

.....2.4.....

最尤法

確率ベクトル $\mathbf{X} = (X_1, X_2, \dots, X_n)$ の統計モデルを $\mathcal{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ とする。 \mathcal{P} に含まれる P_θ に対応する確率密度関数もしくは確率関数を $p(\mathbf{x}|\theta)$ と記すことにする。ただし、 $\mathbf{x} = (x_1, x_2, \dots, x_n)$ である。 $\mathbf{X} = \mathbf{x}$ が観測されたときの尤度関数を

$$L_n(\theta|\mathbf{x}) = p(\mathbf{x}|\theta), \quad \theta \in \Theta$$

で定めることにする。 $L_n(\cdot|\mathbf{x})$ は標本空間から $\{\theta \mapsto p(\mathbf{x}|\theta) : \mathbf{x} \in S\}$ なる関数族への対応となる。 \mathbf{x} があたられたとき、 $L_n(\theta|\mathbf{x})$ は θ の関数とみなす。これを簡単に $L_n(\theta)$ と書くことにする。 $L_n(\theta)$ は \mathbf{x} が与えられたとき、いろいろな θ の「確からしさ」もしくは「尤もらしさ」を表現するものである。

特に、 X_1, X_2, \dots, X_n が独立同一に確率密度関数 $f(x|\theta)$ に従うならば、尤度関数は

$$L_n(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$$

で与えられる。

最尤法とは、与えられたデータを実現させるために「尤もらしい」母数の値を母数の推定値として用いる手法である。すなわち、 $\mathbf{X} = \mathbf{x}$ が与えられたとき、尤度関数を最大にする値 $\hat{\theta}(\mathbf{x})$ を見つけることである：

$$L_n(\hat{\theta}(\mathbf{x})|\mathbf{x}) = p(\mathbf{x}|\hat{\theta}(\mathbf{x})) = \max\{p(\mathbf{x}|\theta) : \theta \in \Theta\} = \max\{L_n(\theta|\mathbf{x}) : \theta \in \Theta\}$$

$\hat{\theta}(\mathbf{x})$ を θ の最尤推定値といい、 $\hat{\theta}(\mathbf{X})$ を θ の最尤推定量という。

例 2.13 確率変数 X が正規分布 $N(\theta, \sigma^2)$ に従うとする。ただし、 σ^2 は既知とする。このとき、尤度関数は

$$L_1(\theta|x) = \frac{1}{\sigma} \psi\left(\frac{x-\theta}{\sigma}\right)$$

となる。ただし、 $\psi(x) = (1/\sqrt{2\pi})e^{-x^2/2}$ である。このとき、最大は

$$\hat{\theta}(x) = x$$

のとき唯一達成される。したがって、 $\hat{\theta}(\mathbf{X}) = \mathbf{X}$ は最尤推定量となる。

つぎに、 X_1, X_2, \dots, X_n は独立同一に正規分布 $N(\theta, \sigma^2)$ に従うとする。ここでも σ^2 は既知とする。このとき、尤度関数は

$$\begin{aligned} L_n(\theta|\mathbf{x}) &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2} \sum_{i=1}^n \frac{x_i - \theta}{\sigma^2}\right] \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left[-\frac{1}{2} \frac{(\bar{x}_n - \theta)^2}{\sigma^2/n}\right] \exp\left[-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \bar{x}_n)^2}{\sigma^2}\right] \end{aligned}$$

となる．よって，最大は

$$\hat{\theta}(\mathbf{x}) = \bar{x}_n, \quad \bar{x}_n = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

で達成される．したがって，最尤推定量は $\hat{\theta}(\mathbf{X}) = \bar{X}_n$ となる．ただし， $\bar{X}_n = (1/n)(X_1 + X_2 + \cdots + X_n)$ である．

