DDA 4010 Assignment 4 赵松霖 120090346

Problem 1:

(a) $p(y_i|\theta,\Sigma) = (2\pi)^{-\frac{p}{2}} |\Sigma|^{-\frac{1}{2}} \exp(-\frac{1}{2}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta))$

$\psi = \Sigma^{-1}$

$\therefore p(y_i|\theta,\psi) = (2\pi)^{-\frac{p}{2}} |\psi|^{\frac{1}{2}} \exp(-\frac{1}{2}(y_i-\theta)^T\psi(y_i-\theta))$

$l(\theta,\psi;Y) = \sum_{i=1}^{n}\log p(y_i|\theta,\psi)$

$= \sum_{i=1}^{n}(-\frac{p}{2}\log(2\pi) + \frac{1}{2}\log|\psi| - \frac{1}{2}(y_i-\theta)^T\psi(y_i-\theta))$

$= \frac{n}{2}\log|\psi| - \frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^T\psi(y_i-\theta) + c$

Since $(y_i-\theta)^T\psi(y_i-\theta) = tr[(y_i-\theta)^T\psi(y_i-\theta)]$

$= tr[\psi(y_i-\theta)(y_i-\theta)^T]$

$= tr[(y_i-\theta)(y_i-\theta)^T\psi]$

$= tr[(y_i-\bar{y}+\bar{y}-\theta)(y_i-\bar{y}+\bar{y}-\theta)^T\psi]$

$= tr\{[(y_i-\bar{y})(y_i-\bar{y})^T + (\bar{y}-\theta)(\bar{y}-\theta)^T + (y_i-\bar{y})(\bar{y}-\theta)^T + (\bar{y}-\theta)(y_i-\bar{y})^T]\psi\}$

$\therefore l(\theta,\psi;Y) = \frac{n}{2}\log|\psi| - \frac{1}{2}\sum_{i=1}^{n}tr\{[(y_i-\bar{y})(y_i-\bar{y})^T + (\bar{y}-\theta)(\bar{y}-\theta)^T + (y_i-\bar{y})(\bar{y}-\theta)^T + (\bar{y}-\theta)(y_i-\bar{y})^T]\psi\} + c$

$= \frac{n}{2}\log|\psi| - \frac{1}{2}tr[\sum_{i=1}^{n}(y_i-\bar{y})(y_i-\bar{y})^T\psi + n(\bar{y}-\theta)(\bar{y}-\theta)^T\psi + \sum_{i=1}^{n}(y_i-\bar{y})(\bar{y}-\theta)^T\psi + \sum_{i=1}^{n}(\bar{y}-\theta)(y_i-\bar{y})^T\psi] + c$

$= \frac{n}{2}\log|\psi| - \frac{1}{2}[tr(\sum_{i=1}^{n}(y_i-\bar{y})(y_i-\bar{y})^T\psi) + tr(n(\bar{y}-\theta)(\bar{y}-\theta)^T\psi)] + c$

$= \frac{n}{2}\log|\psi| - \frac{1}{2}[tr(nS\psi) + tr(n(\bar{y}-\theta)(\bar{y}-\theta)^T\psi)] + c$

$\therefore \frac{l(\theta,\psi;Y)}{n} + c = \frac{1}{2}\log|\psi| - \frac{1}{2}[tr(S\psi) + tr((\bar{y}-\theta)(\bar{y}-\theta)^T\psi)]$

$\therefore p(\theta,\psi) = \exp(\frac{l(\theta,\psi;Y)}{n} + c)$

$\propto |\psi|^{\frac{1}{2}}\exp\{-\frac{1}{2}[tr(S\psi) + tr((\bar{y}-\theta)(\bar{y}-\theta)^T\psi)]\}$

$= |\psi|^{\frac{1}{2}}\exp\{-\frac{1}{2}(\bar{y}-\theta)^T\psi(\bar{y}-\theta)\}\exp(-\frac{1}{2}tr(S\psi))$

$= |\psi|^{\frac{1}{2}}\exp(-\frac{1}{2}(\bar{y}-\theta)^T\psi(\bar{y}-\theta)) \cdot \exp(-\frac{tr(S\cdot\psi)}{2})$

$\propto |\psi|^{\frac{1}{2}}\exp(-\frac{1}{2}(\bar{y}-\theta)^T\psi(\bar{y}-\theta)) \cdot S^{-\frac{p+1-p-1}{2}}\exp(-\frac{tr(S\cdot\psi)}{2})$

$\propto p(\theta|\psi)p(\psi)$

$\therefore p(\theta,\psi)$ and $p(\theta|\psi)p(\psi)$ have the same kernel.

