Chase Joyner

882 Homework 5

December 9, 2016

Problem 1:

a) Notice that for model j, the likelihood function is

$$\mathbf{Y} \mid \mathbf{X}_j \sim MVN(\mathbf{X}_j\boldsymbol{\beta}, \sigma_j^2\mathbf{I})$$

where $\mathbf{I}$ is an $n \times n$ identity matrix.

$$\begin{aligned}
m(\mathbf{y} \mid M = j) &= \int\int f\big(\mathbf{y} \mid \boldsymbol{\beta}_j, \sigma_j^2\big)\pi\big(\boldsymbol{\beta}_j \mid \sigma_j^2\big)\pi\big(\sigma_j^2\big)d\boldsymbol{\beta}_j d\sigma_j^2 \\
&= \int\int \big(2\pi\sigma_j^2\big)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_j^2}(\mathbf{Y} - \mathbf{X}_j\boldsymbol{\beta}_j)'(\mathbf{Y} - \mathbf{X}_j\boldsymbol{\beta}_j)\right\} \\
&\qquad \big|2\pi\sigma_j^2\boldsymbol{\Sigma}_{j0}\big|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma_j^2}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0})'\boldsymbol{\Sigma}_{j0}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0})\right\} \\
&\qquad \frac{b_{j0}^{a_{j0}}}{\Gamma(a_{j0})}\big(\sigma_j^2\big)^{-a_{j0}-1} \exp\left\{-\frac{b_{j0}}{\sigma_j^2}\right\} d\boldsymbol{\beta}_j d\sigma_j^2.
\end{aligned}$$

Define C_j to be

$$C_j = (2\pi)^{-\frac{n+p_j}{2}} \frac{b_{j0}^{a_{j0}}}{\Gamma(a_{j0})}|\boldsymbol{\Sigma}_{j0}|^{-\frac{1}{2}}.$$

Then, we have that

$$\begin{aligned}
m(\mathbf{y} \mid M = j) = C_j \int \Bigg(&\big(\sigma_j^2\big)^{-\frac{n+p_j}{2}-a_{j0}-1} \exp\left\{-\frac{b_{j0}}{\sigma_j^2}\right\} \times \\
&\int \exp\left\{-\frac{1}{2\sigma_j^2}\Big[(\mathbf{Y} - \mathbf{X}_j\boldsymbol{\beta}_j)'(\mathbf{Y} - \mathbf{X}_j\boldsymbol{\beta}_j) + (\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0})'\boldsymbol{\Sigma}_{j0}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\beta}_{j0})\Big]\right\}\Bigg) d\boldsymbol{\beta}_j d\sigma_j^2 \\
= C_j \int \Bigg(&\big(\sigma_j^2\big)^{-\frac{n+p_j}{2}-a_{j0}-1} \exp\left\{-\frac{b_{j0}}{\sigma_j^2} - \frac{1}{2\sigma_j^2}\Big[\mathbf{Y}'\mathbf{Y} + \boldsymbol{\beta}_{j0}'\boldsymbol{\Sigma}_{j0}^{-1}\boldsymbol{\beta}_{j0} - \boldsymbol{\mu}'\mathbf{A}\Big]\right\} \times \\
&\int \exp\left\{-\frac{1}{2\sigma_j^2}(\boldsymbol{\beta}_j - \boldsymbol{\mu})'\mathbf{C}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\mu})\right\}\Bigg) d\boldsymbol{\beta}_j d\sigma_j^2,
\end{aligned}$$

where we define

$$\begin{aligned}
\boldsymbol{\mu} &= \mathbf{C} \times \left(\boldsymbol{\Sigma}_{j0}^{-1}\boldsymbol{\beta}_{j0} + \mathbf{X}_j'\mathbf{Y}\right) \\
\mathbf{C} &= \left(\mathbf{X}_j'\mathbf{X}_j + \boldsymbol{\Sigma}_{j0}^{-1}\right)^{-1} \\
\mathbf{A} &= \boldsymbol{\Sigma}_{j0}^{-1}\boldsymbol{\beta}_{j0} + \mathbf{X}_j'\mathbf{Y}.
\end{aligned}$$

