Homework 8

Aakash Kothapally, STA-360

5 PM EDT Monday, November 22

1c (Task 3)

We create a function to help us obtain our posterior predictive distributions.

```
set.seed(42)
library(MASS)
sim_count = 5000

predict <- function(y) {
  X = cbind(rep(1, 6), seq(1, 11, by = 2))
  beta_init = c(23, 0)
  sigma2_init = rbind(c(0.25, 0), c(0, 0.1))
  sigma2 = 0.7^2

  beta_matrix = matrix(nrow = sim_count, ncol = length(beta_init))
  sigma_matrix = numeric(sim_count)

  for (i in 1:sim_count) {
    beta = mvrnorm(1, solve(solve(sigma2_init) + (t(X) %*% X) / sigma2) %*%
                     (solve(sigma2_init) %*% beta_init + (t(X) %*% y) / sigma2),
                     solve(solve(sigma2_init) + (t(X) %*% X) / sigma2))
    sum_squares = (t(y) %*% y) - (2 * t(beta) %*% t(X) %*% y) + (t(beta) %*% t(X) %*%
                                                                    X %*% beta)
    sigma2 = 1 / rgamma(1, (1 + 6) / 2, (1 * 1/4 + sum_squares) / 2)

    beta_matrix[i,] = beta
    sigma_matrix[i] = sigma2
  }

  return(rnorm(sim_count, beta_matrix %*% c(1, 13), sqrt(sigma_matrix)))
}
```

We can use the function to obtain our posterior predictive distribution for each swimmer. We use the swimming dat file's data to do this, which has data for each swimmer by row. Our posterior predictive distributions are stored in a table-like structure with each swimmer getting their own column with 5000 simulations of the Gibbs sampler. Provided is a small selection of the first few simulations for each of the 4 swimmers.

```
swim = read.table("swim.dat")
swim_pred = apply(swim, MARGIN = 1, predict)
head(swim_pred)
```

```
##          [,1]     [,2]     [,3]     [,4]
## [1,] 22.86482 24.20159 22.34956 24.37127
## [2,] 22.22213 23.75341 22.76882 23.45320
## [3,] 22.16597 23.99481 22.67669 23.35086
## [4,] 22.08643 24.00023 23.05705 23.57172
```

```
## [5,] 22.59669 23.59786 23.13615 23.21242
## [6,] 22.34339 23.65125 22.87595 22.30891
```

1d (Task 4)

We're going to use the posterior predictive distributions' values for each of the simulations to help determine which swimmer is the fastest. Since we are looking at the fastest, we want the swimmer with the smallest time in each run. We can use that to construct the probabilities.

```
fastest_times = apply(swim_pred, MARGIN = 1, which.min)
```

This is the probability that swimmer 1 is the fastest in a given simulation.

```
length(fastest_times[fastest_times==1])/5000
```

```
## [1] 0.6392
```

This is the probability that swimmer 2 is the fastest in a given simulation.

```
length(fastest_times[fastest_times==2])/5000
```

```
## [1] 0.0166
```

This is the probability that swimmer 3 is the fastest in a given simulation.

```
length(fastest_times[fastest_times==3])/5000
```

```
## [1] 0.3152
```

This is the probability that swimmer 4 is the fastest in a given simulation.

```
length(fastest_times[fastest_times==4])/5000
```

```
## [1] 0.029
```

Based on these values, in the majority of simulations, swimmer 1 had the quickest time. We would recommend that the coach picks swimmer 1.

EXTRA CREDIT

Extra Credit

a We will show that $(\theta^T T^{-1}\mu)^T = \mu^T T^{-1}\theta$

First, we know for matrices A, B, and C where ABC is defined,
$(ABC)^T = C^T B^T A^T$.

Proof $(ABC)^T = C^T B^T A^T$

We know this because we know for $i \times k$ matrix A and $k \times j$ matrix B, AB exists and $(AB)_{ij} = \sum_k A_{ik} B_{kj}$.
We also know $(AB)^T_{ij} = (AB)_{ji} = \sum_k A_{jk} B_{ki} = \sum_k A^T_{kj} B_{ik}^T$
$= \sum_k B_{ik}^T A_{kj}^T = (B^T A^T)_{ij}$. So $(AB)^T = B^T A^T$.
Similarly, for A, B, and $j \times l$ matrix C, $(ABC)_{il} = \sum_j (AB)_{ij} (C)_{jl}$.
And $(ABC)^T_{il} = (ABC)_{li} = \sum_j (AB)_{lj} (C)_{ji} = \sum_j (AB)^T_{jl} (C)^T_{ij}$
$= \sum_j C^T_{ij} (AB)^T_{jl} = (C^T (AB)^T)_{il} = (C^T B^T A^T)_{il}$. So $(ABC)^T = C^T B^T A^T$.

