Assignment 2

This homework assignment is **due** on **Mo, 2/16.**
Please work in groups. Groups can be as small (1) or as large as you wish and think is practicable. Just write all names on the assignment when you hand it in.

1. Let (x_1, x_2) be a bivariate random vector with p.d.f. $p(x_1, x_2)$. Let $p(x_1 \mid x_2)$ and $p(x_2 \mid x_1)$ denote the conditional densities. Show that

$$p(x_2) = \left(\int \frac{p(x_1 \mid x_2)}{p(x_2 \mid x_1)} \, dx_1 \right)^{-1}$$

2. Consider the following Markov chain. Let Be(α, β) denote a beta distribution. Starting with any intial value θ_0, iterate over $t = 1, 2, \ldots$

1. **With probability θ_{t-1} set $\theta_t \sim$ Be$(a + 1, 1)$**
2. **Otherwise $\theta_t = \theta_{t-1}$.**

2a. Find the transition kernel $K(\theta_t, \theta_{t+1})$ of this Markov chain.

2b. Show that this Markov chain satisfies the detailed balance condition for the invariant distribution $p(\theta) =$ Be$(a, 1)$.

3. Consider the following bivariate distribution on $\Re^2$, indexed with parameter $\nu > 0$.

$$p(x_1, x_2) \propto \frac{\exp(-x_1 - x_2)}{[1 + \exp(-x_1) + \exp(-x_2)]^{\nu+2}}$$

3a. Create a sample of size $m = 1000$ from $p(x_1, x_2)$ with $\nu = 1$, using a random walk Metropolis algorithm with proposal

$$q(\tilde{\mathbf{x}} \mid \mathbf{x}_t) = N(\mathbf{x}_t, 9 \cdot I)$$

Here $\mathbf{x}_t = (x_{t1}, x_{t2})$ is the state of the Markov chain after t iterations, and $9 \cdot I$ is a (2×2) covariance matrix with 9 on the diagonal.
Thin out the sample by only saving every 10-th iteration, i.e., a Monte Carlo sample of size 100. Plot the sample (scatter plot) and report the MCMC estimates of $\overline{x}_1 = E(x_1)$ and $\overline{x}_2 = E(x_2)$.

3b. Create another sample of size 1000 from the above distribution with $\nu = 1$ using a Gibbs sampler. Use the inverse c.d.f. method to generate the samples from the complete conditional distributions $p(x_1 \mid x_2)$ and $p(x_2 \mid x_1)$.

Again thin out the sample by saving $\mathbf{x}_t$ after every 10-th iteration only, plot the sample and report the Monte Carlo estimates for $\overline{x}_1 = E(x_1)$ and $\overline{x}_2 = E(x_2)$.

4. **Consider the following Bayesian model. Let $\mathbf{Ga}(x;\ \alpha, \beta)$ denote the pdf of a $\mathbf{Ga}(\alpha, \beta)$ distribution, $\mathbf{Ga}(x;\ \alpha, \beta) \propto x^{\alpha-1} e^{-\lambda x}$ with mean $E(x) = \alpha/\beta$.**

   Likelihood: **(sampling model)**

$$p(x_i \mid \boldsymbol{\theta}) = \pi \mathbf{Ga}(x;\ \alpha_1, \lambda_1) + (1-\pi)\mathbf{Ga}(x;\ \alpha_2, \lambda_2).$$

   is a mixture of two gamma distributions. Here $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \lambda_1, \lambda_2, \pi)$ is the 5-dimensional parameter vector.
   We have $n = 50$ observations $x_i \sim p(x_i \mid \theta)$, i.i.d., $i = 1, \ldots, n$. The data are linked on the blackboard page.

   Prior: **We know $\alpha_1 = 4$, $\alpha_2 = 6$ and $\pi = 0.5$, leaving only $\boldsymbol{\theta} = (\lambda_1, \lambda_2)$ as the unknown parameter vector. We assume**

$$p(\lambda_1, \lambda_2) = \mathbf{Ga}(\lambda_1;\ 4, \frac{1}{2})\ \mathbf{Ga}(\lambda_2;\ 4, \frac{1}{2})$$

   i.e., independent gamma priors on λ_1 and λ_2.

4a. **Implement Metropolis-Hastings posterior simulation to generate a Monte Carlo sample from the posterior, $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta} \mid \mathbf{x})$, $i, \ldots, m$.**
   Plot the marginal posterior $p(\lambda_j \mid \mathbf{x})$, $j = 1, 2$, as well as the joint posterior $p(\lambda_1, \lambda_2 \mid \mathbf{x})$.

5. **Consider the distribution on $\Re^2$**

$$p(x_1, x_2) \propto \exp(-5 \cdot |x_1^2 + x_2^2 - 1|)$$

5a. **Implement a random walk Metropolis algorithm which has $p(x_1, x_2)$ as stationary distribution. Plot the Monte Carlo sample that you obtain.**

5b. **Can you propose an algorithm that samples from $p(x_1, x_2)$ more efficiently?**

6. Consider data on the chemical content of leaves of turnip greens. Four different types of leaves were randomly selected, and for each type the calcium concentration, in percentage, was recorded for 4 randomly selected leaves of this type.

Leaf	**%Ca**			
1	**3.18**	**3.19**	**3.07**	**3.03**
2	**3.52**	**3.48**	**3.38**	**3.38**
3	**2.88**	**2.80**	**2.81**	**2.76**
4	**3.04**	**3.18**	**3.34**	**3.28**

Let y_{ij} denote the %Ca for leaf j of type i. We assume the sampling model

$$p(y_{ij} \mid \theta) = N(\mu + \alpha_i, 1/\tau^2).$$

Here τ^2 is the precision, i.e., $1/\tau^2$ is the variance of the normal distribution. We use the prior

$$p(\alpha_j \mid \tau_\alpha^2) = N(0, 1/\tau_\alpha^2) \tag{1}$$

$$p(\mu, \tau^2, \tau_\alpha^2) \propto 1/\tau^2 \, e^{-\tau_\alpha^2} \tag{2}$$

i.e., we assume a Ga$(1,1)$ gamma prior for τ_α^2 and a non-informative prior $1/\tau^2$ for τ^2 (it is easier to think in terms of the precisions here – but of course you could write the same in terms of the variances).

6a. Let $\tau^2 = 1/\sigma^2$ and $\tau_\alpha^2 = 1/\sigma_\alpha^2$ denote the precisions. Find the full conditional posterior distributions for α_i, μ, τ_α^2 and τ^2.

6b. Implement Gibbs sampling to generate a Monte Carlo sample $\boldsymbol{\theta}_i$, $i = 1, \ldots, m$, from the posterior distribution $p(\boldsymbol{\theta} \mid \mathbf{y})$. Here $\boldsymbol{\theta} = (\alpha_1, \ldots, \alpha_4, \mu, \tau^2, \tau_\alpha^2)$ and y are the data.