9.1 MLE: The Case of Single Regression Model

The regression model:

$$y_i = \beta_1 + \beta_2 x_i + u_i,$$

1. $u_i \sim N(0, \sigma^2)$ is assumed.

2. The density function of u_i is:

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} u_i^2\right).$$

Because $u_1, u_2, \cdots, u_n$ are mutually independently distributed, the joint density function of $u_1, u_2, \cdots, u_n$ is written as:

$$
\begin{aligned}
f(u_1, u_2, \cdots, u_n) &= f(u_1) f(u_2) \cdots f(u_n) \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} u_i^2\right)
\end{aligned}
$$

141

3. Using the transformation of variable ($u_i = y_i - \beta_1 - \beta_2 x_i$), the joint density function of $y_1, y_2, \cdots, y_n$ is given by:

$$f(y_1, y_2, \cdots, y_n) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(y_i - \beta_1 - \beta_2 x_i)^2\right)$$
$$\equiv L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n).$$

$L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$ is called the likelihood function.

$\log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$ is called the log-likelihood function.

$$\log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$$
$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_t - \beta_1 - \beta_2 x_i)^2$$

4. **Transformation of Variable (   ) — Review**:

Suppose that the density function of a random variable X is $f_x(x)$.

Defining $X = g(Y)$, the density function of Y, $f_y(y)$, is given by:

$$f_y(y) = f_x(g(y)) \left| \frac{\mathrm{d}g(y)}{\mathrm{d}y} \right|.$$

In the case where X and $g(Y)$ are $n \times 1$ vectors, $\left| \frac{\mathrm{d}g(y)}{\mathrm{d}y} \right|$ should be replaced by $\left| \frac{\partial g(y)}{\partial y'} \right|$, which is an absolute value of a determinant of the matrix $\frac{\partial g(y)}{\partial y'}$.

Example: When $X \sim U(0, 1)$, derive the density function of $Y = -\log(X)$.

$$f_x(x) = 1$$

$X = \exp(-Y)$ is obtained.

Therefore, the density function of Y, $f_y(y)$, is given by:

$$f_y(y) = \left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| f_x(g(y)) = |-\exp(-y)| = \exp(-y)$$

5. **[Going back to 3]:** Given the observed data $y_1, y_2, \cdots, y_n$, the likelihood function $L(\beta_1, \beta_2, \sigma^2|y_1, y_2, \cdots, y_n)$, or the log-likelihood function $\log L(\beta_1, \beta_2, \sigma^2|y_1, y_2, \cdots, y_n)$ is maximized with respect to $(\beta_1, \beta_2, \sigma^2)$.

Solve the following three simultaneous equations:

$$\frac{\partial \log L(\beta_1, \beta_2, \sigma^2|y_1, y_2, \cdots, y_n)}{\partial \beta_1} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i) = 0,$$

$$\frac{\partial \log L(\beta_1, \beta_2, \sigma^2|y_1, y_2, \cdots, y_n)}{\partial \beta_2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i) x_i = 0,$$

$$\frac{\partial \log L(\beta_1, \beta_2, \sigma^2|y_1, y_2, \cdots, y_n)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2 = 0.$$

The solutions of $(\beta_1, \beta_2, \sigma^2)$ are called the maximum likelihood estimates, denoted by $(\tilde{\beta}_1, \tilde{\beta}_2, \tilde{\sigma}^2)$.

The maximum likelihood estimates are:

$$\tilde{\beta}_2 = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}, \qquad \tilde{\beta}_1 = \overline{y} - \tilde{\beta}_2 \overline{x}, \qquad \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \tilde{\beta}_1 - \tilde{\beta}_2 x_i)^2.$$

The MLE of σ^2 is divided by n, not $n - 2$.

9.2   MLE: The Case of Multiple Regression Model I

1. Multivariate Normal Distribution: $X : n \times 1$ and $X \sim N(\mu, \Sigma)$

The density function of X is:

$$f(x) = (2\pi)^{n/2} |\Sigma|^{-1/2} \exp\left(-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right).$$

2. Regression model: $y = X\beta + u,$ $u \sim N(0, \sigma^2 I_n)$

146

Transformation of Variables from u to y:

$$f_u(u) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} u'u\right)$$

$$\begin{aligned} f_y(y) &= f_u(y - X\beta)\left|\frac{\partial u}{\partial y'}\right| \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right) \\ &= L(\theta; y, X), \end{aligned}$$

where $\theta = (\beta, \sigma^2)$, because of $\dfrac{\partial u}{\partial y'} = I_n$.

Therefore, the log-likelihood function is:

$$\log L(\theta; y, X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta),$$

Note that $|\Sigma|^{-1/2} = |\sigma^2 I_n|^{-1/2} = \sigma^{-n/2}$.

3. $\max_{\theta} \quad \log L(\theta; y, X)$

(FOC) $\dfrac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$

(SOC) $\dfrac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

We obtain MLE of β and σ^2:

$$\tilde{\beta} = (X'X)^{-1}X'y, \qquad \tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'(y - X\tilde{\beta})}{n},$$

where $\tilde{\sigma}^2$ is divided by n, not $n - k$.

4. Fisher's information matrix is:

$$I(\theta) = -\mathrm{E}\Big(\frac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}\Big)$$

148

The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the variance - covariance matrix for unbiased estimators of θ .

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}$$

For large n, we approximately obtain: $\begin{pmatrix} \tilde{\beta} \\ \tilde{\sigma}^2 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \sigma^2(X'X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix} \right).$

9.3   MLE: The Case of Multiple Regression Model II

1. Regression model: $y = X\beta + u, \quad u \sim N(0, \sigma^2\Omega)$

Transformation of Variables from u to y:

$$f_u(u) = (2\pi\sigma^2)^{-n/2}|\Omega|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}u'\Omega^{-1}u\right)$$

$$f_y(y) = f_u(y - X\beta)\left|\frac{\partial u}{\partial y'}\right|$$

149

$$= (2\pi\sigma^2)^{-n/2}|\Omega|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta)\right)$$

$$= L(\theta; y, X),$$

where $\theta = (\beta, \sigma^2)$, because of $\dfrac{\partial u}{\partial y'} = I_n$.

The log-likelihood function is:

$$\log L(\theta; y, X) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2}\log|\Omega| - \frac{1}{2\sigma^2}(y - X\beta)'\Omega^{-1}(y - X\beta),$$

where $\theta = (\beta, \sigma^2)$.

2. $\max\limits_{\theta} \ \log L(\theta; y, X)$

   (FOC) $\dfrac{\partial \log L(\theta; y, X)}{\partial \theta} = 0$

   (SOC) $\dfrac{\partial^2 \log L(\theta; y, X)}{\partial \theta \partial \theta'}$ is a negative definite matrix.

150

Then, we obtain MLE of β and σ^2:

$$\tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \qquad \tilde{\sigma}^2 = \frac{(y - X\tilde{\beta})'\Omega^{-1}(y - X\tilde{\beta})}{n}$$

3. Fisher's information matrix is defined as:

$$I(\theta) = -E\left(\frac{\partial^2 \log L(\theta; y, X)}{\partial\theta\partial\theta'}\right)$$

The inverse of the information matrix, $I(\theta)^{-1}$, provides a lower bound of the variance - covariance matrix for unbiased estimators of θ, which is given by:

$$I(\theta)^{-1} = \begin{pmatrix} \sigma^2(X'\Omega^{-1}X)^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}$$