

STA 601 Homework 7

Lingyun Shao

Oct. 30, 2018

7.2

Unit information prior: Letting $\Psi = \Sigma^{-1}$, show that a unit information prior for (θ, Ψ) is given by $\theta|\Psi \sim \text{multivariate normal}(\bar{y}, \Psi^{-1})$ and $\Psi \sim \text{Wishart}(p+1, S^{-1})$, where $S = \sum (y_i - \bar{y})(y_i - \bar{y})^T/n$. This can be done by mimicking the procedure outlined in Exercise 5.6 as follows:

- a) Reparameterize the multivariate normal model in terms of the precision matrix $\Psi = \Sigma^{-1}$. Write out the resulting log likelihood, and find a probability density $p_U(\theta, \Psi) = p_U(\theta|\Psi)p_U(\Psi)$ such that $\log p(\theta, \Psi) = l(\theta, \Psi|\mathbf{Y})/n + c$, where c does not depend on θ or Ψ .

Hint: Write $(y_i - \theta)$ as $(y_i - \bar{y} + \bar{y} - \theta)$, and note that $\sum a_i^T B a_i$ can be written as $\text{tr}(AB)$, where $A = \sum a_i a_i^T$.

Say we have samples y_1, \dots, y_n and each y_i is a p by 1 matrix. The log-likelihood of a multivariate normal can be easily derived as below:

$$\begin{aligned} l(\theta, \Psi|\mathbf{Y}) &= \log p(\mathbf{Y}|\theta, \Psi) \\ &= \sum \log p(y_i|\theta, \Psi) \\ &= \sum \log[(2\pi)^{-p/2} |\Psi|^{1/2} \exp(-\frac{1}{2}(y_i - \theta)^T \Psi (y_i - \theta))] \\ &= \frac{n}{2} \log(|\Psi|) - \frac{1}{2} \sum (y_i - \theta)^T \Psi (y_i - \theta) + c \end{aligned}$$

where the multivariate normal model is reparameterized by using $\Psi = \Sigma^{-1}$ instead of Σ in the density.

Notice that in the log-likelihood, the term $(y_i - \theta)^T \Psi (y_i - \theta)$ is actually a 1 by 1 matrix, i.e. a scalar. So we can take the trace of it and it's still the same. Thus we have

$$\begin{aligned} (y_i - \theta)^T \Psi (y_i - \theta) &= \text{tr}((y_i - \theta)^T \Psi (y_i - \theta)) \\ &= \underbrace{\text{tr}((y_i - \theta)(y_i - \theta)^T \Psi)}_{\text{tr}(AB)=\text{tr}(BA) \text{ if dimension matches}} \\ &= \text{tr} \{ (y_i - \bar{y} + \bar{y} - \theta)(y_i - \bar{y} + \bar{y} - \theta)^T \Psi \} \\ &= \text{tr} \{ ((y_i - \bar{y})(y_i - \bar{y})^T + (\bar{y} - \theta)(\bar{y} - \theta)^T + (y_i - \bar{y})(\bar{y} - \theta)^T + (\bar{y} - \theta)(y_i - \bar{y})^T) \Psi \} \end{aligned}$$

Therefore, the log-likelihood can be written as

$$\begin{aligned}
l(\theta, \Psi | \mathbf{Y}) &= \frac{n}{2} \log(|\Psi|) - \frac{1}{2} \sum (y_i - \theta)^T \Psi (y_i - \theta) + c \\
&= \frac{n}{2} \log(|\Psi|) - \frac{1}{2} \sum \underbrace{tr \{ ((y_i - \bar{y})(y_i - \bar{y})^T + (\bar{y} - \theta)(\bar{y} - \theta)^T + (y_i - \bar{y})(\bar{y} - \theta)^T + (\bar{y} - \theta)(y_i - \bar{y})^T) \Psi \}}_{\text{notice that } tr(A+B)=tr(A)+tr(B)} + c \\
&= \frac{n}{2} \log(|\Psi|) - \frac{1}{2} \left\{ tr \left(\sum (y_i - \bar{y})(y_i - \bar{y})^T \Psi \right) + tr \left(\sum (\bar{y} - \theta)(\bar{y} - \theta)^T \Psi \right) \right\} + \\
&\quad \left\{ tr \left(\underbrace{\sum (y_i - \bar{y})(\bar{y} - \theta)^T \Psi}_0 + tr \left((\bar{y} - \theta) \underbrace{\sum (y_i - \bar{y})^T}_{\mathbf{0}^T} \Psi \right) \right\} + c \\
&= \frac{n}{2} \log(|\Psi|) - \frac{1}{2} \left\{ tr \left(\sum (y_i - \bar{y})(y_i - \bar{y})^T \Psi \right) + tr(n(\bar{y} - \theta)(\bar{y} - \theta)^T \Psi) \right\} + c
\end{aligned}$$

So we have $l(\theta, \Psi | \mathbf{Y})/n = \frac{1}{2} \log(|\Psi|) - \frac{1}{2} \{ tr(\mathbf{S}\Psi) + tr((\bar{y} - \theta)(\bar{y} - \theta)^T \Psi) \} + c$, where $\mathbf{S} = \frac{1}{n} \sum (y_i - \bar{y})(y_i - \bar{y})^T$.