$\therefore p_u(\theta,\psi) = p_u(\theta|\psi)p_u(\psi)$

$\therefore$ a unit information prior for (θ,ψ) is $\theta|\psi \sim MVN(\bar{y},\psi^{-1})$ and $\psi \sim Wishart(p+1, S^{-1})$

(b) Since $\psi_u \sim W_p(p+1, S^{-1})$

$\therefore \quad \Sigma \sim W_p^{-1}(p+1, S^{-1})$

$\therefore \quad p_u(\Sigma) \propto |\Sigma|^{-\frac{p+1+p+1}{2}} \exp\left(-\frac{tr(S\Sigma^{-1})}{2}\right)$

$\therefore \quad p_u(\theta, \Sigma | y_1, \cdots, y_n) \propto p(y_1, \cdots, y_n | \theta, \Sigma)\, p_u(\theta|\Sigma)\, p_u(\Sigma)$

$$\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(y_i-\theta)^T\Sigma^{-1}(y_i-\theta)\right\} \cdot |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\bar{y}-\theta)^T\Sigma^{-1}(\bar{y}-\theta)\right) |\Sigma|^{-p+1} \exp\left(-\frac{tr(S\Sigma^{-1})}{2}\right)$$

$$= |\Sigma|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(y_i-\bar{y})^T\Sigma^{-1}(y_i-\bar{y})\right) \exp\left(-\frac{1}{2} n(\bar{y}-\theta)^T\Sigma^{-1}(\bar{y}-\theta)\right) \cdot$$

$$|\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\bar{y}-\theta)^T\Sigma^{-1}(\bar{y}-\theta)\right) |\Sigma|^{-p+1} \exp\left(-\frac{tr(S\Sigma^{-1})}{2}\right)$$

$$= |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\theta-\bar{y})^T(n+1)\Sigma^{-1}(\theta-\bar{y})\right) |\Sigma|^{-\frac{n+p+1+p+1}{2}} \exp\left\{-\frac{1}{2}tr\left[(n+1)S\Sigma^{-1}\right]\right\}$$

$$\propto dmvn\left(\theta : \bar{y}, \frac{\Sigma}{n+1}\right) \cdot dW_p^{-1}\left(\Sigma : n+p+1, \left[(n+1)S\right]^{-1}\right)$$

Although unit information prior contains information of the sample data, the posterior can be interpreted as a posterior distribution for θ and Σ, since it's derived from $p(\theta, \Sigma | y_1, \cdots, y_n) \propto p(u, \Sigma)\, p(y_1, \cdots, y_n | u, \Sigma)$, so it is still the distribution after observing sample data based on prior belief.

Problem 2:

(a) prior: $\theta \sim MVN(u_0, \Lambda_0)$

$\Sigma \sim W_p^{-1}(\nu_0, S_0^{-1})$

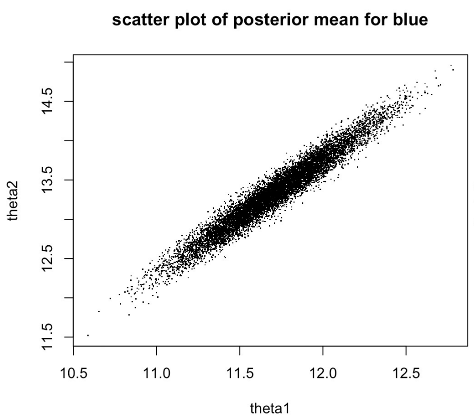
$\theta | y_1, y_2, \cdots, y_n, \Sigma \sim MVN(u_n, \Lambda_n)$

