Homework 5
Due Wednesday, October 2, 2013

Think of each question as a "mini-project." This means that everything must be typed, labeled, and referenced, as appropriate. Your answers for each question should discuss the problem, data, model, method, conclusions. If you develop an interesting computational way to solve a problem, feel free to include pseudo (or actual) code, but, in general, code is not required.

Problem 1. *STAT 101 test scores revisited.* Consider the data in in test_scores.dat. These data are 29 test scores from the first exam in STAT 101 (110 points possible). Use a truncated normal distribution for the sampling distribution, where $Y_i \sim \text{Normal}(\mu, \sigma^2)$ with $Y_i \in (a, b)$, $-\infty < a < b < \infty$. Use a prior distribution where μ and σ^2 are assumed independent, $\mu \sim \text{Normal}(\theta, \tau)$ for $\mu \in (a_0, b_0)$ and $\sigma^2 \sim \text{InverseGamma}(\gamma, \delta)$ for $\sigma^2 \in (a_1, b_1)$. You will need to specify values for θ, τ, γ, and δ and the constraints.

Plot the prior distribution and the posterior distribution. For each parameter (μ and σ^2), plot its marginal posterior distribution and calculate appropriate summary statistics.

Be careful. Constraints on the sampling distribution are not the same as constraints on the prior distribution, and normalizing constants matter in this problem.

The truncated normal density has the form

$$f(y \mid \mu, \sigma^2, a, b) = \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{1}{2\sigma^2}(y-\mu)^2)}{\Phi(b, \mu, \sigma^2) - \Phi(a, \mu, \sigma^2)},$$

where $\Phi(x, \mu, \sigma^2)$ is the normal cumulative distribution function evaluated at x.

The expected value and variance of a truncated normal distribution can be written in terms of μ, σ^2, a, and b. In particular, we have

$$E[Y] = \mu + \sigma \frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}$$

and

$$Var[Y] = \sigma^2 \left[1 + \frac{\frac{a-\mu}{\sigma}\phi(\frac{a-\mu}{\sigma}) - \frac{b-\mu}{\sigma}\phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} - \left(\frac{\phi(\frac{a-\mu}{\sigma}) - \phi(\frac{b-\mu}{\sigma})}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})} \right)^2 \right]$$

where $\phi(\cdot)$ is the standard normal probability density function.

Problem 2. *The Behrens-Fisher problem.* Suppose that we observe two independent normal samples, the first distributed according to a Normal(μ_1, σ_1^2) distribution and the second according to a Normal(μ_2, σ_2^2) distribution. Denote the first sample by $x_1, \ldots, x_m$ and the second sample by $y_1, \ldots, y_n$. The prior distribution for $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ is

$$g(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2) \propto \frac{1}{\sigma_1^2 \sigma_2^2} \quad .$$

- Find the posterior density for $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. Show that the vectors (μ_1, σ_1^2) and (μ_2, σ_2^2) have independent posterior distributions.
- Describe how to simulate from the joint posterior distribution of $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$. (Algorithm, code, and pseudo-code are all fine.)
- The following data are mandible lengths (in millimeters) for 10 male and 10 female golden jackals in the collection of the British Museum. Starting with the model discussed in the previous two bullets, find the posterior density of the different in mean mandible length between the sexes. (Be sure to describe what you did.) Is there sufficient evidence to conclude that the males have a larger average?

Males: 120, 107, 110, 116, 114, 111, 113, 117, 114, 112

Females: 110, 111, 107, 108, 110, 105, 107, 106, 111, 111

Problem 3. Suppose that we are interested in learning about the probability that a coin lands heads when tossed. We model the data we collect as Binomial(n, p). However, instead of working with p, we decide to work with $\theta = \text{logit}(p) = \log(\frac{p}{1-p})$.

A priori, we believe that the coin is fair, so we assign a Normal$(0, 0.25)$ prior distribution to θ. We toss the coin five times and observe five heads.

- Determine the posterior distribution for θ (up to a normalizing constant).

- Using a normal approximation to the posterior density, compute the probability that the coin is biased towards heads (i.e., that θ is positive). Provide sample code and plots of the actual posterior density and the normal approximation.

- Using the prior density as a proposal density, design a rejection algorithm for sampling from the posterior distribution. Approximate the posterior probability that the coin is biased towards heads. Provide sample code and plots of the actual posterior density and the posterior density as determined using the rejection algorithm.

- Using the prior density as a proposal density, simulate values from the posterior distribution using the SIR algorithm. Approximate the posterior probability that the coin is biased towards heads. Provide sample code and plots of the actual posterior density and the posterior density as determined using the SIR algorithm.

Problem 4. *Estimation for the two-parameter exponential distribution.* Martz and Waller (1982) describe the analysis of a "type I/time-truncated" life testing experiment. Fifteen reciprocating pumps were tested for a prespecified time; any failures among the pumps were replaced. One assumes that the failure times follow the two-parameter exponential distribution

$$f(y \,|\, \beta, \mu) = \frac{1}{\beta} \exp(-\frac{(y-\mu)}{\beta}) \;, y \geq \mu \;.$$

Suppose one places a uniform prior on (μ, β). Then Martz and Waller show that the posterior density is given by

$$\pi(\beta, \mu \,|\, \mathbf{y}) \propto \frac{1}{\beta^s} \exp\{-(t - n\mu)/\beta\} \;, \mu \leq t_1 \;,$$

where n is the number of items placed on test, t is the total time on test, t_1 is the smallest failure time, and s is the observed number of failures in a sample of size n. In the example, data were reported in cycles to failure; $n = 15$ pumps were tested for a total time of $t = 15{,}962{,}289$. Eight failures ($s = 8$) were observed and the smallest failure time was $t_1 = 237{,}217$.

- Suppose one transforms the parameters to the real line by the transformations $\theta_1 = \log(\beta)$, $\theta_2 = \log(t_1 - \mu)$. Write down the posterior density of (θ_1, θ_2).

- Construct a normal approximation for the posterior distribution of (θ_1, θ_2).

- Use a multivariate t proposal density and the SIR algorithm to simulate draws from the posterior distribution.

- Suppose one is interested in estimating the reliability at time t_0 defined by
$$R(t_0) = \exp(-(t_0 - \mu)/\beta) \ .$$
Discuss the distribution of $R(t_0)$ when $t_0 = 10^6$ cycles.