Exponentiate $l(\theta, \Psi | \mathbf{Y})/n$, we will get the kernel of the Unit Information Prior we want.

$$\begin{aligned}
p(\theta, \Psi) &= \exp(\log p(\theta, \Psi)) \\
&= \exp(l(\theta, \Psi | \mathbf{Y})/n + c) \\
&\propto \exp\left(\frac{1}{2} \log(|\Psi|) - \frac{1}{2} \{ tr(\mathbf{S}\Psi) + tr((\bar{y} - \theta)(\bar{y} - \theta)^T \Psi) \}\right) \\
&= |\Psi|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (tr(\mathbf{S}\Psi) + tr((\bar{y} - \theta)(\bar{y} - \theta)^T \Psi))\right) \\
&= |\Psi|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (tr(\mathbf{S}\Psi) + (\bar{y} - \theta)^T \Psi (\bar{y} - \theta))\right) \\
&= |\Psi|^{\frac{p+1-p-1}{2}} \exp\left(-\frac{1}{2} tr(\mathbf{S}\Psi)\right) \times |\Psi|^{\frac{1}{2}} \exp\left(-\frac{1}{2} (\theta - \bar{y})^T \Psi (\theta - \bar{y})\right) \\
&\propto p(\Psi) p(\theta | \Psi)
\end{aligned}$$

From the form of the terms, we can see that the first term is the kernel of a Wishart($p+1, \mathbf{S}^{-1}$) and the second term is a multivariate normal(\bar{y}, Ψ^{-1}).

Thus we may have $p_U(\theta, \Psi) = p_U(\theta | \Psi) p_U(\Psi)$, where $p_U(\Psi)$ is the p.d.f. of Wishart($p+1, \mathbf{S}^{-1}$) and $p_U(\theta | \Psi)$ is the p.d.f. of $N_p(\bar{y}, \Psi^{-1})$, thus showing that a unit information prior for (θ, Ψ) is given by $\Psi \sim Wishart(p+1, \mathbf{S}^{-1})$ and $\theta | \Psi \sim N_p(\bar{y}, \Psi^{-1})$.

- b) Let $p_U(\Sigma)$ be the inverse-Wishart density induced by $p_U(\Psi)$. Obtain a density $p_U(\theta, \Sigma | y_1, \dots, y_n) \propto p_U(\theta | \Sigma) p_U(\Sigma) p(y_1, \dots, y_n | \theta, \Sigma)$. Can this be interpreted as a posterior distribution for θ and Σ ?

Since we have $\Psi \sim Wishart(p+1, \mathbf{S}^{-1})$, we know $\Sigma = \Psi^{-1} \sim InvWishart(p+1, \mathbf{S}^{-1})$.

$$\begin{aligned}
p_U(\theta, \Sigma | y_1, \dots, y_n) &\propto p_U(\theta | \Sigma) p_U(\Sigma) p(y_1, \dots, y_n | \theta, \Sigma) \\
&\propto |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta - \bar{y})^T \Sigma^{-1}(\theta - \bar{y})\right) \\
&\quad \underbrace{\hspace{10em}}_{\text{kernel of } N_p(\bar{y}, \Sigma)} \\
&\times |\Sigma|^{-\frac{p+1+p+1}{2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{S}\Sigma^{-1})\right) \\
&\quad \underbrace{\hspace{10em}}_{\text{kernel of inverse Wishart}(p+1, \mathbf{S})} \\
&\times |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum (y_i - \theta)^T \Sigma^{-1}(y_i - \theta)\right) \\
&= |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta - \bar{y})^T \Sigma^{-1}(\theta - \bar{y})\right) \\
&\times |\Sigma|^{-\frac{p+1+p+1}{2}} \exp\left(-\frac{1}{2}\text{tr}(\mathbf{S}\Sigma^{-1})\right) \\
&\times \underbrace{|\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2} \sum (y_i - \bar{y})^T \Sigma^{-1}(y_i - \bar{y})\right) \times \exp\left(-\frac{1}{2}n(\theta - \bar{y})^T \Sigma^{-1}(\theta - \bar{y})\right)}_{\text{verified in part (a)}} \\
&= |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta - \bar{y})^T [(n+1)\Sigma^{-1}](\theta - \bar{y})\right) \\
&\times |\Sigma|^{-\frac{(n+p+1)+p+1}{2}} \exp\left(-\frac{1}{2}\text{tr}([(n+1)\mathbf{S}]\Sigma^{-1})\right)
\end{aligned}$$

From the derivation above, we can see that there are both kernel of $InvWishart(n+p+1, (n+1)\mathbf{S}^{-1})$ and kernel of $N_p(\bar{y}, \Sigma/(n+1))$ in the posterior. So we know that posterior of $\theta, \Sigma | y_1, \dots, y_n$, given a Unit Information prior follows a Normal-inverse-Wishart distribution, that is $(\theta, \Sigma | y_1, \dots, y_n) \sim NIW(\bar{y}, n+1, \mathbf{S}, n+p+1)$.

