## BAYESIAN METHODS: FINAL EXAM, MARCH 5 2002

This homework assignment is to be considered a **take-home** test. Thus, you **may not** collaborate with any other person (whether in the class or not), **nor** may you consult with *anyone*. You **may** use any reading material (class notes, books, etc) you wish. This assignment is due **March 19 at 5:00 pm**

1. Suppose we are given the likelihood $p(y \mid \theta)$ for the univariate parameter $\theta$, and adopt the two-component mixture prior

$$p(\theta) = \alpha p_1(\theta) + (1-\alpha)p_2(\theta)$$

   where $p_1(\theta)$ and $p_2(\theta)$ belong to the family of priors conjugate for $p(y \mid \theta)$.

   (a) find $p(\theta \mid y)$

   (b) Does a mixture of conjugate priors leads a mixture of conjugate posteriors?

2. Show that the probability of accepting a given candidate $\theta$ in the rejection sampling is $c/M$, where $c = \int p(y \mid \theta)p(\theta)d\theta$ is the normalizing constant for the posterior distribution $p(\theta \mid y)$ and $M$ is such that $L(\theta)p(\theta) < Mg(\theta)$.

3. Identifiability and MCMC convergence: consider the two-parameter likelihood model

$$Y \sim N(\theta_1 + \theta_2, 1)$$

   with prior distributions $\theta_1 \sim N(a_1, b_1^2)$ and $\theta_2 \sim N(a_2, b_1^2)$, $\theta_1$ and $\theta_2$ independent.

   (a) Are $\theta_1$ and $\theta_2$ identified by the likelihood? Find the full conditional distributions $p(\theta_1 \mid \theta_2, y)$ and $p(\theta_2 \mid \theta_1, y)$ and define the Gibbs sampler for this problem,

   (b) In this simple problem we can also obtain the marginal posterior distributions $p(\theta_1 \mid y)$ and $p(\theta_2 \mid y)$ in closed form. Find these two distributions. Do the data update the prior distributions for these parameters?

   (c) Set $a_1 = a_2 = 50$, $b_1 = b_2 = 1000$, and suppose we observe $y = 0$. Run the Gibbs sampler defined in part (a) for $t = 100$ iterations, starting each of your sampling chains near the prior mean (say between 40 and 60), and monitoring the progress of $\theta_1$, $\theta_2$ and $\mu = \theta_1 + \theta_2$. Does this algorithm converge? Estimate the posterior mean of $\mu$. Does your answer change using $t = 1000$ iterations?

   (d) Now keep the same values for $a_1$ and $a_2$, but set $b_1 = b_2 = 10$. Again run 100 iterations using the same starting values as in part (b). What is the effect on convergence? Again repeat your analysis using 1000 iterations; is your estimate of $E[\mu \mid y)$ unchanged?

   (e) Summarize your findings, and make recommendations for running and monitoring convergence of samplers running on "partially unidentified" and "nearly partially unidentified" parameter space.

4. Bayesian Logistic Regression: download the flour bettle data set from

   `http://biosun01.biostat.jhsph.edu/~fdominic/teaching/BM/beetles.dat`

   and let $y_i$ and $n_i$ the number of killed and exposed bettles at the dosage $w_i$, respectively. Under the following logistic regression model:

$$\begin{array}{lcl} y_i & \sim & \text{Binomial}(\pi_i, n_i) \\ \text{logit}(\pi_i) & = & \beta_0 + \beta_1 w_i \end{array} \qquad (1)$$

(a) Find the maximum likelihood estimates of $\beta_0$ and $\beta_1$;

(b) Approximate the posterior distributions of $\beta_0$ and $\beta_1$ by using the normal approximation of the likelihood function and a flat prior for $\beta_0$ and $\beta_1$;

(c) Approximate the posterior distributions of $\beta_0$ and $\beta_1$ by using simulation based methods and a flat prior for $\beta_0$ and $\beta_1$;

(d) Compare inferences between (b) and (c);

(e) Approximate the marginal posterior distribution of a future outcome $\tilde{y}$ for $\tilde{w} = 1.9$ and $\tilde{n} = 60$, and calculate the posterior predictive probabilities that $\tilde{y}$ is larger than $30,\ 50$ and $70$.