STA 601/360 Homework 7

Yi Mi

23, October, 2019

Exercise 1

Hoff 7.2

Part (a)

$$\begin{aligned}
\prod_i p\,(y_i \mid \theta, \Psi) &= (2\pi)^{-np/2}|\Psi|^{n/2} exp\{-\frac{1}{2}\Sigma(y_i-\theta)^T\Psi(y_i-\theta)\} \\
&= (2\pi)^{-np/2}|\Psi|^{n/2} exp\{-\frac{1}{2}\Sigma(y_i-\overline{y}+\overline{y}-\theta)^T\Psi(y_i-\theta)\} \\
&= (2\pi)^{-np/2}|\Psi|^{n/2} exp\{-\frac{1}{2}\Sigma[(y_i-\overline{y})^T\Psi(y_i-\overline{y})+(y_i-\overline{y})^T\Psi(\overline{y}-\theta)+ \\
&\quad (\overline{y}-\theta)^T\Psi(y_i-\overline{y})+(\overline{y}-\theta)^T\Psi(\overline{y}-\theta)]\} \\
&= (2\pi)^{-np/2}|\Psi|^{n/2} exp\{-\frac{1}{2}[\Sigma(y_i-\overline{y})^T\Psi(y_i-\overline{y})+n(\overline{y}-\theta)^T\Psi(\overline{y}-\theta)]\} \\
&\because \mathbf{S} = \Sigma(y_i-\overline{y})(y_i-\overline{y})^T/n \\
&= (2\pi)^{-np/2}|\Psi|^{n/2} exp\{-\frac{1}{2}[\mathrm{tr}(n\mathbf{S}\Psi)+n(\overline{y}-\theta)^T\Psi(\overline{y}-\theta)]\}
\end{aligned}$$

$$\begin{aligned}
l(\theta,\Psi \mid y) &= \log\,(\prod_i p\,(y_i \mid \theta,\Psi)) \\
&= \frac{n}{2}log|\Psi| - \frac{1}{2}\mathrm{tr}(n\mathbf{S}\Psi) - \frac{n}{2}(\overline{y}-\theta)^T\Psi(\overline{y}-\theta) + c;
\end{aligned}$$

$$\begin{aligned}
\theta \mid \Psi &\sim \text{multivariate normal}(\overline{y}, \Psi^{-1}) \\
\Psi &\sim \text{Wishart}(p+1, \mathbf{S}^{-1}) \\
p(\theta \mid \Psi) &\propto |\Psi|^{\frac{1}{2}} exp\{-\frac{1}{2}(\theta-\overline{y})^T\Psi(\theta-\overline{y})\} \\
p(\Psi) &\propto exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{S}\Psi)\};
\end{aligned}$$

$$\begin{aligned}
p(\theta,\Psi) &= p(\theta \mid \Psi)p(\Psi) \\
\log p(\theta,\Psi) &= \log p(\theta \mid \Psi) + \log p(\Psi) \\
&= \frac{1}{2}log|\Psi| - \frac{1}{2}(\overline{y}-\theta)^T\Psi(\overline{y}-\theta) - \frac{1}{2}\mathrm{tr}(\mathbf{S}\Psi) + c' \\
&= l(\theta,\Psi \mid y)/n + c''
\end{aligned}$$

Part (b)

$$
\begin{aligned}
p(\Sigma) &\propto |\Sigma|^{-\frac{2p+2}{2}} * exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{S}\Sigma^{-1})\} \\
p(\theta \mid \Sigma) &\propto |\Sigma|^{-\frac{1}{2}} * exp\{-\frac{1}{2}(\theta-\overline{y})^T\Sigma^{-1}(\theta-\overline{y})\} \\
p(y_1,\ldots,y_n \mid \theta,\Sigma) &\propto |\Sigma|^{-\frac{n}{2}} exp\{-\frac{1}{2}[\mathrm{tr}(n\mathbf{S}\Sigma^{-1})+n(\overline{y}-\theta)^T\Sigma^{-1}(\overline{y}-\theta)]\} \\
p(\theta,\Sigma \mid y_1,\ldots,y_n) &\propto p(\Sigma)p(\theta \mid \Sigma)p(y_1,\ldots,y_n \mid \theta,\Sigma) \\
&\propto |\Sigma|^{-\frac{1}{2}(2p+n+2)} exp\{-\frac{1}{2}\mathrm{tr}((n+1)\mathbf{S}\Sigma^{-1})\} \\
&\quad * |\Sigma|^{-\frac{1}{2}} exp\{-\frac{1}{2}(n+1)(\theta-\overline{y})^T\Sigma^{-1}(\theta-\overline{y})\} \\
&\propto \mathrm{inverse-Wishart}(n+p+1, \frac{\mathbf{S}^{-1}}{n+1}) \\
&\quad * \mathrm{mvn}(\overline{y}, \frac{\Sigma}{n+1})
\end{aligned}
$$