For the prior:

The prior here is somewhat using the information of data. UI priors weakly concentrate around a data-based estimator. Such a prior leads to “re-use” of the information in your sample.

However, we can still interpret $p_U(\theta, \Sigma | y_1, \dots, y_n)$ as the posterior distribution for θ and Σ , because it can be derived from a prior belief (although based on data) and our sampling model using $p_U(\theta, \Sigma | y_1, \dots, y_n) \propto p_U(\theta, \Sigma) p(y_1, \dots, y_n | \theta, \Sigma)$, which also has some interpretation that a posterior has.

7.4

Marriage data: The file `agehw.dat` contains data on the ages of 100 married couples sampled from the U.S. population.

- a) Before you look at the data, use your own knowledge to formulate a semiconjugate prior distribution for $\boldsymbol{\theta} = (\theta_h, \theta_w)^T$ and Σ , where θ_h, θ_w are mean husband and wife ages, and Σ is the covariance matrix.

Based on my knowledge, I think both means of ages should be normally distributed. The mean of husband ages, θ_h , should have a prior mean of 40, which is often referred to as middle age. And the mean of wife ages, θ_w , should be 2 years younger (it's so as least for my parents), so the prior mean for θ_w is 38. And the θ_h and θ_w should be highly correlated and the variance of θ_h, θ_w can be the same. I choose the variance to be 100 so that 95% of ages are within (20,60) for husbands and (18,58) for wives.

Thus, we have the prior

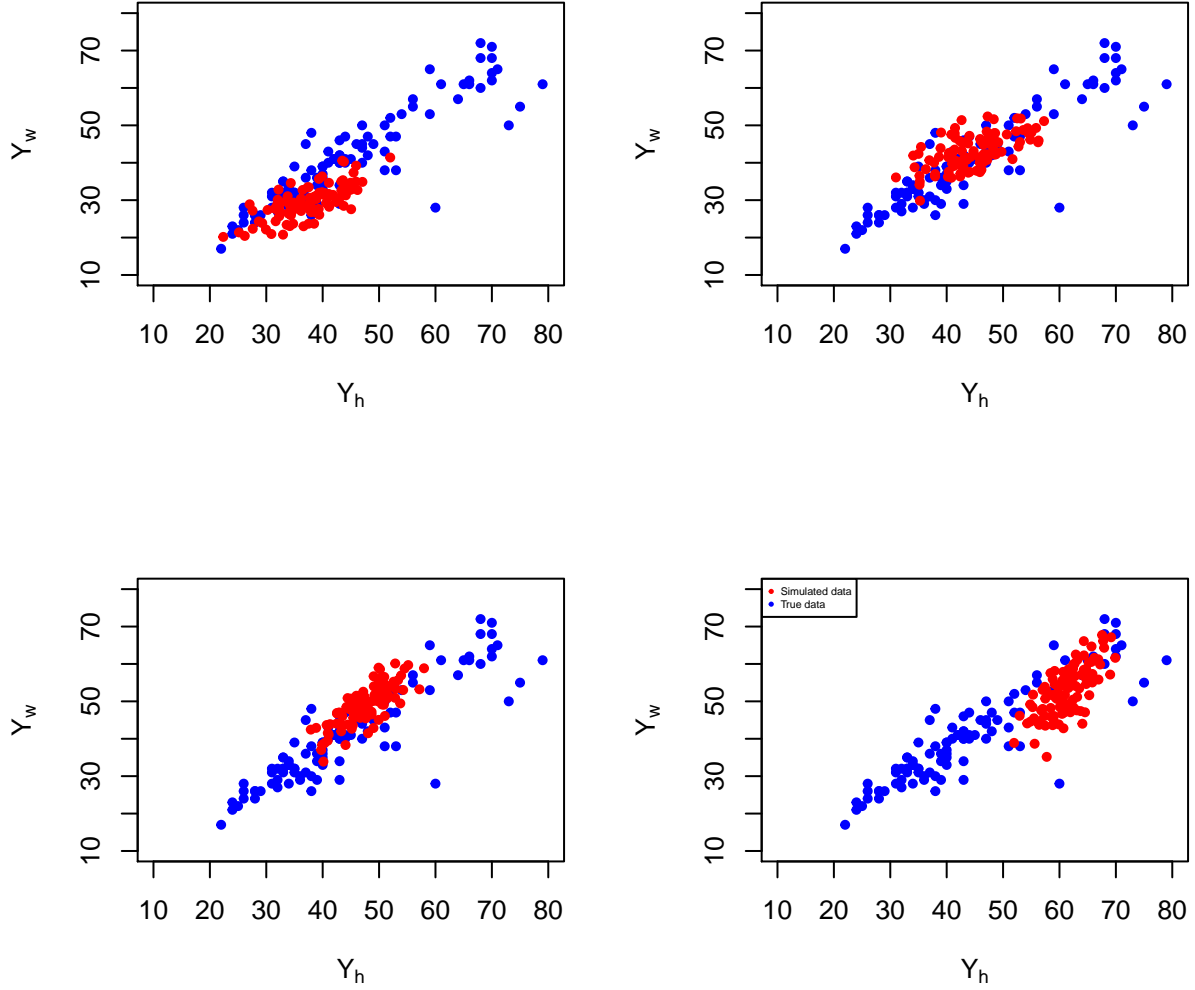
$$\boldsymbol{\theta} = (\theta_h, \theta_w)^T \sim N(\boldsymbol{\mu}_0, \Lambda_0) = N((45, 42)^T, \begin{pmatrix} 100 & 60 \\ 60 & 100 \end{pmatrix})$$

