

Homework 7

Samuel Eure

10/28/2018

Problem 7.2

7.2a

The probability of $p(y_{1:n} \mid \theta, \Sigma)$ is as follows

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

thus, substituting $\Sigma^{-1} = \Psi$ and knowing that for a nonsingular matrix A , $|A| = \frac{1}{|A^{-1}|}$, the reparametarized likelihood is

$$p(y_{1:n} \mid \theta, \Psi) = (2\pi)^{-np/2} |\Psi|^{n/2} e^{-\frac{1}{2} \sum_i^n (y_i - \theta)^T \Psi (y_i - \theta)}$$

Resulting in a log-likelihood of

$$l(\theta, \Psi \mid y_{1:n}) \propto \frac{n}{2} \text{Log}|\Psi| - \frac{1}{2} \sum_i^n (y_i - \theta)^T \Psi (y_i - \theta)$$

Letting $(y_i - \theta) = ([y_i - \bar{y}] + [\bar{y} - \theta])$,

$$l(\theta, \Psi \mid y_{1:n}) \propto \frac{n}{2} \text{Log}|\Psi| - \frac{1}{2} \sum_i^n ([y_i - \bar{y}] + [\bar{y} - \theta])^T \Psi ([y_i - \bar{y}] + [\bar{y} - \theta])$$

Simplifying

$$\sum_i^n ([y_i - \bar{y}] + [\bar{y} - \theta])^T \Psi ([y_i - \bar{y}] + [\bar{y} - \theta]) = \sum_i^n [y_i - \bar{y}]^T \Psi [y_i - \bar{y}] + \sum_i^n [\bar{y} - \theta]^T \Psi [\bar{y} - \theta] + 2 \sum_i^n [y_i - \bar{y}]^T \Psi (\bar{y} - \theta)$$

and also

$$\sum_i^n [y_i - \bar{y}]^T \Psi (\bar{y} - \theta) = \sum_i^n (y_i^T \Psi \bar{y} + \bar{y}^T \Psi \theta - y_i^T \Psi \theta - \bar{y}^T \Psi \bar{y}) = n \bar{y}^T \Psi \bar{y} + n \bar{y}^T \Psi \theta - n \bar{y}^T \Psi \theta - n \bar{y}^T \Psi \bar{y} = 0$$

$$\implies l(\theta, \Psi \mid y_{1:n}) \propto \frac{n}{2} \text{Log}|\Psi| - \frac{1}{2} \left(\sum_i^n [y_i - \bar{y}]^T \Psi [y_i - \bar{y}] + \sum_i^n [\bar{y} - \theta]^T \Psi [\bar{y} - \theta] \right)$$

Now, since Ψ is symmetric, $\sum_i^n [y_i - \bar{y}]^T \Psi [y_i - \bar{y}] = \text{tr}(\sum_i^n [y_i - \bar{y}][y_i - \bar{y}]^T \Psi)$, where $\text{tr}()$ is the trace function. Thus,

$$\implies l(\theta, \Psi \mid y_{1:n}) \propto \left(\frac{n}{2} \text{Log}|\Psi| - \frac{n}{2} [\theta - \bar{y}]^T \Psi [\theta - \bar{y}] \right) + \left(-\frac{n}{2} \text{tr} \left[\sum_i^n \frac{[y_i - \bar{y}][y_i - \bar{y}]^T}{n} \Psi \right] \right)$$

Letting $S = \sum_i^n \frac{[y_i - \bar{y}][y_i - \bar{y}]^T}{n}$,

$$\implies l(\theta, \Psi \mid y_{1:n}) \propto \left(\frac{n}{2} \text{Log}|\Psi| - \frac{n}{2} [\theta - \bar{y}]^T \Psi [\theta - \bar{y}] \right) + \left(-\frac{n}{2} \text{tr} [S \Psi] \right)$$

Thus, if $\frac{l(\theta, \Psi \mid y_{1:n})}{n} \propto \log(p(\theta \mid \Psi)p(\Psi))$, then since the above implies

$$e^{l(\theta, \Psi \mid y_{1:n})/n} \propto p(\theta \mid \Psi)p(\Psi) \propto \left(|\Psi|^{1/2} e^{-\frac{1}{2}(\theta - \bar{y})^T \Psi (\theta - \bar{y})} \right) \left(e^{-\frac{1}{2} \text{tr}(S \Psi)} \right)$$

Which implies that

$$p(\theta \mid \Psi) \propto |\Psi|^{1/2} e^{-\frac{1}{2}(\theta - \bar{y})^T \Psi (\theta - \bar{y})}$$

which implies that

$$(\theta \mid \Psi) \sim MVN(\bar{y}, \Psi^{-1})$$

Furthermore, the above implies

$$p(\Psi) \propto e^{-\frac{1}{2} \text{tr}(S \Psi)} = |\Phi|^{\frac{0}{2}} e^{-\frac{1}{2} \text{tr}(S \Psi)} = |\Phi|^{\frac{p+1-p-1}{2}} e^{-\frac{1}{2} \text{tr}(S \Psi)}$$

implying

$$\Psi \sim \text{Wishart}(p+1, S^{-1})$$

7.2b

Given that

$$\Sigma \sim \text{Inv-Wishart}(p+1, S^{-1})$$

with $S^{-1} = \sum_i^n \frac{[y_i - \bar{y}][y_i - \bar{y}]^T}{n}$, $p_U(\theta, \Sigma \mid y_{1:n})$ can be found as follows

