PSY757 Advanced Topics in Statistical Analysis: Bayesian Statistics

Spring 2012

Time: 4:30 pm – 7:10 pm Mondays
Classroom: Innovation Hall 209

Instructor of Record: **James Thompson**
2056 David King Hall
Ph: 703-993-9356
Email: jthompsz@gmu.edu
(please remember to **always** use your gmu account)
Co-Instructor: **Patrick McKnight**
Office Hours: 3.00pm - 4.00pm Mon or by appt.

Objectives:

This Advanced Topics in Statistical Analysis class will provide an introduction to Bayesian statistical analysis for use in the social sciences. Patrick McKnight and I co-teach this class and alternate primary responsibilities each year. We will provide an introduction to probability theory and Bayes theorem, and then move on to Bayesian statistical analysis. The overall goal of this class is to provide you with enough background so that you understand the principles behind Bayesian analysis, and to provide practical information about how you can apply these principles to examine questions relevant to the social sciences. We will spend a considerable amount of time dedicated to "hands on" examples and exercises. These examples and exercises will be performed in the statistical software package, R (see prerequisites). We will not spend very much time presenting the argument in favor of why you would to use Bayesian statistics, as we will assume that you have shown at least some appreciation of this simply by signing up for the class. We will, however, include some supplementary readings and materials for those interested in such issues.

Prerequisites.

Due to the nature of the material and the relevance to research, we assume all students will have successfully completed the introductory graduate course sequence in statistics (PSYC 611/612 or its equivalent). Some understanding of the basics of probability theory, such as what a distribution is, the difference between discrete and continuous distributions, and what is a probability density, will be expected (see **Before You Start This Class**). We do not intend to cover in great detail the statistical models underlying Frequentist models so you may want to reread some material on ANOVA and regression if you feel weak in those areas. In addition, this is not a math class. Understanding of basic linear algebra, and some familiarity with integral notion is sufficient.

Required Readings:

Kruschke, J.K. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS.* Burlington, MA: Academic Press.

Additional Readings:

Gonick, L. and Smith, W. (1993). *The cartoon guide to statistics.* New York: Harper Collins.
Albert, J. (2009). *Bayesian computation with R (2nd Ed.).* New York: Springer.
Berry, D.A. (1995). *Statistics: A Bayesian Perspective.* Belmont, CA: Duxbury Press.
Lynch, S.M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* New York: Springer.

Before You Start This Class:

If you feel you might be a little rusty on the basic concepts of statistics and probability, then I highly recommend the book by Gonick and Smith (1993). A slightly more advanced introduction, including an introduction to Bayes theorem, can be found in Berry (1995), and I encourage everyone to read at least chapters 4 to 8 of this book. You will be expected to have both R and RBugs installed and working on your computer* **before** the first class. We will begin working on examples in the first week and you will need to have R up and running before then.

*If you do not have access to a laptop you must make arrangements with the instructor **prior** to the first class.

Assessment:

Assessment will be based on performance of homework assignments assigned each week, as well as in-class exercises and discussions. There will be no exams or essays. All the assignments are mandatory, and penalties will apply for late work unless you have a legitimate excuse. The material covered and assignments are cumulative, so it is important that you keep up. If you are going to be late with an assignment, you must notify the instructor **prior** to its due date.

Attendance Policy:

Although you will not be graded on attendance, this is a graduate level course and I expect to see you in class each week.

GMU Honor Code:

George Mason University has a code of Honor that each of you accepts by enrolling as a student. You should read and become familiar with this code at http://mason.gmu.edu/%7Emontecin/plagiarism.htm. The expectation is that all of the work you do for this class will be the work of one individual. However, you are fully encouraged to discuss the readings and topics raised in this class with your fellow students.

Disabilities:

If you are a student with a disability and you need academic accommodations, please see me and contact the Disability Resource Center (DRC) at 703-993-2474. All academic accommodations must be arranged through that office.

Grades:

Total 100 points, letter grades as follows:

A: 90-100	B-: 77-79
A-: 87-89	C: 70-76
B+: 84-86	F: 0-69
B: 80-83	

Important Dates

Last day to drop without penalty January 31; Last day to drop February 24; Spring Break March 12-18.

Schedule of Classes

(Schedule is subject to change, changes announced in class or via email)

Week 1. **Probability, Conditional Probability, and Bayes Rule.**

Kruschke (2010) Chapters 2-4.

Key points: Models of observation and models of belief; The sum of all probabilities equals one; Conditional probabilities are simply the probability of an event, *given* that we know another event is true; Posterior is prior beliefs updated with evidence; Contribution of prior and evidence to posterior depends on their respective precisions.

