# 12　推定量の求め方

## 12.1　最小二乗法

・$n$ 個のデータ (実現値)： $x_1, x_2, \cdots, x_n$
・背後に対応する確率変数を仮定： $X_1, X_2, \cdots, X_n$
・$\mathrm{E}(X_i) = \mu$，$\mathrm{V}(X_i) = \sigma^2$ を仮定
母数 $(\mu, \sigma^2)$ を推定する。

観測データ $x_1, x_2, \cdots, x_n$ をもとにして，$\mu$ の最小二乗推定値を求める。

$$\min_{\mu} \sum_{i=1}^{n} (x_i - \mu)^2$$

$\mu$ の解を $\widehat{\mu}$ とすると，

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

となり，$\widehat{\mu} \equiv \overline{x}$ を得る。

すなわち，

$$\frac{\mathrm{d} \sum_{i=1}^{n} (x_i - \mu)^2}{\mathrm{d}\mu} = 0$$

を $\mu$ について解く。

$\mu$ の最小二乗推定量はデータ $x_i$ を対応する確率変数 $X_i$ で置き換えて，

$$\widehat{\mu} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

となり，$\widehat{\mu} \equiv \overline{X}$ を得る ($\widehat{\mu}$ について，推定値と推定量は同じ記号を使っている)。

以上を回帰分析に応用すると，

$$\min_{\alpha,\beta} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2$$

を解くことになる。

すなわち，

$$\frac{\partial \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2}{\partial \alpha} = 0$$
$$\frac{\partial \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2}{\partial \beta} = 0$$

の連立方程式を $\alpha, \beta$ について解く。

## 12.2　最尤法

$n$ 個の確率変数 $X_1, X_2, \cdots, X_n$ は互いに独立で，同じ確率分布 $f(x) \equiv f(x;\theta)$ とする。ただし，$\theta$ は母数で，例えば，$\theta = (\mu, \sigma^2)$ である。

$X_1, X_2, \cdots, X_n$ の結合分布は，互いに独立なので，

$$f(x_1, x_2, \cdots, x_n; \theta) \equiv \prod_{i=1}^{n} f(x_i;\theta)$$

と表される。

観測データ $x_1, x_2, \cdots, x_n$ を与えたもとで，$\prod_{i=1}^{n} f(x_i;\theta)$ は $\theta$ の関数として表される。すなわち，

$$l(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

となる。
$l(\theta)$ を尤度関数と呼ぶ。

$$\max_{\theta}\ l(\theta)$$

となる $\theta$ を最尤推定値 $\widehat{\theta} = \widehat{\theta}(x_1, x_2, \cdots, x_n)$ と呼ぶ。

データ $x_1, x_2, \cdots, x_n$ を確率変数 $X_1, X_2, \cdots, X_n$ で置き換えて，$\widehat{\theta} = \widehat{\theta}(X_1, X_2, \cdots, X_n)$ を最尤推定量と呼ぶ。

$$\max_{\theta}\ l(\theta)$$

と

$$\max_{\theta}\ \log l(\theta)$$

の $\theta$ の解はともに同じものであることに注意。$\log l(\theta)$ を対数尤度関数と呼ぶ。

**最尤推定量の性質**：　$n$ が大きいとき，

$$\widehat{\theta} \sim N(\theta, \sigma_\theta^2)$$

ただし，

$$\sigma_\theta^2 = \frac{1}{\sum_{i=1}^{n} \mathrm{E}\Big[\Big(\frac{\mathrm{d} \log f(X_i;\theta)}{\mathrm{d}\theta}\Big)^2\Big]} = -\frac{1}{\sum_{i=1}^{n} \mathrm{E}\Big[\frac{\mathrm{d}^2 \log f(X_i;\theta)}{\mathrm{d}\theta^2}\Big]}$$

$\theta$ がベクトル $(k \times 1)$ の場合，$n$ が大きいとき，

$$\widehat{\theta} \sim N(\theta, \Sigma_\theta)$$

ただし，

$$\Sigma_\theta = \Big(\sum_{i=1}^{n} \mathrm{E}\Big[\Big(\frac{\partial \log f(X_i;\theta)}{\partial \theta}\Big)\Big(\frac{\partial \log f(X_i;\theta)}{\partial \theta}\Big)'\Big]\Big)^{-1}$$
$$= -\Big(\sum_{i=1}^{n} \mathrm{E}\Big[\frac{\partial^2 \log f(X_i;\theta)}{\partial \theta \partial \theta'}\Big]\Big)^{-1}$$