Then, by integrating the two integrals, we find

$$\begin{aligned}
m(\mathbf{y} \mid M = j) &= C_j \int \left((\sigma_j^2)^{-\frac{n+p_j}{2} - a_{j0} - 1} \exp\left\{ -\frac{b_{j0}}{\sigma_j^2} - \frac{1}{2\sigma_j^2}\Big[\mathbf{Y}'\mathbf{Y} + \boldsymbol{\beta}_{j0}'\boldsymbol{\Sigma}_{j0}^{-1}\boldsymbol{\beta}_{j0} - \boldsymbol{\mu}'\mathbf{A}\Big] \right\} \times \right. \\
&\qquad \left. \int \exp\left\{ -\frac{1}{2\sigma_j^2}(\boldsymbol{\beta}_j - \boldsymbol{\mu})'\mathbf{C}^{-1}(\boldsymbol{\beta}_j - \boldsymbol{\mu}) \right\} \right) d\boldsymbol{\beta}_j d\sigma_j^2 \\
&= C_j^\star \int (\sigma_j^2)^{-\frac{n}{2} - a_{j0} - 1} \exp\left\{ -\frac{b_{j0}}{\sigma_j^2} - \frac{1}{2\sigma_j^2}\Big[\mathbf{Y}'\mathbf{Y} + \boldsymbol{\beta}_{j0}'\boldsymbol{\Sigma}_{j0}^{-1}\boldsymbol{\beta}_{j0} - \boldsymbol{\mu}'\mathbf{A}\Big] \right\} d\sigma_j^2 \\
&= C_j^\star \int (\sigma_j^2)^{-\frac{n}{2} - a_{j0} - 1} \exp\left\{ -\frac{b_{j0}}{\sigma_j^2} - \frac{1}{2\sigma_j^2}\Big[\mathbf{Y}'\mathbf{Y} + \boldsymbol{\beta}_{j0}'\boldsymbol{\Sigma}_{j0}^{-1}\boldsymbol{\beta}_{j0} - \boldsymbol{\mu}'\mathbf{A}\Big] \right\} d\sigma_j^2 \\
&= C_j^\star \frac{\Gamma(\alpha_j)}{B_j^{\alpha_j}}
\end{aligned}$$

where

$$\begin{aligned}
C_j^\star &= (2\pi)^{-\frac{n}{2}} \frac{b_{j0}^{a_{j0}}}{\Gamma(a_{j0})} |\boldsymbol{\Sigma}_{j0}|^{-\frac{1}{2}} |\mathbf{C}|^{\frac{1}{2}} \\
B_j &= \frac{1}{2}\Big[\mathbf{Y}'\mathbf{Y} + \boldsymbol{\beta}_{j0}'\boldsymbol{\Sigma}_{j0}^{-1}\boldsymbol{\beta}_{j0} - \boldsymbol{\mu}'\mathbf{A}\Big] + b_{j0} \\
\alpha_j &= \frac{n}{2} + a_{j0}.
\end{aligned}$$

b) Now we compute the Bayes factor for

$$H_0\colon \text{ Model 1} \quad \text{and} \quad H_1\colon \text{ Model 2},$$

which is given by the following quantity

$$B_{01} = \frac{m_0(\mathbf{y})}{m_1(\mathbf{y})} = \frac{m(\mathbf{y} \mid M = 1)}{m(\mathbf{y} \mid M = 2)}.$$

c) There are many methods to approximating B_{01}, which can be achieved by approximating $m(\mathbf{y} \mid M = j)$ with $\widehat{m}(\mathbf{y} \mid M = j)$. Here, we discuss three methods to finding $\widehat{m}(\mathbf{y} \mid M = j)$. Here, we will denote $\boldsymbol{\theta}_j = (\boldsymbol{\beta}_j, \sigma_j^2)'$.

1) Monte Carlo approximation. We will compute

$$\widehat{m}(\mathbf{y} \mid M = j) = \frac{1}{S}\sum_{t=1}^{S} f\big(\mathbf{y} \mid \boldsymbol{\theta}_j^{(t)}\big)$$

where we sample $\boldsymbol{\theta}_j^{(t)} \sim \pi_j(\boldsymbol{\theta})$, i.e. sample $\boldsymbol{\beta}_j^{(t)} \mid \sigma_j^{2(t-1)} \sim N\big(\boldsymbol{\beta}_{j0}, \sigma_j^{(t-1)}\boldsymbol{\Sigma}_{j0}\big)$ and $\sigma_j^{(t)} \sim IG(a_{j0}, b_{j0})$. Then, by the law of large numbers,

$$\widehat{m}(\mathbf{y} \mid M = j) \to E\big[f(\mathbf{y} \mid \boldsymbol{\theta}_j)\big].$$

A major problem with this approach is that if the prior distribution is not very informative, then this process is extremely inefficient and will require S to be very large.

