boca Raton, FL: Chapman & Hall/CRC.

Gill, Jeff. 2007. *Bayesian Methods: A Social and Behavioral Sciences Approach.* 2nd edition. Boca Raton, FL: Chapman & Hall/CRC.

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

The first book by Gelman et al. is the standard textbook in the field and should be owned by all serious Bayesians. However, some may find the treatment to be a bit complex. The Gill textbook is written at a slightly lower level for political scientists and is a good reference. The Jackman textbook was recently published so I am not very familiar with it, but he is a well-known Bayesian methodologist in the field and generally has a great knack for explaining Bayesian concepts in his writing.

Computing

For the most part, we will be using R (`http://cran.r-project.org/`), so knowledge of the R programming language is important. Please speak to me if you do not have experience in R.

I may also do an introduction to the BUGS language that is popular amongst those who do Bayesian statistics. Although WinBUGS is the most popular program for the BUGS language, we will most likely be using JAGS (`http://calvin.iarc.fr/~martyn/software/jags/`), which is available across multiple platforms and uses virtually the same language as WinBUGS. You are not expected to know the BUGS language at this point.

Course Schedule

The pace of the class will depend entirely on how much progress we are making. Because of the nature of the class, you are highly encouraged to stop the class at any time to ask any and all questions. We will hopefully be covering the following topics in approximate order:

1. Introduction

2. Probability Review

3. Maximum Likelihood Review

4. Introducing and Comparing the Bayesian Approach

5. Conjugate Models

   - One Parameter Models
   - Multi-Parameter Models

6. Computational Approaches: EM and MCMC

   - The EM Algorithm
   - Gibbs Sampling
   - Metropolis-Hastings Algorithm
   - Checking for MCMC Convergence

7. Data Augmentation Models

   - Bayesian Ideal Point Estimation

8. Model Checking and Model Comparison

9. Hierarchical Models

Depending on how fast we go, we may talk about other topics or go through a few substantive papers that use Bayesian statistics.