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

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

(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).

Likelihood

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

Prior:

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

Posterior is proportional to:

$$\prod_{i=1}^{n} \left[\frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}(x_i^2 - 2\rho x_i y_i + y_i^2)\right)\right] \frac{1}{(1-\rho^2)^{3/2}} \mathbf{1}(-1 \le \rho \le 1)$$

$$\left(\frac{1}{\sqrt{1-\rho^2}}\right)^n \exp\left(-\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n}(x_i^2 - 2\rho x_i y_i + y_i^2)\right) \frac{1}{(1-\rho^2)^{3/2}} \mathbf{1}(-1 \le \rho \le 1)$$

$$(1-\rho^2)^{-\frac{1}{2}(n+3)} \exp\left(-\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n}(x_i^2 - 2\rho x_i y_i + y_i^2)\right) \mathbf{1}(-1 \le \rho \le 1)$$

$$(1-\rho^2)^{-\frac{1}{2}(n+3)} \exp\left(-\frac{1}{2(1-\rho^2)} (\sum_{i=1}^{n} x_i^2 - 2\rho \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2)\right) \mathbf{1}(-1 \le \rho \le 1)$$

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

(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?

Our proposal is $U(\rho_i - .1, \rho_i + .1)$ with a density of $\frac{1}{(\rho_i+.1)-(\rho_i-.1)} = \frac{1}{.2} = 5$. This will cancel out in the acceptance ratio expression $\frac{q(\rho_i|\rho_*)}{q(\rho_*|\rho_i)}$ since it doesn't depend on any ρ. Then let the rest of our acceptance ratio be:

$$\min\left(1, \frac{\pi(\rho_*)}{\pi(p_i)}\right)$$

where ρ_* is the proposed ρ. We now have everything we need to code our sampler.

(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?

Figure 1 shows the density of ρ and the trace plot of the last 1,000 iterations. The Bayes estimator (mean) of ρ is given by $\hat{\rho} = 0.677$. The 90% equitailed credible set for ρ based on the simulated chain is $(0.599, 0.743)$.

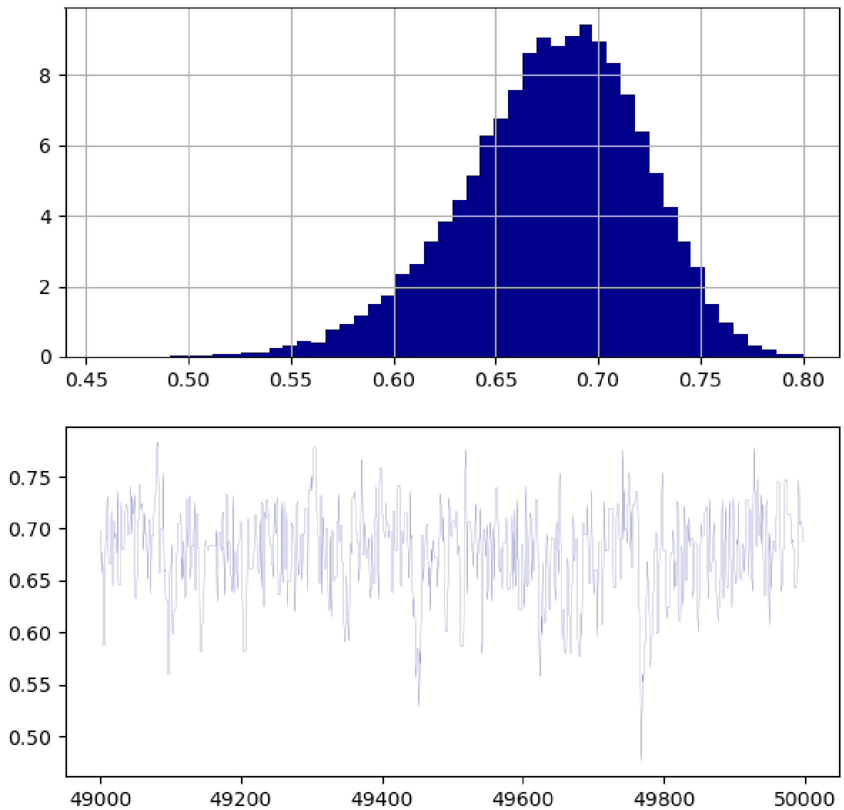

Figure 1: Density of ρ and trace plot of the last 1,000 iterations (Q1c).

