$$
\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta} - \mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big)}{\sqrt{\mathrm{V}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big)}}
$$

$$
= \frac{\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta} - \mathrm{E}\Big(\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)}{\sqrt{\mathrm{V}\Big(\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)}} = \frac{\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta} - \frac{1}{n}\mathrm{E}\Big(\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)}{\frac{1}{\sqrt{n}}\sqrt{\frac{1}{n}\mathrm{V}\Big(\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)}}
$$

$$
= \frac{\frac{1}{n}\frac{\partial \log L(\theta;X)}{\partial\theta}}{\frac{1}{\sqrt{n}}\sqrt{\frac{1}{n}I(\theta)}} = \frac{\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta}}{\sqrt{\frac{1}{n}I(\theta)}} \longrightarrow N(0,1)
$$

We have assumed that $\lim_{n\to\infty}\frac{1}{n}I(\theta) = \sigma^2$.

Therefore,

$$\frac{1}{\sqrt{n}} \frac{\partial \log L(\theta; X)}{\partial \theta} \longrightarrow N(0, \sigma^2)$$

In the case where θ is a vector, we have:

$$\frac{1}{\sqrt{n}} \frac{\partial \log L(\theta; X)}{\partial \theta} \longrightarrow N(0, \Sigma),$$

where $\Sigma = \lim_{n\to\infty} \frac{1}{n} I(\theta)$, which is a $k \times k$ matrix when θ is a $k \times 1$ vector, and $X = (X_1, X_2, \cdots, X_n)$.

Now, replacing θ by $\tilde{\theta}$, consider the asymptotic distribution of

$$\frac{1}{\sqrt{n}} \frac{\partial \log L(\tilde{\theta}; X)}{\partial \theta},$$

which is expanded around $\tilde{\theta} = \theta$ as follows:

$$0 = \frac{1}{\sqrt{n}} \frac{\partial \log L(\tilde{\theta}; X)}{\partial \theta} \approx \frac{1}{\sqrt{n}} \frac{\partial \log L(\theta; X)}{\partial \theta} + \frac{1}{\sqrt{n}} \frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'} (\tilde{\theta} - \theta).$$

Therefore,

$$-\frac{1}{\sqrt{n}}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}(\tilde{\theta}-\theta) \approx \frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta} \longrightarrow N(0,\Sigma).$$

The left-hand side is rewritten as:

$$-\frac{1}{\sqrt{n}}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}(\tilde{\theta}-\theta) = \sqrt{n}\left(-\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\right)(\tilde{\theta}-\theta).$$

Then,

$$\sqrt{n}(\tilde{\theta}-\theta) \approx \Big(-\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\Big)^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)$$

$$\longrightarrow N(0,\Sigma^{-1}\Sigma\Sigma^{-1}) = N(0,\Sigma^{-1}).$$

Using the law of large number, note that

$$-\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'} \longrightarrow \lim_{n\to\infty}\frac{1}{n}\left(-\mathrm{E}\Big(\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\Big)\right)$$

$$= \lim_{n\to\infty}\frac{1}{n}\left(\mathrm{V}\Big(\frac{\partial \log L(\theta;X)}{\partial}\Big)\right) = \lim_{n\to\infty}\frac{1}{n}I(\theta) = \Sigma,$$

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:

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

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

$\Longrightarrow$ Optimization procedure

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

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:

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

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)**

3　変数変換

確率変数 X の密度関数を $f(x)$，分布関数を $F(x) \equiv \mathrm{P}(X < x)$ とする。$Y = aX + b$ とするとき，Y の密度関数 $g(y)$ を求める。

Y の分布関数を $G(y)$ として，次のように変形できる。

$$
\begin{aligned}
G(y) &= \mathrm{P}(Y < y) = \mathrm{P}(aX + b < y) \\
&= \begin{cases} \mathrm{P}\left(X < \dfrac{y-b}{a}\right), & a > 0 \text{ のとき} \\ \mathrm{P}\left(X > \dfrac{y-b}{a}\right), & a < 0 \text{ のとき} \end{cases} \\
&= \begin{cases} \mathrm{P}\left(X < \dfrac{y-b}{a}\right), & a > 0 \text{ のとき} \\ 1 - \mathrm{P}\left(X < \dfrac{y-b}{a}\right), & a < 0 \text{ のとき} \end{cases} \\
&= \begin{cases} F\left(\dfrac{y-b}{a}\right), & a > 0 \text{ のとき} \\ 1 - F\left(\dfrac{y-b}{a}\right), & a < 0 \text{ のとき} \end{cases}
\end{aligned}
$$

分布関数と密度関数との関係は,

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x) \qquad \frac{\mathrm{d}G(x)}{\mathrm{d}x} = g(x)$$

であるので，Y の密度関数は,

$$g(y) = \frac{\mathrm{d}G(y)}{\mathrm{d}y} = \begin{cases} \frac{1}{a} f\left(\frac{y-b}{a}\right), & a > 0 \text{ のとき} \\ -\frac{1}{a} f\left(\frac{y-b}{a}\right), & a < 0 \text{ のとき} \end{cases}$$
$$= \left|\frac{1}{a}\right| f\left(\frac{y-b}{a}\right)$$

と表される。

一般に，確率変数 X の密度関数を $f(x)$ とする。単調変換 $X = h(Y)$ とするとき，Y の密度関数 $g(y)$ は,

$$g(y) = |h'(y)| f(h(y))$$

となる。

4　回帰分析への応用

回帰モデル

$$Y_i = \alpha + \beta X_i + u_i \qquad i = 1, 2, \cdots, n$$

$u_1, u_2, \cdots, u_n$ は互いに独立で，すべての i について $u_i \sim \mathrm{N}(0, \sigma^2)$ を仮定する。

u_i の密度関数は,

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

となる。

Y_i の密度関数 $g(Y_i)$ は,

$$g(Y_i) = |h'(Y_i)| f(h(Y_i))$$

によって求められる。

この場合，$h(Y_i) = Y_i - \alpha - \beta X_i$ なので，$h'(Y_i) = 1$ となる。

したがって，Y_i の密度関数は,

$$
\begin{aligned}
g(Y_i) &= |h'(Y_i)| f(h(Y_i)) = f(h(Y_i)) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Bigl(-\frac{1}{2\sigma^2}(Y_i - \alpha - \beta X_i)^2\Bigr)
\end{aligned}
$$

となる。

$u_1, u_2, \cdots, u_n$ は互いに独立であれば，$Y_1, Y_2, \cdots, Y_n$ も互いに独立になるので，$Y_1, Y_2, \cdots, Y_n$ の結合密度関数は,

$$
g(Y_1, Y_2, \cdots, Y_n) = \prod_{i=1}^{n} g(Y_i) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Bigl(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2\Bigr)
$$

となる。これは α, β, σ^2 の関数となっている。

よって，尤度関数は,

$$
l(\alpha, \beta, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Bigl(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(Y_i - \alpha - \beta X_i)^2\Bigr)
$$