**例 1**：　正規母集団 $\mathrm{N}(\mu, \sigma^2)$ からの標本値 $x_1$, $x_2$, $\cdots$, $x_n$ を用いて，
(1)　$\sigma^2$ が既知のとき，$\mu$ の最尤推定値と最尤推定量
(2)　$\sigma^2$ が未知のとき，$\mu$ と $\sigma^2$ の最尤推定値と最尤推定量
をそれぞれ求める。

［**解**］$\mathrm{N}(\mu, \sigma^2)$ の密度関数は，

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big)$$

となる。したがって，互いに独立な $X_1$, $X_2$, $\cdots$, $X_n$ の結合分布は，

$$f(x_1, x_2, \cdots, x_n; \mu, \sigma^2) \equiv \prod_{i=1}^{n} f(x_i; \mu, \sigma^2)$$
$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x_i-\mu)^2\Big)$$
$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\Big)$$

となる。

(1) $\sigma^2$ が既知のとき，尤度関数 $l(\mu)$ は，

$$l(\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\Big)$$

となる。
$l(\mu)$ を最大にする $\mu$ と $\log l(\mu)$ を最大にする $\mu$ は同じになる。

したがって，対数尤度関数は，

$$\log l(\mu) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

となり，

$$\frac{\mathrm{d}\log l(\mu)}{\mathrm{d}\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$$

となる $\mu$ を求める。$\mu$ の解を $\widehat{\mu}$ とすると，$\mu$ の最尤推定値は，

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i \equiv \overline{x}$$

を得る。

さらに，観測値 $x_1$, $x_2$, $\cdots$, $x_n$ をその確率変数 $X_1$, $X_2$, $\cdots$, $X_n$ で置き換えて，$\mu$ の最尤推定量は，

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i \equiv \overline{X}$$

となる。
$\widehat{\mu}$ の分散を求めるために，

$$\log f(X_i;\mu) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(X_i-\mu)^2$$
$$\frac{\mathrm{d}\log f(X_i;\mu)}{\mathrm{d}\mu} = \frac{1}{\sigma^2}(X_i-\mu)$$
$$\Big(\frac{\mathrm{d}\log f(X_i;\mu)}{\mathrm{d}\mu}\Big)^2 = \frac{1}{\sigma^4}(X_i-\mu)^2$$
$$\mathrm{E}\Big[\Big(\frac{\mathrm{d}\log f(X_i;\mu)}{\mathrm{d}\mu}\Big)^2\Big] = \frac{1}{\sigma^4}\mathrm{E}[(X_i-\mu)^2] = \frac{1}{\sigma^2}$$

と計算される。

最尤推定量の性質から，$n$ が大きいとき，

$$\widehat{\mu} \sim N(\mu, \sigma_\mu^2)$$

ただし，

$$\sigma_\mu^2 = \frac{1}{\sum_{i=1}^{n}\mathrm{E}\Big[\Big(\frac{\mathrm{d}\log f(X_i;\mu)}{\mathrm{d}\mu}\Big)^2\Big]} = \frac{\sigma^2}{n}$$

この場合は，$n$ の大きさに関わらず，$\widehat{\mu} \sim N(\mu, \sigma_\mu^2)$ が成り立つ。

(2) $\sigma^2$ が未知のとき，$\mu$ と $\sigma^2$ の尤度関数は，

$$l(\mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\Big(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\Big)$$

となる。

対数尤度関数は，

$$\log l(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2 \\ -\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2$$

と表される。
$\mu$ と $\sigma^2$ について，最大化するためには，

$$\frac{\partial \log l(\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu) = 0$$

$$\frac{\partial \log l(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i-\mu)^2 = 0$$

の連立方程式を解く。

$\mu$, $\sigma^2$ の解を $\widehat{\mu}$, $\widehat{\sigma}^2$ とすると，最尤推定値は，

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i \equiv \overline{x}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\widehat{\mu}) \equiv \frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})$$

となる。

観測値 $x_1, x_2, \cdots, x_n$ をその確率変数 $X_1, X_2, \cdots, X_n$ で置き換えて，$\mu$, $\sigma^2$ の最尤推定量は，

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^n X_i \equiv \overline{X}$$

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\widehat{\mu}) \equiv \frac{1}{n}\sum_{i=1}^n (X_i-\overline{X})$$

