Biostat 250C HW3

Elvis Cui
Han Cui
Dept. of Biostat
UCLA

Q1: Verify the following.

(1): $X_*^T V_*^{-1} X_* = V_\beta + X^T V_y X = M^{-1}$

Pf: $y_* = X_* \beta + e_*$ where

$$y_* = \begin{pmatrix} y \\ \mu_\beta \end{pmatrix},\ X_* = \begin{pmatrix} X \\ I \end{pmatrix},\ e_* = \begin{pmatrix} e_y \\ e_\beta \end{pmatrix}$$

& $e_* \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma^2 \begin{pmatrix} V_y & 0 \\ 0 & V_\beta \end{pmatrix} \right)$

So $V_* = \begin{pmatrix} V_y & 0 \\ 0 & V_\beta \end{pmatrix}$ and $V_*^{-1} = \begin{pmatrix} V_y^{-1} & 0 \\ 0 & V_\beta^{-1} \end{pmatrix}$

Since $V_* V_*^{-1} = I$.

Thus,

$$X_*^T V_*^{-1} X_* = \begin{pmatrix} X \\ I \end{pmatrix}^T \begin{pmatrix} V_y^{-1} & 0 \\ 0 & V_\beta^{-1} \end{pmatrix} \begin{pmatrix} X \\ I \end{pmatrix}$$

$$= (X^T V_y^{-1}\quad V_\beta^{-1}) \begin{pmatrix} X \\ I \end{pmatrix}$$

$$= X^T V_y^{-1} X + V_\beta^{-1} \quad \square$$

(2) $\hat{\beta} = Mm$, $m = V_\beta^{-1} \mu_\beta + X^T V_y^{-1} y$

Pf: $\hat{\beta}$ solves the normal equation

$$X_*^T V_* X_* \hat{\beta} = X_*^T V_*^{-1} y_*$$

$$\Rightarrow \hat{\beta} = (X_*^T V_* X_*)^{-1} (X_*^T V_*^{-1}) y_*$$

But $M^{-1} = X_*^T V_* X_*$ by (1) &

$$X_*^T V_*^{-1} y_* = (X^T V_y^{-1}\quad V_\beta^{-1}) \begin{pmatrix} y \\ \mu_\beta \end{pmatrix}$$

$$= X^T V_y^{-1} y + V_\beta^{-1} \mu_\beta$$

$\Rightarrow \hat{\beta} = Mm$, $m = V_\beta^{-1} \mu_\beta + X^T V_y^{-1} y$. $\square$

(3): $RSS(y_*, X_*, V_*) = y^T V_y^{-1} y + \mu_\beta^T V_\beta^{-1} \mu_\beta - m^T M m$

Pf: $RSS(y_*, X_*, V_*)$

$$= (y_* - X_* \hat{\beta})^T V_*^{-1} (y_* - X_* \hat{\beta})$$

① $y_*^T V_*^{-1} y_*$

$$= (y^T\ \mu_\beta^T) \begin{pmatrix} V_y^{-1} & 0 \\ 0 & V_\beta^{-1} \end{pmatrix} \begin{pmatrix} y \\ \mu_\beta \end{pmatrix}$$

$$= y^T V_y^{-1} y + \mu_\beta^T V_\beta^{-1} \mu_\beta$$

② $y_*^T V_*^{-1} X_* \hat{\beta}$

$$= (y^T\ \mu_\beta^T) \begin{pmatrix} V_y^{-1} & 0 \\ 0 & V_\beta^{-1} \end{pmatrix} \begin{pmatrix} X \\ I \end{pmatrix} Mm$$

$$= (y^T V_y^{-1} X + \mu_\beta^T V_\beta^{-1}) Mm = m^T M m$$

③ $\hat{\beta}^T X_*^T V_*^{-1} X_* \hat{\beta} = m^T M m$

$\Rightarrow RSS = ① - 2② + ③$

$$= y^T V_y^{-1} y + \mu_\beta^T V_\beta^{-1} \mu_\beta - m^T M m$$

$\square$

Q2: $p(y|x)=\mathcal{N}(y|x,1)$ & $p(x|y)=\mathcal{N}(x|y,1)$.

What is $p(x,y)$?

