Bayesian Modeling in Biostatistics

University of Kentucky
BST 701, Spring 2013
Credit: 3.0

Lecture: 3:30 p.m - 4:45 p.m., Tuesdays & Thursdays
Room 206, Multidisciplinary Sciences Building (MDS 206)

Instructor: Patrick Breheny, Ph.D.
Office: MDS 205D
Phone: 218-2077
e-mail: patrick.breheny@uky.edu
Office hours: Whenever I'm in my office, or by appointment

Course description: This course provides a practical introduction to Bayesian modeling. Students will learn how to analyze data and build models within the Bayesian framework, with a special emphasis placed on hierarchical models. The primary emphasis will be on understanding modeling concepts and the modeling process, and on learning how to carry out these analyses using `R` and `BUGS`. Relatively speaking, less emphasis will be placed on the theoretical aspects of Bayesian statistics and the technical details of Markov Chain Monte Carlo methods. Hopefully, at the end of the class, you will:

- Understand the basic ideas, terms, and principles of Bayesian statistics, which in turn will allow you to more easily read articles and understand research involving Bayesian statistics
- Understand the similarities and differences between Bayesian statistics and frequentist approaches
- Be able to fit and debug Bayesian models using `R` and `BUGS`
- Be able to build hierarchical Bayesian models to accurately reflect multiple sources of variability

Textbooks:

- Lunn, D., *et al.* (2012): *The BUGS Book: A Practical Introduction to Bayesian Analysis.* Chapman & Hall/CRC.
- Gelman, A. and Hill, J. (2007): *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge University Press.

Prerequisite: BST 760 (or equivalent), BST 676 (or equivalent), or consent of instructor.

Course website: The course notes, assignments, data sets, and other relevant materials will be made available on the course web site: `http://web.as.uky.edu/statistics/users/pbreheny/701/S13/index.html`

Grading: Your grade will be based on five assignments, each worth 16% of your grade, and one final project worth 20%. The majority of the questions on the assignments will involve analyzing data from real studies — often, analyzing it several different ways and commenting on how the analyses differ from one another. However, problems may also involve mathematical derivations, simulations, and implementation of computational techniques.

The grading scale will be as follows:

85-100	A
70-84	B
55-69	C
Below 54	E

These assignments comprise your entire grade; consequently, they are longer and more involved than a typical homework assignment and you should expect to spend more time on them. I will distribute the projects well in advance of their due date and let you know when we have covered sufficient material to start on certain problems, as I would not advise waiting until the last minute to work on the entire assignment.

Assignments may be turned in either electronically or as a physical copy (preferred). If they are turned it electronically, they must be turned in as a `.pdf` file (*i.e.*, no Word documents). Each assignment will also involve writing code (to analyze data, make plots, etc.). Please turn in this code **separately and electronically**.

All electronic submissions should be made via DROPitTOme at `http://www.dropitto.me/pbreheny` and should follow the following naming convention: last name, assignment number, proper extension. So, for example, if Adam Smith was taking this course and turning in assignment 1, he would name the file `Smith1.pdf`. The associated code would be `Smith1.R` or `Smith1.bug` for `BUGS` code.

Final project: The final project for this class is quite open-ended. Any sort of project that involves Bayesian analysis, modeling, or computing is acceptable. Oral reports over these projects will be given during finals week; a short ($\approx$ 2-page) specification of the model and main results must be turned in at that time as well.

Computing: Unless you are working with extremely simple models, it is virtually impossible to carry out Bayesian modeling without a computer. We will discuss the use of the programs `BUGS` (Bayesian inference Using Gibbs Sampling) and `JAGS` (Just Another Gibbs Sampler). We will also be discussing the use of `R` as an interface to these programs. Knowledge of the basics of `R` (vectors, matrices, lists, functions, etc.) will be helpful. If you are entirely unfamiliar with `R`, you are still welcome to take the course, but please be aware that you may have to put in a bit of extra time at the beginning of the course to become familiar with the language. Throughout, I will be sharing code and demonstrating how to accomplish a variety of tasks in `R/BUGS/JAGS`.

Proofreading: Despite my best efforts, my notes occasionally have typographical errors. This will no doubt be especially true this semester, as this is the first time I have taught this course. If you see them, please tell me about them! Doing so will benefit not only you, but also your classmates and any future students of this course. I will announce corrections on the course webpage, or if the mistake is critical, send out an e-mail.

Electronic communication: I will occasionally send e-mails to the class through e-mail (to the account listed for you in the campus directory), so please check that account regularly.

Academic honesty: Academic honesty is highly valued at the University of Kentucky. You must always submit work that represents your original words or ideas. You may discuss the assignments and your solutions with other students, but your writeup must be your own. Specifically, copying and pasting another student's words or code is strictly forbidden.

Complaints: Students with suggestions or complaints should see me first, and if we cannot come to an agreement, I will direct you to the head of the department.

Disabilities: If anyone has a disability requiring special accommodations, please let me know as soon as possible, so that these arrangements can be made.

Religious observances: If a religious observance prevents you from taking an exam or finishing an assignment or project, please let me know in advance so that we can make arrangements for you to make up the work.

Inclement weather: The University of Kentucky has a detailed policy for decisions to close in inclement weather. The snow policy is described in detail at `http://www.uky.edu/MicroLabs/documents/temp/policies-weather.htm` or you can call (859) 257-5684.

Bayesian modeling is a very useful addition to a statistician's set of tools, particular when faced with multiple sources of information or hierarchical sources of variability. I hope that you will find both this course both interesting in its own right and useful as a set of tools that you can use in future modeling and research.

Course outline:

1. Part I: Single-level models
   (a) Basic concepts of Bayesian statistics: priors, posteriors, Bayes rule, Monte Carlo integration
   (b) Simple models and comparisons
      i. Single parameter models: binomial, Poisson
      ii. Multi-parameter distributions (*e.g*, normal, gamma) and nuisance parameters
      iii. Two-group comparisons
   (c) Bayesian regression models
      i. Linear regression
      ii. Generalized linear models
      iii. Model comparison
   (d) Advantages of Bayesian statistics
   (e) MCMC Sampling
      i. Gibbs sampling; the Metropolis-Hastings algorithm
      ii. Assessing convergence

2. Part II: Hierarchical models
   (a) Concepts in Bayesian hierarchical models
      i. Pooling and shrinkage
      ii. Latent variables
      iii. Comparison with mixed models
      iv. Empirical Bayes
   (b) Hierarchical linear models
   (c) Hierarchical generalized linear models
   (d) Debugging hierarchical models
      i. Identifiability
      ii. Speeding convergence
   (e) Understanding, summarizing, and checking the fitted models
   (f) Further topics, time permitting: analysis of variance, missing and censored data, causal inference