Solutions of the Homework #1 [*]

Kenta KUDO [†]

December 24th, 2019

This solution key was revised on January 17th. Because we correct the definition of ϕ_i, the related solutions, (8)–(14), are rewritten.

Question 1

(1)

Suppose that u_i, the disturbance term, follows the normal distribution $u_i \sim N(0, \sigma^2)$. The expectation of y_i is representsed as

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

where $\beta^* = \frac{\beta}{\sigma}$ and $u_i^* = \frac{u_i}{\sigma}$. Note that if we assume the normal distribution for the cumulative distribution function, $F(\cdot)$ becomes symmetric, and henceforce we have $1 - F(-X_i\beta^*) = F(X_i\beta^*)$.

(2)

The probability which y_i takes zero is calculated by

$$\begin{aligned}
\mathbb{P}(y_i = 0) &= \mathbb{P}(y_i^* \leq 0) \\
&= \mathbb{P}(X_i\beta + u_i \leq 0) \\
&= \mathbb{P}(u_i \leq -X_i\beta) \\
&= \mathbb{P}(u_i^* \leq -X_i\beta^*) \\
&= F(-X_i\beta^*) = 1 - F(X_i\beta^*).
\end{aligned} \tag{2}$$

[*] All comments welcome!

[†] E-mail: vge011kk@student.econ.osaka-u.ac.jp

Then, we can obtain the probability density function referring to Eq. (1) and Eq. (2) as follows:

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

Therefore, the likelihood function is

$$
\begin{aligned}
L(\beta^*) &= f(y_1, y_2, \cdots, y_n) \\
&= \prod_{i=1}^{n} f(y_i) \\
&= \prod_{i=1}^{n} [F(X_i\beta^*)]^{y_i}[1 - F(X_i\beta^*)]^{1-y_i}. \qquad (3)
\end{aligned}
$$

(3)

We consider the log likelihood function of Eq. (3) and derive the first-order condition as follows:

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

(4)

From the above equation, β^* can be estimated. Remind that β and σ^2 are NOT identified, so we can not estimate separately. One of the way to derive the estimator of β^* is the **method of scoring** as follows:

$$
\beta^{*(j+1)} = \beta^{*(j)} - \left(\frac{\partial^2 \log L(\beta^{(j)})}{\partial \beta \partial \beta'} \right)^{-1} \frac{\partial \log L(\beta^{(j)})}{\partial \beta^*}.
$$

(5)

We represent the c.d.f. and the p.d.f. as $F_i = F(X_i\beta^*)$ and $f_i = f(X_i\beta^*)$. The second derivative of the log-likelihood function is

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

Then the information matrix is given by

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

Note that $\mathbb{E}(y_i) = F_i$, because of the question (1). By using the CLT,

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

we can establish the asymptotic distribution of β^* as follows:

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

(6)

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

where ϕ and Φ are the p.d.f. and the c.d.f. of the standard normal distribution, respectively.

(7)

The likelihood function is

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

Also, the log likelihood function is

$$\begin{aligned}
l(\beta,\sigma^2) &= \sum_{i=1}^{n}\left(\log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\right] - \log\left[1-\Phi\left(-\frac{X_i\beta}{\sigma}\right)\right]\right) && (4) \\
&= \sum_{i=1}^{n}\left\{\left(-\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(y_i - X_i\beta)^2\right) - \log\left[1-\Phi\left(-\frac{X_i\beta}{\sigma}\right)\right]\right\}. && (5)
\end{aligned}$$

(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 $\phi_i = \phi\left(\frac{-X_i\beta}{\sigma}\right) = \phi\left(\frac{X_i\beta}{\sigma}\right)$ and $\Phi_i = \Phi\left(\frac{X_i\beta}{\sigma}\right) = 1 - \Phi\left(-\frac{X_i\beta}{\sigma}\right)$.

(9)

Use the iterative method such as the scoring method.

(10)

The second derivatives of the log-likelihood function are

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

The information matrix is derived from the above relationships. In addition, by using the CLT, we have

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

where $\theta = (\beta, \sigma^2)$ represents a parameter vector. Thus, the asymptotic distribution of $\hat{\theta}$ is

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

(11)

Recall that Φ which represents the c.d.f. of the standard normal distribution. Therefore, the probability which y_i takes zero is

$$\begin{aligned}\mathbb{P}(y_i = 0) =& \mathbb{P}(y_i^* \leq 0) = \mathbb{P}(u_i \leq -X_i\beta) \\ =& \mathbb{P}\left(\frac{u_i}{\sigma} \leq -\frac{X_i\beta}{\sigma}\right) = \Phi\left(-\frac{X_i\beta}{\sigma}\right) = 1 - \Phi_i,\end{aligned}$$

where $\Phi_i = \Phi\left(\frac{X_i\beta}{\sigma}\right) = 1 - \Phi\left(-\frac{X_i\beta}{\sigma}\right)$. Therefore, the likelihood function is,

$$L(\beta, \sigma^2) = \left[\frac{1}{\sigma}\phi\left(\frac{y_i - X_i\beta}{\sigma}\right)\right]^{d_i} \times \left[\Phi\left(-\frac{X_i\beta}{\sigma}\right)\right]^{1-d_i},$$

where d_i is the indicator variable for $y_i > 0$. Also, the log likelihood function is

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

Note that $\phi_i = \phi\left(\frac{X_i\beta}{\sigma}\right)$.

(12)

By the previous question, the first order conditions are given as

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

(13)

Use the iterative method such as the scoring method.

(14)

We can derive the second derivatives as

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

Then, the information matrix is given as follows:

$$I(\beta, \sigma^2) = \begin{pmatrix} \frac{\partial^2 l(\beta,\beta))}{\partial\beta\partial\beta} & \frac{\partial^2 l(\beta,\sigma^2))}{\partial\beta\partial\sigma^2} \\ \frac{\partial^2 l(\sigma^2,\beta))}{\partial\sigma^2\partial\beta} & \frac{\partial^2 l(\beta,\sigma^2))}{\partial\sigma^2\partial\sigma^2} \end{pmatrix}$$

Here, by using the CLT and we have

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

where θ implies a parameter vector, then the asymptotic distribution of $\hat{\theta}$ is

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