Sol. Let $(x_0,y_0)=(0,0)$, then

by Brook's Lemma,

$$p(x,y)=\frac{p(x|y)}{p(x_0|y)}\frac{p(y|x_0)}{p(y_0|x_0)}p(x_0,y_0)$$

$$\propto p(x|y)p(y|x_0)/p(x_0|y)$$

$$\propto e^{-\frac{(x-y)^2}{2}}e^{-\frac{y^2}{2}}/e^{-\frac{y^2}{2}}$$

$$=e^{-\frac{(x-y)^2}{2}} \qquad (*)$$

By $(*)$, we can get marginals of both x,y:

$$p(x)\propto 1 \text{ and } p(y)\propto 1.$$

But they are improper densities (i.e. integrates to $+\infty$). $\square$.

NOTE: I will first consider them independently, then combine the results.

Q3: Consider the BHM,

$$y|\beta,\sigma^2\sim\mathcal{N}(X\beta,\sigma^2 I_n)$$

$$\beta|\sigma^2\sim\mathcal{N}(\mu_\beta,\sigma^2 V_\beta)$$

$$\sigma^2\sim IG(a,b)$$

Assume $\tilde{y}|\beta,\sigma^2,y\sim\mathcal{N}(\tilde{X}\beta,\sigma^2 I_m)$

Find a,b,V_β^{-1} s.t.

(1) $p(\beta|\sigma^2,y)=\mathcal{N}(\hat\beta,\sigma^2(X^TX)^{-1})$ where $\hat\beta$ solves $X^TX\hat\beta=X^Ty$.

Sol. By Mm-formula,

$$\beta|\sigma^2,y\sim\mathcal{N}(Mm,\sigma^2M)$$

where $M=(V_\beta^{-1}+X^TX)^{-1}$,

$m=V_\beta^{-1}\mu_\beta+X^Ty$

& $\hat\beta=(X^TX)^{-1}X^Ty$.

If $\hat\beta=Mm$, $(X^TX)^{-1}=M$, then

$$\begin{cases}(X^TX)^{-1}X^Ty=(V_\beta^{-1}+X^TX)^{-1}(V_\beta^{-1}\mu_\beta+X^Ty)\\(V_\beta^{-1}+X^TX)^{-1}=(X^TX)^{-1}\end{cases}$$

The solution is: $V_\beta^{-1}=0$, μ_β arbitrary.

Such V_β does not exist but can be approximated by the following seq.

$$V_\beta^n=P\begin{bmatrix}\lambda_1^n&&\\&\ddots&\\&&\lambda_p^n\end{bmatrix}P^T \text{ where}$$

$PP^T=I=P^TP$ and $\lambda_1^n\geq\cdots\geq\lambda_p^n>0$.

next let $n\uparrow\infty$ & $\lambda_p^n\uparrow\infty$

we have

$$(V_\beta^n)^{-1}=P\begin{bmatrix}\frac{1}{\lambda_1^n}&&\\&\ddots&\\&&\frac{1}{\lambda_p^n}\end{bmatrix}P^T\xrightarrow{n\uparrow\infty}0.\ \square$$

& $a,b>0$.

(2): $\mathbb{E}(\sigma^2|y)=\hat{\sigma}^2$ from classical analysis.

Sol. Note in BHM,

$$\sigma^2|y \sim IG(a^*, b^*)$$

$$\begin{cases} a^* = a+\frac{n}{2} \\ b^* = b+\frac{1}{2}(y^Ty+\mu_\beta^T V_\beta^{-1}\mu_\beta - m^T M m) \end{cases}$$

$$\mathbb{E}(\sigma^2|y) = \frac{b^*}{a^*-1},$$

$$= \frac{b+\frac{1}{2}(y^Ty+\mu_\beta^T V_\beta^{-1}\mu_\beta - m^T M m)}{a+\frac{n}{2}-1}$$

$$= \frac{1}{2a+n-2}(2b+y^Ty+\mu_\beta^T V_\beta^{-1}\mu_\beta - m^T M m)$$

$$\hat{\sigma}^2 = \frac{1}{n}\|y-X\hat{\beta}\|_2^2 \quad \text{(Maximum likelihood est.)}$$