$$\begin{aligned} p_U(\theta, \Sigma \mid y_{1:n}) &\propto p_U(\theta \mid \Sigma) p_U(\Sigma) p_U(y_{1:n} \mid \theta, \Sigma) \\ &\propto \left(|\Sigma|^{-1/2} e^{-\frac{1}{2}(\theta - \bar{y})^T \Sigma^{-1} (\theta - \bar{y})} \right) \left(|\Sigma|^{-(p+1+p+1)/2} e^{-\frac{1}{2} \text{tr}(S \Sigma^{-1})} \right) \left(|\Sigma|^{-n/2} e^{-\frac{1}{2} \sum_i^n (y_i - \theta)^T \Sigma^{-1} (y_i - \theta)} \right) \\ &\propto |\Sigma|^{-\frac{n+2p+3}{2}} e^{-\frac{1}{2} [(\theta - \bar{y})^T \Sigma^{-1} (\theta - \bar{y}) + \text{tr}(S \Sigma^{-1}) + \sum_i^n (y_i - \theta)^T \Sigma^{-1} (y_i - \theta)]} \end{aligned}$$

but since

$$\begin{aligned}
& (\theta - \bar{y})^T \Sigma^{-1} (\theta - \bar{y}) + \text{tr}(S \Sigma^{-1}) + \sum_i^n (y_i - \theta)^T \Sigma^{-1} (y_i - \theta) \\
&= \text{tr}((\theta - \bar{y})(\theta - \bar{y})^T \Sigma^{-1}) + \text{tr}(S \Sigma^{-1}) + \text{tr}\left(\sum_i^n (y_i - \theta)(y_i - \theta)^T \Sigma^{-1}\right) \\
&= \text{tr}\left((\theta - \bar{y})(\theta - \bar{y})^T \Sigma^{-1} + S \Sigma^{-1} + \sum_i^n (y_i - \theta)(y_i - \theta)^T \Sigma^{-1}\right) \\
&= \text{tr}\left([\theta - \bar{y})(\theta - \bar{y})^T + S + \sum_i^n (y_i - \theta)(y_i - \theta)^T] \Sigma^{-1}\right) = \text{tr}\left((n+1)[\theta - \bar{y})(\theta - \bar{y})^T + S] \Sigma^{-1}\right) \\
&\implies p_U(\theta, \Sigma \mid y_{1:n}) \propto \left(|\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(\theta - \bar{y})^T \frac{\Sigma^{-1}}{(n+1)}(\theta - \bar{y})}\right) \left(|\Sigma|^{-\frac{n+2p+2}{2}} e^{-\frac{1}{2}\text{tr}((n+1)S \Sigma^{-1})}\right)
\end{aligned}$$

Or, perhaps more clearly put

$$p_U(\theta \mid \Sigma, y_{1:n}) p_U(\Sigma \mid y_{1:n}) \propto \left(|\Sigma|^{-\frac{1}{2}} e^{-\frac{1}{2}(\theta - \bar{y})^T (n+1) \Sigma^{-1} (\theta - \bar{y})}\right) \left(|\Sigma|^{-\frac{(n+p+1)+p+1}{2}} e^{-\frac{1}{2}\text{tr}((n+1)S \Sigma^{-1})}\right)$$

Since for a scalar a , $(A = a\Sigma) \Rightarrow A^{-1} = \frac{1}{a}\Sigma^{-1}$, this implies

$$(\theta \mid \Sigma, y_{1:n}) \sim MVN\left(\bar{y}, \frac{\Sigma}{(n+1)}\right)$$

$$(\Sigma \mid y_{1:n}) \sim \text{Inv-Wishart}(p+1+n, \frac{S^{-1}}{n+1})$$

So the joint distribution of θ, Σ conditional on the data is a Multivariate Normal-Inverse Wishart distribution. They are proper distributions, they they could be interpreted as the posterior distribution, however we used posterior data to obtain our prior, so if we are interpreting it this way we should understand that very little prior knowledge has been used. One interesting feature here is that the prior is a conjugate prior for likelihood measure we are using.

Problem 7.4

7.4a

For this problem, the units are in years or years squared. For most married couples I know, the age of the husband is typically about the same as the age of the wife, so I think setting $\theta_h = \theta_w = \theta_{prior}$ is a reasonable idea. Moreover, since my experiences have given me the impression that most married couples are roughly the same age, I will have $\Sigma_h = \Sigma_w = \Sigma_{prior}$. To get more specific, I have not heard of any people getting married before the age of 18, and I have met very few married couples who are over the age of 70, so I think having $\theta_{prior} \in [18, 70]$ is a reasonable goal. If I want my distribution to be symmetric, I may as well pick $\theta_{prior} = \frac{70+18}{2} = 44$. Moreover, if using an normal distribution, I will want $\theta_{prior} \pm 2\sigma_{prior} \subset [18, 70]$, which means I will need $(70 - 44)/2 = \sigma_{prior} = 13$ to have about 95% of my distribution between $[18, 70]$. Thus, I'll set $\sigma_h^2 = \sigma_w^2 = 169$. Since I expect the ages of the two individuals to be strongly reasonably correlated with each other based on life experience, Ill set $corr_{h,w} = .8$, resulting in a $COV(Age_h, Age_w) = 135.2$. Thus

$$\mu_0 = (44, 44)^T$$

```
library(MASS)
up = 70; low = 18
mid = mean(c(up, low))
mu0 = c(mid, mid)
sd = (up-mid)/2
p = 0.8;
Lambda0 = matrix(c(sd**2,sd**2*p,sd**2*p,sd**2), nrow = 2)
print('Lambda_0')

## [1] "Lambda_0"
Lambda0

##      [,1] [,2]
## [1,] 169.0 135.2
## [2,] 135.2 169.0
```

Hence the following prior for θ

$$(\theta) \sim \text{MVN}(\mu_0, \Lambda_0)$$

Moreover, I'll create the prior for my covariance matrix using the same logic from above and setting $\Lambda_0 = \Sigma_0$. However, since I've met a good amount of married couples in my life, I'm not completely oblivious to what these parameters should be, so I'll choose

$$\Sigma \sim \text{Inv-Wishart}(2 + 10, S_0^{-1})$$

with $S_0 = 9\Sigma_0$

Where 2 came from the fact that my θ is two-dimensional, 10 from my confidence, and $9 = 12 - 2 - 1$, i.e. $\nu_0 - p - 1$