As for Σ , a reasonable prior distribution might be Inverse Wishart distribution. I let the prior parameter $S_0 = \Lambda_0$ and $\nu_0 = 5$ in the prior, so we now have

$$\Sigma \sim \text{InvWishart}(\nu_0, S_0^{-1}) = \text{InvWishart}(5, \begin{pmatrix} 100 & 60 \\ 60 & 100 \end{pmatrix})$$

- b) Generate a *prior predictive dataset* of size $n = 100$, by sampling $(\boldsymbol{\theta}, \Sigma)$ from your prior distribution and then simulating $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim i.i.d.$ multivariate normal $(\boldsymbol{\theta}, \Sigma)$. Generate several such datasets, make bivariate scatterplots for each dataset, and make sure they roughly represent your prior beliefs about what such a dataset would actually look like. If your prior predictive datasets do not conform to your beliefs, go back to part a) and formulate a new prior. Report the prior that you eventually decide upon, and provide scatterplots for at least three prior predictive datasets.

```
Y = read.table(file = 'http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/agehw.dat',
header = TRUE)
mu0 = c(45,42)
Lmd0 = matrix(c(100,60,60,100),nrow=2)
nu0 = 5
S0 = Lmd0
nsim = 4
set.seed(1000)
Yhs = matrix(NA, nrow = nrow(Y), ncol = nsim)
Yws = matrix(NA, nrow = nrow(Y), ncol = nsim)
for(i in 1:nsim) {
  theta = rmvnorm(1, mu0, Lmd0)
  Sigma = solve(rWishart(1, nu0, solve(S0))[, , 1])
  Y.sim = rmvnorm(nrow(Y), theta, Sigma)
  Yhs[, i] = Y.sim[, 1]
  Yws[, i] = Y.sim[, 2]
}
par(mfrow = c(2, 2))
for (i in 1:4) {
  plot(Y[, 1], Y[, 2], ylim = c(10,80), xlim = c(10,80), xlab = expression(Y[h]),
       ylab = expression(Y[w]), pch = 20, col = 'blue')
  points(Yhs[, i], Yws[, i], pch = 20, col = 'red')
}
legend("topleft", pch = 20, col = c(2,4),
       legend = c("Simulated data", "True data"), cex = 0.4)
```



As presented in the plots, the simulated data do conform to my prior belief that the ages of couples should be highly correlated. Also by comparing to the true data, we can see the prior belief somewhat matches the structure of the data.

- c) Using your prior distribution and the 100 values in the dataset, obtain an MCMC approximation to $p(\theta, \Sigma | \mathbf{y}_1, \dots, \mathbf{y}_{100})$. Plot the joint posterior distribution of θ_h and θ_w , and also the marginal posterior density of the correlation between Y_h and Y_w , the ages of a husband and wife. Obtain 95% posterior confidence intervals for θ_h , θ_w and the correlation coefficient.

Given the prior $\theta \sim N(\mu_0, \Lambda_0)$, $\Sigma \sim \text{InvWishart}(\nu_0, S_0^{-1})$, we can easily derive the full conditional of each parameter:

$$p(\theta | Y, \Sigma) \propto p(Y | \theta, \Sigma) p(\theta) \\ \implies \theta | Y, \Sigma \sim \text{MVN}(A_n^{-1} b_n, A_n^{-1})$$

where $A_n = \Lambda_0^{-1} + n\Sigma^{-1}$, $b_n = \Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\bar{y}$

$$p(\Sigma|Y, \theta) \propto p(Y|\theta, \Sigma)p(\Sigma) \\ \implies \Sigma|Y, \theta \sim \text{InvWishart}(\nu_0 + n, (S_0 + S_\theta)^{-1})$$

where $S_\theta = \sum (y_i - \theta)(y_i - \theta)^T$.

