Simon Jackman
Department of Political Science
Stanford University
e-mail: jackman@stanford.edu
http://jackman.stanford.edu

A Practical Introduction to Bayesian Statistical Modeling

IPSA Summer School,
University of São Paulo
São Paulo, Brazil

February 6-10, 2012

This class provides a practical introduction to Bayesian statistical inference, with an emphasis on applications in the social sciences.

We will begin with a brief consideration of how Bayesian statistical inference differs from classical or frequentist inference. We will examine these differences in the context of simple, familiar statistical procedures and models: e.g., inference for proportions, regression, etc.

But the bulk of the class will focus on simulation-based Bayesian inference. Vast increases in desktop computing power now make simulation-based, Bayesian approaches attractive for more complex models. Specifically, the set of algorithms known as Markov chain Monte Carlo (MCMC) allow researchers to tackle classes of problems that used to fall in the "too hard" basket. MCMC is now well and truly part of the statistical computing toolkit available to social scientists, implemented in various forms in many different software packages (we will survey some of these, see below).

We will examine how MCMC algorithms make Bayesian inference feasible, their strengths and weaknesses, and some of the pitfalls to avoid when deploying MCMC algorithms.

The applications to be considered in this part of the course include

- generalized linear models for binary and ordinal data
- multinomial choice models
- models for latent variables, e.g., factor-analytical models, including structural equation model; item-response models;
- classification and clustering, both cross-sectionally and dynamically (i.e., change-point or "structural breaks")
- dynamic latent state models (e.g., tracking public opinion over time),

- hierarchical models of various flavors (appropriate to many forms of data in the social sciences),
- predictive inference

Prerequisites. The class presumes that students have an intermediate knowledge of probability and statistical modeling; i.e., knowing what a likelihood function is and the role it plays in classical (non-Bayesian) inference. We will cover many different applications, so familiarity with generalized linear models or measurement modeling would also be helpful.

Text

With only a tinge of embarrassment I will assign my own text, *Bayesian Analysis for the Social Sciences* (Wiley, 2009), which I refer to below as "BASS."

Slides for much of what I will present are on my Stanford web site (scroll down to "class material"). I will also distribute a lot of the necessary computer code.

Software

I use and recommend R. We will also consider the free, general purpose Bayesian inference program OpenBugs (formerly `WinBugs`). We will also consider JAGS, another free, open-source, general purpose Bayesian analysis program, that runs natively on the Macintosh platform; many of the examples in my book use `JAGS` and the `R` package `rjags`; see also my own `R` package called `pscl`. Both `OpenBugs` and `JAGS` require some programming; if you can handle `R` then you can handle this level of programming.

Outline

1. **Morning Session, Monday February 6**:
   - Overview; Bayesian inference contrasted with frequentist inference. Summarizing probability densities; loss functions; Bayes estimates; highest density regions.
     BASS: Introduction and Ch 1.
   - Bayesian inference for simple problems. Asymptotic properties of Bayes estimates. Cromwell's rule. Conjugacy. Improper, reference priors. Learning about proportions and rates. Bayesian regression analysis under normality.
     BASS: Ch 2.
2. **Afternoon Session, Monday February 6**.
   - The limits of conjugate, "pre-simulation" Bayesian inference; e.g., Bayesian analysis of the difference of two proportions; logistic regression.

- The Monte Carlo Method. Metropolis, Ulam, von Neumann and the Manhattan Project. Simulation consistency. Marginalization/integration via Monte Carlo methods. Random number generation.

  BASS: Ch 3.

3. **Morning Session, Tuesday February 7**.

   - Markov chain Monte Carlo. Metropolis algorithm. The Gibbs sampler. Metropolis-within-Gibbs. Models as directed acyclic graphs (DAGs); deriving conditional distributions.

     BASS: Ch 5.

4. **Afternoon session, Tuesday February 7**.

   - Practical MCMC. Software. Introduction to `JAGS`. Calling `JAGS` from `R` via `rjags`.
   - Assessing convergence and run length; trace plots; autocorrelation function; diagnostics via `CODA`; Raftery-Lewis, Heidelberger-Welch; Geweke; effective sample size.

     BASS: Ch 6.

5. **Morning session, Wednesday February 8**.

   - Tricks of the trade. Thinning. Blocking. Re-parameterization. Over-parameterizations.
   - Examples.

     BASS: Ch 6.

6. **Afternoon session, Wednesday February 8**.

   - Hierarchical and multi-level models. "Random effects" as a hierarchical model. Hyper-parameters.
   - Panel data analysis; clustered data; hierarchical models for discrete data.

     BASS: Ch 7.

7. **Morning session, Thursday February 9**.

   - Discrete data. Binary, ordinal and multinomial data. Multinomial logit; multinomial probit.

     BASS: Ch 8. `MNP` and `MCMCpack` packages in `R`.

8. **Afternoon Session, Thursday February 9**.

   - Latent variables. Factor analysis with normal data; item-response theory models for binary and ordinal data; hybrid models with mixes of indicators.

     BASS: Ch 9.

9. **Morning Session, Friday February 10**.

- Analysis of roll call data; estimating legislative ideal points, cut-points.
  BASS: Ch 9.

10. **Afternoon Session, Friday February 10**.

    - Dynamic latent variable models; state-space representation and the dynamic linear model. Bayesian inference via the FFSB algorithm.
      BASS: Ch 9.