Homework 4

Q1

Pairs $(X_i, Y_i), i = 1, \ldots, n$ consist of correlated standard normal random variables (mean 0, variance 1) forming a sample from a bivariate normal $\mathcal{MVN}_2(\mathbf{0}, \Sigma)$ distribution, with covariance matrix

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

The density of $(X, Y) \sim \mathcal{MVN}_2(\mathbf{0}, \Sigma)$ is[1]

$$f(x, y|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right\},$$

with ρ as the only parameter. Take prior on ρ by assuming Jeffreys' prior on Σ as $\pi(\Sigma) = \frac{1}{|\Sigma|^{3/2}} = \frac{1}{(1-\rho^2)^{3/2}}$, since the determinant of Σ is $1 - \rho^2$. Thus

$$\pi(\rho) = \frac{1}{(1-\rho^2)^{3/2}} \mathbf{1}(-1 \leq \rho \leq 1).$$

(a) If $(X_i, Y_i), i = 1, \ldots, n$ are observed, write down the likelihood for ρ. Write down the expression for the posterior, up to the proportionality constant (that is, un-normalized posterior as the product of likelihood and prior).

(b) Since the posterior for ρ is complicated, develop a Metropolis-Hastings algorithm to sample from the posterior. Assume that $n = 100$ observed pairs (X_i, Y_i) gave the following summaries:

$$\sum_{i=1}^{100} x_i^2 = 113.5602, \quad \sum_{i=1}^{100} y_i^2 = 101.6489, \quad \text{and} \quad \sum_{i=1}^{100} x_i y_i = 75.1491.$$

In forming a Metropolis-Hastings chain take the following proposal distribution for ρ: At step i generate a candidate ρ' from the uniform $\mathcal{U}(\rho_{i-1} - 0.1, \rho_{i-1} + 0.1)$ distribution. Why does the proposal distribution cancel in the acceptance ratio expression?

(c) Simulate 51000 samples from the posterior of ρ and discard the first 1000 samples (burn in). Plot two figures: the histogram of ρs and the realizations of the chain for the last 1000 simulations (known as a trace plot). What is the Bayes estimator and 90% equitailed credible set of ρ based on the simulated chain?

(d) Replace the proposal distribution from (b) by the uniform $\mathcal{U}(-1, 1)$ (independence proposal). Comment on the results.

[1] See (6.1) on page 243 in `http://statbook.gatech.edu`.

Q2

Imagine your statistics professor made you watch him flip a coin one hundred times and record the results. He then tells you that, at some point, he switched the coin. Both of the coins had different biases for the probability of landing on heads.

He challenges you to use a Bayesian change point model to estimate at which flip he started using the second coin. You should assume that there were exactly two coins used and that the change point was equally likely to have happened at any flip.

The results of the coin flips can be found in `flips.csv`, where a value of 1 means heads and 0 means tails.

(a) Set up a Gibbs sampler for your model. Put Beta(2, 2) priors on the probability of each coin coming up heads. What likelihood is appropriate?

(b) In your report, include a point estimate, the 94% HPD credible set, and a density plot for the probability of each coin coming up heads and for the change point.

(c) The professor then says that he actually can't remember if he switched the coin or not. Use the posterior odds ratio for the change point to help evaluate whether the professor actually switched the coin. Based on this information, was the coin likely switched or not?

Q3

In a study of mating calls in the gray tree frogs *Hyla hrysoscelis* and *Hyla versicolor*, Gerhart (1994)[2] reports that in a location in Lousiana the following data on the length of male advertisement calls have been collected:

	Sample size	Average duration	SD of duration
Hyla chrysoscelis	43	0.65	0.18
Hyla versicolor	12	0.54	0.14

The two species cannot be distinguished by external morphology, but *H. chrysoscelis* (Fig. 1) are diploids while *H. versicolor* are tetraploids. The triploid crosses exhibit high mortality in larval stages, and if they attain sexual maturity, they are sterile. Females responding to the mating calls try to avoid mismatches.

Assume that duration observations are normally distributed with means μ_1 and μ_2, and precisions τ_1 and τ_2, for the two species respectively. For $i = 1, 2$, assume normal priors on μ_i's as $\mathcal{N}(0.6, 1)$ and gamma priors on τ_i's as $\mathcal{G}a(20, 0.5)$, where 0.5 is a rate hyperparameter.

(a) Based on observations and given priors, in the same loop construct two Gibbs samplers, one for (μ_1, τ_1) and the other for (μ_2, τ_2).

(b) Form a sequence of differences $\mu_{1,j} - \mu_{2,j}, j = 1, \ldots, 11000$, and after rejecting the initial 1000 differences, from the remaining simulations estimate 95% equitailed credible set for $\mu_1 - \mu_2$. Does this set contain zero? What can you say about the hypothesis $H_0 : \mu_1 = \mu_2$ based on this credible set? Elaborate on whether the

[2] Gerhardt, H. C. (1994). Reproductive character displacement of female mate choice in the grey treefrog, *Hyla chrysoscelis*. *Anim. Behav.*, **47**, 959–969.

Figure 1: Hyla chrysoscelis

length of call is a discriminatory characteristic.

Hint: When no raw data are given, that is, when data are summarized via sample size, sample mean, and sample standard deviation, the following identity may be useful:

$$\sum_{i=1}^{n}(y_i - \mu)^2 = (n-1)s^2 + n(\overline{y} - \mu)^2,$$

where $n, \overline{y}$, and s are sample size, sample mean, and sample standard deviation, respectively.