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

Figure 2 shows the density of ρ and the trace plot of the last 1,000 iterations. Based on the plots and the accepted percentage, the $U(-1, 1)$ proposal is much less efficient and so has fewer useful samples, although the mean and credible set remain the same with some allowance for sampling error. We can see the lack of efficiency reflected in the trace plot, as well. Some people may experience a few divide by zero errors depending on seed and initial value.

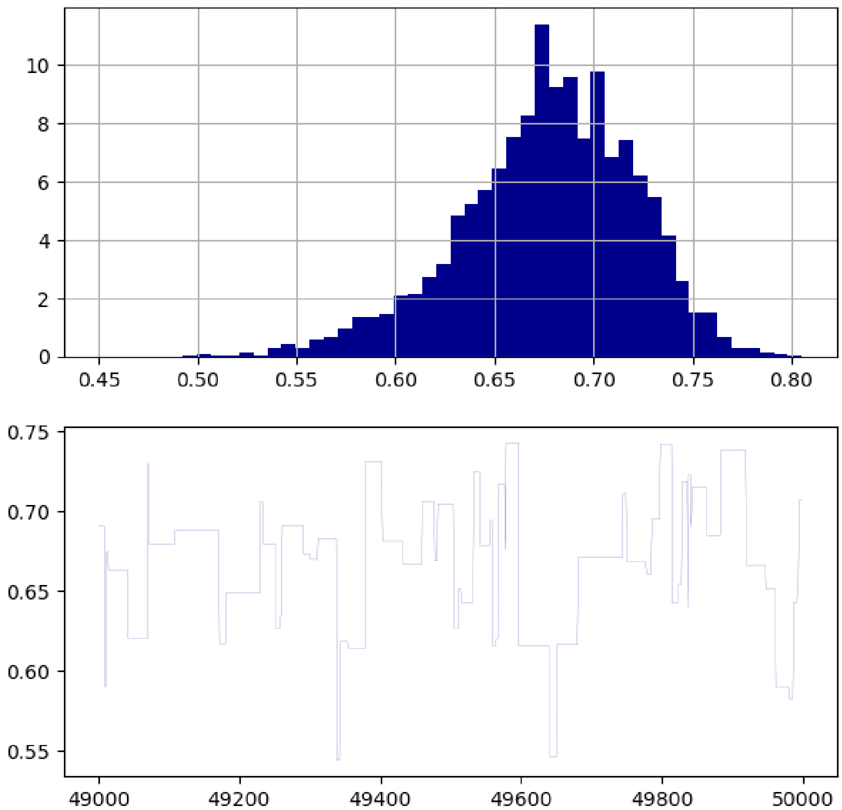

Figure 2: Density of ρ and trace plot of the last 1,000 iterations under fixed uniform proposal (Q1d).

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?

The Bernoulli distribution is most appropriate for our data:

$$P(X = x) = \begin{cases} p^x(1-p)^{1-x}, & x \in \{0, 1\} \\ 0, & \text{otherwise} \end{cases}$$

The discrete uniform distribution is the prior for m:

$$P(m = k) = \frac{1}{n}, \quad k \in \{1, 2, \ldots, n\}$$

And finally the Beta distribution priors are given by:

