

and $\left(\frac{1}{n} \frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\right)^{-1} \left(\frac{1}{\sqrt{n}} \frac{\partial \log L(\theta; X)}{\partial \theta}\right)$ has the same asymptotic distribution as $\Sigma^{-1} \left(\frac{1}{\sqrt{n}} \frac{\partial \log L(\theta; X)}{\partial \theta}\right)$.

11. Optimization (最適化):

MLE of θ results in the following maximization problem:

$$\max_{\theta} \log L(\theta; x).$$

We often have the case where the solution of θ is not derived in closed form.

⇒ Optimization procedure

$$0 = \frac{\partial \log L(\theta; x)}{\partial \theta} = \frac{\partial \log L(\theta^*; x)}{\partial \theta} + \frac{\partial^2 \log L(\theta^*; x)}{\partial \theta \partial \theta'} (\theta - \theta^*).$$

Solving the above equation with respect to θ , we obtain the following:

$$\theta = \theta^* - \left(\frac{\partial^2 \log L(\theta^*; x)}{\partial \theta \partial \theta'}\right)^{-1} \frac{\partial \log L(\theta^*; x)}{\partial \theta}.$$

Replace the variables as follows:

$$\theta \longrightarrow \theta^{(i+1)}, \quad \theta^* \longrightarrow \theta^{(i)}.$$

Then, we have:

$$\theta^{(i+1)} = \theta^{(i)} - \left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial \theta \partial \theta'} \right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial \theta}.$$

⇒ **Newton-Raphson method** (ニュートン・ラフソン法)

Replacing $\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial \theta \partial \theta'}$ by $E \left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial \theta \partial \theta'} \right)$, we obtain the following optimization algorithm:

$$\begin{aligned} \theta^{(i+1)} &= \theta^{(i)} - \left(E \left(\frac{\partial^2 \log L(\theta^{(i)}; x)}{\partial \theta \partial \theta'} \right) \right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial \theta} \\ &= \theta^{(i)} + \left(I(\theta^{(i)}) \right)^{-1} \frac{\partial \log L(\theta^{(i)}; x)}{\partial \theta} \end{aligned}$$

⇒ **Method of Scoring** (スコア法)

2 Qualitative Dependent Variable (質的従属変数)

1. **Discrete Choice Model** (離散選択モデル)
2. **Limited Dependent Variable Model** (制限従属変数モデル)
3. **Count Data Model** (計数データモデル)

Usually, the regression model is given by:

$$y_i = X_i\beta + u_i, \quad u_i \sim N(0, \sigma^2), \quad i = 1, 2, \dots, n,$$

where y_i is a continuous type of random variable within the interval from $-\infty$ to ∞ .

When y_i is discrete or truncated, what happens?

2.1 Discrete Choice Model (離散選択モデル)

2.1.1 Binary Choice Model (二値選択モデル)

Example 1: Consider the regression model:

$$y_i^* = X_i\beta + u_i, \quad u_i \sim (0, \sigma^2), \quad i = 1, 2, \dots, n,$$

where y_i^* is unobserved, but y_i is observed as 0 or 1, i.e.,

$$y_i = \begin{cases} 1, & \text{if } y_i^* > 0, \\ 0, & \text{if } y_i^* \leq 0. \end{cases}$$

Consider the probability that y_i takes 1, i.e.,

$$\begin{aligned} P(y_i = 1) &= P(y_i^* > 0) = P(u_i > -X_i\beta) = P(u_i^* > -X_i\beta^*) = 1 - P(u_i^* \leq -X_i\beta^*) \\ &= 1 - F(-X_i\beta^*) = F(X_i\beta^*), \quad (\text{if the dist. of } u_i^* \text{ is symmetric.}) \end{aligned}$$

where $u_i^* = \frac{u_i}{\sigma}$, and $\beta^* = \frac{\beta}{\sigma}$ are defined.

(*) β^* can be estimated, but β and σ^2 cannot be estimated separately (i.e., β and σ^2 are not identified).

The distribution function of u_i^* is given by $F(x) = \int_{-\infty}^x f(z)dz$.

If u_i^* is standard normal, i.e., $u_i^* \sim N(0, 1)$, we call **probit model**.

$$F(x) = \int_{-\infty}^x (2\pi)^{-1/2} \exp(-\frac{1}{2}z^2)dz, \quad f(x) = (2\pi)^{-1/2} \exp(-\frac{1}{2}x^2).$$

If u_i^* is logistic, we call **logit model**.

$$F(x) = \frac{1}{1 + \exp(-x)}, \quad f(x) = \frac{\exp(-x)}{(1 + \exp(-x))^2}.$$

We can consider the other distribution function for u_i^* .

