ENGR 206: Bayesian Statistics

Prof. David Draper, Department of Applied Mathematics and Statistics (AMS)  
Baskin School of Engineering

Winter 2004: MWF 3.30–4.40pm in Baskin 156 (plus some extra meetings (to be arranged) to make sure we cover everything that needs to be covered (I have to be out of town on research business during several of our usual MWF slots)).

Contact information: David Draper, Baskin 147, 459–1295, draper@ams.ucsc.edu  
TA: Milovan Krnjajic, Baskin 137, 459–1308, milovan@ams.ucsc.edu

Office hours (for both of us): To be arranged soon

Course web page: www.cse.ucsc.edu/classes/engr206/winter04/ (this includes background information and news about the course, and will include all handouts in PDF format, typically posted no later than one day after the class they were handed out in)

1 Catalog description


2 Prerequisites

Permission of the instructor. Basically I’d like to be able to assume that everybody (a) has had at least some sort of introduction to probability up through and including the use of density functions for continuous random variables (including joint and conditional densities) and (b) is comfortable with basic calculus up through and including multiple integration. All subsequent ideas in the course will be presented in a self-contained fashion.

3 Overview and week-by-week course plan

This course will provide an introduction to Bayesian statistical methods for inference and prediction, at a level suitable for graduate students both from statistics and from disciplines
other than statistics itself (advanced undergraduates are welcome as well, with permission of instructor).

Statistics is the study of uncertainty—how to measure it, and what to do about it. As such, it’s of potential interest in many (virtually all?) aspects of science and decision-making. Of the two main ways to quantify uncertainty—involving relative frequency and subjective (Bayesian) notions of probability—the second way is more flexible and general, but for a long time the Bayesian approach was limited in applications by an inability to perform high-dimensional numerical integrations (the key technical challenge in Bayesian work). With the advent of more powerful computers and new simulation-based techniques over the past 15 years, the computing problem is now essentially solved in a wide variety of interesting situations, and there has been a revolution in Bayesian methods and applications.

The course will be methodological but will be guided by a series of real-world case studies. The first half of the course will involve symbolic mathematical calculations in the computer package Maple, and statistical analyses and graphics in the package R; contemporary Bayesian computation using the MCMC packages BUGS and WinBUGS will feature prominently in the second half.

I will survey the background, in mathematics and probability, of the initial participants in the course to decide how it should be run for maximal benefit of everyone involved. I intend to provide a course that will be interesting and profitable for a variety of students at the graduate level, whether or not they intend to study statistics as their main subject.

The week-by-week breakdown of topics to be covered is as follows.

- **Week 1:** Quantification of uncertainty: classical, frequentist, and Bayesian definitions of probability. Subjectivity and objectivity. Sequential learning; Bayes’ Theorem. Inference (science) and decision-making (policy and business). Bayesian decision theory; coherence. Maximization of expected utility. **Case study:** Diagnostic screening for HIV.

- **Week 2:** Probability as quantification of uncertainty about observables. Binary outcomes. Review of frequentist modeling and maximum-likelihood inference. **Case Study:** Hospital-specific prediction of patient-level mortality rates.

- **Week 3:** Prior, posterior, and predictive distributions. Inference and prediction. Coherence and calibration. Conjugate analysis. Comparison with frequentist modeling. **Software:** Maple.

- **Week 4:** Integer-valued outcomes; Poisson modeling. The exponential family; conjugate priors. **Case Study:** Hospital length of stay for birth of premature babies. **Software:** R.

- **Week 5:** Continuous outcomes; Gaussian modeling. Multivariate unknowns; marginal posterior distributions. **Case Study:** Measurement of physical constants (NB10).

- **Week 6:** Introduction to Markov chain Monte Carlo (MCMC) methods. IID sampling; rejection sampling. Markov chains.
• **Week 7:** The Metropolis-Hastings algorithm. Gibbs sampling. Model expansion.  
  **Case Study:** the NB10 data revisited. **Software:** BUGS and WinBUGS.

• **Week 8:** MCMC implementation strategies. Practical monitoring and convergence diagnostics. Bayesian model uncertainty.

• **Week 9:** Hierarchical models. Poisson fixed-effects modeling. Additive and multiplicative treatment effects. Random-effects Poisson regression. Predictive model diagnostics. Model selection as a decision problem. 3CV: 3-fold cross-validation. Log score and DIC. **Case study:** A controlled experiment to assess the effectiveness of in-home geriatric assessment.

• **Week 10:** Hierarchical modeling for meta-analysis. Formulating hierarchical models for quantitative outcomes from scientific context. Approximate fitting of Gaussian hierarchical models: maximum likelihood and empirical Bayes. Incorporating study-level covariates. Shrinkage estimation. **Case studies:** Meta-analyses of (i) aspirin for heart-attack prevention and (ii) the effects of teacher expectancy on pupil performance.

4 Reading List

I'll be distributing extensive handouts which will essentially form the text for the class. Beyond this, the following will (or may be) helpful.

• Draper D (200x). *Bayesian Hierarchical Modeling*. New York: Springer-Verlag (forthcoming; I'll hand out several chapters from this book).

Supplementary reading will be taken from

• Gilks WR, Richardson S, Spiegelhalter DJ (1996). *Markov Chain Monte Carlo in Practice*. New York: Chapman & Hall (I'll also hand out several chapters from this).

Two years ago I used the following as a supplemental textbook:


Many people thought that the first edition was {unnecessary, not cost-effective, a bit hard to read, ...}, so I'm not formally assigning it this year; I'll try to ensure that several copies are available in the library, and you can try reading from it as well if you want.
5 Evaluation

There will be homework assignments (more like small take-home tests) given out in weeks 1, 3, 5, and 7 and due one week later; these will blend paper-and-pen, symbolic computing, statistical and MCMC calculations. A final project will be assigned in week 9 and due at the end of the final examination period.

If you have ideas for what you’d like to do your project on, please talk with me about it well in advance of the ninth week of class. I’ll provide a project for everyone who does not select one for her/himself (in practice the default project is typically selected by about 90% of the class).