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 < x < 1$$

Our model is:

$$\begin{aligned} x_i \mid p_1 &\sim \text{Bernoulli}(p_1) && \text{for } i = 1, \ldots, m, \\ x_i \mid p_2 &\sim \text{Bernoulli}(p_2) && \text{for } i = m+1, \ldots, n, \\ m &\sim \text{Uniform}\{1, 2, \ldots, n\}, \\ p_1 &\sim \text{Beta}(2, 2), \\ p_2 &\sim \text{Beta}(2, 2). \end{aligned}$$

Joint density:

$$P(p_1, p_2, m \mid \mathbf{X}) \propto \overbrace{\left(\prod_{i=1}^{m} p_1^{x_i}(1-p_1)^{1-x_i}\right)\left(\prod_{i=m+1}^{n} p_2^{x_i}(1-p_2)^{1-x_i}\right)}^{\text{Likelihood}} \overbrace{\left(\frac{1}{n}\right)}^{\text{Prior for } m} \overbrace{\left(p_1^{\alpha_1-1}(1-p_1)^{\beta_1-1}\right)}^{\text{Prior for } p_1} \overbrace{\left(p_2^{\alpha_2-1}(1-p_2)^{\beta_2-1}\right)}^{\text{Prior for } p_2}$$

Full conditional for p_1:

$$\begin{aligned} P(p_1 \mid p_2, m, \mathbf{X}) &\propto \overbrace{\left(\prod_{i=1}^{m} p_1^{x_i}(1-p_1)^{1-x_i}\right)}^{\text{Likelihood for the first } m \text{ flips}} \overbrace{\left(p_1^{\alpha_1-1}(1-p_1)^{\beta_1-1}\right)}^{\text{Prior for } p_1} \\ &\propto p_1^{\left(\sum_{i=1}^{m} x_i\right)+\alpha_1-1}(1-p_1)^{\left(m-\sum_{i=1}^{m} x_i\right)+\beta_1-1} \\ &= \text{Beta}(\alpha_1 + \sum_{i=1}^{m} x_i, \beta_1 + m - \sum_{i=1}^{m} x_i) \end{aligned}$$

And similarly for p_2:

$$
\begin{aligned}
P(p_2 \mid p_1, m, \mathbf{X}) &\propto \overbrace{\left(\prod_{i=m+1}^{n} p_2^{x_i}(1-p_2)^{1-x_i}\right)}^{\text{Likelihood for the last } n-m \text{ flips}} \overbrace{\left(p_2^{\alpha_2-1}(1-p_2)^{\beta_2-1}\right)}^{\text{Prior for } p_2} \\
&\propto p_2^{\left(\sum_{i=m+1}^{n} x_i\right)+\alpha_2-1}(1-p_2)^{\left((n-m)-\sum_{i=m+1}^{n} x_i\right)+\beta_2-1} \\
&= \text{Beta}\Big(\alpha_2 + \sum_{i=m+1}^{n} x_i,\ \beta_2 + (n-m) - \sum_{i=m+1}^{n} x_i\Big)
\end{aligned}
$$

Finally, m:

$$
\begin{aligned}
P(m \mid p_1, p_2, \mathbf{X}) &\propto \overbrace{\left(\prod_{i=1}^{m} p_1^{x_i}(1-p_1)^{1-x_i}\right)\left(\prod_{i=m+1}^{n} p_2^{x_i}(1-p_2)^{1-x_i}\right)}^{\text{Likelihood}} \\
&\propto\ p_1^{\sum_{i=1}^{m} x_i}\,(1-p_1)^{m-\sum_{i=1}^{m} x_i}\ p_2^{\sum_{i=m+1}^{n} x_i}\,(1-p_2)^{(n-m)-\sum_{i=m+1}^{n} x_i}
\end{aligned}
$$

Let this un-normalized result be a function $f(m)$. Then the normalized distribution will be:

$$
P(m \mid p_1, p_2, \mathbf{X}) = \frac{f(m)}{\sum_{k=1}^{n} f(k)}
$$

(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.