7.4b

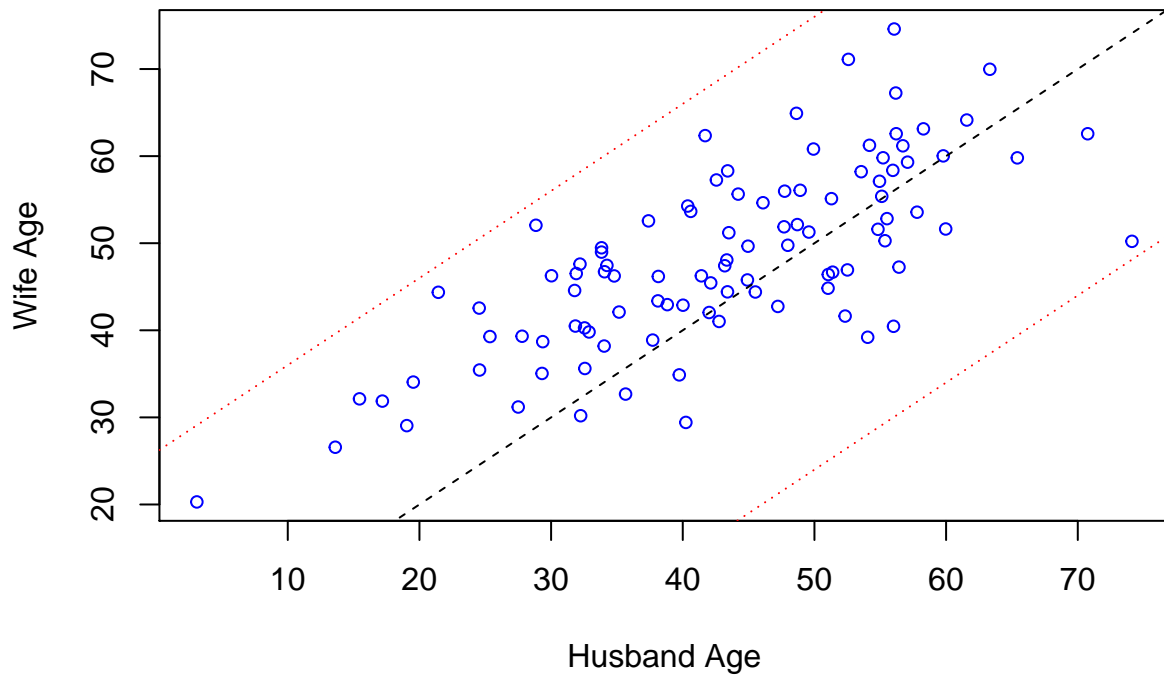
Below I've plotted four different prior predictive datasets. I've added a black dashed line of slope 1 which represents my expected location for pairs of husband and wife ages, and two dashed red lines to show where I expect the majority of my distribution to fall between.

```
set.seed(360)
Sigma0 = Lambda0
nu0 = length(mu0)+10
S0 = Sigma0*(nu0 - length(mu0) - 1)
S0.inverse = solve(S0)
N = 4;
SIGMAS = array(0, dim=c(2,2,N))
THETAS = array(0, dim=c(1,2,N))
n = 100;
Couples = array(0, dim=c(n,2,N))
for(i in 1:N){
  SIGMAS[,i] = solve(rWishart(1, nu0, S0.inverse)[,1])
  THETAS[,i] = mvrnorm(1, mu0, Lambda0)
  Couples[,i] = mvrnorm(n, THETAS[,i], SIGMAS[,i])
}
for(k in 1:N){
  plot(Couples[,k], col = 'blue', cex = .8,
       xlab = "Husband Age", ylab = "Wife Age", main = "Husband vs. Wife Age")
  abline(0,1, lty = 2)
  abline(2*sd,1, lty = 3, col = 'red')
  abline(-2*sd,1, lty = 3, col = 'red')
}
```

