AMS 206B: Intermediate Bayesian Inference
HOMEWORK #4 (MCMC)

1. Consider a model of the form $x|\theta \sim Bin(n, \theta)$ and $\theta \sim Be(1/2, 1/2)$. Assume that you observe $n = 10$ and $x = 1$.

   (a) Report an exact 95% (symmetric) posterior credible interval for θ (for example, you can use the `qbeta` function in `R`).

   (b) Report an approximate credible interval for θ using the Laplace approximation.

   (c) Report an approximate credible interval for θ using Monte Carlo simulation.

   (d) Repeat the previous calculations with $n = 100, x = 10$ and with $n = 1000, x = 100$. Comment on the difference between all 9 situations.

2. Consider a model where $x_1, \ldots, x_n$ is an i.i.d. sample where $x_i|\theta \sim N(\theta, 1)$ and $\theta \sim Cauchy(0, 1)$.

   (a) Develop importance sampling schemes to compute $E(\theta|x_1, \ldots, x_n)$ using the prior (i.e., the Cauchy(0, 1) distribution) as your importance function. Apply this algorithm to the datasets `hw3.1.dat` and `hw3.2.dat`, and construct a histogram of the importance weights associated to each datasets.

   (b) Repeat the same exercise using a $N(\bar{x}, 1/n)$ distribution as your importance distribution. How could you justify this choice of importance function?

   (c) Repeat the same exercise using a Student-t distribution with mean $\bar{x}$ and variance $1/n$ as your importance distribution. How could you justify this choice of importance function?

   (d) Compare all plots and comment.

3. (Wasserman, 2003) A random variable Z has an *inverse Gaussian distribution* if it has density

$$f(x) \propto z^{-3/2} \exp\left\{-\theta_1 z - \frac{\theta_2}{z} + 2\sqrt{\theta_1\theta_2} + \log(\sqrt{2\theta_2})\right\}, \quad z > 0,$$

   where $\theta_1 > 0$ and $\theta_2 > 0$ are parameters. It can be shown that $E(Z) = \sqrt{\theta_2/\theta_1}$ and $E(1/Z) = \sqrt{\theta_1/\theta_2} + 1/(2\theta_2)$.

   (a) Let $\theta_1 = 1.5$ and $\theta_2 = 2$. Draw a sample of size 1,000 using the independence-Metropolis-Hastings method with a Gamma distribution as the proposal density (note that in an independence-Metropolis-Hastings $q(\theta^*|\theta) = q(\theta^*)$). To assess the accuracy of the method, compare the mean of Z and $1/Z$ from the sample to the theoretical means. Try different Gamma distributions to see if you can get an accurate sample.

   (b) Draw a sample of size 1,000 using the random-walk Metropolis method. Since $z > 0$ we cannot just use a Normal density. Let $W = \log(Z)$. Find the density of W. Use the random-walk Metropolis method to get a sample $W_1, \ldots, W_M$ and let $Z_i = e^{W_i}$. Assess the accuracy of the simulation as in the previous part.

4. Consider i.i.d. data $x_1, \dots, x_n$ such that $x_i|\nu, \theta \sim Gamma(\nu, \theta)$ where $E(x_i) = \nu/\theta$, and assign priors $\nu \sim Gamma(3, 1)$ and $\theta \sim Gamma(2, 2)$.

   (a) Develop a Metropolis-within-Gibbs algorithm to sample from $p(\nu, \theta|x_1, \dots, x_n)$ using the full conditional distributions $p(\theta|\nu, x_1, \dots, x_n)$ and $p(\nu|\theta, x_1, \dots, x_n)$. For the second full conditional, use a random walk proposal on $\log(\nu)$.

   (b) Develop a Metropolis-Hastings algorithm that jointly proposes $\log(\nu)$ and $\log(\theta)$ using a Gaussian random walk centered on the current value of the parameters. Tune the variance-covariance matrix of the proposal using a test run that proposes the parameters independently (but evaluates acceptance jointly).

   (c) Develop a Metropolis algorithm that jointly proposes $\log(\nu)$ and $\log(\theta)$ using independent proposals based on the Laplace approximation of the posterior distribution of $\log(\nu)$ and $\log(\theta)$.