となる。

$\sigma^2$ の最尤推定量 $\widehat{\sigma}^2$ は，$\sigma^2$ の不偏推定量 $S^2 = \dfrac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2$ とは異なることに注意。

$\theta = (\mu,\sigma^2)'$ とする。$n$ が大きいとき，

$$\widehat{\theta} \sim N(\theta, \Sigma_\theta)$$

ただし，

$$\Sigma_\theta = -\Big(\sum_{i=1}^n \mathrm{E}\Big[\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big]\Big)^{-1}$$

$$\log f(X_i;\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(X_i-\mu)^2$$

$$\frac{\partial \log f(X_i;\theta)}{\partial\theta} = \begin{pmatrix} \dfrac{\partial \log f(X_i;\theta)}{\partial\mu} \\ \dfrac{\partial \log f(X_i;\theta)}{\partial\sigma^2} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma^2}(X_i-\mu) \\ -\dfrac{1}{2\sigma^2} + \dfrac{1}{2\sigma^4}(X_i-\mu)^2 \end{pmatrix}$$

$$\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'} = \begin{pmatrix} \dfrac{\partial^2 \log f(X_i;\theta)}{\partial\mu^2} & \dfrac{\partial^2 \log f(X_i;\theta)}{\partial\mu\partial\sigma^2} \\ \dfrac{\partial^2 \log f(X_i;\theta)}{\partial\sigma^2\partial\mu} & \dfrac{\partial^2 \log f(X_i;\theta)}{\partial(\sigma^2)^2} \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{\sigma^2} & -\dfrac{1}{\sigma^4}(X_i-\mu) \\ -\dfrac{1}{\sigma^4}(X_i-\mu) & \dfrac{1}{2\sigma^4} - \dfrac{1}{\sigma^6}(X_i-\mu)^2 \end{pmatrix}$$

$$\mathrm{E}\Big[\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big] = \begin{pmatrix} -\dfrac{1}{\sigma^2} & -\dfrac{1}{\sigma^4}\mathrm{E}(X_i-\mu) \\ -\dfrac{1}{\sigma^4}\mathrm{E}(X_i-\mu) & \dfrac{1}{2\sigma^4} - \dfrac{1}{\sigma^6}\mathrm{E}[(X_i-\mu)^2] \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{\sigma^2} & 0 \\ 0 & -\dfrac{1}{2\sigma^4} \end{pmatrix}$$

よって，

$$\Sigma_\theta = -\Big(\sum_{i=1}^n \mathrm{E}\Big[\frac{\partial^2 \log f(X_i;\theta)}{\partial\theta\partial\theta'}\Big]\Big)^{-1} = \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}$$

まとめると，$\mu$, $\sigma^2$ の最尤推定量 $\widehat{\mu} = (1/n)\sum_{i=1}^n X_i$, $\widehat{\sigma}^2 = (1/n)\sum_{i=1}^n (X_i-\overline{X})^2$ の分布は，$n$ が大きいとき，

$$\begin{pmatrix} \widehat{\mu} \\ \widehat{\sigma}^2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \dfrac{\sigma^2}{n} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}\right)$$

となる。

**例 2**：　$X_1, X_n, \cdots, X_n$ は互いに独立で，それぞれパラメータ $p$ を持ったベルヌイ分布に従うものとする。すなわち，$X_i$ の確率関数は，

$$f(x;p) = p^x(1-p)^{1-x} \qquad x = 0, 1$$

となる。

このとき尤度関数は，

$$l(p) = \prod_{i=1}^{n} f(x_i;p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

となり，対数尤度関数は，

$$\begin{aligned}\log l(p) &= \sum_{i=1}^{n} \log f(x_i;p) \\ &= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)\sum_{i=1}^{n}(1-x_i) \\ &= \log(p)\sum_{i=1}^{n} x_i + \log(1-p)(n - \sum_{i=1}^{n} x_i)\end{aligned}$$

となる。
$\log l(p)$ を最大にする $p$ を求める。

$$\frac{\mathrm{d}\log l(p)}{\mathrm{d}p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}(n - \sum_{i=1}^{n} x_i) = 0$$