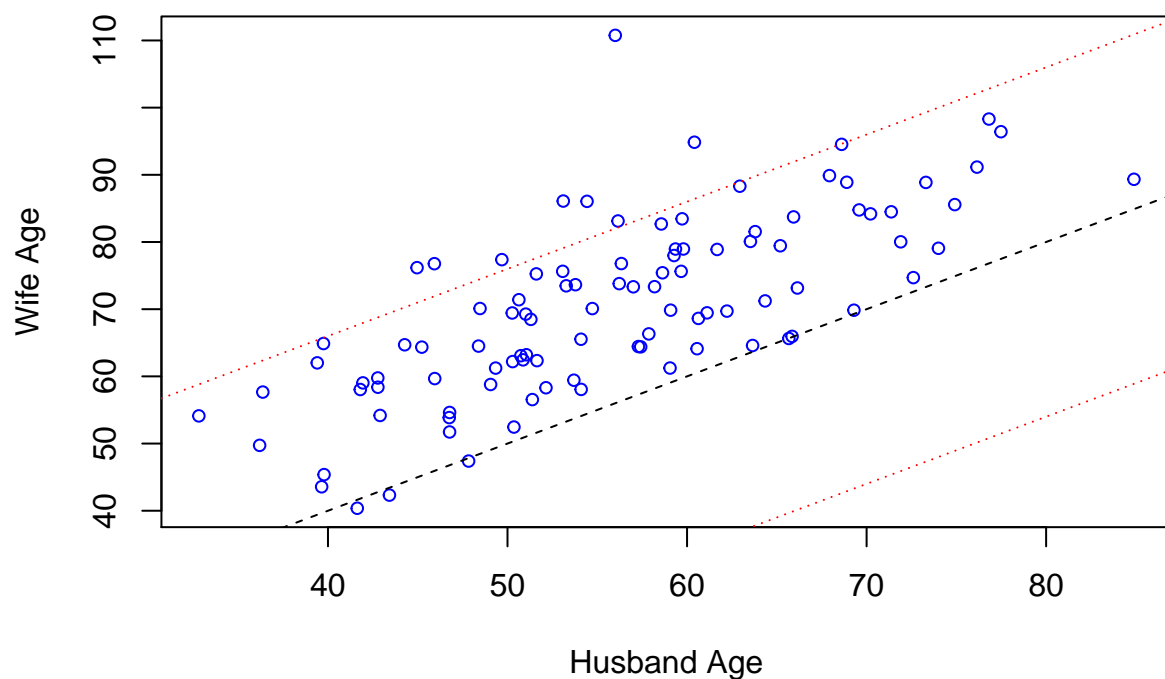
Husband vs. Wife Age



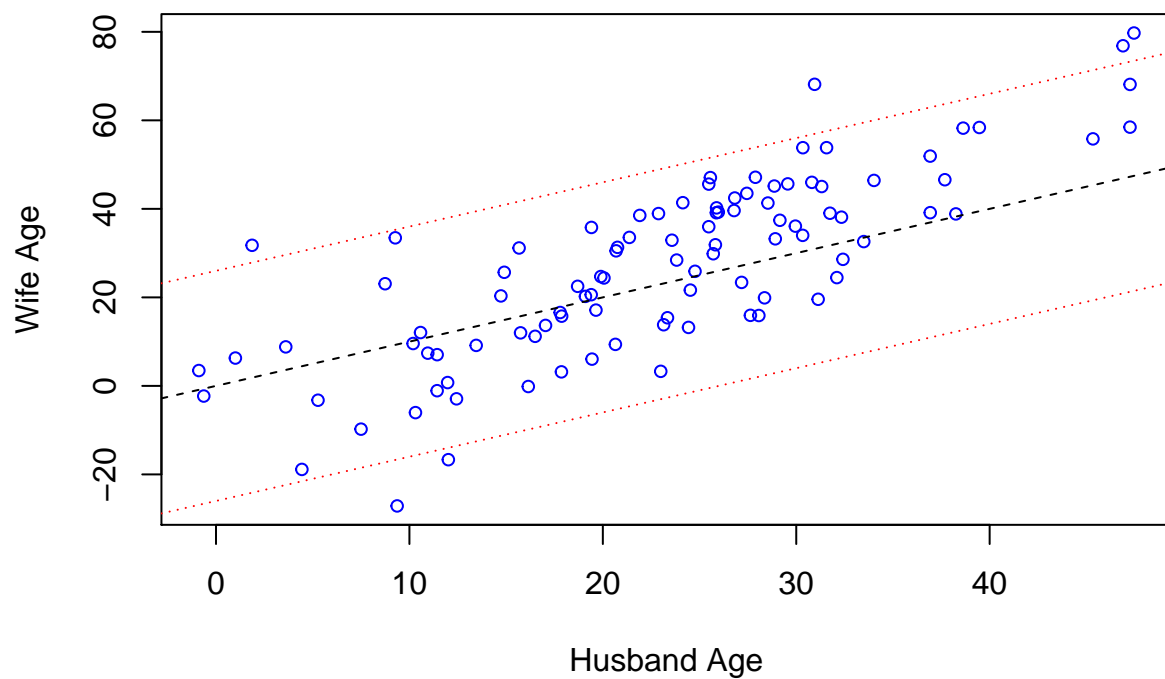
Husband vs. Wife Age



Husband vs. Wife Age



Husband vs. Wife Age



7.4c

```
Y = read.table("agehw.dat", header = T)
n = length(Y[,1])
```

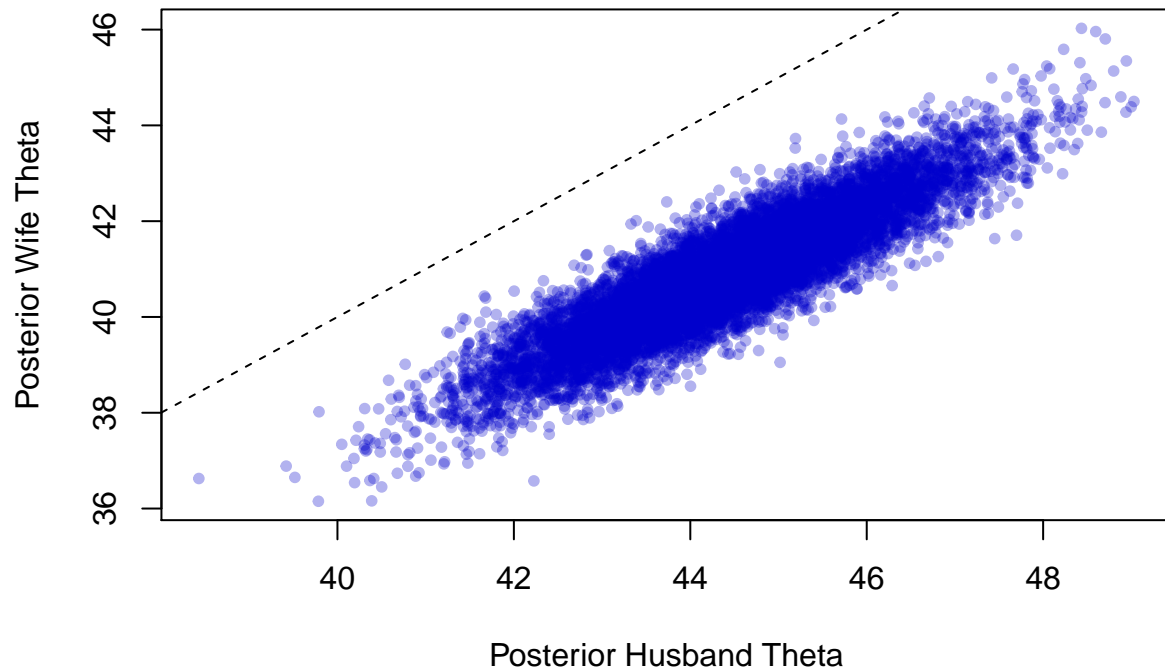
```

yBar = apply(Y,2,mean)
Size = 10000
T.mcmc = array(0, dim = c(Size,2))
Sigma.mcmc = array(0, dim = c(2,2,Size))
CurrentSigma = Sigma0
Lambda0.inverse = solve(Lambda0)
nu100 = nu0+n
CorrY = numeric(Size)
for(i in 1:Size){
  #Sample new theta
  SigmaPrevious.inverse = solve(CurrentSigma)
  Lambda100 = solve(Lambda0.inverse + n*SigmaPrevious.inverse)
  mu100 = Lambda100%*%(Lambda0.inverse%*%mu0 + n*SigmaPrevious.inverse%*%yBar)
