Hw4 for 6448 (due April 20) (Debdeep Pati)

1. Consider the linear regression model $Y = X\beta + \epsilon$ where $X \in \mathbb{R}^{n\times p}$, $\beta \in \mathbb{R}^p$, $\epsilon \sim \mathrm{N}(0, I_n)$. Consider a spike and slab prior on β

$$\pi(\beta) = \prod_{j=1}^{p} \{\delta_0(\beta_j)p_{0j} + (1 - p_{0j})N(\beta_j; 0, c_j^2)\}$$

   where p_{0j} is the prior probability of excluding the j-th predictor by setting its coefficient to 0. Show that the conditional posterior of β_j , for $j = 1, \ldots, p$, is given by

$$\pi(\beta_j \mid \beta_{-j}, Y, X) = \hat{p}_j\delta_0(\beta_j) + (1 - \hat{p}_j)\mathrm{N}(\beta_j; E_j, V_j)$$

   where $V_j = (c_j^{-2} + X_j'X_j)^{-1}, E_j = V_jX_j'(Y - X_{-j}\beta_{-j}), X_j = j$ th column of X, $X_{-j} = X$ with j th column excluded, $\beta_{-j} = \beta$ with j th element excluded, and

$$\hat{p}_j = \frac{p_{0j}}{p_{0j} + (1 - p_{0j})\frac{N0;0,c_j^2)}{N(0;E_j,V_j)}}$$

   is the conditional probability of $\beta_j = 0$. Here $N(x; \mu, \sigma^2)$ denotes the normal density with mean μ, variance σ^2 evaluated at x.

2. Generate observations from the previous model with $n = 100$, $p = 200$, and elements of X i.i.d from Unif$(0, 1)$, first 5 coefficients of β are set to 10 and the remaining entries are all 0. Run the MCMC for 5,000 iterations using the spike and slab prior with $c_j = 4$ and $p_{0j} = 0.975$. Compare the results with Horseshoe (code provided on the course webpage). Perform variable selection in both the cases using HPPM, marginal and median probability models. Compare the results.