Exercise 2

Hoff 7.4

Part (a)

```r
age = as.matrix(read.table(file=url("http://www2.stat.duke.edu/~pdh10/F
CBS/Exercises/agehw.dat")))
age = age[-1,]
age = matrix(as.numeric(age), nrow = 100)
```

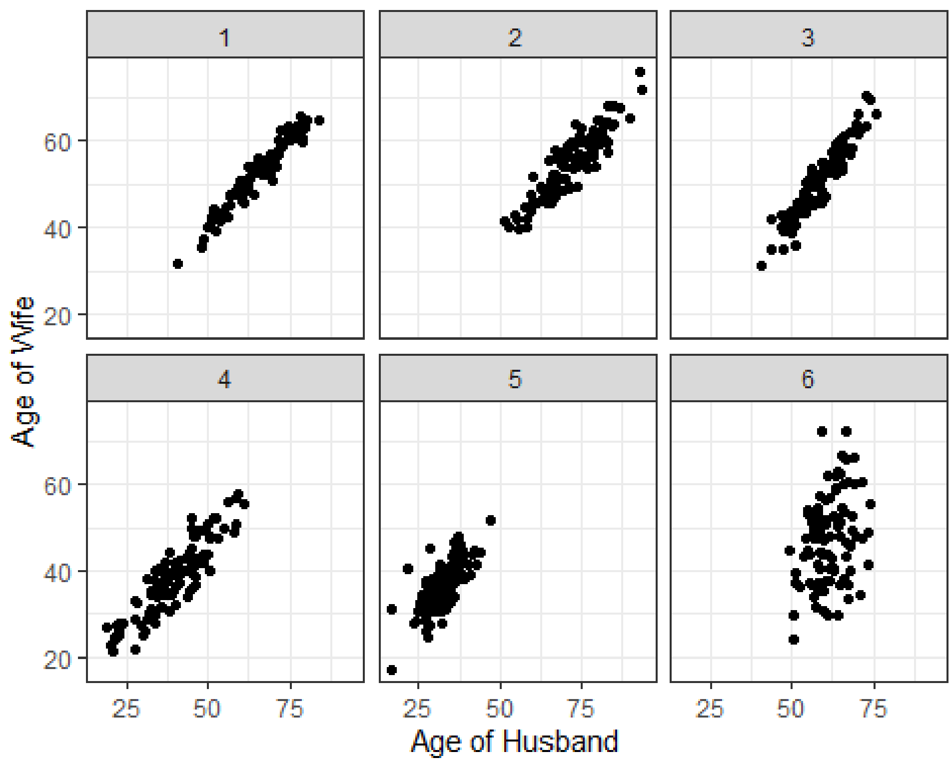
I will set $\mu_h = 50$ and $\mu_w = 47$. I have the prior belief that the age of 95% people will fall in to the range between 25 and 75, so $50 + 1.96 * \sigma_h = 75$, inducing that $\sigma_h = 12.75$. And the age of 95% of the wife will fall in to the range between 24 and 70, so $47 + 1.96 * \sigma_w = 70$, inducing that $\sigma_w = 11.73$. Besides, I will set $\rho_{hw} = 0.8$, so that $\sigma_{hw} = 119.65$.