  newTheta = mvrnorm(1, mu100, Lambda100)
  T.mcmc[i,] = newTheta
  #Get S_theta
  S_theta = matrix(0, nrow = 2, ncol = 2)
  S_theta = (t(Y)-newTheta)%*%t(t(Y)-newTheta)
  S100.inverse = solve(S0 + S_theta)
  newSigma = solve(rWishart(1,nu100, S100.inverse)[,1])
  Sigma.mcmc[,i] = newSigma
  CurrentSigma = newSigma
  SamplesY = mvrnorm(n, newTheta, newSigma)
  CorrY[i] = cor(SamplesY[,1], SamplesY[,2])
}

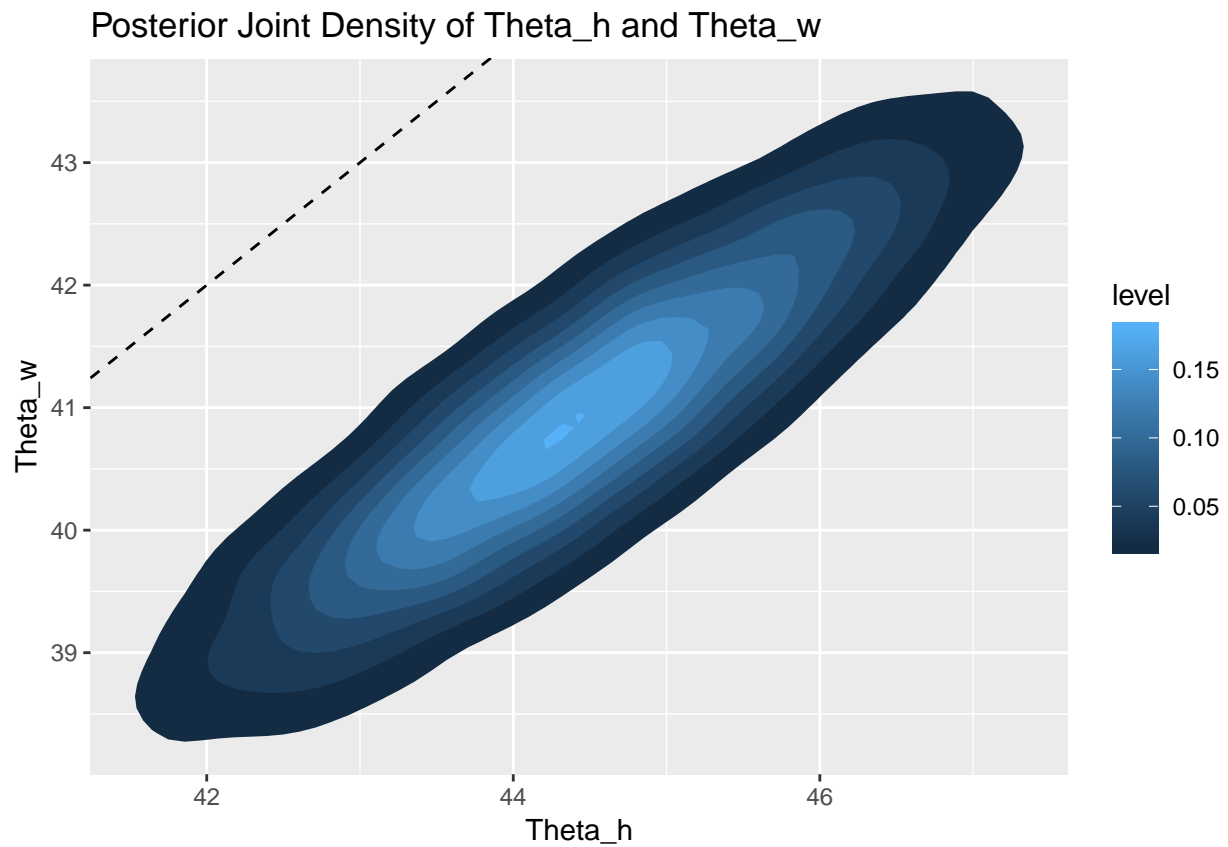
library(ggplot2)
T.mcmc = data.frame(T.mcmc)
names(T.mcmc) = c("H","W")
plot(T.mcmc$H, T.mcmc$W, xlab = "Posterior Husband Theta",
      ylab = "Posterior Wife Theta", main = "Posterior Samples of Theta",
      pch = 20, col = rgb(0,0,.8, .3),cex = 1)
abline(0, 1,lty = 2 )

