Homework 6

Problem 1

1.a

In HW5 Problem 5 I have shown that:

$$
\begin{aligned}
\pi(\mu_1, \mu_2) &\sim \ell(\mu_1, \mu_2) \cdot p(\mu_1, \mu_2) \\
&= \ell(\mu_1, \mu_2) \cdot 1 \\
&= \ell(\mu_1, \mu_2) \\
&= N_n(y_1, \ldots y_n \mid \mu, \Sigma) \\
&\sim \prod_{i=1}^{n} \exp\left\{ -\frac{1}{2}(y_i - \mu)^\top \Sigma^{-1}(y_i - \mu) \right\} \\
&\sim \prod_{i=1}^{n} \exp\left\{ -\frac{1}{2}\left[y_i^\top \Sigma^{-1} y_i - 2\mu^\top \Sigma^{-1} y_i + \mu^\top \Sigma^{-1}\mu \right] \right\} \\
&\sim \prod_{i=1}^{n} \exp\left\{ -\frac{1}{2}\left[-2\mu^\top \Sigma^{-1} y_i + \mu^\top \Sigma^{-1}\mu \right] \right\} \\
&= \exp\left\{ -\frac{1}{2} \sum_{i=1}^{n} \left[-2\mu^\top \Sigma^{-1} y_i + \mu^\top \Sigma^{-1}\mu \right] \right\} \\
&= \exp\left\{ -\frac{n}{2} \left[-2\mu^\top \Sigma^{-1} \bar{y} + \mu^\top \Sigma^{-1}\mu \right] \right\} \\
&\sim \exp\left\{ -\frac{n}{2} \left[(\mu - \bar{y})^\top \Sigma^{-1}(\mu - \bar{y}) \right] \right\} \\
&\sim N_2(\mu \mid \bar{y}, \frac{\Sigma}{n})
\end{aligned}
$$

$$
\pi(\mu) \sim N_2(\mu \mid \bar{y}, \frac{\Sigma}{n}) \\
\text{where } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i
$$

So that is the target distribution. The proposal distribution is:

$$
q(\theta, \phi) \sim N_2(\phi \mid \theta, \Sigma)
$$

Acceptance probability of MH is defined as follows:

$$
a(\theta, \phi) = \min\left(1, \frac{\pi(\phi))q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)} \right)
$$

Proposal is symmetric $q(\phi, \theta) = q(\theta, \phi)$, as shown below:

$$q(\theta, \phi) = N(\phi \mid \theta, \Sigma)$$
$$= exp\left(\frac{-\frac{1}{2}(\phi - \theta)^\top \Sigma^{-1}(\phi - \theta)}{\sqrt{(2\pi)^k |\Sigma|}}\right)$$
$$= exp\left(\frac{-\frac{1}{2}(\theta - \phi)^\top \Sigma^{-1}(\theta - \phi)}{\sqrt{(2\pi)^k |\Sigma|}}\right)$$
$$= N(\theta \mid \phi, \Sigma)$$
$$= q(\phi, \theta)$$

Then our term for acceptance probability simplifies to:

$$a(\theta, \phi) = \min\left(1, \frac{\pi(\phi))q(\phi, \theta)}{\pi(\theta)q(\theta, \phi)}\right)$$
$$= \min\left(1, \frac{\pi(\phi))}{\pi(\theta)}\right)$$
$$= \min\left(1, \frac{N_2(\phi \mid \bar{y}, \frac{\Sigma}{n})}{N_2(\theta \mid \bar{y}, \frac{\Sigma}{n})}\right)$$
$$= \min\left(1, exp\left\{-\frac{n}{2}\left((\phi - \bar{y})^\top \Sigma^{-1}(\phi - \bar{y}) - (\theta - \bar{y})^\top \Sigma^{-1}(\theta - \bar{y})\right)\right\}\right)$$

Metropolis Hastings Algorithm:

1. Set iteration counter $j = 1$ and choose arbitrary initial value $\boldsymbol{\mu}^{(0)}$
2. Generate a value ϕ from the proposal density $q\left(\boldsymbol{\mu}^{(j-1)}, \cdot\right)$ (as defined above)
3. Evaluate the acceptance probability of the move given by $\alpha\left(\boldsymbol{\mu}^{(j-1)}, \phi\right)$ (as defined above). If the move is accepted, then set the new value of the Markov chain as $\boldsymbol{\mu}^{(j)} = \phi$. If the move is not accepted, then set $\boldsymbol{\mu}^{(j)} = \boldsymbol{\mu}^{(j-1)}$, i.e. Markov chain does not move.
4. Change the counter from j to $j+1$ and return to Step 2 until convergence is reached or until R samples are generated.