$$
\begin{aligned}
\theta &\sim \mathrm{mvn}(\mu_0, \Lambda_0) \\
\sigma^2 &\sim \mathrm{inverse-Wishart}(v_0, S_0^{-1}) \\
\mu_0 &= (50,47)^T \\
\Lambda_0 &= \begin{bmatrix} 12.75^2 & 119.65 \\ 119.65 & 11.73^2 \end{bmatrix} \\
v_0 &= p + 2 = 4 \\
S_0 &= \Lambda_0
\end{aligned}
$$

Part (b)

```r
mu.0 = c(50, 47)
L.0 = S.0 = rbind(c(12.75^2, 119.65), c(119.65, 11.73^2))
v.0 = 4
```

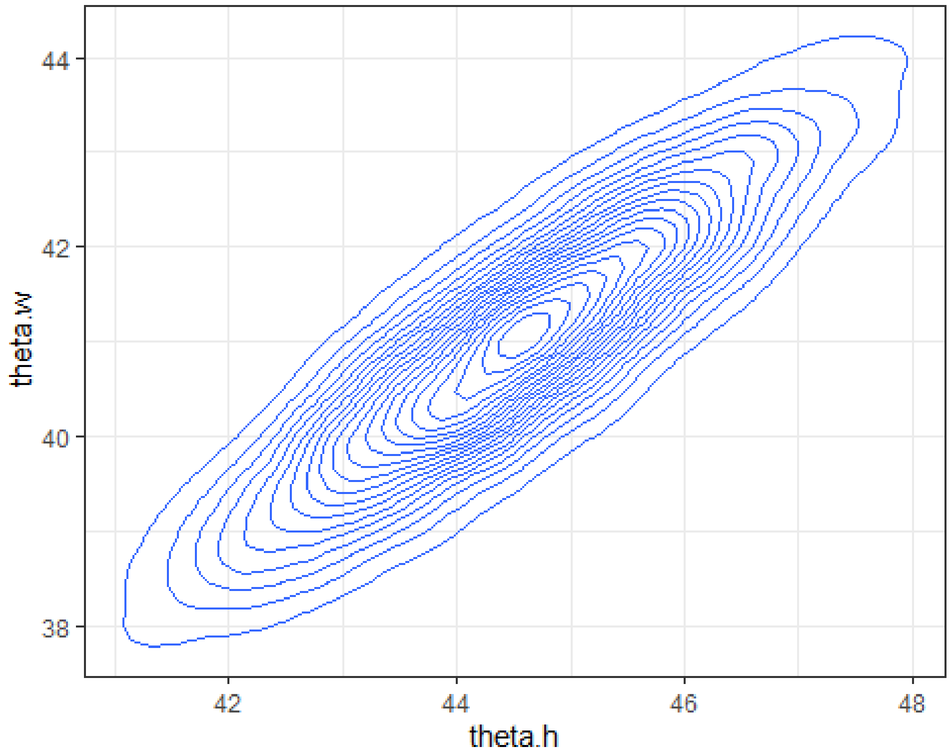
```r
# sample from prior distribution
Y.h = c()
Y.w = c()
set = c()

solve = Matrix::solve
for(i in 1:6){
    theta = mvrnorm(1, mu.0, L.0)
    sigma = solve(rwish(v.0, solve(S.0)))
    Y = mvrnorm(100, theta, sigma)
    Y.h = c(Y.h, Y[,1])
    Y.w = c(Y.w, Y[,2])
    set = c(set, rep(i, 100))
}

df1 = data.frame(cbind(Y.h, Y.w, set))
ggplot(df1, aes(x = Y.h, y = Y.w)) + geom_point() + facet_wrap(~set) +
labs(x = 'Age of Husband', y = 'Age of Wife')
```



