Homework Solutions

Jun Sakamoto

January 24,2019

1 Homework1

1

(1)

$$\begin{aligned} E(y_i) &= 0 \times P(y_i = 0) + 1 \times P(y_i = 1) \\ &= P(y_i = 1) \\ &= P(y_i^* > 0) \\ &= P(X_i\beta + u_i > 0) = P(u_i > -X_i\beta) \\ &= P(u_i^* > -X_i\beta^*) = 1 - P(u_i^* \le -X_i\beta^*) \\ &= 1 - F(-X_i\beta^*) = F(X_i\beta^*), \end{aligned}$$

where $\beta^* = \frac{\beta}{\sigma}, u_i^* = \frac{u_i}{\sigma}$. Since the normal distribution is symmetric, then $1 - F(-X_i\beta^*) = F(X_i\beta^*)$
(2)

$$\begin{aligned} &= P(y_i^* \le 0) \\ &= P(X_i\beta + u_i \le 0) = P(u_i \le -X_i\beta) \\ &= P(u_i^* \le -X_i\beta^*) \\ &= F(-X_i\beta^*) = 1 - F(X_i\beta^*). \end{aligned}$$

Then,

$$\begin{aligned} f(y_i) &= [P(y_i = 1)]^{y_i}[P(y_i = 0)]^{1-y_i} \\ &= [F(X_i\beta^*)]^{y_i}[1 - F(X_i\beta^*)]^{1-y_i}. \end{aligned}$$

Therefore, the likelihood function is,

$$\begin{aligned} L(\beta^*) &= f(y_1, y_2, ..., y_n) \\ &= \prod_{i=1}^{n} f(y_i) = \prod_{i=1}^{n}[F(X_i\beta^*)]^{y_i}[1 - F(X_i\beta^*)]^{1-y_i}. \end{aligned}$$

(3)
We consider the log likelihood function of (2)

$$\begin{aligned} \frac{\partial l(\beta^*)}{\partial \beta^*} &= \sum_{i=1}^{n}\left(\frac{y_i X_i' f(X_i\beta^*)}{F(X_i\beta^*)} - \frac{(1-y_i)X_i' f(X_i\beta^*)}{1 - F(X_i\beta^*)}\right) \\ &= \sum_{i=1}^{n}\frac{[y_i - F(X_i\beta^*)]X_i' f(X_i\beta^*)}{[1 - F(X_i\beta^*)]F(X_i\beta^*)} = 0 \end{aligned}$$

(4)
From the above equation, β^* can be estimated, but β and σ^2 can not be estimated separately.
(5)
We define as $F_i = F(X_i\beta^*), f_i = f(X_i\beta^*)$. The second derivatives is,

$$\begin{aligned}\frac{\partial^2 l(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}} &= \sum_{i=1}^n \frac{X_i' \frac{\partial f_i}{\partial\beta^*}(y_i - F_i)}{[1-F_i]F_i} + \sum_{i=1}^n \frac{X_i' f_i \frac{\partial(y_i - F_i)}{\partial\beta^*}}{[1-F_i]F_i} \\ &+ \sum_{i=1}^n X_i' f_i (y_i - F_i) \frac{\partial[F_i(1-F_i)]^{-1}}{\partial\beta^*} \\ &= \sum_{i=1}^n \frac{X_i' X_i f_i'(y_i - F_i)}{[1-F_i]F_i} - \sum_{i=1}^n \frac{X_i' X_i f_i^2}{[1-F_i]F_i} \\ &+ \sum_{i=1}^n X_i' f_i (y_i - F_i) \frac{x_i f_i (1-2F_i)]}{[F_i(1-F_i)]^2}.\end{aligned}$$

Then the information matrix is

$$I(\beta^*) = -E\left[\frac{\partial^2 l(\beta^*)}{\partial\beta^*\partial\beta^{*\prime}}\right] = \sum_{i=1}^n \frac{X_i'[X_i f_i^2}{[1-F_i]F_i}.$$