1.b

```r
MH <- function(y, Sigma, R=1000){
  n <- dim(y)[1]

  # variables where generated values will be saved
  mu_chain      <- matrix(NA, R, 2)
  mu_chain[1,]  <- runif(2,0,10)

  y.mean1 <- mean(y[,1])
  y.mean2 <- mean(y[,2])
  y.mean <- c(y.mean1, y.mean2)

  Sigma_inv = solve(Sigma)

  # create parameter chain
  for (i in 2:(R)){
```

```r
    # old value
    old <- mu_chain[i-1,]
    # porposal
    new <- mvtnorm::rmvnorm(n=1, mean=old, sigma=Sigma)

    v1 = matrix(new-y.mean, nrow=2, ncol=1)
    v2 = matrix(old-y.mean, nrow=2, ncol=1)
    term1 = t(v1) %*% Sigma_inv %*% v1
    term2 = t(v2) %*% Sigma_inv %*% v2

    rate = exp(-n/2 * (term1-term2))
    a <- min(1, rate)

    if (a > runif(1)){
        mu_chain[i,] <- new
    }
    else{
        mu_chain[i,] <- old
    }
  }

  return(mu_chain)
}
```

1.c

```r
set.seed(42)
mu_real = c(6,2)
sigma1 = sqrt(2)
sigma2 = sqrt(0.5)
rho = 0.05
n = 100
Sig = matrix(c(sigma1^2, sigma1*sigma2*rho,
               sigma1*sigma2*rho, sigma2^2), nrow = 2, ncol = 2)

Y <- MASS::mvrnorm(n, mu_real, Sig)


R1 = 100
R2 = 100000


mu_chain_1 <- MH(Y, Sig, R=R1)
mu_chain_2 <- MH(Y, Sig, R=R2)
```

1.d

```r
loglikelihood=function(mu1,mu2){
    sum=0
    for(i in 1:n){ sum=sum+sigma2^2*(Y[i,1]-mu1)^2-2*rho*sigma1*sigma2*(Y[i,1]-mu1)*
```

```
        (Y[i,2]-mu2)+sigma1^2*(Y[i,2]-mu2)^2 }
    return(-sum/(2*sigma1^2*sigma2^2*(1-rho^2))) }

Dev=function(mu1,mu2) {
    return(-2*loglikelihood(mu1,mu2))
}

#DIC with (4.12) from the script
DIC1=function(mu1,mu2) {
    Dbar=mean(Dev(mu1,mu2))
    return(2*Dbar-Dev(mean(mu1),mean(mu2)))
}

#DIC with (4.13) from the script
DIC2=function(mu1,mu2) {
    Dbar=mean(Dev(mu1,mu2))
    R=length(mu1)
    return(R/(2*(R-1))*mean((Dev(mu1,mu2)-Dbar)^2)+Dbar)
}
```

```
DIC1(mu_chain_1[,1], mu_chain_1[,2])
```

```
## [1] 612.1953
```

```
DIC2(mu_chain_1[,1], mu_chain_1[,2])
```

```
## [1] 282610.4
```

```
DIC1(mu_chain_2[,1], mu_chain_2[,2])
```

```
## [1] 192.9054
```

```
DIC2(mu_chain_2[,1], mu_chain_2[,2])
```

```
## [1] 807.9504
```

For both DIC from eq. 4.12 and eq. 4.13, MH with more iterations obtained a lower score, which indicates that longer sampling is beneficial.

The results from HW5 with Gibbs where as follows: - 197.3104 - 4447.148 - 179.0809 - 179.5949

We notice that Gibbs has lower DIC scores for all chain lengths and both DIC1 and DIC2. Which let's us assume, that Gibbs sampler was a better choice for sampling from this distribution.

It is worth noting, that proposal density wasn't optimal for MH sampler. Lower variance, for example, would be very beneficial, at least for the R=100000 chain.

Problem 2

The chain is reversible and $\pi(\cdot)$ is it's stationary distribution, if the detailed balance equation is satisfied (script, p42). The detailed balance equation is defined as follows:

$$\pi(\boldsymbol{\theta})p(\boldsymbol{\theta},\phi) = \pi(\phi)p(\phi,\boldsymbol{\theta}) \quad \forall \boldsymbol{\theta}, \phi$$

Then, using transition kernel of MH algorithm:

$$\pi(\theta)q(\theta,\phi)a(\theta,\phi) = \pi(\phi)q(\phi,\theta)a(\phi,\theta) \quad \forall \theta, \phi$$

Now let's solve the equation

$$\frac{a(\theta,\phi)}{a(\phi,\theta)} = \frac{\pi(\phi)}{\pi(\theta)} \cdot \frac{q(\phi,\theta)}{q(\theta,\phi)}$$

$$\text{if } \frac{\pi(\phi)q(\phi,\theta)}{\pi(\theta)\cdot q(\theta,\phi)} > 1:$$
$$\frac{1}{\frac{\pi(\theta)q(\theta,\phi)}{\pi(\phi)q(\phi,\theta)}} = \frac{\pi(\phi)}{\pi(\theta)}\frac{q(\phi,\theta)}{q(\theta,\phi)}$$
$$1 = 1$$

$$\text{if } \frac{\pi(\phi)q(\phi,\theta)}{\pi(\theta)\cdot q(\theta,\phi)} \leq 1:$$
$$\frac{\frac{\pi(\phi)q(\phi,\theta)}{\pi(\theta)q(\theta,\phi)}}{1} = \frac{\pi(\phi)}{\pi(\theta)}\frac{q(\phi,\theta)}{q(\theta,\phi)}$$
$$1 = 1$$

So the detailed balance equation holds for all θ, ϕ and for an arbitrary kernel $q(\phi, \theta)$:

$$\pi(\theta)q(\theta,\phi)a(\theta,\phi) = \pi(\phi)q(\phi,\theta)a(\phi,\theta) \quad \forall \theta, \phi$$

Therefore, the chain is reversible and $\pi(\cdot)$ is it's stationary distribution.