Based on these two full conditional distribution, we can define our functions to update parameters in the Gibbs sampler as the following:

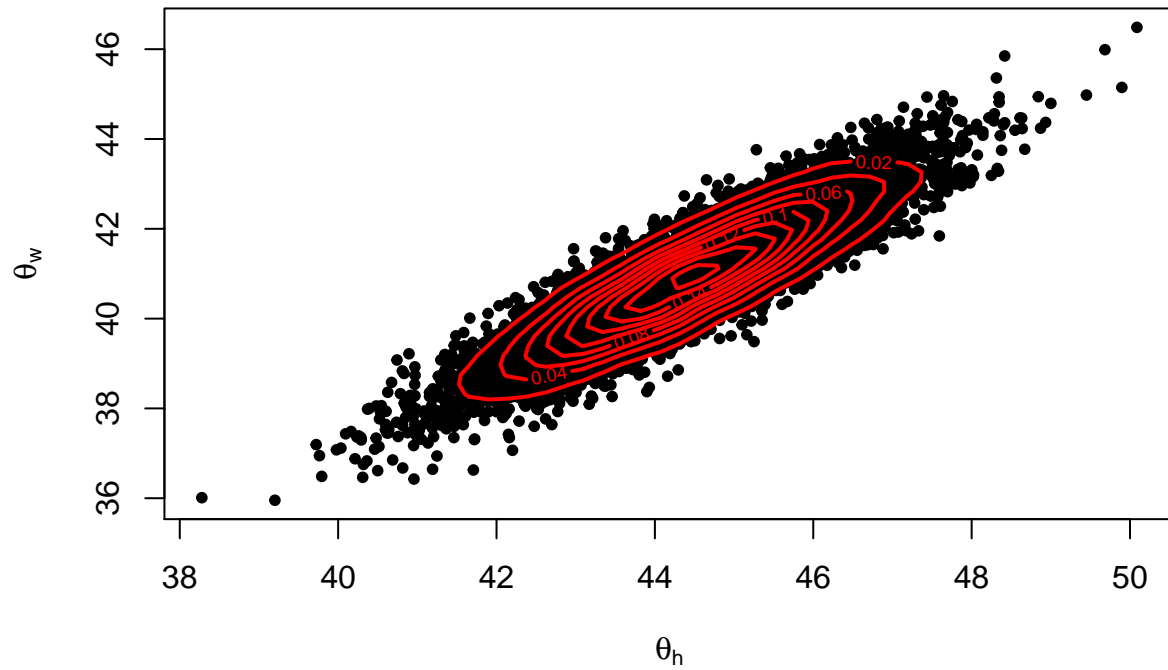
```
update_theta = function(Y, Sigma, mu0, Lmd0) {
  n = nrow(Y)
  y.bar = colMeans(Y)
  theta.var = solve(solve(Lmd0) + n*solve(Sigma))
  theta.mean = theta.var%*(solve(Lmd0)%*mu0 + n*solve(Sigma)%*y.bar)
  return(rmvnorm(1, theta.mean, theta.var))
}

update_Sigma = function(Y, theta, S0, nu0) {
  n = nrow(Y)
  nun = nu0 + n
  Sn = S0 + (t(Y) - c(theta))%*t(t(Y) - c(theta))
  return(solve(rWishart(1, nun, solve(Sn))[, , 1]))
}

# Initial values
theta = colMeans(Y)
Sigma = cov(Y)
Nmc = 10000
theta.mc = NULL
Sigma.mc = NULL
for(i in 1:Nmc) {
  theta = update_theta(Y, Sigma, mu0, Lmd0)
  Sigma = update_Sigma(Y, theta, S0, nu0)
  theta.mc = rbind(theta.mc, theta)
  Sigma.mc = rbind(Sigma.mc, as.vector(Sigma))
}

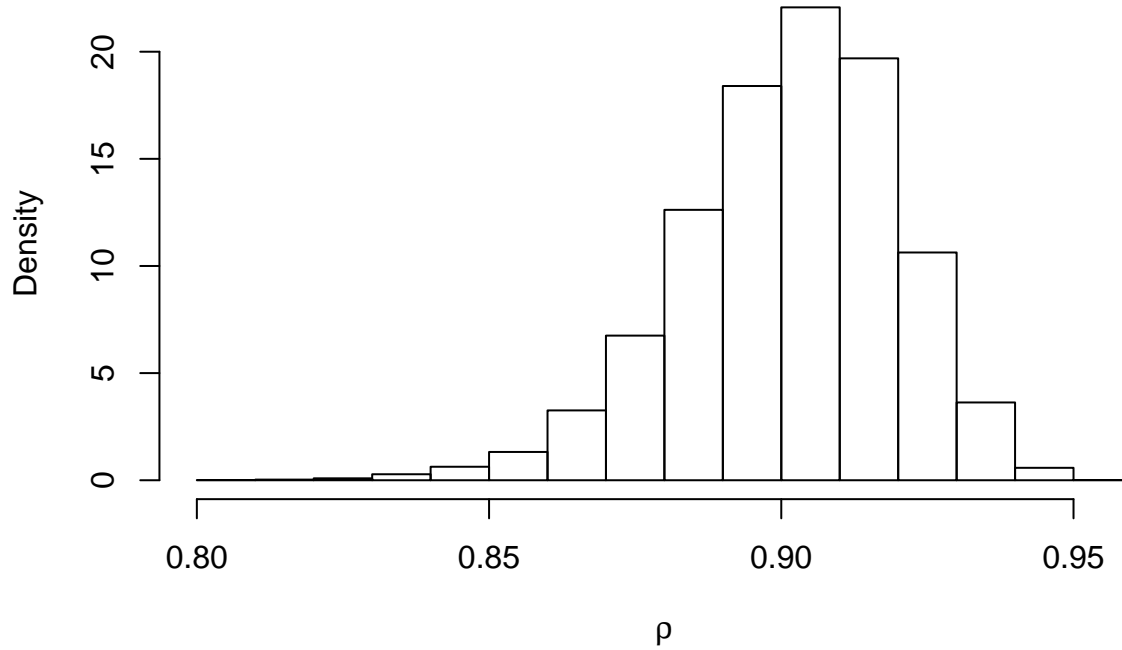
theta.mc.kde = MASS::kde2d(theta.mc[,1], theta.mc[,2], n = 50)
x = theta.mc.kde$x
y = theta.mc.kde$y
z = theta.mc.kde$z
plot(theta.mc[, 1], theta.mc[, 2], xlab = expression(theta[h]),
      ylab = expression(theta[w]), pch = 20,
      main = 'Joint Posterior Distribution')
contour(x,y,z, add=TRUE, col = 2, lwd = 2)
```

