EPPS 7390: Bayesian Analysis for Social and Behavioral Sciences
Fall 2017
Tuesday 4:00-6:45, GR 3.402A

Prof. Patrick T. Brandt
E-mail: pbrandt@utdallas.edu
Web: http://www.utdallas.edu/~pbrandt
Office: Green 2.802
Hours: Monday and Tuesday 1-4pm, and by appointment

TA: Lola Arowolo
E-mail: lola@utdallas.edu
Office: Green 2.816
Hours: Wednesday and Thursday 9am-12pm

Description

This course will cover advanced Bayesian methods in political economy and social science statistics. Initial weeks will cover basic maximum likelihood methods, methods of moments and non-linear optimization methods. Interpretation and application of these methods will be emphasized. Later weeks will address Bayesian methods. Readings and applications will be based on current political science, economic, and political economy research applications. Use of different software packages and other tools will be addressed. Students will be required to present current research in the course. Specific course topics will be developed based on the research interests and projects of the students enrolled.

The emphasis of this course is on the acquisition and understanding of Bayesian analytical techniques. It is very much an applied course, but for applications to be informed, there is a fundamental level of theoretical knowledge required. Many mistakes will be made by those who only know how to interpret findings and use "canned" computer packages. We emphasize that students know exactly what is going on when they do statistical analysis.

Student Learning Objectives

On completing this course, students will be able to:

- Apply and develop maximum likelihood and Bayesian statistical models for social science data.
- Recognize the basic assumptions, as well as strengths and weaknesses, of maximum likelihood and Bayesian methods.
- Implement and interpret maximum likelihood and Bayesian analysis in their own independent research project.

Required texts and course materials

Texts

The following book has been ordered:

- Gill, Jeff. 2006. *Essential Mathematics for Political and Social Research*, Cambridge University Press. or "EMPSR".
- Gill, Jeff. 2014. *Bayesian Methods: A Social and Behavioral Sciences Approach, Third Edition* or "BMSBSA".

In general you should use all the books as reference materials. If I have not indicated a reading from a book it does not mean that a topic is not covered in it. Thus, use the texts as reference materials for each week and take the indicated readings as a guide. It is my expectation that you will find additional texts and resources to supplement the assigned course materials. Some sources will be indicated in course lectures and notes.

Students will be expected to use other advanced sources for data and statistical reference, including JSTOR, journal databases, Econlit, Statlib, POLMETH, and other resources to which you will be directed.

eLearning

ELearning is used in this class for distributing course materials (notes, datasets, code, etc) and to check your progress in the class. You can access the course page at `http://elearning.utdallas.edu`. Additional information about how to use eLearning is available at this site as well. You will need a UTD net-id to access this site. You are responsible for announcements made through eLearning.

Lecture notes

I will make available copies of my lecture notes / slides, and code examples via eLearning. I will typically provide these in advance of covering the material, usually the night before or the day of the lecture.

Computing and Course Materials

A major component of applied statistics is using computers and data to implement models and test hypotheses. Toward that end, students should be expected to utilize a variety of statistical packages for their work. In this course we will primarily use R (`http://www.R-project.org`) and JAGS (Just Another Gibbs Sampler, `http://mcmc-jags.sourceforge.net`). You should install both of these software programs on your local or personal computer for this course. If you use a PC, there is another Bayesian modeling software called WinBugs (`https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-winbugs/`) that you might want to check out. If you have these already installed, make sure you have the most up to date versions.

Unlike my previous courses, I will begin this course with an overview of statistical computing. The reason is twofold. First, at this stage if you are still pursuing methodology training you need to develop some facility with various computing tools. Second, learning these tools will facilitate your work on the course material. The goal here is to get you doing replicable and well documented social science data analysis with minimal effort. *You should invest in these skills now since a) they are low cost at this stage in your careers, b) you will not have time to do easily it later (as a professor, researcher, consultant), and c) the tools I am going to show you have served me well for over fifteen years and will for the next ten at least.*

I will also make programs and code available on an as needed basis in several statistical packages to demonstrate techniques. Note that some things are easier to do in different packages and then convert the data to another format for analysis (I often construct variables and datasets in Stata and then move them to other software for analysis.) Your best resource for learning and implementing new methods is your peers and the voluminous manuals that come with statistical software!