```


Posterior Samples of Theta

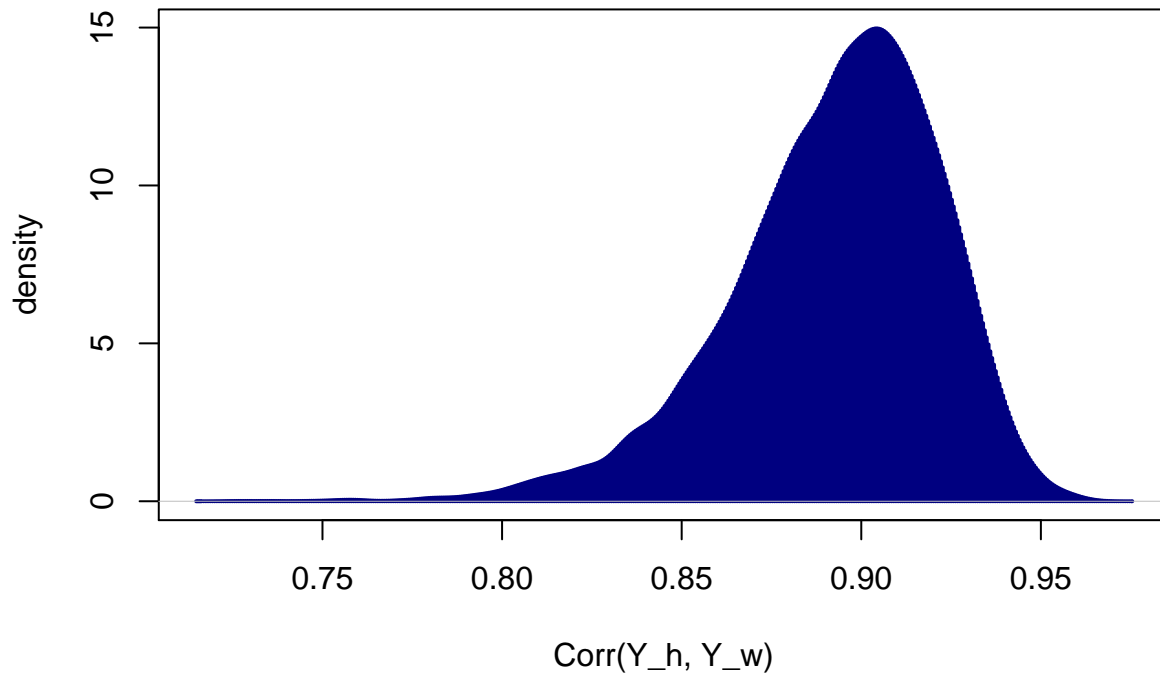


```
ggplot(T.mcmc, aes(x=H, y=W) ) +  
  stat_density_2d(aes(fill = ..level..), geom = "polygon")+  
  ggtitle("Posterior Joint Density of Theta_h and Theta_w")+  
  xlab("Theta_h")+  
  ylab("Theta_w")+  
  geom_abline(linetype = 2)
```



```
plot(density(CorrY), col = 'navy', lwd = 2, type = "h",  
     main = "Marginal Posterior Density of Corr(Y_h, Y_w)",  
     xlab = "Corr(Y_h, Y_w)", ylab = "density")
```

Marginal Posterior Density of $\text{Corr}(Y_h, Y_w)$



```
print("Theta_h 95% Posterior Confidence Interval")
```

```
## [1] "Theta_h 95% Posterior Confidence Interval"
```

```
quantile(T.mcmc[,1], prob = c(.025, .975))
```

```
##      2.5%      97.5%
```

```
## 41.77354 47.12468
```

```
print("Theta_w 95% Posterior Confidence Interval")
```

```
## [1] "Theta_w 95% Posterior Confidence Interval"
```

```
quantile(T.mcmc[,2], prob = c(.025, .975))
```

```
##      2.5%      97.5%
```

```
## 38.41173 43.39586
```

```
print("Corr(Y_h, Y_w) 95% Posterior Confidence Interval")
```

```
## [1] "Corr(Y_h, Y_w) 95% Posterior Confidence Interval"
```

```
quantile(CorrY, prob = c(.025, .975))
```

```
##      2.5%      97.5%
```

```
## 0.8251926 0.9385977
```

7.4d

Part i Jeffreys' Prior

Using a prior of $p_J(\theta, \Sigma) \propto |\Sigma|^{-(p+2)/2}$, this implies $p_J(\theta, \Sigma) \propto p_J(\theta)p_J(\Sigma)$ with $p_J(\Sigma) \propto |\Sigma|^{-(p+2)/2}$ and $p_J(\theta) \propto 1$. Thus,

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

$$\implies (\theta \mid \Sigma, y_{1:n})_J \sim MVN(\bar{y}, \Sigma/n)$$

$$p_J(\Sigma \mid \theta, y_{1:n}) \propto |\Sigma|^{-(n+1+p+1)/2} e^{-\frac{1}{2}tr(S_\theta \Sigma)} \implies (\Sigma \mid \theta, y_{1:n})_J \sim \text{InvWishart}(n+1, S_\theta^{-1})$$

```
Size = 10000
T.mcmc = array(0, dim = c(Size,2))
Sigma.mcmc = array(0, dim = c(2,2,Size))
n = length(Y[,1])
CorrY = numeric(Size)
Sigma = cov(Y)
for(i in 1:Size){
  #Sample new theta
  Theta = mvrnorm(1, yBar, Sigma/n)
  T.mcmc[i,] = Theta
  #Get S_theta
  S.theta = (t(Y)-Theta)%*%t(t(Y)-Theta)
  Sigma = solve(rWishart(1,n+1, solve(S.theta))[, ,1])
  Sigma.mcmc[, ,i] = Sigma
  SamplesY = mvrnorm(n, Theta, Sigma)
  CorrY[i] = cor(SamplesY[,1], SamplesY[,2])
}

print("Theta_h 95% Posterior Confidence Interval")

## [1] "Theta_h 95% Posterior Confidence Interval"
quantile(T.mcmc[,1], prob = c(.025, .975))

##      2.5%      97.5%
## 41.68734 47.11159

print("Theta_w 95% Posterior Confidence Interval")

## [1] "Theta_w 95% Posterior Confidence Interval"
quantile(T.mcmc[,2], prob = c(.025, .975))

##      2.5%      97.5%
## 38.35619 43.39855

print("Corr(Y_h, Y_w) 95% Posterior Confidence Interval")

## [1] "Corr(Y_h, Y_w) 95% Posterior Confidence Interval"
quantile(CorrY, prob = c(.025, .975))
```

```
##          2.5%          97.5%
## 0.8395811 0.9454284
```