Joint Posterior Distribution



```
rho.mc = Sigma.mc[, 2]/sqrt(Sigma.mc[, 1]*Sigma.mc[, 4])  
hist(rho.mc, xlab = expression(rho), freq = FALSE,  
     main = 'Marginal Posterior  
Density of the Correlation')
```

Marginal Posterior Density of the Correlation



```
CI = rbind(quantile(theta.mc[, 1], c(0.025, 0.975)),
           quantile(theta.mc[, 2], c(0.025, 0.975)),
           quantile(rho.mc, c(0.025, 0.975)))
row.names(CI) = c("$\\theta_h$", "$\\theta_w$", "$\\rho$")
kable(CI, digits=3, escape = FALSE, caption = 'My Prior in (a)')
```

Table 1: My Prior in (a)

	2.5%	97.5%
θ_h	41.819	47.075
θ_w	38.415	43.339
ρ	0.861	0.933

d) Obtain 95% posterior confidence intervals for θ_h , θ_w and the correlation coefficient using the following prior distributions:

i. Jeffreys' prior, described in Exercise 7.1;

According to 7.1 Jefferys' prior is given as $p_J(\theta, \Sigma) \propto |\Sigma|^{-(p+2)/2}$, so if we treat θ, Σ as independent, we have $P_J(\theta) \propto 1$, $P_J(\Sigma) \propto |\Sigma|^{-(p+2)/2}$. Then the full conditionals can be written as

$$p(\theta|Y, \Sigma) \propto p(Y|\theta, \Sigma)p(\theta)$$

$$\implies \theta|Y, \Sigma \sim MVN(\bar{y}, \Sigma/n)$$

$$p(\Sigma|Y, \theta) \propto p(Y|\theta, \Sigma)p(\Sigma)$$

$$\implies \Sigma|Y, \theta \sim \text{InvWishart}(n+1, S_\theta^{-1})$$

where $S_\theta = \sum (y_i - \theta)(y_i - \theta)^T$.

Based on these two full conditional distribution, we can define our functions to update parameters in the Gibbs sampler as the following:

```
update_theta.1 = function(Y, Sigma) {
  n = nrow(Y)
  y.bar = colMeans(Y)
  return(rmvnorm(1, y.bar, Sigma/n))
}

update_Sigma.1 = function(Y, theta) {
  n = nrow(Y)
  S_th = (t(Y) - c(theta))%*%t(t(Y) - c(theta))
  return(solve(rWishart(1, n+1, solve(S_th))[, , 1]))
}

n = nrow(Y)
p = ncol(Y)
# Initial values
theta = colMeans(Y)
Sigma = cov(Y)
Nmc = 10000
theta.mc.1 = NULL
Sigma.mc.1 = NULL
for(i in 1:Nmc) {
  theta = update_theta.1(Y, Sigma)
  Sigma = update_Sigma.1(Y, theta)
  theta.mc.1 = rbind(theta.mc.1, theta)
  Sigma.mc.1 = rbind(Sigma.mc.1, as.vector(Sigma))
}
rho.mc.1 = Sigma.mc.1[, 2]/sqrt(Sigma.mc.1[, 1]*Sigma.mc.1[, 4])
CI.1 = rbind(quantile(theta.mc.1[, 1], c(0.025, 0.975)),
             quantile(theta.mc.1[, 2], c(0.025, 0.975)),
             quantile(rho.mc.1, c(0.025, 0.975)))
row.names(CI.1) = c("$\\theta_h$", "$\\theta_w$", "$\\rho$")
kable(CI.1, digits=3, escape = FALSE, caption = 'Jeffery's Prior')
```