So we know $(\theta^T T^{-1}\mu)^T = (\mu)^T (T^{-1})^T (\theta^T)^T$. We can simplify further.
First, since for any matrix X, $(X^T)^T = X$, $(\theta^T)^T = \theta$.
Next, we know the inverse of a symmetric matrix is also symmetric, so $(T^{-1})^T = T^{-1}$!
We can derive this result: $TT^{-1} = I$, and $I = I^T$, so $(TT^{-1}) = (TT^{-1})^T$,
or $(TT^{-1}) = (T^{-1})^T T^T$.

Also, $T^{-1}T = TT^{-1} = I$, so $T^{-1}T = (T^{-1})^T T^T$.
And $T = T^T$, since T is symmetric, so $T^{-1}T = (T^{-1})^T T$
↳ $T^{-1} T T^{-1} = (T^{-1})^T T T^{-1}$
↳ $T^{-1} I = (T^{-1})^T I$
$T^{-1} = (T^{-1})^T$.

So since T is symmetric, $(T^{-1})^T = T^{-1}$. So $\boxed{(\theta^T T^{-1}\mu)^T = \mu^T T^{-1}\theta}$

Also, $\theta^T T^{-1}\mu$ is a scalar (because of the dimensions being 1×1), and the transpose of a scalar is the same scalar, so $\theta^T T^{-1}\mu = \mu^T T^{-1}\theta$.

b $$p(\theta) = (2\pi)^{-p/2}(T^{-1})^{-1/2} e^{-\frac{1}{2}(\theta-\mu)^T (T^{-1})(\theta-\mu)}$$

$$\propto e^{-\frac{1}{2}(\theta^T T^{-1}\theta + \theta^T T^{-1}\mu - \underbrace{\mu^T T^{-1}\theta} + \mu^T T^{-1}\mu)}$$

Based on a, $\mu^T T^{-1}\theta = \theta^T T^{-1}\mu$.

$$\propto e^{-\frac{1}{2}(\theta^T T^{-1}\theta \underbrace{- \theta^T T^{-1}\mu - \theta^T T^{-1}\mu}_{\text{Combine like terms}} + \mu^T T^{-1}\mu)}$$

$$\propto e^{-\frac{1}{2}(\theta^T T^{-1}\theta - 2\theta^T T^{-1}\mu + \underbrace{\mu^T T^{-1}\mu})}$$

Can be canceled out, no θ term

$$\propto \boxed{e^{-\frac{1}{2}(\theta^T T^{-1}\theta - 2\theta^T T^{-1}\mu)}}$$

c $$p(\theta \mid \Sigma, y_{1:n}) \propto p(y_{1:n} \mid \theta, \Sigma) \cdot p(\theta)$$

We can use b: $p(\theta) \propto e^{-1/2(\theta^T T^{-1}\theta - 2\theta^T T^{-1}\mu)}$

$$p(y_{1:n} \mid \theta, \Sigma) = \prod_{i=1}^{n} (2\pi)^{-p/2}(\Sigma)^{-1/2} e^{-\frac{1}{2}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)}$$

$$\propto e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)}$$

$$\propto e^{-\frac{1}{2}\sum_{i=1}^{n}(y_i^T\Sigma^{-1}y_i \underbrace{- y_i^T\Sigma^{-1}\theta - \theta^T\Sigma^{-1}y_i}_{\text{Equal}} + \theta^T\Sigma^{-1}\theta)}$$

$$\propto e^{-\frac{1}{2}\sum_{i=1}^{n}(-2\theta^T\Sigma^{-1}y_i + \theta^T\Sigma^{-1}\theta)}$$

$$\propto e^{-\frac{1}{2}(-2\theta^T\Sigma^{-1}n\bar{y} + n\theta^T\Sigma^{-1}\theta)}$$