Grading, Assignments, and Course Policies

Grading

There will be approximately five to six assignments, an exam, presentations, and an original research paper. Theses graded assignments compose the following percentages of your final grade:

- Assignments: 25%
- Exam: 25%
- Presentations: 25%
- Paper: 25%

The exam will be a take-home exam, approximately during week 8.

Late assignments and papers will be penalized. Late assignments and papers will be penalized 10 points per day. Papers or assignments submitted via e-mail will be considered received by the date and time stamp on the e-mail received in my e-mail inbox. The paper will be due around the date of the last day of classes (December 9th).

My strong preference is to not allow incompletes in this course, since you need most of the material completed for later course work or dissertation research. If you believe you will need to take an incomplete you 1) must follow university policy (as stated below) and 2) should contact me as soon as practicable (i.e., not the week of the final exam or the day before the final paper is due).

Grades are based on the standard grading scale: A = 100-90, B = 89-80, C = 79-70, etc.

If you have any questions about your grade on an assignment, please wait until 24 hours after receiving your assignment before discussing the grade with me. There are no exceptions to this policy.

If you wish to have an assignment re-graded, it must be returned to the me within two days of the day it was returned (if I am not available that day, ask the staff of the School of Economic, Political and Policy Sciences to leave it in my box). Assignments to be re-graded must include a memo stating the reason why you believe they assignment should be re-graded. Finally, re-graded assignments can be graded higher, lower, or the same as the initial grade.

Assignments

Assignments and exams will cover applied and theoretical problems. Assignments should be typewritten as much as possible. I realize that may require setting mathematical text or typographical symbols. This can be done in standard word processing software. I will be covering the use of LaTeX for this task in the initial class sessions. I know some of you have mastered this already, but you can always learn more tricks.

Any statistical output or data analysis you do should be fully interpreted and presented as though it were being sent for publication to a journal. *This means that regression output from your statistical package of choice that is copied into a word processor document is unacceptable. You should take the time to typeset the results into a meaningful table or present a well documented and coherent graphical summary of any results.* If you have any questions about what to include in your data output and assignments, consult empirical work in standard journals (APSR, ASR, Econometrica, ISQ, AJPS, JOP, etc.) or ask. Assignments that do not meet this requirement will be returned and not graded until revised.

You may work together on assignments, but each person must turn in their own work. Working together has two benefits. First, it can help you see if you really understand the material (if you can explain it to someone else and convince them that you are right, you probably are). Second, it gets you in the habit of working with others to solve problems (and remember that most research is coauthored!) You do need to be

careful about two issues in the course of working together. The first is plagiarism. The second is letting the person who "gets it" do all the work. Just because someone looks like they have solved a problem does not mean it is the correct answer. One of the things I have seen before is that working together can help on the "easy" problems, but on some of the harder problems, a "group" will often come up with the wrong answer. Beware of "groupthink"!

Finally, if you have questions about the assignments, I encourage you to come and ask **me** about them. Pounding your head on a desk for 6 days and then coming to get help on day seven (when the assignment is invariably due) is poor form and is not going to help you learn the material. One of the best methods I have found for asking and answering questions for this course is e-mail. The benefit of e-mail is that it forces you to compose your question(s) very specifically and to think through what you are asking logically. In addition, I can generally offer a faster response via e-mail than if you wait for office hours.

Presentations

Each student will do one presentation in the course. The presentations will cover a topic (about which you will have some choice) in the latter weeks of the course. I will "model" one of these presentations before you have to do one. The presentations are planned to do several things:

- Make you comfortable presenting technical / methods topics (albeit to a very friendly audience).
- Get you prepared for teaching this kind of material in the future (even if you just want to teach at a liberal arts college, as a junior faculty member you will be called upon as the "methods" person either in the classroom, with students, or by colleagues since you are the most recently trained person!)
- Get you collaborative research experience. Much if not all, sophisticated empirical work is done by "teams". You will do your presentations as a team project. It is good to learn now what your strengths and weaknesses are as part of a research team.