したがって，$p$ について解くと，$p$ の最尤推定値 $\widehat{p}$ は，

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

となる。

さらに，$x_i$ を $X_i$ で置き換えて，$p$ の最尤推定量 $\widehat{p}$ は，

$$\widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

となる。

$\widehat{p}$ の分布を求める。

$$\log f(X_i;p) = X_i \log(p) + (1 - X_i)\log(1-p)$$

$$\frac{\mathrm{d}\log f(X_i;p)}{\mathrm{d}p} = \frac{X_i}{p} - \frac{1-X_i}{1-p} = \frac{X_i - p}{p(1-p)}$$

$$\mathrm{E}\Big[\Big(\frac{\mathrm{d}\log f(X_i;p)}{\mathrm{d}p}\Big)^2\Big] = \frac{\mathrm{E}[(X_i-p)^2]}{p^2(1-p)^2}$$

$$\begin{aligned}\mathrm{E}[(X_i-p)^2] &= \sum_{x_i=0}^{1}(x_i-p)^2 f(x_i;p) \\ &= \sum_{x_i=0}^{1}(x_i-p)^2 p^{x_i}(1-p)^{1-x_i} \\ &= p^2(1-p) + (1-p)^2 p = p(1-p)\end{aligned}$$

$$\frac{1}{\sum_{i=1}^{n}\mathrm{E}\Big[\Big(\frac{\mathrm{d}\log f(X_i;p)}{\mathrm{d}p}\Big)^2\Big]} = \frac{p(1-p)}{n}$$

したがって，

$$\widehat{p} \sim N\big(p,\ \frac{p(1-p)}{n}\big)$$

を得る。

**例 3**：　$X_1, X_n, \cdots, X_n$ は互いに独立で，それぞれパラメータ $\lambda$ を持ったポアソン分布に従うものとする。すなわち，$X_i$ の確率関数は，

$$f(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x = 0, 1, 2, \cdots$$

となる。

このとき尤度関数は，

$$l(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

となり，対数尤度関数は，

$$\begin{aligned}\log l(\lambda) &= \sum_{i=1}^{n} \log f(x_i;\lambda) \\ &= \log(\lambda)\sum_{i=1}^{n} x_i - n\lambda - \sum_{i=1}^{n}\log(x_i!)\end{aligned}$$

となる。
$\log l(\lambda)$ を最大にする $p$ を求める。

$$\frac{\mathrm{d}\log l(\lambda)}{\mathrm{d}\lambda} = \frac{1}{\lambda}\sum_{i=1}^{n} x_i - n = 0$$

したがって，$\lambda$ について解くと，$\lambda$ の最尤推定値 $\widehat{\lambda}$ は，

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

となる。

さらに，$x_i$ を $X_i$ で置き換えて，$\lambda$ の最尤推定量 $\widehat{\lambda}$ は,

$$\widehat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

となる。

$\widehat{\lambda}$ の分布を求める。

$$\log f(X_i;\lambda) = X_i\log(\lambda) - \lambda - \log(X_i!)$$

$$\frac{\mathrm{d}\log f(X_i;\lambda)}{\mathrm{d}\lambda} = \frac{X_i}{\lambda} - 1$$

$$\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2} = -\frac{X_i}{\lambda^2}$$

$$\mathrm{E}\Big(\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2}\Big) = \frac{\mathrm{E}(X_i)}{\lambda^2}$$

$$\begin{aligned}
\mathrm{E}(X_i) &= \sum_{x=0}^{\infty} x f(x;\lambda) \\
&= \sum_{x=0}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=1}^{\infty} x\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \sum_{x=1}^{\infty} \lambda\frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} \\
&= \sum_{x=0}^{\infty} \lambda\frac{\lambda^x e^{-\lambda}}{x!} \\
&= \lambda
\end{aligned}$$

$$-\frac{1}{\sum_{i=1}^{n}\mathrm{E}\Big(\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2}\Big)} = \frac{\lambda}{n}$$

したがって,

$$\widehat{\lambda} \sim N\big(\lambda,\ \frac{\lambda}{n}\big)$$

を得る。

**例 4**： $X_1,\, X_n,\, \cdots,\, X_n$ は互いに独立で，それぞれパラメータ $\lambda$ を持った指数分布に従うものとする。すなわち，$X_i$ の密度関数は,

$$f(x;\lambda) = \lambda e^{-\lambda x} \qquad\qquad x > 0$$

となる。

このとき尤度関数は,

$$l(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n}\lambda e^{-\lambda x_i}$$