$$p(\theta \mid \Sigma, y_{1:n}) \propto e^{-\frac{1}{2}(-2\theta^T\Sigma^{-1}n\bar{y} + n\theta^T\Sigma^{-1}\theta)} e^{-\frac{1}{2}(\theta^T T^{-1}\theta - 2\theta^T T^{-1}\mu)}$$

$$\propto e^{-\frac{1}{2}(-2\theta^T(\Sigma^{-1}n\bar{y} + T^{-1}\mu) + \theta^T(n\Sigma^{-1}+T^{-1})\theta)}$$

$$\propto e^{-\frac{1}{2}\theta^T(n\Sigma^{-1}+T^{-1})\theta + \theta^T(n\Sigma^{-1}\bar{y}+T^{-1}\mu)}$$

$$\propto MVN((n\Sigma^{-1}+T^{-1})^{-1}(n\Sigma^{-1}\bar{y}+T^{-1}\mu), (n\Sigma^{-1}+T^{-1})^{-1})$$

So we have showed

$$p(\theta, \Sigma, y) \sim MVN(\mu^*(\Sigma) = T^*(n\Sigma^{-1}\bar{y}+T^{-1}\mu), T^* = (n\Sigma^{-1}+T^{-1})^{-1})$$

We will prove this equality. We can also use the fact that $tr(A+B) = tr(A) + tr(B)$.

d $$tr\left(\Psi\Sigma^{-1}\right) + \underbrace{\sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)}_{\substack{\text{This is a scalar} \\ a = tr(a)}} = tr\left(\Psi\Sigma^{-1}\right) + tr\left(\sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)\right)$$

$$= tr\left(\Psi\Sigma^{-1}\right) + \sum_{i=1}^{n} tr\underbrace{\left(\left((y_i-\theta)^T\Sigma^{-1}(y_i-\theta)\right)\right)}_{tr(ABC) = tr(CAB)}$$

$$= tr\left(\Psi\Sigma^{-1}\right) + \sum_{i=1}^{n} tr\left((y_i-\theta)(y_i-\theta)^T\Sigma^{-1}\right)$$

$$= tr\left(\Psi\Sigma^{-1}\right) + tr\left(\sum_{i=1}^{n}(y_i-\theta)(y_i-\theta)^T\Sigma^{-1}\right)$$

$$= tr\left(\Psi\Sigma^{-1} + \sum_{i=1}^{n}(y_i-\theta)(y_i-\theta)^T\Sigma^{-1}\right)$$

$$= tr\left(\left(\Psi + \sum_{i=1}^{n}(y_i-\theta)(y_i-\theta)^T\right)\Sigma^{-1}\right)$$

So $$tr\left(\Psi\Sigma^{-1}\right) + \sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta) = tr\left(\left(\Psi + \sum_{i=1}^{n}(y_i-\theta)(y_i-\theta)^T\right)\Sigma^{-1}\right)$$ ✓

e $$p(\Sigma \mid y, \theta) \propto p(y \mid \theta, \Sigma) \cdot p(\Sigma)$$

$$p(y \mid \theta, \Sigma) \propto \det(\Sigma)^{-n/2} e^{-\frac{1}{2}\left(\sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)\right)}$$

$$p(\Sigma) \propto \det(\Sigma)^{-\frac{1}{2}(v+p+1)} e^{-\frac{1}{2}\left(tr\left(\Psi\Sigma^{-1}\right)\right)}$$

$$p(\Sigma \mid y, \theta) \propto \det(\Sigma)^{-\frac{1}{2}(v+p+1+n)} e^{-\frac{1}{2}\underbrace{\left(tr\left(\Psi\Sigma^{-1}\right) + \sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)\right)}_{\text{Using the answer to d}}}$$

$$\propto \det(\Sigma)^{-\frac{1}{2}(v+p+1+n)} e^{-\frac{1}{2}\left(tr\left(\left(\Psi + \sum_{i=1}^{n}(y_i-\theta)(y_i-\theta)^T\right)\Sigma^{-1}\right)\right)}$$

$$\Sigma \mid \theta, \text{data} \sim IW\left(n+v, \left(\Psi + \sum_{i=1}^{n}(y_i-\theta)(y_i-\theta)^T\right)^{-1}\right)$$ ✓