PHS 43010: Homework #4 MCMC

1. Two genes are believed to co-express in human bodies. Measurements of gene expression for both genes are standardized to follow a $N(0,1)$ distribution. We want to investigate the correlation ρ between the two genes and obtained the measurements $(X_1, Y_1), \ldots, (X_n, Y_n)$ of both genes in $n = 13$ individuals. Here X_i and Y_i denote the expression of gene one and two, respectively. The data are

x_i	0.92	0.42	3.62	0.89	-0.69	0.45	-0.11	-0.14	-0.47	1.09	-0.34	0.62	0.27
y_i	0.26	1.65	2.10	0.62	-1.16	1.29	-0.82	-0.36	-0.29	0.86	0.19	1.25	0.33

(a) Derive the likelihood function assuming (X_i, Y_i) follows a bivariate normal distribution with marginal distribution $N(0,1)$ and corelation ρ.

(b) Suppose that ρ follows a $U(0,1)$ prior. Consider the a Metropolis-Hastings sampling algorithm for ρ. Suppose the proposal density of ρ is

(i) Given the current value $\rho^{(s)}$, sample $\rho^* \sim U(\rho^{(s)} - 0.2, \rho^{(s)} + 0.2)$.

(ii) If the sampled $\rho^* < 0$, then set $\rho^* = |\rho^*|$.

(iii) If the sampled $\rho^* > 1$, then set $\rho^* = 2 - \rho^*$.

Argue that this is a symmetric proposal density.

(c) Sample from the posterior distribution of ρ using M-H algorithm with the above proposal density. Make trace plots and check autocorrelation. See if thinning will reduce autocorrelation. Summarize the posterior distribution of ρ using the MCMC samples.

2. The Gibbs sampler can take a long time to converge if the target distribution is multi- modal. Suppose we are trying to sample from a posterior density that is the mixture of three normal densities: θ has density $0.45N(3, 1/3) + 0.10N(0, 1/3) + 0.45N(-3, 1/3)$, where $N(\mu, \sigma^2)$ represents the normal density with mean μ and variance σ^2. Consider using Gibbs sampling to sample from it.

(a) Plot the marginal density of θ. If the indicator $\delta \in \{1, 2, 3\}$, argue that the full conditional distribution for θ is $\theta \mid \delta \sim N(\mu_\delta, \sigma^2_\delta)$. What are μ_δ and σ^2_δ here, for $\delta = 1, 2, 3$?

(b) Use Bayes' theorem to show that the full conditional for δ is

$$Pr(\delta = k \mid \theta) = \frac{Pr[\delta = k] \times N(\mu_k, \sigma^2_k)}{\sum_{j=1}^{3} Pr[\delta = j] \times N(\mu_j, \sigma^2_j)},$$

for $k \in \{1, 2, 3\}$.

(c) Write a Gibbs sampling algorithm to sample from the joint density of (θ, δ). Begin the chain with the initial values $\delta^0 = 2$ and $\theta^0 = 0$, and generate 1000 values of θ. Give a plot of a relative frequency histogram of the θ values – using a command like hist(theta.values, freq=F) in R and comment on how it compares to the true marginal density of θ plotted in (a).

(d) Repeat part (c), but generate 20000 values of θ. Again give a plot of a relative frequency histogram of the θ values and comment on how it compares to the true marginal density of θ plotted in (a).

3 (Univariate MH). Suppose that $D = 5$ doses of a medication are being tested for toxicity. When a patient is put on a dose he/she might expreience toxicity event. Suppose n_i patients have been treated using dose i and x_i of them experience toxicity. The data are

Dose d_i	1	2	3	4	5
n_i	3	3	3	6	3
x_i	0	0	0	3	2

Assume $x_i \mid n_i \sim Bin(n_i, p_i)$ and apply the logistic regression

$$p_i = \frac{\exp(\beta_0 + \beta_1 d_i)}{1 + \exp(\beta_0 + \beta_1 d_i)}.$$

Suppose $\beta_0 = -3$ and let β_1 follow a prior $N(0, 2)$ distribution. Use Metropolis-Hastings to sample the posterior $p(\beta_1 \mid \{n_i, x_i\})$ where $\{n_i, x_i\}$ represent the data on all five doses. Plot the histogram of β_1 posterior samples, and obtain the posterior mean and variance of β_1. What is the value of

$$\hat{p}_4 = \frac{\exp(\beta_0 + \hat{\beta}_1 d_4)}{1 + \exp(\beta_0 + \hat{\beta}_1 d_4)}$$

where $\hat{\beta}_1$ is the posterior mean?