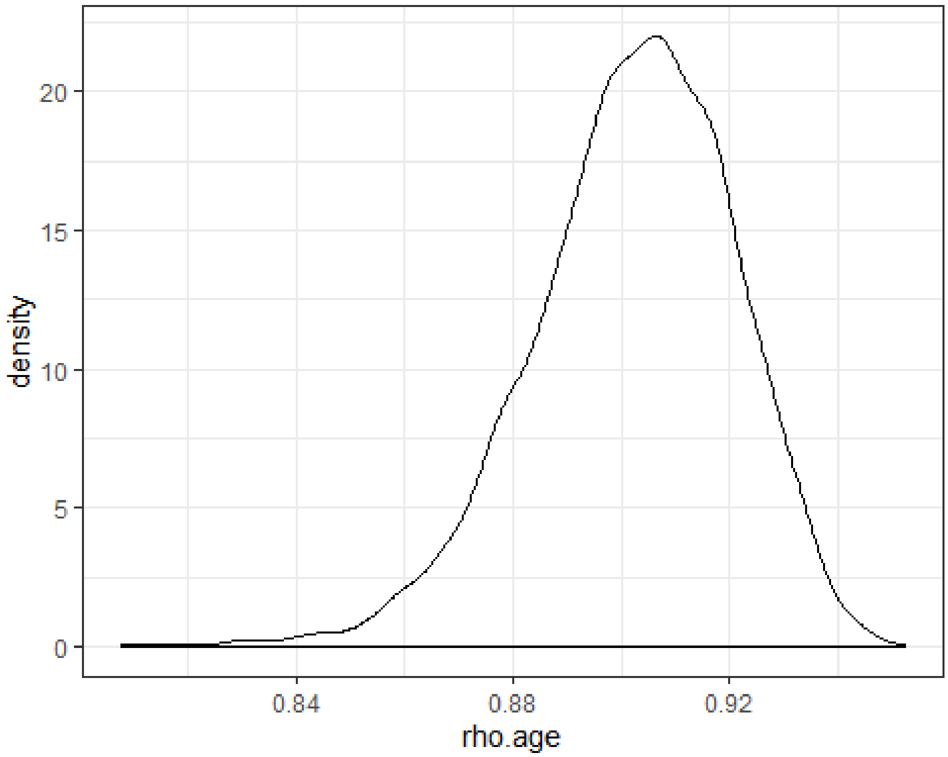
The plots above show that the age of husband and wife are highly positive correlated and wife is younger than husband which are consistent with my prior belief. Even if the points sampled are not always centered around my prior belief, the uncertainty is not that intolerable. So the prior I eventually decide upon is the original prior.

Part (c)

```
# MCMC for my own prior
set.seed(1)
mcmc = function(age){

  n = dim(age)[1]
  ybar = apply(age, 2, mean)
  sigma = cov(age)
  THETA = SIGMA = NULL

  S=5000

  for(s in 1:S){
    L.n = solve(solve(L.0) + n*solve(sigma))
    mu.n = L.n %*% (solve(L.0) %*% mu.0 + n*solve(sigma) %*% ybar)
    theta = rmvnorm(1, mu.n, L.n)

    S.theta = (t(age) - c(theta)) %*% t(t(age) - c(theta))
    S.n = S.0 + S.theta
    sigma = solve(monomvn::rwish(v.0 + n, solve(S.n)))

    THETA = rbind(THETA, theta)
    SIGMA = rbind(SIGMA, c(sigma))
  }
  return(list(THETA, SIGMA))
}

age.mcmc = mcmc(age)
THETA = age.mcmc[[1]]
SIGMA = age.mcmc[[2]]

# joint posterior distribution of theta.h and theta.w
df2 = data.frame(THETA)
colnames(df2) = c('h', 'w')
gp1 = qplot(df2$h, df2$w, geom='density2d', bins=20,
      xlab = expression(theta.h), ylab=expression(theta.w))
gp1
```




```r
# marginal posterior density of the correlation between Y.h and Y.w
rho = function(sigma){
  sigma[2]/(sqrt(sigma[1]*sigma[4]))
}
rho.age = apply(SIGMA, MARGIN = 1, rho)
gp2 = ggplot(data.frame(rho.age), aes(x = rho.age))+geom_density()
gp2
```




```
quantile(THETA[, 1], c(0.025, 0.975))

##     2.5%    97.5%
## 41.77714 47.17880

quantile(THETA[, 2], c(0.025, 0.975))

##     2.5%    97.5%
## 38.47317 43.43669

quantile(rho.age, c(0.025, 0.975))

##      2.5%     97.5%
## 0.8615058 0.9340716
```

Part (d)