Likelihood Function: y_i is the following Bernoulli distribution:

$$f(y_i) = (P(y_i = 1))^{y_i}(P(y_i = 0))^{1-y_i} = (F(X_i\beta^*))^{y_i}(1 - F(X_i\beta^*))^{1-y_i}, \quad y_i = 0, 1.$$

[Review — Bernoulli Distribution (ベルヌイ分布)]

Suppose that X is a Bernoulli random variable. the distribution of X , denoted by $f(x)$, is:

$$f(x) = p^x(1 - p)^{1-x}, \quad x = 0, 1.$$

The mean and variance are:

$$\mu = E(X) = \sum_{x=0}^1 xf(x) = 0 \times (1 - p) + 1 \times p = p,$$

$$\sigma^2 = V(X) = E((X - \mu)^2) = \sum_{x=0}^1 (x - \mu)^2 f(x) = (0 - p)^2(1 - p) + (1 - p)^2 p = p(1 - p).$$

[End of Review]

The likelihood function is given by:

$$L(\beta^*) = f(y_1, y_2, \dots, y_n) = \prod_{i=1}^n f(y_i) = \prod_{i=1}^n (F(X_i\beta^*))^{y_i} (1 - F(X_i\beta^*))^{1-y_i},$$

The log-likelihood function is:

$$\log L(\beta^*) = \sum_{i=1}^n (y_i \log F(X_i\beta^*) + (1 - y_i) \log(1 - F(X_i\beta^*))),$$

Solving the maximization problem of $\log L(\beta^*)$ with respect to β^* , the first order condition is:

$$\begin{aligned} \frac{\partial \log L(\beta^*)}{\partial \beta^*} &= \sum_{i=1}^n \left(\frac{y_i X_i' f(X_i\beta^*)}{F(X_i\beta^*)} - \frac{(1 - y_i) X_i' f(X_i\beta^*)}{1 - F(X_i\beta^*)} \right) \\ &= \sum_{i=1}^n \frac{X_i' f(X_i\beta^*) (y_i - F(X_i\beta^*))}{F(X_i\beta^*) (1 - F(X_i\beta^*))} = \sum_{i=1}^n \frac{X_i' f_i (y_i - F_i)}{F_i (1 - F_i)} = 0, \end{aligned}$$

where $f_i \equiv f(X_i\beta^*)$ and $F_i \equiv F(X_i\beta^*)$. Remember that $f(x) \equiv \frac{dF(x)}{dx}$.

The second order condition is:

$$\begin{aligned}
 \frac{\partial^2 \log L(\beta^*)}{\partial \beta^* \partial \beta^{*'}} &= \sum_{i=1}^n \frac{X'_i \frac{\partial f_i}{\partial \beta^*} (y_i - F_i)}{F_i(1 - F_i)} + \sum_{i=1}^n \frac{X'_i f_i \frac{\partial (f_i - F_i)}{\partial \beta^*}}{F_i(1 - F_i)} \\
 &\quad + \sum_{i=1}^n X'_i f_i (y_i - F_i) \frac{\partial (F_i(1 - F_i))^{-1}}{\partial \beta^*} \\
 &= \sum_{i=1}^n \frac{X'_i X_i f'_i (y_i - F_i)}{F_i(1 - F_i)} - \sum_{i=1}^n \frac{X'_i X_i f_i^2}{F_i(1 - F_i)} + \sum_{i=1}^n X'_i f_i (y_i - F_i) \frac{X_i f_i (1 - 2F_i)}{(F_i(1 - F_i))^2}
 \end{aligned}$$

is a negative definite matrix.

For maximization, the method of scoring is given by:

$$\begin{aligned}
 \beta^{*(j+1)} &= \beta^{*(j)} + \left(-E \left(\frac{\partial^2 \log L(\beta^{*(j)})}{\partial \beta^* \partial \beta^{*'}} \right) \right)^{-1} \frac{\partial \log L(\beta^{*(j)})}{\partial \beta^*} \\
 &= \beta^{*(j)} + \left(\sum_{i=1}^n \frac{X'_i X_i (f_i^{(j)})^2}{F_i^{(j)} (1 - F_i^{(j)})} \right)^{-1} \sum_{i=1}^n \frac{X'_i f_i^{(j)} (y_i - F_i^{(j)})}{F_i^{(j)} (1 - F_i^{(j)})},
 \end{aligned}$$

where $F_i^{(j)} = F(X_i\beta^{*(j)})$ and $f_i^{(j)} = f(X_i\beta^{*(j)})$. Note that

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

because of $E(y_i) = F_i$.