Note that $E(y_i) = F_i$. Therefore, by using the fact that,

$$\sqrt{n}(\hat{\beta}^* - \beta^*) \to N\left(0, \lim_{n\to\infty}\left(\frac{1}{n}I(\beta^*\right)^{-1}\right),$$

so the asymptotic distribution of β^* is

$$\beta^* \sim N(\beta^*, I(\hat{\beta}^*)^{-1}).$$

(6)

$$\begin{aligned}E(y_i^*|y_i^* > 0) &= E(X_i\beta + u_i | X_i\beta + u_i > 0) \\ &= X_i\beta + E(u_i|u_i > -X_i\beta) \\ &= X_i\beta + \int_{-X_i\beta}^{\infty} \frac{u_i}{\sigma}\frac{f(u_i)}{1 - F(-X_i\beta)}du_i \\ &= \int_{-X_i\beta}^{\infty} \frac{u_i}{\sigma}\frac{\phi(\frac{u_i}{\sigma})}{1-\Phi(\frac{-X_i\beta}{\sigma})}du_i \\ &= \frac{\sigma\phi\left(\frac{-X_i\beta}{\sigma}\right)}{\Phi\left(\frac{-X_i\beta}{\sigma}\right)},\end{aligned}$$

where ϕ and Φ are the density function and the cumulative density function of the standard normal distribution, respectively.
(7)
Before truncation, the density function of y is,

$$f(y_i) = \frac{1}{\sigma}\phi\left(\frac{y_i - X_i\beta}{\sigma}\right).$$

and the probability of y_i^* is observed is,

$$\begin{aligned}P(y_i^* \text{is observed}) &= P(y_i > 0) \\ &= P(X_i\beta + u_i > 0) = P(u_i > -X_i\beta) \\ &= 1 - P(u_i \le -X_i\beta) = P(u_i \le X_i\beta) \\ &= \Phi\left(\frac{X_i\beta}{\sigma}\right)\end{aligned}$$

Dealing with the truncation, the likelihood function is,

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - X_i\beta}{\sigma}\right)}{\Phi\left(\frac{X_i\beta}{\sigma}\right)}.$$

Also, the log likelihood function is

$$\begin{aligned} l(\beta, \sigma^2) &= \sum_{i=1}^{n} \left(\log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\right] - \log\Phi\left(\frac{X_i\beta}{\sigma}\right)\right) \\ &= \sum_{i=1}^{n} \left\{\left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y_i - X_i\beta)^2\right) - \log\Phi\left(\frac{X_i\beta}{\sigma}\right)\right\}. \end{aligned}$$

(8)

$$\begin{aligned} \frac{\partial l(\beta, \sigma^2)}{\partial\beta} &= \sum_{i=1}^{n} \left(\frac{1}{\sigma^2}(y_i - X_i\beta) - \frac{\frac{1}{\sigma}\phi\left(\frac{X_i\beta}{\sigma}\right)}{\Phi\left(\frac{X_i\beta}{\sigma}\right)}\right) X_i' \\ &= \sum_{i=1}^{n} \left(\frac{1}{\sigma^2}(y_i - X_i\beta) - \frac{\frac{1}{\sigma}\phi_i}{\Phi_i}\right) X_i' = 0, \\ \frac{\partial l(\beta, \sigma^2)}{\partial\sigma^2} &= \sum_{i=1}^{n} \left(-\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y_i - X_I\beta)^2 + \frac{X_i\beta}{2\sigma^3}\frac{\phi\left(\frac{X_i\beta}{\sigma}\right)}{\Phi\left(\frac{X_i\beta}{\sigma}\right)}\right) \\ &= \sum_{i=1}^{n} \left(-\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y_i - X_i\beta)^2 + \frac{X_i\beta}{2\sigma^3}\frac{\phi_i}{\Phi_i}\right) = 0, \end{aligned}$$

where $\phi_i = \phi\left(\frac{X_i\beta}{\sigma}\right)$ and $\Phi_i = \Phi\left(\frac{X_i\beta}{\sigma}\right)$.

(9)

Use the iterative method such as the Newton-Raphson method.

(10)

What are the asymptotic distributions of the estimators of β and σ^2?