Week 2. **An Example with a Single Binomial Distribution: Analytic & Grid Approximation Approaches.**

Kruschke (2010) Chapters 5 & 6

Key points: For a simple example such as a coin-toss, and a beta distribution for a prior, we can use a mathematical formula to calculate the posterior; For non-beta priors and a single parameter model, discretizing continuous priors is a simple way to approximate the posterior.

Week 3. **An Example with a Binomial Distribution: The Metropolis Algorithm**

Kruschke (2010) Chapter 7

Key points: For multiple parameter models, we need to approximate the posterior by randomly sampling; The Metropolis algorithm is a heuristic for generating random samples from a target distribution; In MCMC, *Monte Carlo* refers to random sampling, like rolling die at the casino; *Markov Chain* refers to each successive step in the sampling algorithm being independent of all previous steps.

Week 4. **An Example with Two Binomial Distributions: Gibbs Sampling**

Kruschke (2010) Chapter 8

Key points: The Metropolis algorithm can be inefficient under certain circumstances; Gibbs sampling allows us to sample multiple parameter models by stepping though the resampling one parameter at a time; To use Gibbs sampling, it must be possible to generate samples from the posterior conditioned on each individual parameter; When it can be applied, Gibbs sampling is more efficient than Metropolis.

Week 5. **Hierarchical Models I: Introducing Hyperparameters**

Kruschke (2010) Chapter 9

Key points: *Hyperparameters* describe dependencies between other parameters; Evidence affects our beliefs about hyperparameters, and our beliefs about the dependence of parameters on the hyperparameters; If multiple parameters are dependent on a single hyperparameter, evidence that speaks to one parameter can affect the other parameters, reducing uncertainty.

Week 6. **Hierarchical Models II: Model Comparison**

Kruschke (2010) Chapter 10

Key points: Hierarchical modeling allows for model comparison; One can think of different models being dependent on a categorical hyperparameter; Different models can have different numbers of parameters; Nested model comparison is a useful way of testing models, but care is needed in the specification of models; Bayesian model comparison is especially useful for non-nested models.

Week 7. **NHST, Testing Point Hypothesis**

Kruschke (2010) Chapters 11 & 12

Key points: The outcome of null hypothesis significance testing (NHST) is dependent on the intentions of the experimenter, as it depends on the space of all possible (unobserved) data; Bayesian analysis requires the expression of prior knowledge; The highest density interval (HDI) or the range of practical equivalence (ROPE) can be used for making inferences; Model comparison can also be used for making inferences.

Week 8. **Revision of the Generalized Linear Model**

Kruschke (2010) Chapter 14

Key points: A GLM consists of predictor and predicted variables; These variables can be nominal, ordinal, or metric; Depending on whether predictor and/or predicted variables are nominal, ordinal, or metric, we can perform the equivalent of t-tests, simple and multiple regression, ANOVA, logistic regression, etc, etc, etc.

Week 9. **Single Group Means, Precision, Repeated Measures.**

Kruschke (2010) Chapter 15

Key points: For a metric predicted variable from a single group, we can estimate the posterior mean using a normal likelihood function; The posterior mean is the weighted sum of the prior mean and the data, with weighting from the relative precisions; For a normal likelihood, the prior mean is usually given a normal distribution and prior precision a gamma distribution; However, BUGS can handle arbitrary priors.

Week 10. **Simple Linear Regression**

Kruschke (2010) Chapter 16

Key points: A simple linear regression consists of a metric predictor and a metric predicted variable; We can use a hierarchical model to estimate the posterior distributions of the intercept, slope, and residuals; Priors are given for each of the three parameters, usually normal for intercept and slope and gamma for residuals; Using MCMC to estimate posteriors works best if data are standardized; HDI or ROPE, based on the posterior distribution, can be used for inferences about the slope.

Week 11. **Multiple Linear Regression**

Kruschke (2010) Chapter 17

Key points: Simple regression can be easily extended to include multiple metric predictors; Bayesian multiple regression copes well with correlated predictors; When testing a large number of predictors, one can set up a hierarchical model with a prior for a hyperparameter describing the relations between predictor regression coefficients.

Week 12. **Oneway ANOVA**

Kruschke (2010) Chapter 18

Key points: Use ANOVA when the predictor variable is nominal and predicted variable is metric; The ANOVA model includes predictors that have one component per predictor level; Coefficients indicate the size of predicted variable deflections when predictor goes from zero to one; Again, a hierarchical model is used with a hyperprior over the distribution of deflections; We allow the precision of the deflections to be determined by the data.

Week 13. **Multifactor ANOVA**

Kruschke (2010) Chapter 19

Key points: Multiple factors and their interactions can be incorporated into an ANOVA model; hyperpriors and their distributions are specified for each factor and interaction; Metric and nominal predictors can be included in ANCOVA models; Repeated measures models can be designed with varying complexity, with the simplest modeling the influence of subjects on the baseline.

Week 14. **Revision/Catch Up/Outstanding Questions.**