STA 630 - Homework 3

Anthony Bernardi

March 25th, 2025

1 Problem 1

Consider the multivariate normal model with a known variance-covariance matrix $\sum$ and unknown mean vector μ. Derive the posterior distribution and show all steps.

We can specify a normal prior, and begin in the following way.

$$\pi(\theta) = N(\mu_0, \Sigma_0) \tag{1}$$

Which gives the following density function.

$$\pi(\theta) = (2\pi)^{-p/2}|\Sigma_0|^{-1/2} \exp\left\{-\frac{1}{2}(\theta - \mu_0)^T \Sigma_0^{-1}(\theta - \mu_0)\right\} \tag{2}$$

Which we can then simplify to get the following.

$$\pi\theta = (2\pi)^{-p/2}|\Sigma_0|^{-1/2} \exp\left\{-\frac{1}{2}\theta^T \Sigma_0^{-1}\theta + \mu_0^T \Sigma_0^{-1}\theta^T - \frac{1}{2}\mu_0^T \Sigma_0^{-1}\mu_0\right\} \tag{3}$$

Which approximates to the following.

$$\pi(\theta) \propto \exp\left(-\frac{1}{2}\theta^T \Sigma_0^{-1}\theta + \theta^T \Sigma_o^{-1}\mu_0\right) \tag{4}$$

Given our sampling model is multivariate normal, we can now imput the likelihood and proceed.

$$p(Y|\theta, \Sigma) = \prod_{i=1}^{n}(2\pi)^{-p/2}|\Sigma^{-1/2}| \exp(-(y_i - \theta)^T \Sigma^{-1}(y_i - \theta)) \tag{5}$$

Which we can simplify to the following.

$$p(Y|\theta,\Sigma) = (2\pi)^{np/2}|\Sigma|^{-n/2}\exp(-\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)) \quad (6)$$

$$p(Y|\theta,\Sigma) \propto \exp(-\frac{1}{2}\theta^T \cdot n\Sigma^{-1}\theta + \theta^T n\Sigma^{-1}(y_i-\theta)) \quad (7)$$

This now gives us the chance to get the posterior.

$$p(\theta|Y,\Sigma) \propto \exp(-\frac{1}{2}\theta^T\Sigma_0^{-1}\theta + \theta^T\Sigma_0^{-1}\mu_0) \times \exp(-\frac{1}{2}\theta^T n\Sigma^{-1}\theta + \theta^T n\Sigma^{-1}\bar{y}) \quad (8)$$

We can simplify this if we want.

$$p(\theta|Y,\Sigma) \propto \exp(-\frac{1}{2}\theta^T A_n\theta + \theta^T B_n) \quad (9)$$

Where we have the following.

$$A_n = \Sigma_0^{-1} + n\Sigma^{-1} \quad (10)$$

$$B_n = \Sigma_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y} \quad (11)$$

2 Problem 2

Problem 7.1 from Hoff.

2.1 Part A

Given the Jeffrey's prior for the multivariate normal distribution, why is this not a valid probability distribution?

We start by examining the Jeffrey's prior in this case, given below.

$$p_j(\theta,\Sigma) = |\Sigma|^{-(p+2)/2} \quad (12)$$

We can display that this is not a valid probability distribution, as it does not integrate to 1, and violates one of the axioms of probability.

$$\int_\theta |\Sigma|^{-(p+2)/2} d\theta d\Sigma \neq 1 \quad (13)$$

2.2 Part B

Obtain the form of the posterior distribution for the multivariate normal using the Jeffrey's prior that is proportional to the posterior for the individual parameters θ and Σ.

We can start by examining the likelihood function for the multivariate normal distribution, which we went over earlier.

$$p(Y|\theta, \Sigma) = (2\pi)^{np/2}|\Sigma|^{-n/2} \exp(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^T\Sigma^{-1}(y_i - \theta)) \tag{14}$$

We can now multiply this by the Jeffrey's prior to get the posterior, following this case.

$$p(\theta, \Sigma|Y) \propto |\Sigma|^{-(p+2)/2} \times |\Sigma|^{-n/2} \exp(-\frac{1}{2}\sum_{i=1}^{n}(y_i - \theta)^T\Sigma^{-1}(y_i - \theta)) \tag{15}$$

Ultimately, we want this to factor into the following, aligning with the principles of conditional independence.

$$p(\theta, \Sigma|Y) = p(\theta|Y, \Sigma)p(\Sigma|Y) \tag{16}$$

We know that the Sigma value follows an inverse Wishart distribution, and the theta value follows a multivariate normal distribution.

We use these two facts to guide our factoring and find an answer that is proportional to the posterior for the individual parameters.