Part ii Unit Information Prior

From 7.2b we know that $(\theta \mid \Sigma, y_{1:n}) \sim MVN(\bar{y}, \Sigma/(n+1))$

and $(\Sigma)_U \sim \text{Inv-Wishart}(p+1, (\sum_i^n (y_i - \bar{y})(y_i - \bar{y})^T/n)^{-1})$. Thus,

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

$$\propto |\Sigma|^{-n/2} e^{-\frac{1}{2} \text{tr}(\sum_i^n (y_i - \theta)(y_i - \theta)^T \Sigma^{-1})} |\Sigma|^{-1/2} e^{-\frac{1}{2} \text{tr}((\theta - \bar{y})(\theta - \bar{y})^T \Sigma^{-1})} |\Sigma|^{-(p+1+p+1)/2} e^{-\frac{1}{2} \text{tr}(\sum_i^n (y_i - \bar{y})(y_i - \bar{y})^T/n) \Sigma^{-1}}$$

And since

$$\sum_i^n (y_i - \theta)(y_i - \theta)^T = \sum_i^n (y_i - \bar{y})(y_i - \bar{y})^T + n(\bar{y} - \theta)(\bar{y} - \theta)^T,$$

$$p(\Sigma \mid \theta, y_{1:n})_U \propto |\Sigma|^{-(n+p+2+p+1)/2} e^{-\frac{1}{2} \text{tr}((n+1)A \Sigma^{-1})}$$

Where $A = \sum_i^n (y_i - \bar{y})(y_i - \bar{y})^T + (\bar{y} - \theta)(\bar{y} - \theta)^T$. Thus,

$$(\Sigma \mid \theta, y_{1:n})_U \sim \text{Inv-Wishart}(n+p+2, A^{-1}/(n+1))$$

```
Size = 10000
T.mcmc = array(0, dim = c(Size,2))
Sigma.mcmc = array(0, dim = c(2,2,Size))
n = length(Y[,1])
CorrY = numeric(Size)
Sigma = cov(Y)
S_y = (t(Y)-yBar)%*%t(t(Y)-yBar)/n
for(i in 1:Size){
  #Sample new theta
  Theta = mvrnorm(1, yBar, Sigma/(n+1))
  T.mcmc[i,] = Theta
  #Get S_theta
  S.ybar.theta = (yBar-Theta)%*%t(yBar-Theta)
  A = (S.ybar.theta+S_y)*(n+1)
  Sigma = solve(rWishart(1,n+1, solve(A))[, ,1])
  Sigma.mcmc[,i] = Sigma
  SamplesY = mvrnorm(n, Theta, Sigma)
  CorrY[i] = cor(SamplesY[,1], SamplesY[,2])
}

print("Theta_h 95% Posterior Confidence Interval")

## [1] "Theta_h 95% Posterior Confidence Interval"
quantile(T.mcmc[,1], prob = c(.025, .975))
```

```
##      2.5%      97.5%
## 41.73359 47.19448
print("Theta_w 95% Posterior Confidence Interval")

## [1] "Theta_w 95% Posterior Confidence Interval"
quantile(T.mcmc[,2], prob = c(.025, .975))

##      2.5%      97.5%
## 38.38512 43.49657
print("Corr(Y_h, Y_w) 95% Posterior Confidence Interval")

## [1] "Corr(Y_h, Y_w) 95% Posterior Confidence Interval"
quantile(CorrY, prob = c(.025, .975))

##      2.5%      97.5%
## 0.8395165 0.9441518
```

Part iii Diffuse Prior

```
Size = 10000
I = matrix(c(1,0,0,1), nrow = 2, byrow = T)
mu0 = c(0,0); Lambda0 = I*10**(5);
S0 = 1000*I; nu0 = 3
T.mcmc = array(0, dim = c(Size,2))
Sigma.mcmc = array(0, dim = c(2,2,Size))
Sigma = cov(Y)
Lambda0.inverse = solve(Lambda0)
nu100 = nu0+n

CorrY = numeric(Size)

for(i in 1:Size){
  #Sample new theta
  Sigma.inverse = solve(Sigma)
  Lambda100 = solve(Lambda0.inverse + n*Sigma.inverse)
  mu100 = Lambda100%*(Lambda0.inverse%*mu0 + n*Sigma.inverse%*yBar)
  Theta = mvrnorm(1, mu100, Lambda100)
  T.mcmc[i,] = Theta
  #Get S_theta
  S_theta = matrix(0, nrow = 2, ncol = 2)
  S_theta = (t(Y)-Theta)%*%t(t(Y)-Theta)
  S100.inverse = solve(S0 + S_theta)
  Sigma = solve(rWishart(1,nu100, S100.inverse)[,1])
  Sigma.mcmc[,i] = Sigma
  SamplesY = mvrnorm(n, Theta, Sigma)
  CorrY[i] = cor(SamplesY[,1], SamplesY[,2])
}

print("Theta_h 95% Posterior Confidence Interval")

## [1] "Theta_h 95% Posterior Confidence Interval"
```

```

quantile(T.mcmc[,1], prob = c(.025, .975))

##      2.5%      97.5%
## 41.72867 47.15722

print("Theta_w 95% Posterior Confidence Interval")

## [1] "Theta_w 95% Posterior Confidence Interval"
quantile(T.mcmc[,2], prob = c(.025, .975))

##      2.5%      97.5%
## 38.31736 43.47825

print("Corr(Y_h, Y_w) 95% Posterior Confidence Interval")

## [1] "Corr(Y_h, Y_w) 95% Posterior Confidence Interval"
quantile(CorrY, prob = c(.025, .975))

##      2.5%      97.5%
## 0.7624955 0.9162095