- p_1 mean: 0.360, 94% HDI: [0.249, 0.476]
- p_2 mean: 0.647, 94% HDI: [0.428, 0.847]
- m mode: 77, 94% HDI: [43, 98]

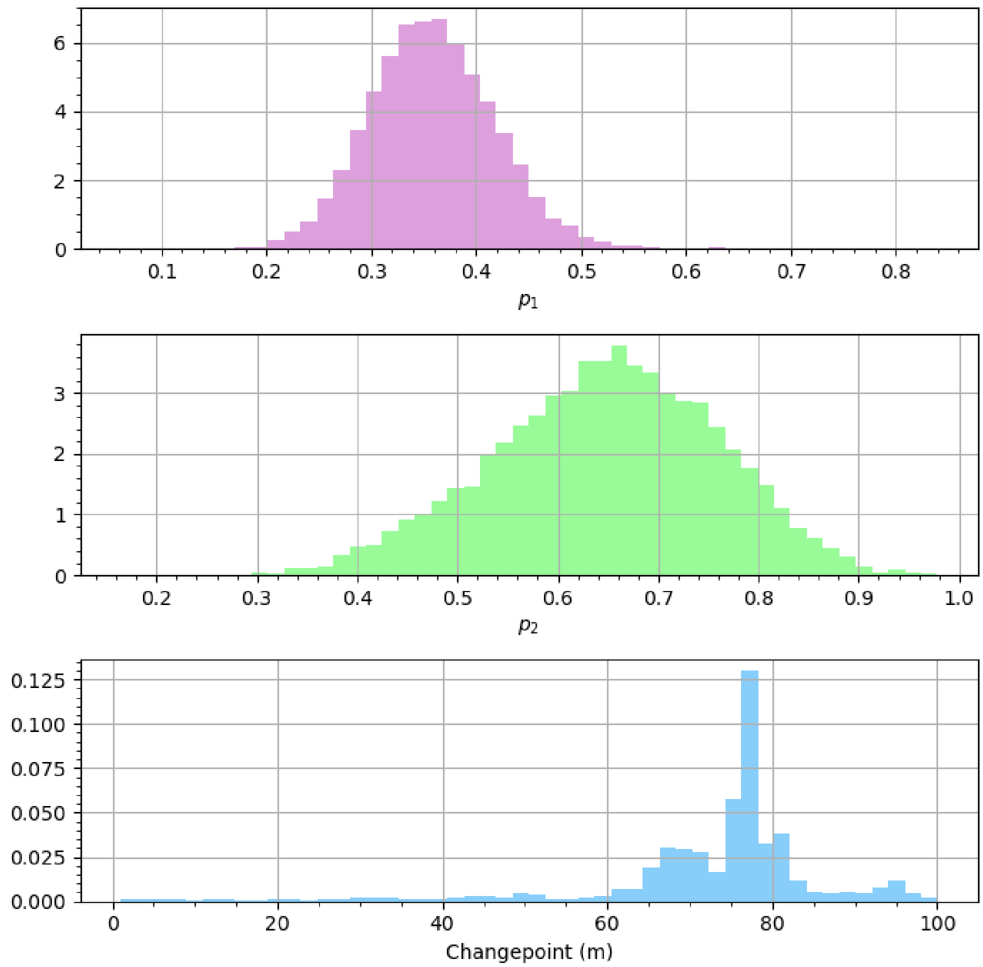

Figure 3: Densities of p_1, p_2, and m.

(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?

- Posterior odds ratio that $m = n$: 0.0017
- Posterior odds ratio that $m = 77$: 0.2085

There are different ways to look at this. We could compare to the prior odds of $m = n$, which are 100 to 1. Then the posterior odds here have reduced by an order of magnitude, indicating that based on our sample data, we've moved quite a ways away from the prior. Also, we could compare our posterior odds to the posterior odds of the mode, which is at the 77th flip. This point has the highest posterior odds at 0.2085.

The true values for the generated data were:

- p_1: 0.35

- p_2: 0.65
- m: 75

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. 4) 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.

