

# Introduction to Bayesian Data Analysis & Cognitive Modeling

Instructor: Adrian Brasoveanu, [abrsvn@ucsc.edu](mailto:abrsvn@ucsc.edu)

Fall 2012

## 1 Organizational matters

- Class: Tue 2:00PM–5:45PM, The Cave
- Office hours: Thur 2:30PM–3:30PM and by email appointment ([abrsvn@ucsc.edu](mailto:abrsvn@ucsc.edu)), Stevenson 259

### Web access

- There is an eCommons site for this course: <https://ecommons.ucsc.edu/xsl-portal>; check fairly regularly for announcements and/or new materials.
- The syllabus, readings, handouts, scripts etc. are/will be posted under Resources.

### Important note

If you qualify for classroom accommodations because of a disability, please get an Accommodation Authorization from the Disability Resource Center (DRC) and submit it to me in person outside of class (e.g., office hours) as soon as possible. Contact DRC at 459-2089 (voice), 459-4806 (TTY), or <http://drc.ucsc.edu> for more information on the requirements and/or process.

## 2 Goals of the course

Providing solid empirical foundations for increasingly sophisticated linguistic theories requires increasingly sophisticated methods of empirical investigation and statistical analysis of the resulting data. In addition, linguistic theories should be complemented and further constrained by cognitive theories of (a) how the richly structured, abstract representations and operations argued for in theoretical linguistics can be learned/induced from ‘raw’ observed data, and (b) the kinds of mechanisms that underlie the processing of such representations and operations in actual natural language usage.

The seminar will introduce tools that can address both of these ‘interface’ issues, i.e., the ‘interfaces’ of theoretical linguistics with both experimental data and formal cognitive models: we will introduce modern Bayesian methods of data analysis, as well as some basic cognitive models (based on Bayesian ideas) of learning abstract, highly structured representations of the kind deployed in theoretical linguistics.

The course will be very hands-on, by the end of it you should at the very least be able to analyze your own experimental data in fairly sophisticated ways and if the data is lacking, synthesize it and analyze the resulting synthetic data set. You should also have the very basic tools needed to pursue questions that require integrating linguistic theories and broader cognitive models.

**The overarching goal of the course** is to show how Bayesian probabilistic models provide a very flexible framework that can accommodate plausible reasoning about both noisy data and highly structured theoretical constructs of the kind used in theoretical linguistics and cognitive science in general.

The two extensive quotes below provide more details about both aspects.

### Why Bayesian Data Analysis?

“Traditional data analysis has many well-documented problems that make it a feeble foundation for science, especially now that Bayesian methods are readily accessible. Chief among the problems is that the basis for declaring a result to be ‘statistically significant’ is ill-defined: the so-called  $p$  value has no unique value for any set of data. Another problem with traditional analyses is that they produce impoverished estimates of parameter values, with no indication of trade-offs among parameters and with confidence intervals that are ill-defined because they are based on  $p$  values. Traditional methods also often impose many computational constraints and assumptions into which data must be inappropriately squeezed.

[...]

It is important to understand that Bayesian methods for data analysis are distinct from Bayesian models of mind. In Bayesian data analysis, any useful descriptive model of the data has parameters estimated by normative, rational methods. The descriptive models have no necessary relation or commitment to particular theories of the natural mechanisms that actually generated the data. Thus, every cognitive scientist, regardless of his or her preferred model of cognition, should use Bayesian methods for data analysis. Even if Bayesian models of mind lose favor, Bayesian data analysis remains appropriate.” (Kruschke 2010: 293)

### Why Bayesian Cognitive Modeling?

“Because Bayesian statistics provides a formal framework for making inferences, there are different ways it can be applied in cognitive modeling. One way is to use Bayesian methods as a statistician would, as a method for conducting standard analyses of data. Traditionally, the framework for statistical inference based on sampling distributions and null hypothesis significance testing has been used. Calls for change, noting the clear superiority of Bayesian methods, date back at least to the seminal paper of Edwards, Lindman, and Savage (1963), and have grown more frequent and assertive in the past few years [...]. It seems certain Bayesian statistics will play a progressively more central role in the way cognitive science analyzes its data.

A second possibility is to apply Bayesian methods to cognitive modeling as a theoretician would, as a working assumption about how the mind makes inferences. This has been an influential theoretical position for the last decade or so in the cognitive sciences [...]. These uses of Bayesian statistics as theoretical analogies have led to impressive new models, and raised and addressed a range of important theoretical questions. As with all theoretical metaphors – including previous ones like information processing and connectionist metaphors – “Bayes in the head” constitutes a powerful theoretical perspective, but leaves room for other complementary approaches.