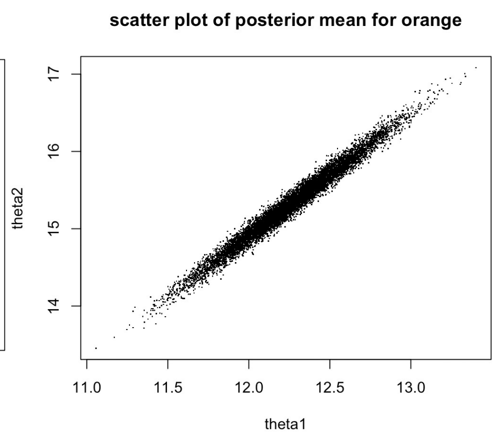
$$\begin{cases} u_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}u_0 + n\Sigma^{-1}\bar{y}) \\ \Lambda_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1} \end{cases}$$

$\Sigma | y_1, \cdots, y_n \sim W_p^{-1}(\nu_n, S_n^{-1})$

$$\begin{cases} \nu_n = \nu_0 + n \\ S_n = S_0 + \sum_{i=1}^{n}(y_i-\theta)(y_i-\theta)^T \end{cases}$$

(b)



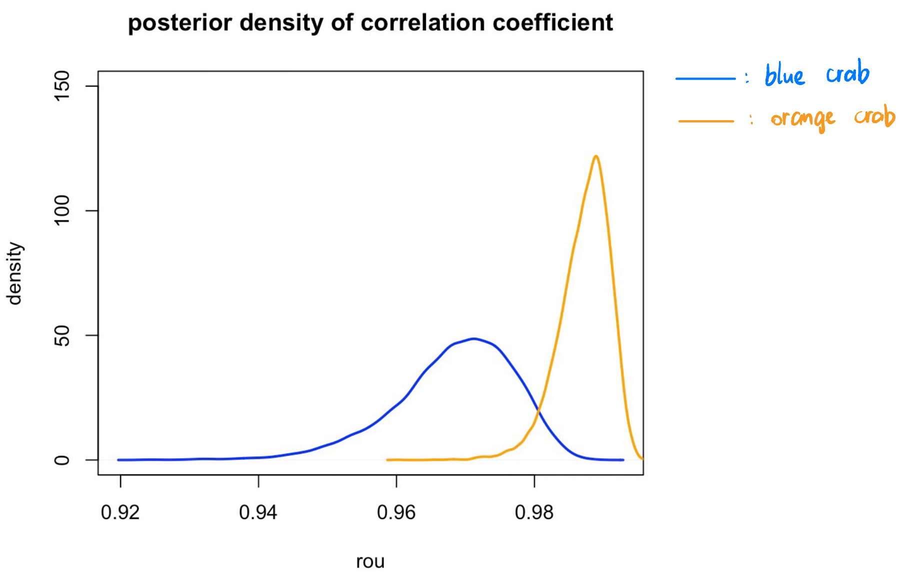



```
> x_posterior_mean
[1] 11.71574 13.34695
> y_posterior_mean
[1] 12.26290 15.32447
```

```
> quantile(x_THETA[,1]-y_THETA[,1], prob=c(.025,.975))
      2.5%      97.5%
-1.3726379  0.2783448
> quantile(x_THETA[,2]-y_THETA[,2], prob=c(.025,.975))
      2.5%      97.5%
-3.2702363 -0.6752497
```

For body depth and rear width, orangecrab is larger than bluecrab. The 95% credible of bluecrab - oragecrab is (-1.37, 0.28) for body depth and (-3.27, -0.68) for rear width.

(c)




```
> mean(y_rou > x_rou)
[1] 0.99
```

on average, the posterior correlation coefficient is larger for orange crab than blue crab.

$P(\rho_{blue} < \rho_{orange} \mid y_{blue}, y_{orange}) = 0.99$, the difference is significant.

the variance of ρ is larger for blue crab than orange crab, i.e. the ρ of orange crab is more concentrated on the mean, while ρ of blue crab is less concentrated.

This suggests that the body depth and rear width is more correlated for orange crab, and we are more certain about the true posterior correlation coefficient of orange crab.