3 Problem 3

Devise and implement an Importance sampler for estimating the expected value of a mixture of two beta distributions. Use that sample to estimate the probability that this random variable is included in the interval [0.45, 0.55]

The mixture of Beta distributions can be represented by the following in this case.

$$\delta = 0.3 \times Beta(5, 2) + 0.7 \times Beta(2, 8) \tag{17}$$

We then use the following R code to implement the Importance sampler.

```
# implementing importance sampling with beta mixture model
X = seq(0, 1, length.out = 1e2)
plot(X, dbeta(X, 5, 2))
plot(X, dbeta(X, 2, 8))

# trying to implement the mixture
b1 <- dbeta(X, 5, 2)
b2 <- dbeta(X, 2, 8)
mix_beta = 0.3*b1 + 0.7*b2
plot(mix_beta) # mixture model, our f(x) in this case

p <- dgamma(X, 2, 5) # picking another
plot(density(mix_beta), col = "cyan") # f(x)
lines(density(p), col = "pink") # p(x)

# Now just have to multiply and find a q(x) function for importance sampling
# taking the proudct
pq <- p * mix_beta
plot(density(pq), col = "orange") # p(x)q(x)
# how to sample from this?

q <- dgamma(X, 3, 6)
plot(density(q))

# now we can try to get the ratio that we'll sample from
sampling_ratio <- (p / q) * mix_beta
sampling_ratio[1] <- 0
plot(density(sampling_ratio), col = "green")

# now can do sampling...
# just to pull values from q
# and then compute these importance values with the samplign ratio
set.seed(100)
T = 10000
imp_vals <- rep(NA, T)

for (i in 1:T){
```

```
  y <- sample(1:100,1)
  imp_vals[i] <- sampling_ratio[y]
}

# now can get the mean
mean_imp <- mean(imp_vals)

# probability that the random variable is included in the interval 0.45,0.55
# can do this the naive way
target_prob <- sum(imp_vals > .45 & imp_vals < .55) / length(imp_vals)
print(paste0("Our target probability is: ", round(target_prob, 2)))
```

The code above implements the mixture distribution, and then uses the Gamma density function for an easier function to sample from, in order to implement Importance sampling in a somewhat proper way.

We then see that our mean value is 1.781, and the probability that the random variable is included in the interval [0.45, 0.55] is 0.17.

4 Problem 4

Consider a univariate normal model with mean μ and variance σ^2. If we use a Beta prior for μ and a log normal prior for σ^2, and we assume that these two parameters are independent, implement a Metropolis-Hastings algorithm to evaluate the posterior distribution of μ and σ^2.

We start by defining the likelihood function for the normal distribution as it is pertinent to our eventual algorithm, along with the paramters and their corresponding priors.

To implement the Metropolis-Hastings algorithm, we need to define the likelihood, and then consider a few candidate functions to sample from instead of the posterior.

More specifically, we need a markov chain that has a stationary distribution that is the posterior distribution, and fits the requirements to have a stationary distribution along with the detailed balance condition.

We attempt to implement the Metropolis-Hastings algorithm in the following R code.

```
# Want to investigate the posterior to approximate from
# product of a beta and log-normal random variables
```

```
# Try simluating this in R

mu <- dbeta(X, 2, 2)
plot(mu, type = 'l')
sigma <- dlnorm(X, 1, 10)
plot(sigma, type = 'l')

# entering the data
Y <- c(2.3656 , 2.4952 , 1.0837 , 0.7586 ,
       0.8780 , 1.2765 , 1.4598 , 0.1801 ,
       -1.0009 , 1.4870 , -0.1193 , 0.2578)

# trying the product
fx <- mu * sigma
plot(fx, type = 'l') # how to approximate this...
# envelope candidate function
env <- dbeta(X, 2, 6)

# propsing this more formally
prop_env <- function(n, p1, p2){
  return(rbeta(n, p1, p2))
}

metropolis_MCMC <- function(startval, iterations){
  chain = array(dim = c(iterations+1, 12)) # establishing chain
  chain[1,] = startvalue # length 12 along w data
  for (i in 1:iterations){
    proposal = prop_env(chain[i,], 2, 6)

    prob = exp(prop_env(12, 2, 6) - prop_env(chain[i,],2, 6))
    if (runif(1) < prob){
      chain[i+1, ] = proposal
    }else{
      chain[i+1, ] = chain[i,]
    }
  }
  return(chain)
}
```

```
start = Y
chain = metropolis_MCMC(startval = start ,
                        iterations = 100000)
```

This is where I am, I know I am missing the data implementation and the likelihood, but I wanted to outline my intuitive understanding of Metropolis-Hastings along with my ideas for implementation in R.