   (d) Develop a Metropolis algorithm that jointly proposes $\log(\nu)$ and $\log(\theta)$ using independent proposals based on a modified version of the Laplace approximation of the posterior distribution of $\log(\nu)$ and $\log(\theta)$ in which the normal distribution is replaced by a heavy tailed distribution (such as a multivariate Cauchy).

   (e) Run each of the algorithms for the dataset in `hw4.1.dat` and compute the effective sample sizes associated with each parameter under each of the samplers. Also, construct trace and autocorrelation plots. Report posterior means for each of the parameters of interest, along with 95% symmetric credible intervals. Discuss.

5. (Robert and Casella) Consider a random effects model,

$$y_{i,j} = \beta + u_i + \epsilon_{i,j}, \quad i = 1 : I, j = 1 : J,$$

   where $u_i \sim N(0, \sigma^2)$ and $\epsilon_{i,j} \sim N(0, \tau^2)$. Assume a prior of the form

$$\pi(\beta, \sigma^2, \tau^2) \propto \frac{1}{\sigma^2 \tau^2}.$$

   (a) Find the full conditional distributions: (i) $\pi(u_i|y, \beta, \tau^2, \sigma^2)$; (ii) $\pi(\beta|y, u, \tau^2, \sigma^2)$; $\pi(\sigma^2|y, u, \beta, \tau^2)$; (iii) $\pi(\tau^2|y, u, \beta, \sigma^2)$.

   (b) Find $\pi(\beta, \tau^2, \sigma^2|y)$ up to a proportionality constant.

   (c) Find $\pi(\sigma^2, \tau^2|y)$ up to a proportionality constant and show that this posterior is not integrable since, for $\tau \neq 0$, it behaves like σ^{-2} in a neighborhood of 0.

   This problem shows that even though the full conditional posteriors exist and the Gibbs sampling could be easily implemented, the joint posterior distribution does not exist. Users should be aware of the risks of using the Gibbs sampler in situations like this!

6. (Carlin, Gelfand and Smith, 1992) Let $y_1, \dots, y_n$ be a sample from a Poisson distribution for which there is a suspicion of a change point m along the observation process where the means change, $m = 1, \dots, n$. Given m $y_i \sim Poisson(\theta)$, for $i = 1, \dots, m$ and $y_i \sim Poisson(\phi)$, for $i = m + 1, \dots, n$. The model is completed with independent

prior distributions $\lambda \sim Gamma(\alpha, \beta)$, $\phi \sim Gamma(\gamma, \delta)$ and m uniformly distributed over $\{1, \ldots, n\}$ where α, β, γ and δ are known constants. Implement a Gibbs sampling algorithm to obtain samples from the joint posterior distribution. Run the Gibbs sampler to apply this model to the data `mining.r` which consists of counts of coal mining disasters in Great Britain by year from 1851 to 1962.

7. Souza (1999) considers a number of hierarchical models to describe the nutritional pattern of pregnant women. One of the models adopted was a hierarchical regression model where

$$\begin{aligned} y_{i,j} &\sim N(\alpha_i + \beta_i t_{i,j}, \sigma^2), \\ (\alpha_i, \beta_i)' | \alpha, \beta &\sim N((\alpha, \beta)', diag(\tau_\alpha^{-1}, \tau_\beta^{-1})), \\ (\alpha, \beta)' &\sim N((0,0)', diag(P_\alpha^{-1}, P_\beta^{-1}), \end{aligned}$$

prior independent scale parameters σ^{-2}, τ_α and $\tau_\beta \sim Gamma(a, b)$, and $y_{i,j}$ and $t_{i,j}$ are the jth weight measurement and visit time of the ith woman with $j = 1 : n_i$ and $i = 1 : I$ for $I = 68$ pregnant women. Here $n = \sum_{i=1}^{I} n_i = 427$, $P_\alpha = P_\beta = 1/1000$ and $a = b = 0.001$. Find the full conditional distributions of $\alpha, \beta, \tau_\alpha, \tau_\beta$, σ^{-2}, α_i, β_i, and (α_i, β_i).