

Proof of the above equality:

$$\int L(\theta; x) dx = 1$$

Take a derivative with respect to θ .

$$\int \frac{\partial L(\theta; x)}{\partial \theta} dx = 0$$

(We assume that (i) the domain of x does not depend on θ and (ii) the derivative $\frac{\partial L(\theta; x)}{\partial \theta}$ exists.)

Rewriting the above equation, we obtain:

$$\int \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x) dx = 0,$$

i.e.,

$$E\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right) = 0.$$

Again, differentiating the above with respect to θ , we obtain:

$$\begin{aligned}
 & \int \frac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'} L(\theta; x) dx + \int \frac{\partial \log L(\theta; x)}{\partial \theta} \frac{\partial L(\theta; x)}{\partial \theta'} dx \\
 &= \int \frac{\partial^2 \log L(\theta; x)}{\partial \theta \partial \theta'} L(\theta; x) dx + \int \frac{\partial \log L(\theta; x)}{\partial \theta} \frac{\partial \log L(\theta; x)}{\partial \theta'} L(\theta; x) dx \\
 &= E\left(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\right) + E\left(\frac{\partial \log L(\theta; X)}{\partial \theta} \frac{\partial \log L(\theta; X)}{\partial \theta'}\right) = 0.
 \end{aligned}$$

Therefore, we can derive the following equality:

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

where the second equality utilizes $E\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right) = 0$.

3. Cramer-Rao Lower Bound (クラメール・ラオの下限): $(I(\theta))^{-1}$

Suppose that an estimator of θ is given by $s(X)$.

The expectation of $s(X)$ is:

$$E(s(X)) = \int s(x)L(\theta; x)dx.$$

Differentiating the above with respect to θ ,

$$\begin{aligned}\frac{\partial E(s(X))}{\partial \theta} &= \int s(x) \frac{\partial L(\theta; x)}{\partial \theta} dx = \int s(x) \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x) dx \\ &= \text{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)\end{aligned}$$

For simplicity, let $s(X)$ and θ be scalars.

Then,

$$\begin{aligned}\left(\frac{\partial E(s(X))}{\partial \theta}\right)^2 &= \left(\text{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)\right)^2 = \rho^2 V(s(X)) V\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right) \\ &\leq V(s(X)) V\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),\end{aligned}$$

where ρ denotes the correlation coefficient between $s(X)$ and $\frac{\partial \log L(\theta; X)}{\partial \theta}$, i.e.,

$$\rho = \frac{\text{Cov}\left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta}\right)}{\sqrt{V(s(X))} \sqrt{V\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right)}}.$$

Note that $|\rho| \leq 1$.

Therefore, we have the following inequality:

$$\left(\frac{\partial E(s(X))}{\partial \theta}\right)^2 \leq V(s(X)) V\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

i.e.,

$$V(s(X)) \geq \frac{\left(\frac{\partial E(s(X))}{\partial \theta}\right)^2}{V\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right)}$$

Especially, when $E(s(X)) = \theta$,

$$V(s(X)) \geq \frac{1}{-E\left(\frac{\partial^2 \log L(\theta; X)}{\partial \theta^2}\right)} = (I(\theta))^{-1}.$$

Even in the case where $s(X)$ is a vector, the following inequality holds.

$$V(s(X)) \geq (I(\theta))^{-1},$$

where $I(\theta)$ is defined as:

$$\begin{aligned} I(\theta) &= -E\left(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\right) \\ &= E\left(\frac{\partial \log L(\theta; X)}{\partial \theta} \frac{\partial \log L(\theta; X)}{\partial \theta'}\right) = V\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right). \end{aligned}$$

The variance of any unbiased estimator of θ is larger than or equal to $(I(\theta))^{-1}$.

4. Asymptotic Normality of MLE:

Let $\tilde{\theta}$ be MLE of θ .

As n goes to infinity, we have the following result:

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

where it is assumed that $\lim_{n \rightarrow \infty} \left(\frac{I(\theta)}{n}\right)$ converges.

That is, when n is large, $\tilde{\theta}$ is approximately distributed as follows:

$$\tilde{\theta} \sim N\left(\theta, (I(\theta))^{-1}\right).$$

Suppose that $s(X) = \tilde{\theta}$.

When n is large, $V(s(X))$ is approximately equal to $(I(\theta))^{-1}$.

5. 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.

$\implies$ 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:

$$\theta \longrightarrow \theta^{(i+1)}$$

$$\theta^* \longrightarrow \theta^{(i)}$$

Then, we have:

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

$\Rightarrow$ **Newton-Raphson method** (ニュートン・ラプソン法)

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:

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

$\Rightarrow$ **Method of Scoring** (スコア法)

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

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

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

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

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

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

$$\begin{aligned} \log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \dots, y_n) \\ = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i)^2 \end{aligned}$$

4. Transformation of Variable (変数変換) — 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:

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

In the case where X and $g(Y)$ are $n \times 1$ vectors, $\left| \frac{dg(y)}{dy} \right|$ 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:

$$f_y(y) = \left| \frac{dx}{dy} \right| f_x(g(y)) = | -\exp(-y) | = \exp(-y)$$

5. Given the observed data $y_1, y_2, \dots, y_n$, the likelihood function $L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \dots, y_n)$, or the log-likelihood function $\log L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \dots, y_n)$ is maximized with respect to $(\beta_1, \beta_2, \sigma^2)$.

Solve the following three simultaneous equations:

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

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

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

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:

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

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

Transformation of Variables from u to y :

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

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