となり，対数尤度関数は,

$$\begin{aligned}
\log l(\lambda) &= \sum_{i=1}^{n}\log f(x_i;\lambda) \\
&= n\log\lambda - \lambda\sum_{i=1}^{n} x_i
\end{aligned}$$

となる。
$\log l(\lambda)$ を最大にする $p$ を求める。

$$\frac{\mathrm{d}\log l(\lambda)}{\mathrm{d}\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

したがって，$\lambda$ について解くと，$\lambda$ の最尤推定値 $\widehat{\lambda}$ は,

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i}$$

となる。

さらに，$x_i$ を $X_i$ で置き換えて，$\lambda$ の最尤推定量 $\widehat{\lambda}$ は,

$$\widehat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i}$$

となる。

$\widehat{\lambda}$ の分布を求める。

$$\log f(X_i;\lambda) = \log\lambda - \lambda X_i$$

$$\frac{\mathrm{d}\log f(X_i;\lambda)}{\mathrm{d}\lambda} = \frac{1}{\lambda} - X_i$$

$$\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2} = -\frac{1}{\lambda^2}$$

$$-\frac{1}{\sum_{i=1}^{n}\mathrm{E}\Big(\frac{\mathrm{d}^2\log f(X_i;\lambda)}{\mathrm{d}\lambda^2}\Big)} = \frac{\lambda^2}{n}$$

したがって,

$$\widehat{\lambda} \sim N\big(\lambda,\ \frac{\lambda^2}{n}\big)$$

を得る。

### 12.2.1 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。

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

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，

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

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，

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

となる。

### 12.2.2 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，

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

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，

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

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，

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

となる。これは $\alpha, \beta, \sigma^2$ の関数となっている。

よって，尤度関数は，

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

となる。
対数尤度関数は，

$$
\begin{aligned}
\log l(\alpha, \beta, \sigma^2) = &-\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) \\
&-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2
\end{aligned}
$$

となる。

$\log l(\alpha,\beta,\sigma^2)$ を最大にするために，

$$\frac{\partial \log l(\alpha,\beta,\sigma^2)}{\partial\alpha} = \frac{1}{\sigma^2}\sum_{i=1}^n (Y_i-\alpha-\beta X_i) = 0$$
$$\frac{\partial \log l(\alpha,\beta,\sigma^2)}{\partial\beta} = \frac{1}{\sigma^2}\sum_{i=1}^n X_i(Y_i-\alpha-\beta X_i) = 0$$
$$\frac{\partial \log l(\alpha,\beta,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (Y_i-\alpha-\beta X_i)^2 = 0$$

の連立方程式を解く。

上 2 つの式は $\sigma^2$ に依存していない。$\alpha$，$\beta$ の最尤推定量は最小二乗推定量と同じになる。

すなわち，

$$\widehat{\beta} = \frac{\sum_{i=1}^n (X_i-\overline{X})(Y_i-\overline{Y})}{\sum_{i=1}^n (X_i-\overline{X})^2}$$
$$\widehat{\alpha} = \overline{Y} - \widehat{\beta}\overline{X}$$

$\sigma^2$ の最尤推定量は，

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (Y_i-\widehat{\alpha}-\widehat{\beta}X_i)^2$$

となり，$s^2$ とは異なる。

$\widehat{\theta} = (\widehat{\alpha},\widehat{\beta},\widehat{\sigma}^2)'$，$\theta = (\alpha,\beta,\sigma^2)'$ とする。$n$ が大きいとき，

$$\widehat{\theta} \sim N(\theta, \Sigma_\theta)$$

ただし，

$$\Sigma_\theta = \Big(\sum_{i=1}^n \mathrm{E}\Big[\Big(\frac{\partial \log g(Y_i;\theta)}{\partial\theta}\Big)\Big(\frac{\partial \log g(Y_i;\theta)}{\partial\theta}\Big)'\Big]\Big)^{-1}$$
$$= -\Big(\sum_{i=1}^n \mathrm{E}\Big[\frac{\partial^2 \log g(Y_i;\theta)}{\partial\theta\partial\theta'}\Big]\Big)^{-1}$$

$Y_i$ の密度関数 $g(Y_i;\theta)$ の対数は，

$$\log g(Y_i;\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(Y_i-\alpha-\beta X_i)^2$$