The second derivatives are,

$$\begin{aligned} \frac{\partial^2 l(\beta, \sigma^2))}{\partial\beta\partial\beta} &= \sum_{i=1}^{n} \left(-\frac{1}{\sigma^2}X_i + \frac{\frac{X_i\beta}{\sigma^2}\phi_i}{\Phi_i} + \frac{\frac{1}{\sigma^2}\phi_i^2}{\Phi_i^2}X_i\right) X_i' \\ \frac{\partial^2 l(\beta, \sigma^2))}{\partial\beta\partial\sigma^2} &= \sum_{i=1}^{n} \left(-\frac{1}{\sigma^4}(y_i - X_i\beta) + \frac{1}{2\sigma^3}\frac{\phi_i}{\Phi_i} + \frac{X_i\beta}{2\sigma^5}\frac{\phi}{\Phi}X_i\beta - \frac{X_i\beta}{2\sigma^4}\frac{\phi_i^2}{\Phi_i^2}\right) X_i' \\ \frac{\partial^2 l(\beta, \sigma^2))}{\partial\sigma^2\partial\beta} &= \sum_{i=1}^{n} \left(-\frac{1}{\sigma^4}(y_i - X_i\beta) + \frac{1}{2\sigma^3}\frac{\phi_i}{\Phi_i} + \frac{X_i\beta}{2\sigma^5}\frac{\phi}{\Phi}X_i\beta - \frac{X_i\beta}{2\sigma^4}\frac{\phi_i^2}{\Phi_i^2}\right) X_i \\ \frac{\partial^2 l(\beta, \sigma^2))}{\partial\sigma^2\partial\sigma^2} &= \sum_{i=1}^{n} \left(\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(y_i - X_i\beta)^2 - \frac{3X_i\beta}{4\sigma^5}\frac{\phi_i}{\Phi_i} + \frac{(X_i\beta)^3}{4\sigma^7}\frac{\phi_i}{\phi_i} + \frac{(X_i\beta)^2}{3\sigma^6}\frac{\phi_i^2}{\Phi_i^2}\right). \end{aligned}$$

Therefore, the information matrix is

$$I(\beta, \sigma^2) = \begin{pmatrix} \sum_{i=1}^{n} \left(-\frac{1}{\sigma^2}X_i + \frac{\frac{X_i\beta}{\sigma^2}\phi_i}{\Phi_i} + \frac{\frac{1}{\sigma^2}\phi_i^2}{\Phi_i^2}X_i\right) X_i' & \sum_{i=1}^{n} \left(-\frac{1}{\sigma^4}(y_i - X_i\beta) + \frac{1}{2\sigma^3}\frac{\phi_i}{\Phi_i} + \frac{X_i\beta}{2\sigma^5}\frac{\phi}{\Phi}X_i\beta - \frac{X_i\beta}{2\sigma^4}\frac{\phi_i^2}{\Phi_i^2}\right) X_i' \\ \sum_{i=1}^{n} \left(-\frac{1}{\sigma^4}(y_i - X_i\beta) + \frac{1}{2\sigma^3}\frac{\phi_i}{\Phi_i} + \frac{X_i\beta}{2\sigma^5}\frac{\phi}{\Phi}X_i\beta - \frac{X_i\beta}{2\sigma^4}\frac{\phi_i^2}{\Phi_i^2}\right) X_i & \sum_{i=1}^{n} \left(\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(y_i - X_i\beta)^2 - \frac{3X_i\beta}{4\sigma^5}\frac{\phi_i}{\Phi_i} + \frac{(X_i\beta)^3}{4\sigma^7}\frac{\phi_i}{\phi_i} + \frac{(X_i\beta)^2}{3\sigma^6}\frac{\phi_i^2}{\Phi_i^2}\right) \end{pmatrix}$$

Then by using the fact that

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\left(0, \lim_{n\to\infty}\left(\frac{1}{n}I(\theta)\right)^{-1}\right),$$

where $\theta = \beta, \sigma^2$, so the asymptotic distribution of $\hat{\theta}$ is:

$$\hat{\theta} \sim N(\theta, I(\hat{\theta})^{-1}).$$

(11)

$$\begin{aligned} P(y_i = 0) = & \quad P(y_i^* \le 0) = P(u_i \le -X_i\beta) \\ = & \quad P\left(\frac{u_i}{\sigma} \le -\frac{X_i\beta}{\sigma}\right) = F\left(-\frac{X_i\beta}{\sigma}\right) \\ = & \quad 1 - F\left(\frac{X_i\beta}{\sigma}\right) = 1 - \Phi\left(-\frac{X_i\beta}{\sigma}\right). \end{aligned}$$

Therefore, the likelihood function is,

where d_i is the indicator variable for $y_i > 0$.