```

7.4e

All the confidence intervals for $\theta_h, \theta_w, Corr$ seem to convey the same amount of information irregardless of what prior was used for θ and Σ , with the exception of the confidence interval obtained from using the diffuse prior (7.4iii) for the correlation, which is shifted downward by about 0.06. Since the posterior information obtained from my prior doesn't differ from the information obtained (i.e. the confidence intervals) obtained from the other prior (with the exception of the diffuse prior) I don't have a preference for any of the priors. This similarity in the posterior distributions may be due to the fact that I had 100 samples to condition my posterior distribution on, which is much larger than the "prior weights" I used which were between 3 and 10. If the sample size were much smaller, I beleive the priors may have yielded different results since the prior information used would carry a larger relative weight towards the posterior distribution and thus have more influence.

Math Problem

Given that

$$P(Y_b | Y_a) \propto P(Y_a, Y_b) \propto \exp\left((Y - \theta)^T \Sigma^{-1} (Y - \theta)\right) = \exp\left(-\frac{1}{2} \begin{bmatrix} Y_a - \theta_a \\ Y_b - \theta_b \end{bmatrix}^T \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} \begin{bmatrix} Y_a - \theta_a \\ Y_b - \theta_b \end{bmatrix}\right)$$

given that the inverse of matrix Σ is defined as

$$\begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix}^{-1} = \frac{\begin{bmatrix} \Sigma_{bb} & -\Sigma_{ab} \\ -\Sigma_{ba} & \Sigma_{aa} \end{bmatrix}}{\Sigma_{aa}\Sigma_{bb} - \Sigma_{ba}\Sigma_{ab}}$$

Thus,

$$\begin{aligned}
P(Y_b | Y_a) &\propto \exp\left(-\frac{1}{2} \frac{\begin{bmatrix} Y_a - \theta_a \\ Y_b - \theta_b \end{bmatrix}^T \begin{bmatrix} \Sigma_{bb} & -\Sigma_{ab} \\ -\Sigma_{ba} & \Sigma_{aa} \end{bmatrix} \begin{bmatrix} Y_a - \theta_a \\ Y_b - \theta_b \end{bmatrix}}{\Sigma_{aa}\Sigma_{bb} - \Sigma_{ba}\Sigma_{ab}}\right) \\
&\propto \exp\left(-\frac{1}{2} \frac{\begin{bmatrix} (Y_a - \theta_a)\Sigma_{bb} - (Y_b - \theta_b)\Sigma_{ab} \\ (Y_b - \theta_b)\Sigma_{aa} - (Y_a - \theta_a)\Sigma_{ab} \end{bmatrix}^T \begin{bmatrix} Y_a - \theta_a \\ Y_b - \theta_b \end{bmatrix}}{\Sigma_{aa}\Sigma_{bb} - \Sigma_{ba}\Sigma_{ab}}\right) \\
&\propto \exp\left(-\frac{1}{2} \frac{(Y_a - \theta_a)\Sigma_{bb}(Y_a - \theta_a) - (Y_b - \theta_b)\Sigma_{ab}(Y_a - \theta_a) + (Y_b - \theta_b)\Sigma_{aa}(Y_b - \theta_b) - (Y_a - \theta_a)\Sigma_{ab}(Y_b - \theta_b)}{\Sigma_{aa}\Sigma_{bb} - \Sigma_{ba}\Sigma_{ab}}\right) \\
&\propto \exp\left(-\frac{1}{2} \frac{(Y_a - \theta_a)^2\Sigma_{bb} + (Y_b - \theta_b)^2\Sigma_{aa} - (Y_b - \theta_b)\Sigma_{ab}(Y_a - \theta_a) - (Y_b - \theta_b)\Sigma_{ab}(Y_a - \theta_a)}{\Sigma_{aa}\Sigma_{bb} - \Sigma_{ba}\Sigma_{ab}}\right) \\
&\propto \exp\left(-\frac{1}{2} \frac{(Y_b - \theta_b)^2\Sigma_{aa} - (Y_b - \theta_b)\left(2\Sigma_{ab}(Y_a - \theta_a)\right)}{\Sigma_{aa}\Sigma_{bb} - \Sigma_{ba}\Sigma_{ab}}\right) \\
&\propto \exp\left(-\frac{\Sigma_{aa}}{2(\Sigma_{aa}\Sigma_{bb} - \Sigma_{ba}\Sigma_{ab})} \left(\frac{(Y_b - \theta_b)^2 - (Y_b - \theta_b)\left(2\Sigma_{ab}(Y_a - \theta_a)\Sigma_{aa}^{-1}\right)}{1}\right)\right) \\
&\propto \exp\left(-\frac{1}{2(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})} \left(\frac{(Y_b - \theta_b)^2 - 2(Y_b - \theta_b)\left(\Sigma_{ab}\Sigma_{aa}^{-1}(Y_a - \theta_a)\right)}{1}\right)\right) \\
&\propto \exp\left(-\frac{1}{2(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})} \left(Y_b - \theta_b - \Sigma_{ab}\Sigma_{aa}^{-1}(Y_a - \theta_a)\right)^2\right)
\end{aligned}$$

Which implies $Y_b | Y_a$ is a *MVN* with covariance matrix $\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}$ and mean $\theta_b + \Sigma_{ab}\Sigma_{aa}^{-1}(Y_a - \theta_a)$