For unit information prior:

```
# MC for unit information prior
mc.unit <- function(age){
  n = dim(age)[1]
  ybar = apply(age, 2, mean)
  THETA = SIGMA = NULL

  S=5000
  for(s in 1:S){
    sigma = solve(monomvn::rwish(v.0 + n, solve(S.0)/(n+1)))
```

```r
    theta = rmvnorm(1, ybar, sigma/(n+1))

    THETA = rbind(THETA, theta)
    SIGMA = rbind(SIGMA, c(sigma))
  }
  return(list(THETA, SIGMA))
}
age.mc.unit = mc.unit(age)
THETA.unit = age.mc.unit[[1]]
SIGMA.unit = age.mc.unit[[2]]

# calculate rho
rho.age.unit = apply(SIGMA.unit, MARGIN = 1, rho)

quantile(THETA.unit[, 1], c(0.025, 0.975))

##     2.5%    97.5%
## 41.93616 46.92356

quantile(THETA.unit[, 2], c(0.025, 0.975))

##     2.5%    97.5%
## 38.65183 43.18698

quantile(rho.age.unit, c(0.025, 0.975))

##      2.5%     97.5%
## 0.7187766 0.8614268
```

Part (e)

Comparing the posterior confidence interval from d to it obtained in c, I found that the interval for Unit information prior is slightly narrower but it is not a big difference. And the posterior correlation coefficient is more left-skewed for my prior. I think my prior information is somewhat helpful.

```r
sample.25 = sample(1:100, 25, replace = F)
age.mcmc.25 = mcmc(age[sample.25,])
THETA.prior.25 = age.mcmc.25[[1]]
SIGMA.prior.25 = age.mcmc.25[[2]]
rho.prior.25 = apply(SIGMA.prior.25, MARGIN = 1, rho)
age.mc.unit.25 = mc.unit(age[sample.25,])
THETA.unit.25 = age.mc.unit.25[[1]]
SIGMA.unit.25 = age.mc.unit.25[[2]]
rho.unit.25 = apply(SIGMA.unit.25, MARGIN = 1, rho)

quantile(THETA.prior.25[, 1], c(0.025, 0.975))

##     2.5%    97.5%
## 39.03576 49.30042

quantile(THETA.prior.25[, 2], c(0.025, 0.975))
```

```
##     2.5%    97.5% 
## 36.86024 46.59326

quantile(rho.prior.25, c(0.025, 0.975))

##      2.5%     97.5% 
## 0.7849661 0.9506882

quantile(THETA.unit.25[, 1], c(0.025, 0.975))

##     2.5%    97.5% 
## 39.01374 48.89060

quantile(THETA.unit.25[, 2], c(0.025, 0.975))

##     2.5%    97.5% 
## 36.88568 46.04737

quantile(rho.unit.25, c(0.025, 0.975))

##      2.5%     97.5% 
## 0.6276319 0.9019599
```

I randomly sampled 25 samples from the data and calculated the posterior confidence interval for my own prior and unit information prior. As before, I have reason to believe my prior information is helpful but not that helpful for small sample size because the posterior interval derived from my prior information is much wider than unit information prior.

Exercise 3

$$
\begin{aligned}
\Sigma &= \begin{bmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{bmatrix} \\
\Sigma^{-1} &= \begin{bmatrix} \Sigma^{-1}_{(a,a)} & \Sigma^{-1}_{(a,b)} \\ \Sigma^{-1}_{(b,a)} & \Sigma^{-1}_{(b,b)} \end{bmatrix} \\
\text{where} \quad & \Sigma^{-1}_{(a,a)} = \Sigma^{-1}_{aa} + \Sigma^{-1}_{aa}\Sigma_{ab}\Sigma_0\Sigma_{ba}\Sigma^{-1}_{aa}, \\
& \Sigma^{-1}_{(a,b)} = -\Sigma^{-1}_{aa}\Sigma_{ab}\Sigma_0, \\
& \Sigma^{-1}_{(b,a)} = -\Sigma_0\Sigma_{ba}\Sigma^{-1}_{aa}, \\
& \Sigma^{-1}_{(b,b)} = \Sigma_0, \text{ denoting } \Sigma_0 = \Sigma_{bb} - \Sigma_{ba}\Sigma^{-1}_{aa}\Sigma_{ab} \text{ to simplify}
\end{aligned}
$$