尤度関数に対数をとったものを対数尤度とよび，

$$l_n(\theta) = \log L_n(\theta | \mathbf{x})$$

と記す¹⁾．特に， $X_i, i = 1, 2, \dots, n$ ，が独立同一に確率密度関数 $f(x | \theta)$ に従う場合には

$$l_n(\theta) = \log \prod_{i=1}^n f(x_i | \theta) = \sum_{i=1}^n \log f(x_i | \theta)$$

となる．

もし， Θ が開集合で $l_n(\theta)$ が θ に関して微分可能で $\hat{\theta}(\mathbf{x})$ が存在するならば， $\hat{\theta}(\mathbf{x})$ は方程式

$$\frac{\partial}{\partial \theta_i} l_n(\theta) = 0, \quad i = 1, 2, \dots, k$$

をみtas．この方程式を尤度方程式という．

例 2.14 標識 1, 2, 3 のどれかをもつ個体から構成される母集団を考える．それぞれの標識の出現確率は Hardy-Weinberg 比率で与えられるとする：

$$p(1|\theta) = \theta^2, \quad p(2|\theta) = 2\theta(1-\theta), \quad p(3|\theta) = (1-\theta)^2, \quad 0 < \theta < 1$$

たとえば，3 つの個体を観測し， $x_1 = 1, x_2 = 2, x_3 = 1$ を得たとする．このとき

$$L_3(\theta | \mathbf{x}) = p(1|\theta)p(2|\theta)p(1|\theta) = 2\theta^5(1-\theta)$$

となる．尤度方程式は

$$\frac{\partial}{\partial \theta} l_3(\theta) = \frac{5}{\theta} - \frac{1}{1-\theta} = 0$$

となり，唯一の解 $\hat{\theta} = 5/6$ を得る．これは

$$\frac{\partial^2}{\partial \theta^2} l_3(\theta) = -\frac{5}{\theta^2} - \frac{1}{(1-\theta)^2} < 0, \quad 0 < \theta < 1$$

よりわかる．

一般に， n 個の観測 x_1, x_2, \dots, x_n を得たとする．いま

$$n_j = \#\{x_i = j : i = 1, 2, \dots, n\}, \quad j = 1, 2, 3$$

とする．尤度関数は

$$L_n(\theta | \mathbf{x}) = \theta^{2n_1} \{2\theta(1-\theta)\}^{n_2} (1-\theta)^{2n_3} = 2^{n_2} \theta^{2n_1+n_2} (1-\theta)^{n_2+2n_3}$$

¹⁾形式的に， $0/0 = 0, 0 \times \infty = 0$ とする．

より

$$\frac{\partial}{\partial \theta} l_n(\theta) = \frac{2n_1 + n_2}{\theta} - \frac{n_2 + 2n_3}{1 - \theta} = \frac{1}{\theta(1 - \theta)} \{(2n_1 + n_2) - 2(n_1 + n_2 + n_3)\theta\}$$

より, $2n_1 + n_2 > 0$, $n_2 + 2n_3 > 0$ のとき, 最尤推定値は唯一存在して,

$$\hat{\theta}(\mathbf{x}) = \frac{2n_1 + n_2}{2(n_1 + n_2 + n_3)}$$

となる. もし, $2n_1 + n_2 = 0$ のとき, 尤度関数は

$$2^{n_2}(1 - \theta)^{(2n_1 + n_2) + (n_2 + 2n_3)} = 2^{n_2}(1 - \theta)^{2(n_1 + n_2 + n_3)}$$

となり, $\theta = 0$ のとき, 尤度関数は最大になり, $\Theta = (0, 1)$ なので, 最尤推定値は存在しない. また, $n_2 + 2n_3 = 0$ のときは, 尤度関数は $2^{n_2}\theta^{2(n_1 + n_2 + n_3)}$ なり, 最尤推定値は存在しない.