It is known that

$$\sqrt{n}(\hat{\beta}^* - \beta^*) \longrightarrow N\left(0, \lim_{n \rightarrow \infty} \left(-\frac{1}{n} E\left(\frac{\partial^2 \log L(\beta^*)}{\partial\beta^* \partial\beta^{*'}}\right)\right)^{-1}\right),$$

where $\hat{\beta}^* \equiv \lim_{j \rightarrow \infty} \beta^{*(j)}$ denotes MLE of β^* .

Practically, we use the following normal distribution:

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

where $I(\hat{\beta}^*) = -E\left(\frac{\partial^2 \log L(\hat{\beta}^*)}{\partial\beta^* \partial\beta^{*'}}\right) = \sum_{i=1}^n \frac{X_i' X_i \hat{f}_i^2}{\hat{F}_i(1 - \hat{F}_i)}$, $\hat{f}_i = f(X_i\hat{\beta}^*)$ and $\hat{F}_i = F(X_i\hat{\beta}^*)$.

Thus, the significance test for β^* and the confidence interval for β^* can be constructed.

Another Interpretation: This maximization problem is equivalent to the nonlinear least squares estimation problem from the following regression model:

$$y_i = F(X_i\beta^*) + u_i,$$

where $u_i = y_i - F_i$ takes $u_i = 1 - F_i$ with probability $P(y_i = 1) = F(X_i\beta^*) = F_i$ and $u_i = -F_i$ with probability $P(y_i = 0) = 1 - F(X_i\beta^*) = 1 - F_i$.

Therefore, the mean and variance of u_i are:

$$E(u_i) = (1 - F_i)F_i + (-F_i)(1 - F_i) = 0,$$

$$\sigma_i^2 = V(u_i) = E(u_i^2) - (E(u_i))^2 = (1 - F_i)^2 F_i + (-F_i)^2 (1 - F_i) = F_i(1 - F_i).$$

The weighted least squares method solves the following minimization problem:

$$\min_{\beta^*} \sum_{i=1}^n \frac{(y_i - F(X_i\beta^*))^2}{\sigma_i^2}.$$

The first order condition is:

$$\sum_{i=1}^n \frac{X_i' f(X_i \beta^*) (y_i - F(X_i \beta^*))}{\sigma_i^2} = \sum_{i=1}^n \frac{X_i' f_i (y_i - F_i)}{F_i (1 - F_i)} = 0,$$

which is equivalent to the first order condition of MLE.

Thus, the binary choice model is interpreted as the nonlinear least squares.

Prediction: $E(y_i) = 0 \times (1 - F_i) + 1 \times F_i = F_i \equiv F(X_i \beta^*)$.

Example 2: Consider the two utility functions: $U_{1i} = X_i\beta_1 + \epsilon_{1i}$ and $U_{2i} = X_i\beta_2 + \epsilon_{2i}$.

A linear utility function is problematic, but we consider the linear function for simplicity of discussion.

We purchase a good when $U_{1i} > U_{2i}$ and do not purchase it when $U_{1i} < U_{2i}$.

We can observe $y_i = 1$ when we purchase the good, i.e., when $U_{1i} > U_{2i}$, and $y_i = 0$ otherwise.

$$\begin{aligned} P(y_i = 1) &= P(U_{1i} > U_{2i}) = P(X_i(\beta_1 - \beta_2) > -\epsilon_{1i} + \epsilon_{2i}) \\ &= P(-X_i\beta^* < \epsilon_i^*) = P(-X_i\beta^{**} < \epsilon_i^{**}) = 1 - F(-X_i\beta^{**}) = F(X_i\beta^{**}) \end{aligned}$$

where $\beta^* = \beta_1 - \beta_2$, $\epsilon_i^* = \epsilon_{1i} - \epsilon_{2i}$, $\beta^{**} = \frac{\beta^*}{\sigma^*}$ and $\epsilon_i^{**} = \frac{\epsilon_i^*}{\sigma^*}$.

We can estimate β^{**} , but we cannot estimate ϵ_i^* and σ^* , separately.

Mean and variance of ϵ_i^{**} are normalized to be zero and one, respectively.

If the distribution of ϵ_i^{**} is symmetric, the last equality holds.

We can estimate β^{**} by MLE as in Example 1.

Example 3: Consider the questionnaire:

$$y_i = \begin{cases} 1, & \text{if the } i\text{th person answers YES,} \\ 0, & \text{if the } i\text{th person answers NO.} \end{cases}$$

Consider estimating the following linear regression model:

$$y_i = X_i\beta + u_i.$$

When $E(u_i) = 0$, the expectation of y_i is given by:

$$E(y_i) = X_i\beta.$$

Because of the linear function, $X_i\beta$ takes the value from $-\infty$ to ∞ .