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.,

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

Especially, when $\mathrm{E}(s(X)) = \theta$, i.e., when $s(X)$ is an unbiased estimator of θ, the numerator of the right-hand side leads to one.

Therefore, we obtain:

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

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

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

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

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

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

Thus, $(I(\theta))^{-1}$ results in the lower bound of the variance of any unbiased estimator of θ.

6. 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 $\lim_{n\to\infty}\left(\dfrac{I(\theta)}{n}\right)$ converges.

$\longrightarrow$ The proof will be shown later.

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 $(I(\theta))^{-1}$.

7. **Optimization (最適化):**

MLE of θ results in the following maximization problem:

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

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

$\Longrightarrow$ Optimization procedure

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

Solving the above equation with respect to θ, we approximately 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:

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

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)**

Convergence speed might be improved, compared with Newton-Raphson method.

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:

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

   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:

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

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:

$$
\begin{aligned}
f_y(y_1, y_2, \cdots, y_n) &= \prod_{i=1}^{n} f_u(y_i - \beta_1 - \beta_2 x_i) \Big| \frac{\mathrm{d}u_i}{\mathrm{d}y_i} \Big| \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\Big(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta_1 - \beta_2 x_i)^2 \Big) \\
&\equiv L(\beta_1, \beta_2, \sigma^2 | y_1, y_2, \cdots, y_n).
\end{aligned}
$$

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

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

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

Transformation of Variables from u to y:

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

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

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