$$
\begin{aligned}
p(y) &= (2\pi)^{-p/2}|\Sigma|^{-1/2} exp\{-\frac{1}{2}(y-\theta)^T\Sigma^{-1}(y-\theta)\} \\
p(y_a) &= (2\pi)^{-p_a/2}|\Sigma_{aa}|^{-1/2} exp\{-\frac{1}{2}(y_a-\theta_a)^T\Sigma_{aa}^{-1}(y_a-\theta_a)\} \\
p(y_b \mid y_a) &= \frac{p(y)}{p(y_a)} \\
&= (2\pi)^{-p_b/2}|\Sigma_{bb}|^{-1/2} exp\{-\frac{1}{2}[(y-\theta)^T\Sigma^{-1}(y-\theta)-(y_a-\theta_a)^T\Sigma_{aa}^{-1}(y_a-\theta_a)]\} \\
&= (2\pi)^{-p_b/2}|\Sigma_{bb}|^{-1/2} exp\{-\frac{1}{2}[(y_a-\theta_a)^T\Sigma_{(a,a)}^{-1}(y_a-\theta_a)+ \\
&\quad (y_b-\theta_b)^T\Sigma_{(b,a)}^{-1}(y_a-\theta_a)+(y_a-\theta_a)^T\Sigma_{(a,b)}^{-1}(y_b-\theta_b)+ \\
&\quad (y_b-\theta_b)^T\Sigma_{(b,b)}^{-1}(y_b-\theta_b)-(y_a-\theta_a)^T\Sigma_{aa}^{-1}(y_a-\theta_a)]\} \\
&= (2\pi)^{-p_b/2}|\Sigma_{bb}|^{-1/2} exp\{-\frac{1}{2}[(y_a-\theta_a)^T\Sigma_{aa}^{-1}\Sigma_{ab}\Sigma_0\Sigma_{ba}\Sigma_{aa}^{-1}(y_a-\theta_a)- \\
&\quad (y_a-\theta_a)^T\Sigma_{aa}^{-1}\Sigma_{ab}\Sigma_0(y_b-\theta_b)-(y_b-\theta_b)^T\Sigma_0\Sigma_{ba}\Sigma_{aa}^{-1}(y_a-\theta_a)+ \\
&\quad (y_b-\theta_b)^T\Sigma_0(y_b-\theta_b)]\} \\
&= (2\pi)^{-p_b/2}|\Sigma_{bb}|^{-1/2} exp\{-\frac{1}{2}[(y_b-\theta_b)^T\Sigma_0[(y_b-\theta_b)-\Sigma_{ba}\Sigma_{aa}^{-}1(y_a-\theta_a)]+ \\
&\quad (y_a-\theta_a)^T\Sigma_{aa}^{-1}\Sigma_{ab}\Sigma_0[\Sigma_{ba}\Sigma_{aa}^{-1}(y_a-\theta_a)-(y_a-\theta_a)]]\} \\
&= (2\pi)^{-p_b/2}|\Sigma_{bb}|^{-1/2} * \\
&\quad exp\{-\frac{1}{2}[(y_b-\theta_b)^T-(y_a-\theta_a)^T\Sigma_{aa}^{-1}\Sigma_{ab}]\Sigma_0[(y_b-\theta_b)-\Sigma_{ba}\Sigma_{aa}^{-1}(y_a-\theta_a)]\} \\
&= (2\pi)^{-p_b/2}|\Sigma_{bb}|^{-1/2} * \\
&\quad exp\{-\frac{1}{2}[(y_b-\theta_b)-\Sigma_{ba}\Sigma_{aa}^{-1}(y_a-\theta_a)]^T\Sigma_0[(y_b-\theta_b)-\Sigma_{ba}\Sigma_{aa}^{-1}(y_a-\theta_a)]\} \\
\Longrightarrow Y_b &\mid Y_a \sim \text{multivariate normal}(\theta_{b|a},\Sigma_{b|a}), \text{where} \\
&\theta_{b|a} = \theta_b + \Sigma_{ba}\Sigma_{aa}^{-1}(y_a-\theta_a) \\
&\Sigma_{b|a} = \Sigma_0 = \Sigma_{bb} - \Sigma_{ba}\Sigma_{aa}^{-1}\Sigma_{ab}
\end{aligned}
$$