

6 最尤法について

標本 $X_1, X_2, \dots, X_n$ の密度関数 :

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

θ は未知母数 $\Rightarrow \widehat{\theta}_n(x_1, x_2, \dots, x_n)$ によって推定

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

のように, θ の関数として考える。

$l(\theta)$: 尤度関数

尤度関数を最大にする θ を $\widehat{\theta}_n$ とする。

$\widehat{\theta}_n \equiv \widehat{\theta}_n(X_1, X_2, \dots, X_n) \Rightarrow$ 最尤推定量

$\widehat{\theta}_n(x_1, x_2, \dots, x_n) \Rightarrow$ 最尤推定値

すなわち,

$$\frac{\partial l(\theta)}{\partial \theta} = 0$$

を解くことによって, 最尤推定量 $\widehat{\theta}_n \equiv \widehat{\theta}_n(X_1, X_2, \dots, X_n)$ が得られる。

最尤推定量の性質：

小標本について (n が小さいとき)：

- 一般に, 最尤推定量は不偏性を持っていないが, 適当な変換によって, 不偏推定量を作ることが出来る場合が多い。
- 有効推定量が存在すれば (すなわち, クラメール・ラオの不等式の等号を満たすような推定量が存在するならば), 最尤推定量は有効推定量に一致する。
- 十分統計量が存在すれば, 最尤推定量は十分統計量の関数となる。

大標本について (n が大きいとき) :

$n \rightarrow \infty$ のとき,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \rightarrow N(0, \sigma^2)$$

となる。 $\Rightarrow$ 一致性, 漸近的正規性, 漸近的有効性

ただし,

$$\sigma^2 = \sigma^2(\theta) = \frac{1}{\mathbb{E} \left[\left(\frac{\partial \log f(X; \theta)}{\partial \theta} \right)^2 \right]}$$

すなわち, $n \rightarrow \infty$ のとき,

$$\frac{\sqrt{n}(\widehat{\theta}_n - \theta)}{\sigma(\theta)} = \frac{\widehat{\theta}_n - \theta}{\sigma(\theta)/\sqrt{n}} \rightarrow N(0, 1)$$

となる。

したがって, 厳密ではないが, n が大きいとき,

$$\widehat{\theta}_n \sim N\left(\theta, \frac{\sigma^2(\theta)}{n}\right)$$

と近似できる。

すなわち、 $n \rightarrow \infty$ のとき、 $\widehat{\theta}_n$ の分散はクラメール・ラオの不等式の下限 $\frac{\sigma^2(\theta)}{n}$ に近づくことを意味する。

⇒ 漸近的に有効推定量

さらに、分母の θ を最尤推定量 $\widehat{\theta}_n$ で置き換えて、 $n \rightarrow \infty$ のとき、

$$\frac{\widehat{\theta}_n - \theta}{\sigma(\widehat{\theta}_n)/\sqrt{n}} \rightarrow N(0, 1)$$

となる。

実際には、 n が大きいとき、

$$\widehat{\theta}_n \sim N\left(\theta, \frac{\sigma^2(\widehat{\theta}_n)}{n}\right)$$

と近似して用いる。

例：

$X_1, X_2, \dots, X_n$ は互いに独立で、すべて同一の正規分布 (すなわち、平均 μ , 分散 σ^2 ですべて同一の分布) に従うものとする。 μ, σ^2 の最尤推定量を求める。

$$\begin{aligned} f(x_1, x_2, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^n f(x_i; \mu, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i - \mu)^2\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right) = l(\mu, \sigma^2) \end{aligned}$$

対数をとる。(最大化しやすくなる場合が多い)

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

対数尤度関数 $\log l(\mu, \sigma^2)$ を μ と σ^2 について微分して、ゼロと置く。

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

この 2 つの連立方程式を解く。

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

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

μ, σ^2 の最尤推定量は,

$$\bar{X}, \quad S^{**2} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

となる。

$E(\bar{X}) = \mu$ なので, μ の最尤推定量 $\bar{X}$ は不偏推定量である。

$E(S^{**2}) = \frac{n-1}{n} \sigma^2 \neq \sigma^2$ なので, σ^2 の最尤推定量 S^{**2} は不偏推定量でない。

例 :

