Statistics 220 – Bayesian Data Analysis

Instructor:	Mark Irwin	Office Hours:	Monday 11:00 – 12:00,
Office:	Science Center 611		Wednesday 11:00 – 12:00,
Phone:	617-495-5617		or by appointment
e-mail:	irwin@stat.harvard.edu	Web site:	http://www.courses.fas.harvard.edu/~stat220/

Teaching Fellow: Alan Lenarcic (lenarcic@fas.harvard.edu)

Objectives:

Begins with basic Bayesian models, whose answers often appear similar to classical answers, followed by more complicated hierarchical and mixture models with nonstandard solutions. Includes methods for monitoring adequacy of models and examining sensitivity of conclusions to change in models. Throughout, emphasis will be on drawing inferences via computer simulation rather than mathematical analysis.

Prerequisites:

Statistics 110 and 111.

Lectures:

Tuesday and Thursday, 10:00 – 11:30, Science Center 221

Grading:

Course grades will be determined based on homework assignments (4 or 5 during the term) and a term project, the exact form of which is to be determined.

Computing:

Many of the computational problems can be done in BUGS (Bayesian inference Using Gibbs Sampling). Windows and Linux (Intel based) versions are available. Familiarity with BUGS is not assumed. R/S-Plus will also be useful.

Textbooks and References:

Required Text:

Gelman A, Carlin JB, Stern HS, and Rubin DB (2004). Bayesian Data Analysis, (2nd edition). Chapman & Hall / CRC.

Other References:

Gilks WR, Richardson S, and Spiegelhalter DJ (1996). Markov Chain Monte Carlo in Practice. Chapman & Hall / CRC.

Liu J (2001). Monte Carlo Strategies in Scientific Computing. Springer-Verlag.

Other references and papers will be added to the web site during the course.

Topics to be covered:

1. Basics of the Bayesian inference: (a) setting up a probability model and the basics of statistical inferences; (b) using the probability theory and the Bayes rule; (c) simple examples: binomial, normal, and exponential families; (d) non-informative priors; (e) exchangeability and modeling; (f) asymptotic normality (informal introduction).
2. Multi-parameter models, normal with unknown mean and variance, the multivariate normal distribution, multinomial models, Dirichlet process models, and examples. Computation and simulation from arbitrary posterior distributions in two parameters.
3. Inference from large samples and comparison to standard non-Bayesian methods.
4. Hierarchical models, estimating several means, estimating population parameters from data, meta analysis, and some examples. Emphasis on statistical thinking for complicated problems.
5. Model checking, comparison to data and prior knowledge, sensitivity analysis, prior and posterior predictive checking, posterior odds ratio, model likelihood.
6. Computation 1 (non-iterative methods): integration, conjugate analysis, normal and Laplace approximations, importance sampling, rejection sampling, sequential imputation.
7. Computation 2 (iterative sampling methods): Metropolis-Hastings algorithms and the Gibbs sampler. Their variations include Metropolized independent sampling, iterative predictive updating, and simulated tempering; etc.
8. Missing data problems and data augmentation methodology. Latent variable models; Student-t models for robust inference and sensitivity analysis; Bayesian nonparametrics.