2) Importance sampling. Here, we will make use of an importance distribution $\pi_j^\star(\boldsymbol{\theta})$. Sampling $\boldsymbol{\theta}_j^{(t)} \sim \pi_j^\star(\boldsymbol{\theta})$, compute the approximation

$$\widehat{m}(\mathbf{y} \mid M = j) = \frac{1}{S}\sum_{t=1}^{S} w_t f\big(\mathbf{y} \mid \boldsymbol{\theta}_j^{(t)}\big)$$

where the weight w_t is given by

$$w_t = \frac{\pi_j\big(\boldsymbol{\theta}_j^{(t)}\big)}{\pi_j^\star\big(\boldsymbol{\theta}_j^{(t)}\big)} = \frac{\pi\big(\boldsymbol{\beta}_j^{(t)} \mid \sigma_j^{2(t)}\big)\pi\big(\sigma_j^{2(t)}\big)}{\pi_\beta^\star\big(\boldsymbol{\beta}_j^{(t)} \mid \sigma_j^{2(t)}\big)\pi_\sigma^\star\big(\sigma_j^{2(t)}\big)}.$$

It can be shown that the importance distribution in this fashion does not impact inference and thus should be chosen in a good way. We will use the LSE estimates and impose

$$\begin{aligned} \pi_\beta^\star\big(\boldsymbol{\beta}_j \mid \sigma_j^2\big) &= N\big((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \sigma_j^2(\mathbf{X}'\mathbf{X})^{-1}\big) \\ \pi_\sigma^\star\big(\sigma_j^2\big) &= IG(a, b) \end{aligned}$$

where a and b are chosen such that $\dfrac{b}{a-1} = \text{MSE}$ and

3) Importance sampling on steroids. This is the last method and is similar to method 2. The difference is now that

$$\widehat{m}(\mathbf{y} \mid M = j) = \frac{\sum_{t=1}^{S} w_t f\big(\mathbf{y} \mid \boldsymbol{\theta}_j^{(t)}\big)}{\sum_{t=1}^{S} w_t}.$$

d) Here we implement a small simulation study to compare these three techniques. Suppose that model 1 and model 2 are

$$\begin{aligned} &\text{Model 1:} \quad Y_i = 3 + 0.5x_{1i} + 1.25x_{2i} + e_{1i}, \quad e_{1i} \overset{iid}{\sim} N(0, 1) \\ &\text{Model 2:} \quad Y_i = 3 + 0.5x_{1i} + 1.25x_{3i} + e_{2i}, \quad e_{2i} \overset{iid}{\sim} N(0, 1.2^2). \end{aligned}$$

That is, not a full versus reduced model scenario. We will generate our data under model 1. To do this, we generate covariate x_1 from $N(2, 1)$ and covariates x_2 and x_3 from a $\text{Binom}(1, 0.5)$. Before we approximate the Bayes factor, we first compute the exact Bayes factor from problem 1. Here we impose that the prior distributions are

$$\begin{aligned} \boldsymbol{\beta}_1 \mid \sigma_1^2 &\sim N(\mathbf{0}_{p_1}, \mathbf{T}_{p_1}) \qquad & \boldsymbol{\beta}_2 \mid \sigma_2^2 &\sim N(\mathbf{0}_{p_2}, \mathbf{T}_{p_2}) \\ \sigma_1^2 &\sim IG(1, 1) & \sigma_2^2 &\sim IG(1, 1), \end{aligned}$$

where $\mathbf{0}_{p_j}$ is a vector of size p_j whose entries are all ones and $\mathbf{T}_{p_j}$ is a $p_j \times p_j$ diagonal matrix whose entries are all 1000. Using this formulation, the expression in problem 1, and the generated data, the exact Bayes factor is given by

$$B_{01} = \frac{C_1^\star}{C_2^\star} \cdot \frac{\Gamma(\alpha_1)}{B_1^{\alpha_1}} \cdot \frac{B_2^{\alpha_2}}{\Gamma(\alpha_2)} = 8877.709.$$

Now, implementing the three methods, we obtained an estimated Bayes factor of

$$\begin{aligned} \text{Method 1:} \quad & \widehat{B}_{01} = 98876.07 \\ \text{Method 2:} \quad & \widehat{B}_{01} = 8961.766 \\ \text{Method 3:} \quad & \widehat{B}_{01} = 8549.591. \end{aligned}$$

As we expected, method 1 did not work very well. Although it gave the correct decision, it does not estimate the exact Bayes factor well. Methods 2 and 3 work pretty similarly and have an approximation that is very close to the exact Bayes factor.