A third way to use Bayesian statistics in cognitive science is to use them to relate models of psychological processes to data [...]. This is different from the data analysis approach, because the focus is not generic statistical models like the generalized linear model. Instead the goal is to relate a detailed model of some aspect of cognition to behavioral or other observed data. One way to think of the distinction is that data analysis typically does inference on the measured dependent variables from an experimental design – measures of recall, learning, response times, and so on –

whereas modeling applications typically do inference on latent psychological parameters – memory capacities, learning rates, decision criteria, and so on – that control the behavioral predictions of the model. It is also different from the use of Bayesian inference as a metaphor for the mind [...]. There is no requirement that the cognitive models being related to data make Bayesian assumptions. Instead, they are free to make any sort of processing claims about how cognition works. The goal is simply to use Bayesian statistical methods to evaluate the proposed model against available data.

This third approach [...] is an especially interesting, important, and promising approach, precisely because it deals with fully developed models of cognition, without constraints on the theoretical assumptions used to develop the models. The idea is to begin with existing theoretically grounded and empirically successful models of cognition, and embed them within a hierarchical Bayesian framework. This embedding opens a vista of potential extensions and improvements to current modeling, because it provides a capability to model the rich structure of cognition in complicated settings.” (Lee 2011: 1-2)

### 3 Schedule (subject to change)

We will work our way through a series of R scripts. The scripts are/will be posted on the eCommons page for the course. It is both highly recommended and required that you also work through those scripts on your own (R and JAGS required).

In addition, it is highly recommended that you at least skim the readings marked as [fyi only] and it is required that you do all the readings that are not marked as [fyi only].

#### 3.1 Introduction: The Basics (Week 1)

1. [Kruschke \(2011\)](#), chapters 3 & 4
2. additional materials (slides, R scripts) covering the following topics:
  - basic probability theory
  - introduction to bayesian inference

#### 3.2 Fundamentals applied to inference about binomial proportions (Weeks 1–3)

1. [Kruschke \(2011\)](#), chapters 5-9
2. additional materials (R scripts) covering:
  - (a) the fundamentals and basic inference binomial proportions:
    - examples of Beta distributions
    - examples of updating Beta priors with Bernoulli likelihoods
    - sequential update, model comparison, posterior predictive check
    - discrete prior distributions; updating discrete priors with a Bernoulli likelihood
    - general overview: classical inference, Bayesian inference and MCMC
    - introduction to Markov Chains
    - introduction to the Metropolis-Hastings family of sampling algorithms
    - inference for one binomial proportion with Metropolis-Hastings
    - tuning the random-walk Metropolis algorithm
    - inference for two binomial proportions with a Metropolis random walk

- (b) the fundamentals with JAGS and more advanced inference about binomial proportions:
  - introducing JAGS with the mean model: simulated data, R analysis, JAGS analysis
  - the structure of JAGS models: reexpressing parameters, number of chains, number of iterations, burnin, thinning, the Brooks-Gelman-Rubin (BGR) convergence diagnostic (a.k.a. Rhat), graphical summaries of posterior distributions
  - binomial proportion inference with JAGS instead of the Metropolis algorithm we built “by hand” for this purpose
  - comparison of 3 models for the same binomial proportion data with different uniform priors: posterior estimation with JAGS and computing the evidence / marginal likelihood for each model based on the JAGS posterior samples
  - inference for 2 binomial proportions with JAGS instead of the Metropolis algorithm we built “by hand” for this purpose