となる。

$$\frac{\partial \log g(Y_i;\theta)}{\partial\theta} = \begin{pmatrix} \dfrac{\partial \log g(Y_i;\theta)}{\partial\alpha} \\ \dfrac{\partial \log g(Y_i;\theta)}{\partial\beta} \\ \dfrac{\partial \log g(Y_i;\theta)}{\partial\sigma^2} \end{pmatrix}$$
$$= \begin{pmatrix} \dfrac{1}{\sigma^2}(Y_i-\alpha-\beta X_i) \\ \dfrac{1}{\sigma^2}X_i(Y_i-\alpha-\beta X_i) \\ -\dfrac{1}{2\sigma^2}+\dfrac{1}{2\sigma^4}(Y_i-\alpha-\beta X_i)^2 \end{pmatrix}$$

$$\frac{\partial^2 \log g(Y_i;\theta)}{\partial\theta\partial\theta'}$$
$$= \begin{pmatrix} \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\alpha^2} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\alpha\partial\beta} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\alpha\partial\sigma^2} \\ \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\beta\partial\alpha} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\beta^2} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\beta\partial\sigma^2} \\ \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\sigma^2\partial\alpha} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial\sigma^2\partial\beta} & \dfrac{\partial^2 \log g(Y_i;\theta)}{\partial(\sigma^2)^2} \end{pmatrix}$$
$$= \begin{pmatrix} -\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & -\dfrac{u_i}{\sigma^4} \\ -\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & -\dfrac{X_iu_i}{\sigma^4} \\ -\dfrac{u_i}{\sigma^4} & -\dfrac{X_iu_i}{\sigma^4} & \dfrac{1}{2\sigma^4}-\dfrac{u_i^2}{\sigma^6} \end{pmatrix}$$

ただし，$u_i = Y_i-\alpha-\beta X_i$

期待値をとると，

$$\mathrm{E}\Big(\frac{\partial^2 \log g(Y_i;\theta)}{\partial\theta\partial\theta'}\Big) = \mathrm{E}\begin{pmatrix} -\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & -\dfrac{u_i}{\sigma^4} \\ -\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & -\dfrac{X_iu_i}{\sigma^4} \\ -\dfrac{u_i}{\sigma^4} & -\dfrac{X_iu_i}{\sigma^4} & \dfrac{1}{2\sigma^4}-\dfrac{u_i^2}{\sigma^6} \end{pmatrix}$$
$$= \begin{pmatrix} -\dfrac{1}{\sigma^2} & -\dfrac{X_i}{\sigma^2} & 0 \\ -\dfrac{X_i}{\sigma^2} & -\dfrac{X_i^2}{\sigma^2} & 0 \\ 0 & 0 & -\dfrac{1}{2\sigma^4} \end{pmatrix}$$

となる。

$$\Sigma_\theta = -\Big(\sum_{i=1}^n \mathrm{E}\Big[\frac{\partial^2 \log g(Y_i;\theta)}{\partial\theta\partial\theta'}\Big]\Big)^{-1}$$
$$= \begin{pmatrix} \dfrac{n}{\sigma^2} & \dfrac{\sum_{i=1}^n X_i}{\sigma^2} & 0 \\ \dfrac{\sum_{i=1}^n X_i}{\sigma^2} & \dfrac{\sum_{i=1}^n X_i^2}{\sigma^2} & 0 \\ 0 & 0 & \dfrac{n}{2\sigma^4} \end{pmatrix}^{-1}$$
$$= \begin{pmatrix} \sigma^2\begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}^{-1} & 0 \\ 0 & \dfrac{2\sigma^4}{n} \end{pmatrix}$$

したがって，

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} \sim N\Big(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \sigma^2\begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}^{-1}\Big)$$

となる。
$\longrightarrow$ 最小二乗推定量の分布と同じ。

### 12.2.3　誤差項に系列相関がある場合

回帰モデル

$$Y_i = \alpha + \beta X_i + u_i$$
$$u_i = \rho u_{i-1} + \epsilon_i \qquad i = 2, 3, \cdots, n$$

$\epsilon_2$, $\epsilon_3$, $\cdots$, $\epsilon_n$ は互いに独立で，すべての $i$ について $\epsilon_i \sim \mathrm{N}(0, \sigma^2)$ を仮定する。

$u_i$ を消去すると，

$$(Y_i - \alpha - \beta X_i) = \rho(Y_{i-1} - \alpha - \beta X_{i-1}) + \epsilon_i$$

または

$$(Y_i - \rho Y_{i-1}) = \alpha(1-\rho) + \beta(X_i - \rho X_{i-1}) + \epsilon_i$$

