Homework 1 (30%, due 04/07/2017)

Solve the following problems in R under set.seed(1000). Include sufficient amount of details. Attache your R codes.

(1) The Land of Oz is blessed by many things, but not by good weather. They never have two nice (N) days in a row. If they have a nice day, they are just as likely to have snow (S) as rain (R) the next day. If they have snow or rain, they have an even chance of having the same the next day. If there is change from snow or rain, only half of the time is this a change to a nice day. Parts (a)-(e) each deserves 2 points, (f) deserves 5 points.

(a) Form a Markov chain $z_1, z_2, \ldots, z_n$ for the weather of n days in a row, with states $\mathcal{S} \equiv \{N, S, R\}$. Find the transition probabilities $p_{jk} \equiv P(z_i = k | z_{i-1} = j)$ for all $j, k \in \mathcal{S}$.

(b) Generate a Markov chain $z_1, \ldots, z_n$ with $n = 100$ and $z_1 = N$.

(c) Generate $\mu_N, \mu_S, \mu_R \overset{iid}{\sim} N(0,1)$ and $\sigma_N^2, \sigma_S^2, \sigma_R^2 \overset{iid}{\sim} IG(1/2, 1/2)$, and simulate $X_i | z_i \overset{ind}{\sim} N(\mu_{z_i}, \sigma_{z_i}^2)$ for $i = 1, \ldots, n$ with $n = 100$. Here IG means Inverse Gamma distribution.

(d) Now pretend that you don't observe $z_2, \ldots, z_n, \mu_j$'s, σ_j^2's. Propose a Bayes model with parameters $z_2, \ldots, z_n, \mu_j, \sigma_j^2$ for $j = N, S, R$. Write a Gibbs sampler to generate MCMC samples of length $T = 20,000$.

(e) Estimate $z_2, \ldots, z_n$ and compare them with the truth. What is the proportion of errors?

(f) Suppose that p_{jk}'s are unknown. Propose a new Bayes model with parameters $z_2, \ldots, z_n, \mu_j, \sigma_j^2, p_{jk}$ for $j, k = N, S, R$. Write a Gibbs sampler to generate MCMC samples of length $T = 20,000$. Estimate $z_2, \ldots, z_n$ and compare them with the truth. What is the proportion of errors? *Hint*: consider prior $(p_{jN}, p_{jS}, p_{jR}) \sim Dirichlet(2, 2, 2)$, for every j.

(2) (7 points) Consider the Bayesian variable selection problem introduced in lecture, i.e., the following hierarchical model

$$Y_i | X_i, \beta, \sigma^2 \overset{ind}{\sim} N(\beta_1 X_{i1} + \cdots + \beta_p X_{ip}, \sigma^2),$$

with prior distributions

$$\begin{aligned} \beta_j | \gamma_j &\overset{ind}{\sim} (1 - \gamma_j) N(0, \tau_j^2) + \gamma_j N(0, c_j^2 \tau_j^2) \\ \gamma_1, \ldots, \gamma_p &\overset{ind}{\sim} \text{Bernoulli}(1/2), \end{aligned}$$

where $\tau_j^2, c_j^2, p, \sigma^2$ are all known. In this problem try $\tau_j^2 = 1$, $c_j^2 = 9$ for all j, and $\sigma^2 = 1$.

Let $n = 100$, $p = 10$. Generate samples as follows

$$\begin{aligned} X_{11}, \ldots, X_{np} &\overset{iid}{\sim} N(0, 3^2), \\ Y_i &= 3X_{i1} - 2X_{i2} - 0.5X_{i3} + 0.1X_{i4} + N(0,1), i = 1, \ldots, n. \end{aligned}$$

Write a Gibbs sampler and find the proportion of incorrectly selected variables. If your selection result is not good, what should you do? *Hint*: try other values of c_j^2, say $c_j^2 = 4, 16$?

(3) (8 points) Redo Problem (2) with σ^2, τ_j^2 being treated as unknown parameters. Modify the hierarchical model by incorporating the following prior

$$\sigma^2, \tau_1^2, \ldots, \tau_p^2 \overset{iid}{\sim} IG(1/2, 1/2).$$