Figure 4: Hyla chrysoscelis

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).

We denote the observations from Hyla chrysoscelis by $\mathbf{y}_1$ and from Hyla versicolor by $\mathbf{y}_2$. For group $i = 1, 2$, our model is:

$$\begin{aligned} y_{ij} \mid \mu_i, \tau_i &\sim \mathcal{N}(\mu_i, \tau_i) \\ \mu_i &\sim \mathcal{N}(0.6, 1) \\ \tau_i &\sim \mathrm{Ga}(20, 0.5) \end{aligned}$$

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

The density for a single observation is given by

$$f(y_i \mid \mu_i, \tau_i) = \sqrt{\frac{\tau_i}{2\pi}} \exp\left\{-\frac{\tau_i}{2}(y_i - \mu_i)^2\right\}$$

The likelihood for $\mathbf{y}_i$ is

$$\begin{aligned} L(\mu_i, \tau_i) &= \prod_{j=1}^{n_i} f(y_{ij} \mid \mu_i, \tau_i) \\ &= \prod_{j=1}^{n_i} \sqrt{\frac{\tau_i}{2\pi}} \exp\left\{-\frac{\tau_i}{2}(y_{ij} - \mu_i)^2\right\} \\ &= \left(\frac{\tau_i}{2\pi}\right)^{\frac{n_i}{2}} \exp\left\{-\frac{\tau_i}{2}\sum_{j=1}^{n_i}(y_{ij} - \mu_i)^2\right\} \end{aligned}$$

The priors are

$$\begin{aligned} \pi(\mu_i) &\propto \exp\left\{-\frac{1}{2}(\mu_i - 0.6)^2\right\}, \\ \pi(\tau_i) &\propto \tau_i^{20-1}\exp\{-0.5\tau_i\} = \tau_i^{19}\exp\{-0.5\tau_i\} \end{aligned}$$

The joint posterior (up to proportionality) is

$$\begin{aligned} \pi(\mu_i, \tau_i \mid \mathbf{y}_i) &\propto L(\mu_i, \tau_i)\,\pi(\mu_i)\,\pi(\tau_i) \\ &\propto \left(\frac{\tau_i}{2\pi}\right)^{\frac{n_i}{2}} \exp\left\{-\frac{\tau_i}{2}\sum_{j=1}^{n_i}(y_{ij} - \mu_i)^2\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mu_i - 0.6)^2\right\}\tau_i^{19}\exp\{-0.5\tau_i\} \\ &\propto \tau_i^{\frac{n_i}{2}+19}\exp\left\{-\frac{\tau_i}{2}\sum_{j=1}^{n_i}(y_{ij} - \mu_i)^2 - \frac{1}{2}(\mu_i - 0.6)^2 - 0.5\tau_i\right\} \end{aligned}$$

To derive $\pi(\mu_i \mid \tau_i, \mathbf{y}_i)$, we focus on the terms that depend on μ_i. Dropping factors that do not involve μ_i, we have

$$\pi(\mu_i \mid \tau_i, \mathbf{y}_i) \propto \exp\left\{-\frac{\tau_i}{2}\sum_{j=1}^{n_i}(y_{ij} - \mu_i)^2 - \frac{1}{2}(\mu_i - 0.6)^2\right\}$$

Using the identity

$$\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 = (n_i-1)s_i^2 + n_i(\overline{y}_i-\mu_i)^2,$$

and noting that the constant term $(n_i-1)s_i^2$ does not involve μ_i,

$$\begin{aligned}
\pi(\mu_i|\tau_i,\mathbf{y}_i) &\propto \exp\left\{-\frac{\tau_i}{2}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 - \frac{1}{2}(\mu_i-0.6)^2 - 0.5\tau_i\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\tau_i\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 + (\mu_i-0.6)^2 + \tau_i\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\tau_i\left[(n_i-1)s_i^2 + n_i(\overline{y}_i-\mu_i)^2\right] + (\mu_i-0.6)^2 + \tau_i\right)\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[(n_i\tau_i+1)\mu_i^2 - 2(n_i\tau_i\overline{y}_i+0.6)\mu_i\right]\right\} \\
&\propto \exp\left\{-\frac{n_i\tau_i+1}{2}\left(\mu_i^2 - 2\frac{n_i\tau_i\overline{y}_i+0.6}{n_i\tau_i+1}\mu_i\right)\right\}
\end{aligned}$$