$$= \frac{1}{n}(y^Ty - y^T P_X y),\ P_X = X(X^TX)^{-1}X^T$$

In the special case $V_\beta^{-1}=0$,

$\mu_\beta^T V_\beta^{-1}\mu_\beta = 0$ and $m^T M m = y^T P_X y$,

$$\mathbb{E}(\sigma^2|y) = \frac{1}{2a+n-1}(2b+y^Ty-y^TP_Xy)$$

if $\mathbb{E}(\sigma^2|y)=\hat{\sigma}^2$, then

$$\begin{cases} a=1 \\ b=0 \\ V_\beta^{-1}=0 \end{cases}$$

$\square$

(3) $Var(\tilde{y}|\sigma^2, y) = ?$ From classical analysis.

Sol. From classical analysis, we have

$$Var(\tilde{y}) = Var(\tilde{X}\hat{\beta}+\varepsilon)$$

$$= \sigma^2[\tilde{X}(X^TX)^{-1}\tilde{X}^T + I_m]$$

For BHM, if $\beta|\sigma^2, y \sim N(Mm, \sigma^2 M)$, then:

$$p(\tilde{y}|\sigma^2,y) = \int p(\tilde{y}|\beta,\sigma^2,y)\,p(\beta|\sigma^2,y)\,d\beta$$

$$\propto \int e^{-\frac{1}{2\sigma^2}\|\tilde{y}-\tilde{X}\beta\|_2^2} e^{-\frac{1}{2\sigma^2}(\beta-Mm)^T M^{-1}(\beta-Mm)} d\beta$$

$$= \int e^{-\frac{1}{2\sigma^2}\{\tilde{y}^T\tilde{y}+\beta^T\tilde{X}^T\tilde{X}\beta - 2\tilde{y}^T\tilde{X}\beta + \beta^T M^{-1}\beta - 2m^T\beta + m^T M m\}} d\beta$$

$$\propto \int e^{-\frac{1}{2\sigma^2}(\beta-\Sigma(m+\tilde{X}^T\tilde{y}))^T\Sigma^{-1}(\beta-\Sigma(m+\tilde{X}^T\tilde{y}))} d\beta \times e^{-\frac{1}{2\sigma^2}[\tilde{y}^T\tilde{y}-(m+\tilde{X}^T\tilde{y})^T\Sigma(m+\tilde{X}^T\tilde{y})]}$$

$$\propto e^{-\frac{1}{2\sigma^2}(\tilde{y}-\mu_{\tilde{y}})^T(I_m-\tilde{X}\Sigma\tilde{X}^T)(\tilde{y}-\mu_{\tilde{y}})}$$

where

$\Sigma^{-1} = \tilde{X}^T\tilde{X}+M^{-1}$;

$\mu_{\tilde{y}}$ is something that has nothing to do with $Var(\tilde{y}|\sigma^2,y)$ so it is unimportant.

Thus,

$$Var(\tilde{y}|\sigma^2,y) = [I_m - \tilde{X}\Sigma\tilde{X}^T]^{-1}\sigma^2$$

$\Rightarrow$ Next page, see Woodbury's identity or Binomial inversion theorem

We have

$$[I_m - \tilde{X}\Sigma\tilde{X}^T]^{-1}$$

$$= I_m + \tilde{X}[\Sigma^{-1} - \tilde{X}^T\tilde{X}]^{-1}\tilde{X}^T$$

Plug-in $\Sigma^{-1} = M^{-1} + \tilde{X}^T\tilde{X}$

and note $M^{-1} = V_\beta^{-1} + X^TX$

We have

$$Var(\tilde{y} \mid \sigma^2, y) = \sigma^2(I_m + \tilde{X}(V_\beta^{-1} + X^TX)^{-1}\tilde{X}^T)$$

To make it equal to $Var(\tilde{y})$ in the classical analysis:

$$\tilde{X}(V_\beta^{-1} + X^TX)^{-1}\tilde{X}^T = \tilde{X}(X^TX)^{-1}\tilde{X}^T$$

$\Rightarrow V_\beta^{-1} = 0$. & This can be approx. by a sequence of matrices in (1).

(Put it together):

To reproduce results from classical analysis, we need to set:

$$\begin{cases} a = \frac{1}{2} \\ b = 0 \\ V_\beta^{-1} = 0. \end{cases}$$