Presentations will be no more than 15(!) minutes, including time allocated for questions from the class. The presentations should include the following components:

1. The mathematical details of the statistical model you are covering. This does not have to be all of the gore, but it should convey the basic idea.
2. A substantive example(s) that illustrates how the methods is used. The example need not be "new": it can be a replication of previous work (in which case you need to get the data and reproduce published results), it can be an extension (you apply a new method to an already addressed empirical question), a Bayesian implementation of something that was previously modeled with frequentist methods (so showing the differences in Bayesian and non-Bayesian methods, etc.) Odds are fairly high that you will be meeting with me before your specific presentation about this part.
3. Detailed illustration of how to implement the method you are presenting. This means you should provide example code for your application.
4. Relevant citations to the model(s) in the literature for further reading.

For each presentation you will need to turn in the following materials at a minimum two hours before the presentation, so I can make the available to the other students in the course via eLearning:

1. Copies of slides (in either Powerpoint, Keynote, or PDF format the latter being preferred). These items should be e-mailed to me so I can review them and post them for the course.

2. Copies of any data, code or other replication materials to implement your example(s).

3. An annotated bibliography of at least 10 references on the topic you are presenting. This list should *not* include course texts or standard reference texts (e.g., do not include any of the Gill books, Gelman et al.'s standard Bayesian reference book, Geweke's text, McCullagh and Nelder, etc.) References should include a mix of technical and application citations. A BibTeX citation format is preferred for these references so we can compile a master reference file for LaTeX by the end of the course for all of the students.

The following are a list of possible presentation topics. If you are interested in others, please come talk to me.

- Multilevel / hierarchical models
- Robustness versus sensitivity analysis in Bayesian models
- Reversible-jump Markov chain Monte Carlo and exact Bayesian inference methods
- Event count models
- Event history / duration models (with and without frailty or with split populations)
- Multinomial / polychotomous choice or outcome models
- Censored / truncation model (tobit, censored regressions, etc.)
- Panel data / random coefficient models
- Data augmentation
- Compositional data model for proportions
- Penalized ML and profile methods.
- Shrinkage priors, L1 and L2 regularization, lasso
- Dynamic linear models / state space models with unknown parameters and or non-normal measurement equations
- Categorial and regression trees
- Random forest models
- Anything else you think might be interesting to those in the course.

Papers

The paper will require students to analyze and interpret an advanced Bayesian model. The paper you write should be based on your own research and interests—there is no requirement to use a particular dataset or technique. However, the paper should be original (it should include an analysis used in previous courses, conference papers, or other writing only for comparison). The following are suggested models for a paper:

- A conference paper (examples can be seen on my webpage).
- A dissertation or thesis chapter (examples available upon request).

- A replication article (see the replication standard for Political Analysis at `http://www.oxfordjournals.org/polana/for_authors/general.html` or Gary King (and others) comments at `http://gking.harvard.edu/papers`

Before beginning your paper, you are **REQUIRED** to come and talk with me about your paper. At or prior to this meeting, please provide a ONE PAGE research design and lays out the following for your paper:

- Main research question you are addressing.
- Hypothesis (or Hypotheses) you wish to evaluate in your paper.
- Data and variables to be used in the analysis.
- Tests and methods that will be used to evaluate the hypotheses.
- Tentative listing of the techniques and models you may use.

The data you use for the analysis is something that you should already have or to which you have easy access. Data collection is not a topic we are covering in this course and time spent building large complex datasets will detract from your ability to complete the paper adequately.

The meeting to discuss the paper should be scheduled before the 7th week of class (roughly October 10th).

(HINT: YOUR PAPER SHOULD BE ON THE SAME THING AS YOUR PRESENTATION!)

Attendance

It should go without saying that in a class of this size your attendance is easily noted and therefore required. If you are unable to make a class or will be late, advise the instructor as far in advance as possible.