$X_1, X_2, \dots, X_n$ は互いに独立で, すべて同一のベルヌイ分布ですべて同一の分布) に従うものとする。すなわち, X の確率関数は $P(X = x) = f(x; p) = p^x(1-p)^{1-p}$,

$x = 0, 1$, となる。 p の最尤推定量を求める。

$$f(x_1, x_2, \dots, x_n; p) = \prod_{i=1}^n f(x_i; p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum_i x_i} (1-p)^{n-\sum_i x_i} = l(p)$$

対数をとる。

$$\log l(p) = \left(\sum_i x_i \right) \log(p) + (n - \sum_i x_i) \log(1-p)$$

対数尤度関数 $\log l(p)$ を p について微分して，ゼロと置く。

$$\frac{\partial \log l(p)}{\partial p} = \frac{\sum_i x_i}{p} - \frac{n - \sum_i x_i}{1-p} = \frac{\sum_i x_i - np}{p(1-p)} = 0$$

この方程式を解く。

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

p の最尤推定量は，

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

となる。

$E(\bar{X}) = p$ なので、 p の最尤推定量 $\bar{X}$ は不偏推定量である。

X がベルヌイ分布 $f(x; p)$ のとき、 $E(X) = p$ に注意。

例：

$X_1, X_2, \dots, X_n$ は互いに独立で、すべて同一のポアソン分布 (すなわち、平均 λ ですべて同一の分布) に従うものとする。 λ の最尤推定量を求める。

ポアソン分布の確率関数は、

$$P(X = x) = f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \dots$$

なので、尤度関数は、

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

対数尤度関数は,

$$\log l(\lambda) = \log(\lambda) \sum_{i=1}^n x_i - n\lambda - \log\left(\prod_{i=1}^n x_i!\right)$$

となる。

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

これを解いて, λ の最尤推定量 $\widehat{\lambda}$ は,

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

となる。

$\widehat{\lambda}$ は, λ の不偏推定量, 有効推定量, 十分推定量, 一致推定量である。

証明:

X がパラメータ λ のポアソン分布に従うとき,

$$E(X) = V(X) = \lambda$$

となる。

不偏性 :

$$E(\widehat{\lambda}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \sum_{i=1}^n \lambda = \lambda$$

有効性 :

$$V(\widehat{\lambda}) = V\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n V(X_i) = \frac{1}{n^2} \sum_{i=1}^n \lambda = \frac{\lambda}{n}$$

$$\begin{aligned} \frac{1}{nE\left[\left(\frac{\partial \log f(X; \lambda)}{\partial \lambda}\right)^2\right]} &= \frac{1}{nE\left[\left(\frac{\partial (X \log \lambda - \lambda - \log X!)}{\partial \lambda}\right)^2\right]} = \frac{1}{nE\left[\left(\frac{X}{\lambda} - 1\right)^2\right]} \\ &= \frac{\lambda^2}{nE[(X - \lambda)^2]} = \frac{\lambda^2}{nV(X)} = \frac{\lambda^2}{n\lambda} = \frac{\lambda}{n} \end{aligned}$$

したがって,

$$V(\widehat{\lambda}) = \frac{1}{nE\left[\left(\frac{\partial \log f(X; \lambda)}{\partial \lambda}\right)^2\right]}$$

となり, $V(\widehat{\lambda})$ は, クラメル・ラオの下限に一致する。よって, $\widehat{\lambda}$ は有効推定量である。

十分性:

$$\prod_{i=1}^n f(x_i; \lambda) = \frac{\lambda^{\sum_{i=1}^n x_i} e^{-n\lambda}}{\prod_{i=1}^n x_i!} = \frac{\lambda^{n\bar{x}} e^{-n\lambda}}{(n\bar{x})!} \frac{(n\bar{x})!}{\prod_{i=1}^n x_i!} = g(\bar{x}; \lambda) h(x_1, x_2, \dots, x_n)$$

と分解できる。

一致性:

$$E(\bar{X}) = \lambda, \quad V(\bar{X}) = \frac{\lambda}{n}$$

なので, チェビシェフの不等式に当てはめる。

$$P(|\bar{X} - \lambda| > \epsilon) < \frac{\lambda}{n\epsilon^2} \rightarrow \infty$$

したがって，一緻性も成り立つ。