3. [fyi only] [Lee and Wagenmakers \(to appear\)](#), chapters 3, 6 & 9

### 3.3 Inference about (generalized) linear models and extensions (Weeks 4–8)

1. detailed R scripts (many are based on scripts from [Kéry 2010](#) and [Kruschke 2011](#), but pretty heavily modified) covering:
  - (a) basic linear and linear mixed-effects models
    - essentials of linear models
    - t-tests with equal and unequal variances (simulated data, R analysis, JAGS analysis)
    - simple linear regression (simulated data, R analysis, JAGS analysis)
    - goodness-of-fit assessment in Bayesian analyses (posterior predictive distributions and Bayesian p-values)
    - interpretation of confidence vs credible intervals, fixed-effects 1-way ANOVA (simulated data, R analysis, JAGS analysis)
    - random-effects 1-way ANOVA (simulated data, R analysis, JAGS analysis)
    - 2-way ANOVA w/o and w/ interactions (simulated data, R analysis, AGS analysis)
    - [skim/skip] ANCOVA and the importance of covariate standardization (simulated data, R analysis, JAGS analysis)
  - (b) basic generalized linear and generalized linear mixed-effects models
    - intro to generalized linear models
    - ‘Poisson t-test’ (simulated data, R analysis, JAGS analysis)
    - ‘binomial t-test’ (simulated data, R analysis, JAGS analysis)
    - ‘binomial ANCOVA’ (simulated data, R analysis, JAGS analysis)
    - inferring binomial proportions with hierarchical priors (random-effects for ‘coins’, i.e., basically, random-effect ‘binomial ANOVA’)
  - (c) more complex models: linear mixed-effects, generalized linear mixed-effects, extensions
    - linear mixed-effects models—random intercepts only, independent random intercepts and slopes, correlated random intercepts and slopes (simulated data, R analysis, JAGS analysis)
    - binomial GLMM (simulated data, R analysis, JAGS analysis)

- GLMMs that take into account inter-annotator disagreement (simulated data, R analysis, JAGS analysis)
  - ‘ordinal probit t-test’, i.e., an ordinal probit regression with only one predictor, namely a factor with 2 levels (simulated data, R analysis, JAGS analysis)
2. [fyi only] [Kruschke \(2011\)](#), chapters 14-22
  3. [fyi only] [Lee and Wagenmakers \(to appear\)](#), chapters 4, 5, 7 & 8

### 3.4 Applications in Cognitive Science (Weeks 8–10)

1. [Lee and Wagenmakers \(to appear\)](#), chapters 10-15—whichever ones we are most interested in and can get to:
  - memory retention: no individual differences, full individual differences, structured individual differences (ch. 10)
  - Signal Detection Theory and a hierarchical extension (ch. 11)
  - The SIMPLE memory model and a hierarchical extension (ch. 12)
  - Heuristic decision-making: Take-The-Best and various hierarchical extensions—very closely related to Optimality Theory (ch. 13)
  - the Generalized Context Model and various hierarchical extensions (ch. 14)
  - number concept development and the integration of multiple experimental tasks into one hierarchical model (ch. 15)

## 4 Student Evaluation

- reading all the assigned readings and working through all the R scripts on your own outside of class
- 10-minute summaries of the material covered in the previous class (rotates through all registered students)
- a final paper describing an original research project, the preparation and content of which must satisfy the following requirements:
  1. the paper/project will involve:
    - (a) the description of an open issue in theoretical linguistics (pick whatever topic you are interested in on the P side, S side or in between)
    - (b) the description of an experiment and/or corpus study that could make progress with respect to that issue
    - (c) the description of a statistical model (and/or cognitive model, if you want to go to that level) adequate for modeling the experimental/corpus data and extracting the generalizations/results relevant to the semantic issue under consideration
    - (d) simulating at least 2 datasets based on the model (with 2 theoretically relevant sets of parameters for the model), providing graphical summaries of the datasets, running the statistical analyses of the datasets, extracting the generalizations and formulating conclusions that directly bear on the theoretical issue under consideration

2. a 2-page abstract by the end of week 8
3. a fairly detailed handout (6-8 pages) to be presented in week 10
4. final paper (10-15 pages), to be submitted one week after the last class (Dec. 13, 2012)

There will be no incompletes for this class or extensions of any deadlines unless justified by serious unforeseeable events. Please note that lack of adequate planning & preparation for minor perturbations like having a cold does not count as a suitable justification. You will be evaluated based on what you (fail to) accomplish by the specified deadlines.

## References

- Kéry, Marc: 2010, *Introduction to WinBUGS for Ecologists*. Academic Press/Elsevier.
- Kruschke, John K.: 2010, 'Bayesian Data Analysis', *WIREs Cognitive Science* **1**, 658–676.
- Kruschke, John K.: 2011, *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*. Academic Press/Elsevier.
- Lee, Michael D.: 2011, 'How cognitive modeling can benefit from hierarchical Bayesian models', *Journal of Mathematical Psychology* **55**, 1–7.
- Lee, Michael D. and Eric-Jan Wagenmakers: to appear, *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press, Cambridge.