Table 2: Jeffery's Prior

	2.5%	97.5%
θ_h	41.774	47.124
θ_w	38.406	43.452
ρ	0.862	0.935

ii. the unit information prior, described in Exercise 7.2;

According to 7.2 the unit information prior is given as $\Sigma \sim \text{InvWishart}(p+1, \mathbf{S}^{-1}), \theta|\Sigma \sim \text{MVN}(\bar{y}, \Sigma)$. Then the full conditionals can be written as

$$p(\theta|Y, \Sigma) \propto p(Y|\theta, \Sigma)p(\theta|\Sigma) \\ \Rightarrow \theta|Y, \Sigma \sim MVN(\bar{y}, \Sigma/(n+1))$$

$$p(\Sigma|Y, \theta) \propto p(Y|\theta, \Sigma)p(\Sigma) \\ \Rightarrow \Sigma|Y, \theta \sim InvWishart(n+p+1, \mathbf{S}^{-1}/(n+1))$$

where $p = 2$, $\mathbf{S} = \sum (y_i - \bar{y})(y_i - \bar{y})^T/n$.

Based on these two full conditional distribution, we can define our functions to update parameters in the Gibbs sampler as the following:

```
update_theta.2 = function(Y, Sigma) {
  n = nrow(Y)
  y.bar = colMeans(Y)
  return(rmvnorm(1, y.bar, Sigma/(n+1)))
}

update_Sigma.2 = function(Y, theta) {
  n = nrow(Y)
  p = ncol(Y)
  y.bar = colMeans(Y)
  S = (t(Y) - y.bar)%*%t(t(Y) - y.bar)/n
  return(solve(rWishart(1, n+p+1, solve(S)/(n+1))[, , 1]))
}

n = nrow(Y)
p = ncol(Y)
# Initial values
theta = colMeans(Y)
Sigma = cov(Y)
Nmc = 10000
theta.mc.2 = NULL
Sigma.mc.2 = NULL
for(i in 1:Nmc) {
  theta = update_theta.2(Y, Sigma)
  Sigma = update_Sigma.2(Y, theta)
  theta.mc.2 = rbind(theta.mc.2, theta)
  Sigma.mc.2 = rbind(Sigma.mc.2, as.vector(Sigma))
}
rho.mc.2 = Sigma.mc.2[, 2]/sqrt(Sigma.mc.2[, 1]*Sigma.mc.2[, 4])
CI.2 = rbind(quantile(theta.mc.2[, 1], c(0.025, 0.975)),
             quantile(theta.mc.2[, 2], c(0.025, 0.975)),
             quantile(rho.mc.2, c(0.025, 0.975)))
row.names(CI.2) = c("$\\theta_h$", "$\\theta_w$", "$\\rho$")
kable(CI.2, digits=3, escape = FALSE, caption = 'UI Prior')
```

Table 3: UI Prior

	2.5%	97.5%
θ_h	41.788	47.040
θ_w	38.442	43.342
ρ	0.861	0.935

iii. a “diffuse prior” with $\mu_0 = 0$, $\Lambda_0 = 10^5 \times \mathbf{I}$, $S_0 = 1000 \times \mathbf{I}$ and $\nu_0 = 3$.

For this type of priors, we can just use the update functions defined in c).

```
n = nrow(Y)
p = ncol(Y)
mu0 = rep(0, 2)
S0 = 1000*diag(2)
nu0 = 3
Lmd0 = 10^5*diag(2)
theta = colMeans(Y)
Sigma = cov(Y)
Nmc = 10000
theta.mc.3 = NULL
Sigma.mc.3 = NULL
for(i in 1:Nmc) {
  theta = update_theta(Y, Sigma, mu0, Lmd0)
  Sigma = update_Sigma(Y, theta, S0, nu0)
  theta.mc.3 = rbind(theta.mc.3, theta)
  Sigma.mc.3 = rbind(Sigma.mc.3, as.vector(Sigma))
}
rho.mc.3 = Sigma.mc.3[, 2]/sqrt(Sigma.mc.3[, 1]*Sigma.mc.3[, 4])
CI.3 = rbind(quantile(theta.mc.3[, 1], c(0.025, 0.975)),
             quantile(theta.mc.3[, 2], c(0.025, 0.975)),
             quantile(rho.mc.3, c(0.025, 0.975)))
row.names(CI.3) = c("$\\theta_h$", "$\\theta_w$", "$\\rho$")
kable(CI.3, digits=3, escape = FALSE, caption = 'Diffuse Prior')
```

Table 4: Diffuse Prior

	2.5%	97.5%
θ_h	41.686	47.143
θ_w	38.302	43.459
ρ	0.794	0.900

e) Compare the confidence intervals from d) to those obtained in c). Discuss whether or not you think that your prior information is helpful in estimating θ and Σ , or if you think one of the alternatives in d) is preferable. What about if the sample size were much smaller, say $n = 25$?

```
cbind(CI, CI.1, CI.2, CI.3) %>%
kable(digits=3, escape = FALSE, caption = '(1) My Prior\\t(2) Jeffery\\s Prior\\t(3) UI Prior\\t(4) Dif
```