This is the kernel of a normal density:

$$\mu_i \mid \tau_i, \mathbf{y}_i \sim \mathcal{N}\left(\frac{n_i\tau_i\overline{y}_i+0.6}{n_i\tau_i+1}, \frac{1}{n_i\tau_i+1}\right)$$

Next, to derive $\pi(\tau_i \mid \mu_i, \mathbf{y}_i)$, we collect the terms that involve τ_i:

$$\begin{aligned}
\pi(\tau_i \mid \mu_i, \mathbf{y}_i) &\propto \tau_i^{\frac{n_i}{2}+19}\exp\left\{-\frac{\tau_i}{2}\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 - 0.5\tau_i\right\} \\
&\propto \tau_i^{\frac{n_i}{2}+19}\exp\left\{-\frac{\tau_i}{2}\left[\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 + 1\right]\right\}
\end{aligned}$$

Again, applying

$$\sum_{j=1}^{n_i}(y_{ij}-\mu_i)^2 = (n_i-1)s_i^2 + n_i(\overline{y}_i-\mu_i)^2$$

we conclude that

$$\tau_i \mid \mu_i, \mathbf{y}_i \sim \mathrm{Ga}\left(\frac{n_i}{2}+20, \frac{1}{2}\left[(n_i-1)s_i^2 + n_i(\overline{y}_i-\mu_i)^2 + 1\right]\right)$$

(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 length of call is a discriminatory characteristic.

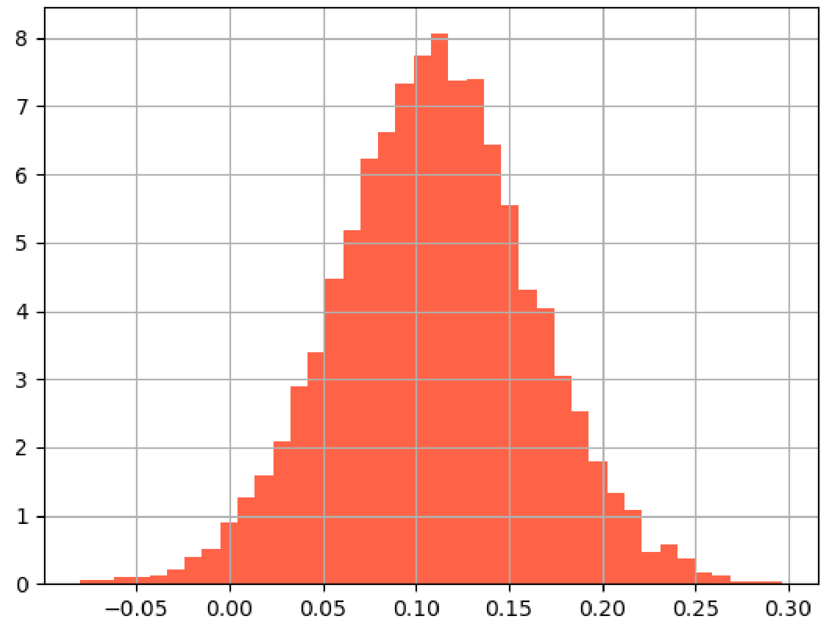

Figure 5: Density of $\mu_1 - \mu_2$ (plot not required).

The 95% equitailed credible set is $[0.007, 0.213]$. The true difference between the means has a 95% credibility of falling with that range, which is strong evidence that the call duration is indeed 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.