Also, the log likelihood function is,

$$\begin{aligned} l(\beta, \sigma^2) = & \quad \textstyle\sum_{i=1}^n \left\{d_i \log\left[\frac{1}{\sigma}\phi_i\right] + (1-d_i)\log(1-\Phi_i)\right\} \\ = & \quad \textstyle\sum_{i=1}^n \left\{d_i\left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log 2\sigma^2 - \frac{1}{2\sigma^2}(y_i - X_i\beta)\right) + (1-d_i)\log(1-\Phi_i)\right\} \end{aligned}$$

(12)

$$\begin{aligned} \frac{\partial l(\beta,\sigma^2)}{\partial \beta} = & \quad \textstyle\sum_{i=1}^n \left\{d_i \frac{1}{\sigma^2}(y_i - X_i\beta) - (1-d_i)\frac{\phi}{\sigma(1-\Phi)}\right\} X_i' = 0 \\ \frac{\partial l(\beta,\sigma^2)}{\partial \sigma^2} = & \quad \textstyle\sum_{i=1}^n \left\{d_i\left(-\frac{1}{2\sigma^2} + \frac{(y_i - X_i\beta)^2}{2\sigma^4}\right) + (1-d_i)\frac{\phi X_i\beta}{2\sigma^3(1-\Phi)}\right\} = 0 \end{aligned}$$

(13)

Use the iterative method such as Newton-Rapson method.

(14)

By using the second derivatives, we can get the information matrix, that is,

$$I(\beta,\sigma^2) = \begin{pmatrix} \sum_{i=1}^n \left(-\frac{1}{\sigma^2}(X_i\beta\phi_i - \left[\frac{\phi^2}{1-\phi_i}\right] - \Phi_i)X_i'X_i\right) & \sum_{i=1}^n \left(-\frac{1}{2\sigma^3}((X_i\beta)^2\phi_i + \phi - \left[\frac{X_i\beta\phi^2}{1-\phi_i}\right])X_i'\right) \\ \sum_{i=1}^n \left(-\frac{1}{2\sigma^3}((X_i\beta)^2\phi_i + \phi - \left[\frac{X_i\beta\phi^2}{1-\phi_i}\right])X_i'\right) & \sum_{i=1}^n \left(-\frac{1}{4\sigma^4}((X_i\beta)^3\phi_i + (X_i\beta)\phi - \left[\frac{X_i\beta\phi^2}{1-\phi_i} - 2\Phi_i\right)\right) \end{pmatrix}$$

Then, by using the fact that,

$$\sqrt{n}(\hat{\theta} - \theta) \to N\left(0, \lim_{n\to\infty}\left(\frac{1}{n}I(\theta)\right)^{-1}\right)$$

where $\theta = \beta, \sigma^2$, so the asymptotic distribution of $\hat{\theta}$ is

$$\hat{\theta} \sim N(\theta, I(\hat{\theta})^{-1})$$

2

(15)

$$\begin{aligned} E(y_i) &= \sum_{y_i=1}^{\infty} y_i \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!} \\ &= \sum_{y_i=1}^{\infty} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{(y_i-1)!} \\ &= \lambda_i e^{-\lambda_i} \sum_{y_i=1}^{\infty} \frac{\lambda_i^{y_i-1}}{(y_i-1)!} \\ &= \lambda_i \end{aligned}$$

Note that $\sum_{y_i=1}^{\infty} \frac{\lambda_i^{y_i-1}}{(y_i-1)!} = e^{\lambda}$ from the Taylor expansion.
(16)
The likelihood function is

$$L(\beta) = \prod_{i=1}^{n} \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}.$$

Then, the log likelihood function is

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} \log \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \\ &= \sum_{i=1}^{n} (-e^{X_i \beta} + y_i X_i \beta - \log y_i!) \end{aligned}$$

(17)

$$\frac{\partial l(\beta)}{\partial \beta} = \sum_{i=1}^{n} (y_i - e^{X_i \beta}) X_i = 0$$

(18)
Use the iterative method such as Newton-Rapson method.
(19)
The Hessian of log likelihood function is

$$\frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} e^{X_i \beta} X_i X_i^{'}$$

Then the information matrix is

$$I(\hat{\beta}) = \frac{\partial^2 l(\beta)}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} e^{X_i \beta} X_i X_i^{'}$$