Table 5: (1) My Prior(2) Jeffery’s Prior(3) UI Prior(4) Diffuse Prior

	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%	2.5%	97.5%
θ_h	41.819	47.075	41.774	47.124	41.788	47.040	41.686	47.143
θ_w	38.415	43.339	38.406	43.452	38.442	43.342	38.302	43.459
ρ	0.861	0.933	0.862	0.935	0.861	0.935	0.794	0.900

After comparing the confidence intervals, I think the prior I proposed is helpful in estimating θ and Σ . For θ , it has a narrower posterior CI than Jeffery’s prior and the ‘diffuse prior’. For Σ , it captures that high correlation between ages of couples very well.

However, we also find that the unit information prior leads to a even narrower posterior CI and the same CI

for ρ . Thus, we might prefer to choose a unit information prior instead.

Also we should notice that the results are very similar from all of the priors since we have a fairly large sample size and the prior information are somewhat overwhelmed by the data information.

However, a smaller sample size would result in wider posterior confidence intervals from the weak priors, since there is less data to improve the precision of the posterior distributions. There was hence a much larger impact on the intervals from the “diffuse” prior when compared to other kinds of priors.

Math Problem:

Verify that if

$$Y = (Y_a, Y_b) \sim N(\theta = (\theta_a, \theta_b), \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix})$$

is a multivariate normal, then conditional distribution is given by

$$Y_b|Y_a \sim N(\theta_b + \Sigma_{ba}\Sigma_{aa}^{-1}(y_a - \theta_a), \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})$$

For the conditional density, we know that

$$p(y_b|y_a) = \frac{p(y_b, y_a)}{p(y_a)} \propto p(y_b, y_a)$$

So we can derive the kernel of $p(y_b|y_a)$ by leaving out all the terms that does not depend on y_b in $p(y_a, y_b)$.

Also there are two notes:

$$(a). \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}, \text{ then } \Sigma^{-1} = \begin{bmatrix} (\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1} & -(\Sigma_{aa} - \Sigma_{ab}\Sigma_{bb}^{-1}\Sigma_{ba})^{-1}\Sigma_{ab}\Sigma_{bb}^{-1} \\ -(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}\Sigma_{ba}\Sigma_{aa}^{-1} & (\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1} \end{bmatrix}$$

(b). If Y and θ are both $p \times 1$ matrices, Σ is $p \times p$ matrix. Then $Y^T\Sigma^{-1}Y, \theta^T\Sigma^{-1}\theta, Y^T\Sigma^{-1}\theta$ are all scalars and they are equal to their transposes.

Then we can derive the conditional distribution:

$$\begin{aligned}
p(y_b|y_a) &\propto p(y_b, y_a) \\
&\propto \exp\left(-\frac{1}{2}[(y_a - \theta_a)^T, (y_b - \theta_b)^T] \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}^{-1} \begin{pmatrix} y_a - \theta_a \\ y_b - \theta_b \end{pmatrix}\right) \\
&\propto \exp\left(\frac{1}{2}((y_b - \theta_b)^T(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}(y_b - \theta_b) \right. \\
&\quad \left. - 2(y_a - \theta_a)^T\Sigma_{aa}^{-1}\Sigma_{ab}(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}(y_b - \theta_b))\right) \quad (\text{from note (a) and (b)}) \\
&\propto \exp\left(\frac{1}{2}(y_b^T(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}y_b - 2y_b^T(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}\theta_b \right. \\
&\quad \left. - 2(y_a - \theta_a)^T\Sigma_{aa}^{-1}\Sigma_{ab}(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}y_b)\right) \quad (\text{from note (b)}) \\
&\propto \exp\left(\frac{1}{2}(y_b^T(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}y_b - 2y_b^T(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}\theta_b \right. \\
&\quad \left. - 2y_b^T(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}\Sigma_{ba}\Sigma_{aa}^{-1}(y_a - \theta_a))\right) \quad (\text{from note (b)}) \\
&\propto \exp\left(\frac{1}{2}(y_b^T(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}y_b - 2y_b^T(\Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab})^{-1}\{\theta_b + \Sigma_{ba}\Sigma_{aa}^{-1}(y_a - \theta_a)\})\right) \\
&\propto \exp\left(\frac{1}{2}((y_b - \theta_{b|a})^T\Sigma_{b|a}^{-1}(y_b - \theta_{b|a}))\right)
\end{aligned}$$

We can see that the kernel of $p(y_b|y_a)$ is exactly the kernel of a $normal(\theta_{b|a}, \Sigma_{b|a})$, where $\theta_{b|a} = \theta_b + \Sigma_{ba}\Sigma_{aa}^{-1}(y_a - \theta_a)$, $\Sigma_{b|a} = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}$. Therefore, we have verified that the conditional distribution $p_{Y_b|Y_a}(y_b|y_a)$ is $N(\theta_{b|a}, \Sigma_{b|a})$.