と書き直すことが出来る。

$\theta = (\alpha, \beta, \sigma^2, \rho)$ とする。

$$\log f(Y_i;\theta) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\Big((Y_i - \rho Y_{i-1}) - \alpha(1-\rho) - \beta(X_i - \rho X_{i-1})\Big)^2$$

尤度関数は，

$$\log l(\theta) = \sum_{i=2}^{n} \log f(Y_i;\theta) = -\frac{n-1}{2}\log(2\pi) - \frac{n-1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=2}^{n}\Big((Y_i - \rho Y_{i-1}) - \alpha(1-\rho) - \beta(X_i - \rho X_{i-1})\Big)^2$$

となる。
尤度関数をそれぞれ $\alpha$，$\beta$，$\sigma^2$，$\rho$ について微分し，ゼロとおく。

$$\frac{\partial \log l(\theta)}{\partial \alpha} = \frac{1-\rho}{\sigma^2}\sum_{i=2}^{n}\Big((Y_i - \rho Y_{i-1}) - \alpha(1-\rho) - \beta(X_i - \rho X_{i-1})\Big) = 0$$

$$\frac{\partial \log l(\theta)}{\partial \beta} = \frac{1}{\sigma^2}\sum_{i=2}^{n}(X_i - \rho X_{i-1})\Big((Y_i - \rho Y_{i-1}) - \alpha(1-\rho) - \beta(X_i - \rho X_{i-1})\Big) = 0$$

$$\frac{\partial \log l(\theta)}{\partial \sigma^2} = -\frac{n-1}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=2}^{n}\Big((Y_i - \rho Y_{i-1}) - \alpha(1-\rho) - \beta(X_i - \rho X_{i-1})\Big)^2 = 0$$

$$\frac{\partial \log l(\theta)}{\partial \rho} = \frac{1}{\sigma^2}\sum_{i=2}^{n}(Y_{i-1} - \alpha - \beta X_{i-1})\Big((Y_i - \alpha - \beta X_i) - \rho(Y_{i-1} - \alpha - \beta X_{i-1})\Big) = 0$$

$(Y_i - \alpha - \beta X_i) - \rho(Y_{i-1} - \alpha - \beta X_{i-1})$ は $(Y_i - \rho Y_{i-1}) - \alpha(1-\rho) - \beta(X_i - \rho X_{i-1})$ を書き直したもの。

4 つの連立方程式を解いて，最尤推定量 $\widehat{\alpha}$，$\widehat{\beta}$，$\widehat{\sigma}^2$，$\widehat{\rho}$ が得られる。
$\longrightarrow$ 下記のように収束計算によって求める。

(i) 初期段階では，$\widehat{\rho} = 0$ とする。

(ii) $X_i^* = X_i - \widehat{\rho} X_{i-1}$
$Y_i^* = Y_i - \widehat{\rho} Y_{i-1}$

(iii) $\begin{pmatrix} \widetilde{\alpha} \\ \widehat{\beta} \end{pmatrix} = \begin{pmatrix} n-1 & \sum_{i=2}^{n} X_i^* \\ \sum_{i=2}^{n} X_i^* & \sum_{i=2}^{n} X_i^{*2} \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=2}^{n} Y_i^* \\ \sum_{i=2}^{n} X_i^* Y_i^* \end{pmatrix}$

(iv) $\widehat{\alpha} = \dfrac{\widetilde{\alpha}}{1-\widehat{\rho}}$

(v) $\widehat{u}_i = Y_i - \widehat{\alpha} - \widehat{\beta} X_i$

(vi) $\widehat{\sigma}^2 = \dfrac{1}{n-1}\sum_{i=2}^{n}(\widehat{u}_i - \widehat{\rho}\widehat{u}_{i-1})^2$

(vii) $\widehat{\rho} = \dfrac{\sum_{i=2}^{n} \widehat{u}_i \widehat{u}_{i-1}}{\sum_{i=2}^{n} \widehat{u}_{i-1}^2}$

(viii) ステップ (ii) 〜 (vii) を，収束するまで繰り返し計算する。

## 12.3　尤度比検定

$n$ 個の確率変数 $X_1$, $X_2$, $\cdots$, $X_n$ は互いに独立で，同じ確率分布 $f(x) \equiv f(x;\theta)$ とする。
尤度関数は，

$$l(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$