Course conduct

The following rules apply in class:

1. Turn off your cell phone. It is VERY distracting to others. "Off" means that it does not ring OR vibrate. I will ask you to leave if your phone rings. (Exceptions to this policy can be made, come to talk to me.)
2. Do not fall asleep. It is rude and distracting. Bring coffee if you need it (I do.)
3. Be polite and courteous to your fellow students.
4. Raise your hand when you want to be recognized to answer or ask a question. If you do not raise your hand I will not recognize you or your answer.
5. You are responsible for things: therefore if you miss something I announced or fail to complete an assignment, my response will be "How is this my problem?"
6. Note that this syllabus is not a contract. It is subject to change at my discretion.
7. Class starts at 4:00pm. Not 4:15. Be here on time.
8. Respect my time and I will respect yours. We are both busy.
9. *You are expected to be pro-active in anticipating and planning for any absences or problems you will have in completing course work.* Make arrangements for possibly missed work prior to the due date is preferable and more likely to be successful than doing it after the due date.

Course Outline

Readings listed for each week are suggestions. My lectures and discussion will parallel these readings. The readings are listed in order of importance (from most to least important). Feel free to ignore or consult them as needed. I will regularly post or hand out lecture notes on the material we are covering.

In general, I have my own "order" to covering the material. This may differ from the texts, or what others might do. My experience is that this layout of the material works well.

Week 1 (August 22): Introduction, Statistical Tools, and Writing about Data

- R, review as necessary. Focus on data and structures of R/S in Chapters 1–3 of the main R Getting Started Manual.
- BMSBSA, Chapter 11 (skim).
- Download and install JAGS on your computer from `http://mcmc-jags.sourceforge.net/`

Week 2 (August 29): Probability, likelihood, and Bayesianism

- EMPSR, Chapters 7 and 8 (and any others as necessary)
- BMSBSA, Chapters 1 and 2

Week 3 (Sept 5): Maximum Likelihood, GLMs, Inference, and Quantities of Interest

- EMPSR, Chapter 6, esp. section 6.1–6.4.
- BMSBSA, Appendices A and B.
- Long, J. S. 1997. *Regression models for Categorical and Limited Dependent Variables*, Chapter 3
- King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2):347-61.
- See also the Clarify website for the King et al. paper at `http://gking.harvard.edu/clarify/docs/clarify.html` and `http://gking.harvard.edu/zelig/`

Week 4 (Sept 12): Subjective and Objective Models of Probability, Bayes Theorem

- BMSBSA, Chapters 2, 3, 5.

Week 5 (Sept 19): Bayesian Priors

- BMSBSA, Chapter 4

Week 6 (Sept 26): Model Quality and Comparison

- BMSBSA, Chapters 6 and 7

Week 7 (Oct 3): Bayesian Estimation, Monte Carlo, and MCMC

- BMSBSA, Chapters 9 and 10
- Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo" *American Journal of Political Science*, 44(2): 375-404.

Week 8 (Oct 10): Bayesian Inference, Reporting, Convergence

- BMSBSA, Chapter 13

Week 9 (Oct 17): Exam, take-home, no class meeting
Week 10 (Oct 24): Advanced MCMC

- BMSBSA, Chapters 12 and 14

Week 11 (Oct 31): Sequential MCMC and Particle Filters

- Lopes, Hedibart and Ruey Tsay. 2010. "Particle Filters and Bayesian Inference in Financial Econometrics." *Journal of Forecasting* 30: 168–209.

Week 12 (Nov 7): Building a new Bayesian model from scratch

- Brandt, Patrick T. and Todd Sandler. 2012. "A Bayesian Poisson Vector Autoregression Model" *Political Analysis*. 20(3): 292–315.

Week 13 (Nov 14): Presentations
Week 14 (Nov 21): No class, Thanksgiving Break
Week 15 (Nov 28): Presentations
Week 16 (Dec 5): Presentations

These descriptions and timelines are subject to change at the discretion of the Professor.
UT Dallas Syllabus Policies and Procedures

The information contained in the following link constitutes the University's policies and procedures segment of the course syllabus.
`https://go.utdallas.edu/syllabus-policies`