例 2.15 X_1, X_2, \dots, X_n は独立同一分布に従い, 各 $X_i, i = 1, 2, \dots, n$, は $1, 2, \dots, k$ の値をとり, その確率は $\theta_j = \mathbb{P}\{X_i = j\}, j = 1, 2, \dots, k$, で与えられるとする. ここで, $n \geq k - 1$ を仮定する. いま, $N_j = \sum_{i=1}^n I\{X_i = j\}$ おく. $\mathbf{X} = \mathbf{x}$ が与えられたとする. このとき, $n_j = \sum_{i=1}^n I\{x_i = j\}$ とおけば, 対数尤度は

$$l_n(\boldsymbol{\theta}) = \log L_n(\boldsymbol{\theta} | \mathbf{x}) = \log p(\mathbf{x} | \boldsymbol{\theta}) = \sum_{j=1}^k n_j \log \theta_j$$

となる. ただし, $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ で

$$\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j \quad (2.6)$$

である.

まず, $n_j > 0, j = 1, 2, \dots, k$, を仮定する. このとき, θ_j のどれかがゼロならば, $p(\mathbf{x} | \boldsymbol{\theta}) = 0$ となる. したがって, 最尤推定値は $\theta_j > 0$ となるので, 上の仮定のもとでは, 最尤推定値は $[0, 1]^k$ の内点である. したがって, 最尤推定値は尤度方程式

$$\frac{\partial}{\partial \theta_j} l_n(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta_j} \sum_{\ell=1}^k n_\ell \log \theta_\ell = \sum_{\ell=1}^k \frac{n_\ell}{\theta_\ell} \frac{\partial \theta_\ell}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, k - 1$$

となる. (2.6) から $\partial \theta_k / \partial \theta_j = -1, j = 1, 2, \dots, k - 1$, となる. よって, 尤度方程式は

$$\frac{\hat{\theta}_k}{\hat{\theta}_j} = \frac{n_k}{n_j}, \quad j = 1, 2, \dots, k - 1$$

となる. これを再度 (2.6) に代入すれば, $\hat{\theta}_k = n_k/n$ となり,

$$\hat{\theta}_j = \frac{n_j}{n}, \quad j = 1, 2, \dots, k$$

となる。ただし、 $n = \sum_{\ell=1}^k n_\ell$ とした。つぎに、 $\theta_j = n_j/n, j = 1, 2, \dots, k$ が実際に $l_n(\theta)$ を最大にしていることを確認するために、 $l_n(\theta)$ は $(\theta_1, \theta_2, \dots, \theta_{k-1})$ に関して concave であることを示す。 $1 \leq r \leq k-1, 1 \leq j \leq k-1$ に対して、

$$\frac{\partial}{\partial \theta_r} \frac{\partial}{\partial \theta_j} l_n(\theta) = \frac{\partial}{\partial \theta_r} \left(\frac{n_j}{\theta_j} - \frac{n_k}{\theta_k} \right) = \begin{cases} -\left(\frac{n_r}{\theta_r^2} + \frac{n_k}{\theta_k^2} \right) < 0, & r = j \\ -\frac{n_k}{\theta_k^2} < 0, & r \neq j \end{cases}$$

となる。

ある j に対して、 $n_j = 0$ のとき、 $\hat{\theta}_j = n_j/n$ が最尤推定値であることも確認することができる。

定理 2.7 $T(X)$ を未知母数 θ の十分統計量とする。このとき、 θ の最尤推定量が一意に存在するならば、 θ の最尤推定量は T の関数である。

証明 $p(x|\theta)$ を確率関数または確率密度関数とする。因子分解定理から θ と T を通してのみ x に依存する関数 g と x のみに依存する関数 h が存在して、

$$p(x|\theta) = h(x)g(T(x)|\theta)$$

と書ける。これより θ に関して $p(x|\theta)$ を最大化することは θ に関して $g(T(x)|\theta)$ を最大化することと同値になる。また、 $T(x) = t$ と与えられたとき、 $g(t|\theta)$ が 2 つ以上 θ で最大になるとすれば、それに対応する x において $g(T(x)|\theta)$ も 2 つ以上の θ で最大化されるので、仮定と矛盾する。したがって、 $g(t|\theta)$ を θ について最大にする点はひとつである。□

