Stat 111 Spring 2010 Week 14 - Bayesian Computations

1. **Approximate Methods**
   Section 5.2 of Carlin and Lewis describes the Normal approximations to a posterior density, and the Laplace method for approximating an integral. Section 4.6 of Rice describes approximations for the mean and variance of a transformation of a random variable.

   a) Find Taylor series approximations to the mean and variance of $Y = \log(X)$, where $X \sim \text{Gamma}(\alpha, 1)$. Also show that the exact mean and variance for Y are $\psi(\alpha)$ and $\psi'(\alpha)$ respectively, where $\psi(\alpha) = \frac{d}{d\alpha} \log \Gamma(\alpha)$ and $\psi'(\alpha) = \frac{d^2}{d\alpha^2} \log \Gamma(\alpha)$ (i.e, ψ and ψ' are the "digamma" and "trigamma" functions, and are implemented in R). Compare the approximation to the true value for various α.

   c) Use a Normal approximation to approximate the posterior density for α, based on a sample $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Gamma}(\alpha, 1)$. Simulate data with $n = 10$ and $\alpha = 5$.

   d) Use the Laplace method to derive Stirling's approximation to $x!$ and to $\log(\Gamma(x))$. Assess this approximation by graphing $\log(\Gamma(x))$ and Stirling's approximation against a grid of x values.

2. **Rejection and Importance Sampling**
   We have looked at some relatively simple Bayesian problems, for which we were able to simulate from the posterior density using standard distributions (or constrained versions of standard distributions). With more complexity, things don't often work out so nice, and we require more computationally intensive methods for simulation. Rejection sampling is one way to generate independent draws from a density f. Candidate draws are made from an appropriate "envelope density" f_o. Then the draw is accepted or rejected using a randomized decision designed to correct for the differences between f and f_o.

   a) Suppose the densities f_o and f are such that $f(x)/f_o(x) \le M$ for some constant $M < \infty$. Draw $X \sim f_o$ and $Y \,|\, x \sim \text{Uniform}(0, M f_o(x))$. Show that this generates points uniformly under the curve $g(x) = M f_o(x)$.

   b) Show that, if you throw away any points where $y > f(x)$ (i.e., reject unless $y < f(x)$) then the resulting x values follow the density function f. That is, the conditional probability density for X, given that $Y < f(X)$, is $f(x)$. See Example D in Section 3.5 of Rice.

   c) For example, we could use the inverse CDF method to simulate values from the tail of a Normal distribution. But the Normal CDF approximation breaks down as you get far in the tails. Use an Exponential distribution as an envelope to generate exact random draws $Z \sim N(0, 1)$, constrained to to have $Z \ge 3.0$.

   d) We were able to use direct simulation techniques to simulate posterior draws for the Normal hierarchical model with equal level-1 sample sizes n for each of the k groups. The problem becomes much more complicated when we allow unequal sample sizes n_i, even if σ is assumed known. There is no longer a single shrinkage factor B; we now have $B_i = V_i/(V_i + A)$, where $V_i = \sigma^2/n_i$ for group i. If we specify a value n_o (e.g., the average n_i) then we can define $B_o = V_o/(V_o + A)$, where $V_o = \sigma^2/n_o$. For a certain non-informative prior specification, the posterior density for B_o is as follows:

$$f_{B_o|y}(B_o \,|\, y, S_1) \;\propto\; |W|^{1/2} |X'WX|^{-1/2} \, \frac{B_o^{(k-q-2)/2-1}}{(S_1 + B_o S_{2(B_o)})^{(m+k-q-2)/2}}, \qquad 0 < B_o \leq 1.$$

This looks very similar to the constrained F density we had for the equal-n case, but there are additional terms that vary with B_o. The matrix W is diagonal with entries that depend on B_o. S_1 is the SSE and $S_{2(B_o)}$ is the SSM, based on a regression estimate that also depends on B_o. This is a messy density, but it is finite (as long as $k - q > 4$) and bounded between 0 and 1. So we could use a Uniform envelope density f_o. An "adaptive rejection sampling" algorithm improves the envelope after each candidate draw. Demonstrate this for simulating from the posterior distribution for B_o using the NBA data (which in fact has unequal n_i's).