Then by using the fact that

$$\sqrt{n}(\hat{\theta} - \theta) \longrightarrow N\left(0, \lim_{n \to \infty} \left(\frac{1}{n} I(\theta)\right)^{-1}\right),$$

where $\theta = \beta$, so the asymptotic distribution of $\hat{\theta}$ is:

$$\hat{\theta} \sim N(\theta, I(\hat{\theta})^{-1}).$$

2 Homework 2

(1)
Define $e_i := (e_{i1}, e_{i2}, \cdots, e_{iT})'$. Then we get

$$\Sigma := E[e_i e_i'] = \begin{bmatrix} \sigma_v^2 + \sigma_u^2 & \sigma_v^2 & \cdots & \sigma_v^2 \\ \sigma_v^2 & \sigma_v^2 + \sigma_u^2 & \cdots & \sigma_v^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_v^2 & \sigma_v^2 & \cdots & \sigma_v^2 + \sigma_u^2 \end{bmatrix} (T \times T).$$

In addition, we also define $e := (e_1, e_2, \cdots, e_N)$. Then we get

$$\Omega = E[ee'] = \begin{bmatrix} \Sigma & 0 & 0 & 0 \\ 0 & \Sigma & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \Sigma \end{bmatrix} (NT \times NT).$$

(2)

Since Ω is a symmetric real matrix, there exists P such that

$$P^{-1}\Omega P == \begin{pmatrix} \lambda_1 & & & O \\ & \lambda_2 & & \\ & & \ddots & \\ O & & & \lambda_n \end{pmatrix} = \Lambda(PP' = I_n),$$

where λ_i is a eigen value. Thus we get

$$\begin{aligned} \Omega &= P\Lambda P^{-1} \\ &= P\Lambda P' \\ &= CC'(C := P\Lambda^{1/2}). \end{aligned}$$

Using above results, we can rewrite the model as

$$\begin{aligned} C^{-1}y &= C^{-1}X\beta + C^{-1}v + C^{-1}u \\ \leftrightarrow y^* &= X^*\beta + v^* + u^*, \end{aligned}$$

where $X := (X_1, X_2, \cdots, X_N)'$, $(NK \times K+1)$. Thus GLS is given by

$$b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

(3)

Just apply transformation of random variables for y. Then we get $u + v \sim N(0, \Omega)$.

(4)

The log-likelihood function is given by

$$l(\beta, \sigma_u^2, \sigma_v^2) = \frac{-nT}{2}\log(2\pi)^- - \frac{1}{2}\log|\Omega| - \frac{1}{2}(y - X\beta)'\Omega^{-1}(y - X\beta).$$

Thus $\overline{\beta}$ is given by

$$\overline{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

(5)

Clearly the are same.

(6)

Since $\overline{\beta}$ is MLE, it is consistent and efficient estimator. In addition, since $\overline{\beta} = b$ and b is also unbiased $\overline{\beta}$ is also unbiased.

(7)

$\hat{\beta}$ is given by

$$\hat{\beta} = \left\{ \sum_i \sum_t (x_{it} - \overline{x}_i)(x_{it} - \overline{x}_i)' \right\}^{-1} \sum_i \sum_t (x_{it} - \overline{x}_i)(y_{it} - \overline{y}_i).$$

(8)

Define

$$\begin{aligned}
\tilde{y}_{it} &:= y_{it} - \overline{y}_i, \\
\tilde{x}_{it} &:= x_{it} - \overline{x}_i, \\
\tilde{y}_i &:= \sum_t \tilde{y}_{it}, \\
\tilde{x}_i &:= \sum_t \tilde{x}_{it}.
\end{aligned}$$

Then $\hat{\beta}$ can be written as

$$\begin{aligned}
\hat{\beta} &= (\sum \tilde{x}_i' \tilde{x}_i)^{-1} \sum_i \tilde{x}_i' \tilde{y}_i : \\
&= (\sum \tilde{x}_i' \tilde{x}_i)^{-1} \sum_i \tilde{x}_i' (\tilde{x}\beta + \tilde{u}_i) \\
&= \beta + (\sum \tilde{x}_i' \tilde{x}_i)^{-1} \sum_i \tilde{x}_i' \tilde{u}_i.
\end{aligned}$$

Since $E[(\sum \tilde{x}_i' \tilde{x}_i)^{-1} \sum_i \tilde{x}_i' \tilde{u}_i] = 0$, $\forall i$, LLM gives $\hat{\beta} \to_p \beta$.

(9)

Under H_0, $\overline{\beta}$ is consistent and efficient. Thus we should use $\overline{\beta}$. Note that $\hat{\beta}$ is also consistent and not efficient.

(10)

Under H_1, only $\hat{\beta}$ is consistent. Thus we should use $\hat{\beta}$.