定理 2.8 関数 $g: \mathbb{R} \rightarrow \mathbb{R}$ が 1 対 1 のとき、 $\hat{\theta}$ が θ の最尤推定量であれば、 $g(\hat{\theta})$ は $g(\theta)$ の最尤推定量である。

証明 g は 1 対 1 だから、逆関数 g^{-1} が存在し、 $\tau = g(\theta)$ のとき、 $\theta = g^{-1}(\tau)$ となる。これより

$$L_n(\theta|\mathbf{x}) = p(\mathbf{x}|\theta) = p(\mathbf{x}|g^{-1}(\tau)) = \tilde{L}_n(\tau|\mathbf{x})$$

と書けるので、

$$\sup_{\tau} \tilde{L}_n(\tau|\mathbf{x}) = \sup_{\tau} L_n(g^{-1}(\tau)|\mathbf{x}) = \sup_{\theta} L_n(\theta|\mathbf{x})$$

よって、 $\hat{\theta} = g^{-1}(\hat{\tau})$ のとき最大化される。したがって、 $\hat{\tau} = g(\hat{\theta})$ のとき最大化される。□

例 2.16 (指数分布族の最尤推定量) $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ は k 母数指数分布族

$$f(x|\boldsymbol{\eta}) = S(x) \exp\left[\sum_{j=1}^k \eta_j T_j(x) - A(\boldsymbol{\eta})\right]$$

からの大きさ n のランダム標本とする。ただし, $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_k)$ とし, 自然母数空間 $\boldsymbol{\eta} \in \mathcal{E}$ は \mathbb{R}^k の開集合とする。である。このとき, \mathbf{X} の同時確率(密度)関数は

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\eta}) &= \prod_{i=1}^n f(x_i|\boldsymbol{\eta}) \\ &= \prod_{i=1}^n S(x_i) \exp\left[\sum_{j=1}^k n\eta_j \bar{T}_j(\mathbf{x}) - nA(\boldsymbol{\eta})\right] \end{aligned}$$

となる。ただし, $\bar{T}_j(\mathbf{x}) = \sum_{i=1}^n T_j(x_i)$, $j = 1, 2, \dots, k$ である。したがって, 対数尤度は

$$l_n(\boldsymbol{\eta}) = n \sum_{j=1}^k \eta_j \bar{T}_j(\mathbf{x}) - nA(\boldsymbol{\eta}) + (\text{定数項})$$

となる。系 2.1 を用いれば,

$$\frac{\partial}{\partial \eta_j} l_n(\boldsymbol{\eta}) = n\bar{T}_j(\mathbf{x}) - n \frac{\partial}{\partial \eta_j} A(\boldsymbol{\eta}) = n\bar{T}_j(\mathbf{x}) - n\mathbb{E}_\eta[T_j(X)]$$

と

$$\frac{\partial^2}{\partial \eta_j \partial \eta_m} l_n(\boldsymbol{\eta}) = -n \frac{\partial^2}{\partial \eta_j \partial \eta_m} A(\boldsymbol{\eta}) = -n \text{COV}_\eta(T_j(X), T_m(X))$$

となるので, 行列 $((\partial^2/\partial \boldsymbol{\eta}' \boldsymbol{\eta})A(\boldsymbol{\eta}))$ は負の定符号となる。したがって, 尤度方程式

$$\frac{1}{n} \sum_{i=1}^n T_j(x_i) = \frac{\partial}{\partial \eta_j} A(\boldsymbol{\eta})$$

または

$$\frac{1}{n} \sum_{i=1}^n T_j(x_i) = \mathbb{E}_\eta[T_j(X)], \quad j = 1, 2, \dots, k$$

は唯一の解を持ち, これは $\boldsymbol{\eta}$ の最尤推定値となる。