3. **Gibbs Sampling (MCMC)**
Gibbs sampling is a method for simulating from a joint posterior density, and is a special type of Markov Chain Monte Carlo (MCMC). MCMC work by constructing a Markov Chain with the desired joint posterior density as the equilibrium distribution. Gibbs sampling alternately draws from conditional distributions for each parameter, given current values of all other parameters and the observed data. These full conditional distributions are often much simpler than the marginal posterior distributions. For example, presentation 2 showed the non-standard marginal posterior density for B_o in the Normal hierarchical model (NHM) with unequal n_i's. It is even more complicated if we allow the σ's for each group to also vary (and treat them as unknown). But it is not hard to specify a Gibbs sampling algorithm for this problem because the conditionals are all standard distributions. Gibbs sampling is currently the method used by most people to make inference for the NHM (not for long, if I can help it, because direct simulation is possible and preferable).

a) Example C in Rice Section 3.3 (Copulas) shows that the marginal distributions do not determine the joint distribution, so it may not be intuitive that a joint posterior distribution is completely specified by all of the conditional distributions. Suppose you know $f_{\lambda|\theta,y}(\lambda \,|\, \theta, y)$ and $f_{\theta|y,\lambda}(\theta \,|\, y, \lambda)$. Show how to identify the joint posterior density $f_{\lambda,\theta}(\lambda, \theta \,|\, y)$ and the marginal densities $f_{\lambda|y}(\lambda \,|\, y)$ and $f_{\theta|y,\lambda}(\theta \,|\, y, \lambda)$. As an example, suppose $\lambda \,|\, \theta, y \sim$ Gamma(α, θ) and $\theta \,|\, \lambda, y \sim$ Gamma$(m + \alpha,\, y + \lambda)$, where α and m are known.

b) Show that if the Gibbs sampling algorithm begins with a draw from the actual posterior distribution of the parameters, then the successive draws will also be from the exact posterior distribution. This establishes that the posterior distribution is an equilibrium state of the Markov Chain. Use the Normal hierarchical model with known σ and $V_i = \sigma^2/n_i$ as an example. The Gibbs algorithm cycles through making a draw from $\mu \,|\, y, \theta, A$ (or $\beta \,|\, y, \theta, A$), from $A \,|\, y, \theta, \mu$, and from $\theta \,|\, y, \mu, A$. Repeating these steps defines a first-order Markov Chain (values at iteration $t + 1$ depend only on the values at iteration t) with $f_{\theta,\mu,A\,|\,y}$ as its equilibrium distribution.

c) Demonstrate the Gibbs sampling algorithm for the Normal hierarchical model with known σ ($V_i = \sigma^2/n_i$). Use the NBA data and the prior density $p_{\mu,A}(\mu, A) \propto c$.

Problems to turn in:

1. We saw in week 13 that, if R is the 1-way ANOVA F statistic, then the sampling distribution for BR is $F(k-1, N-k)$, where $B = V/(V+A)$. Note that R itself follows this F distribution when $B = 1$, meaning the variance A of the level-1 means is 0. Consider the WNBA data with $k = 13$ teams and $n = 5$ games per team (so $N = nk = 5(13) = 65$. What values of R would lead you to reject the null hypothesis of equal means ($B = 1$)? Find the power of this test when $B = 0.5$ (meaning $A = V = \sigma^2/n$). Use qf and pf in R to compute the cutoff value and the power.

2. Suppose you wish to simulate values from the joint distribution $f_{xy}(x, y) = 2$, $0 < y < x < 1$. That is, you want to sample uniformly in the triangular region $0 < y < x < 1$.

   a) One way to get a point in this region is to draw $X \sim U(0,1)$, and then draw $Y \mid X = x \sim U(0, x)$. Show that the resulting joint density is not Uniform.

   b) To use Gibbs sampling in this problem, we would start with some value of x, then draw Y from its conditional distribution given this x, then draw a new x value from the conditional distribution given the value y. Iterating between these conditional distributions results in a Markov chain that has f_{xy} as its equilibrium distribution. Work out the conditional distributions for $X|y$ and for $Y|x$, and outline the steps of the algorithm.

   c) To sample directly from this distribution, we could simulate X from its marginal distribution, and then draw Y from its conditional distribution, given $X = x$. Find the marginal density for X and explain how you could generate a random value from this distribution.

   d) There is an easier direct simulation that requires only that you simulate U_1 and U_2 to be independent $U(0,1)$ variables. Explain this simple approach and verify that it works.