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

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

The expectation of $s(X)$ is:

$$\mathrm{E}(s(X)) = \int s(x) L(\theta; x) \mathrm{d}x.$$

Differentiating the above with respect to θ,

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

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

Then,

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

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

i.e.,

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

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

Therefore, we have the following inequality:

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

i.e.,

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

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

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

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

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

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

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

4. Asymptotic Normality of MLE:

Let $\tilde{\theta}$ be MLE of θ.

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

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

where it is assumed that $\displaystyle\lim_{n\to\infty} \left(\frac{I(\theta)}{n} \right)$ 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, $\mathrm{V}(s(X))$ is approximately equal to $(I(\theta))^{-1}$.

5. Optimization (最適化):

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

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:

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

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

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

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

$\Longrightarrow$ **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 $u_1, u_2, \cdots, u_n$ are mutually independently distributed, the joint den-

sity function of $u_1, u_2, \cdots, u_n$ is written as:

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

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:

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

$L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n)$ 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.

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

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

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. 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 $(\alpha, \beta, \sigma^2)$.

Solve the following three simultaneous equations:

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

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

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

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:

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

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