Problem 3:

(a)

```
> theta_A
[1] 24.20049
> theta_B
[1] 24.80535
> sigmasq_A
[1] 4.0928
> sigmasq_B
[1] 4.691578
> rou
[1] 0.6164509
```

(b)

```
> t.test(YA, YB, paired = TRUE, alternative = "two.sided")

	Paired t-test

data:  YA and YB
t = -3.302, df = 57, p-value = 0.001661
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -0.9765539 -0.2392385
sample estimates:
mean difference
     -0.6078962
```

95% CI for $\theta_A - \theta_B$: $(-0.9765539, -0.2392385)$

(c) Use unit information prior

$\Phi \sim \text{wishart}(p+1, S^{-1})$

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

where $S = \Sigma (y_i - \bar{y})(y_i - \bar{y})^T / n$

c. posterior:

$$p(\Sigma \mid \theta, y_{miss}, y_{obs}) \propto p(\theta, \Sigma \mid y_{miss}, y_{obs})$$

$$\propto |\Sigma|^{-\frac{(n+p+2)+p+1}{2}} \exp\left\{-\frac{1}{2} tr\left[\left((n+1)(\theta-\bar{y})(\theta-\bar{y})^T + (n+1)S\right)\Sigma^{-1}\right]\right\}$$

① $\therefore \Sigma \mid \theta, y_{miss}, y_{obs} \sim$ inverse-wishart $\left(n+p+2, \left[(n+1)\left((\theta-\bar{y})(\theta-\bar{y})^T + S\right)\right]^{-1}\right)$

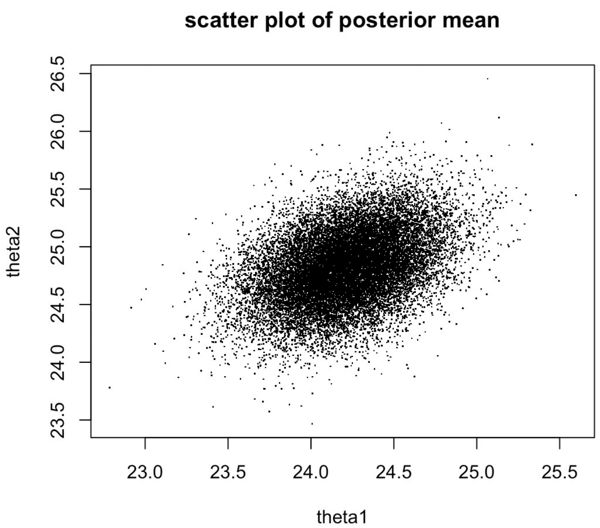
$$p(\theta \mid \Sigma, y_{miss}, y_{obs}) \propto p(\theta, \Sigma \mid y_{miss}, y_{obs})$$

$$\propto \exp\left(-\frac{1}{2}(\theta-\bar{y})^T\left[(n+1)\Sigma^{-1}\right](\theta-\bar{y})\right)$$

② $\therefore \theta \mid \Sigma, y_{miss}, y_{obs} \sim MVN\left(\bar{y}, \frac{\Sigma}{n+1}\right)$

③ $Y_{[b]} \mid Y_{[a]}, \theta, \Sigma \sim MVN(\theta_{b|a}, \Sigma_{b|a})$

$$\begin{cases} \theta_{b|a} = \theta_{[b]} + \Sigma_{[b,a]}(\Sigma_{[a,a]})^{-1}(y_{[a]} - \theta_{[a]}) \\ \Sigma_{b|a} = \Sigma_{[b,b]} - \Sigma_{[b,a]}(\Sigma_{[a,a]})^{-1}\Sigma_{[a,b]} \end{cases}$$




```
> mean(THETA[,1]-THETA[,2])
[1] -0.6164279
> quantile(THETA[,1]-THETA[,2], prob=c(.025,.975))
       2.5%       97.5%
-1.26593358  0.02954628
```

The results of gibbs sampler with unit information prior is similar to the result of (b). (b) is the frequentists' method, while (c) is bayesian's method. (b) predicts the missing value using the information of another data in the same sample with the correlation of Y_A and Y_B. (c) predicts the missing value using similar idea, but uses an iterative method to sample the posterior value, and updates our belief of missing data. So the estimate of $\theta_A - \theta_B$ is similar.

The CI of $\theta_A - \theta_B$ in (c) is wider than (b), since gibbs sampling of posterior $\theta_A - \theta_B$ has larger variance than freqentist's method.