6.2 最尤法

確率ベクトル $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ 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 $p(\boldsymbol{x}|\,\theta)$ と記すことにする．ただし，$\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ である．$\boldsymbol{X} = \boldsymbol{x}$ が観測されたときの尤度関数を

$$L_n(\theta|\,\boldsymbol{x}) = p(\boldsymbol{x}|\,\theta), \qquad \theta \in \Theta$$

で定めることにする．$L_n(\cdot\,|\,\boldsymbol{x})$ は標本空間から $\{\theta \mapsto p(\boldsymbol{x}|\,\theta) : \boldsymbol{x} \in S\}$ なる関数族への対応となる．$\boldsymbol{x}$ があたられたとき，$L_n(\theta|\,\boldsymbol{x})$ 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 $\boldsymbol{x}$ 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

$$L_n(\theta|\,\boldsymbol{x}) = \Pi_{i=1}^n f(x_i|\,\theta)$$

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，$\boldsymbol{X} = \boldsymbol{x}$ が与えられたとき，尤度関数を最大にする値 $\widehat{\theta}(\boldsymbol{x})$ を見つけることである：

$$L_n(\widehat{\theta}(\boldsymbol{x})|\,\boldsymbol{x}) = p(\boldsymbol{x}|\,\widehat{\theta}(\boldsymbol{x})) = \max\{p(\boldsymbol{x}|\,\theta) : \theta \in \Theta\} = \max\{L_n(\theta|\,\boldsymbol{x}) : \theta \in \Theta\}$$

$\widehat{\theta}(\boldsymbol{x})$ を θ の最尤推定値といい，$\widehat{\theta}(\boldsymbol{X})$ を θ の最尤推定量という．

例 6.6 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

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

のとき唯一達成される．したがって，$\widehat{\theta}(X) = X$ 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

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

となる．よって，最大は

$$\widehat{\theta}(\boldsymbol{x}) = \bar{x}_n, \qquad \bar{x}_n = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$

で達成される．したがって，最尤推定量は $\widehat{\theta}(\boldsymbol{X}) = \bar{X}_n$ 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，

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

と記す[(6-4)]．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

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

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 $\widehat{\theta}(\boldsymbol{x})$ が存在するならば，$\widehat{\theta}(\boldsymbol{x})$ は方程式

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

をみたす．この方程式を尤度方程式という．

例 6.7 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：

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

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

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

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 $\widehat{\theta}=5/6$ を得る．これは

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

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

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

とする．尤度関数は

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

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，

$$\widehat{\theta}(\boldsymbol{x})=\frac{2n_1+n_2}{2(n_1+n_2+n_3)}$$

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 6.8 $X_1, X_2, \ldots, X_n$ は独立同一分布に従い，各 $X_i, i = 1, 2, \ldots, n,$ は $1, 2, \ldots, k$ の値をとり，その確率は $\theta_j = \mathbb{P}\{X_i = j\}, j = 1, 2, \ldots, k,$ 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．$\boldsymbol{X} = \boldsymbol{x}$ 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

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

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

$$\theta_k = 1 - \sum_{j=1}^{k-1} \theta_j \tag{6.1}$$

である．

まず，$n_j > 0, j = 1, 2, \ldots, k,$ を仮定する．このとき，θ_j のどれかがゼロならば，$p(\boldsymbol{x}|\,\boldsymbol{\theta}) = 0$ 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

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

となる．(6.1) から $\partial\theta_k/\partial\theta_j = -1, j = 1, 2, \ldots, k-1,$ となる．よって，尤度方程式は

$$\frac{\widehat{\theta}_k}{\widehat{\theta}_j} = \frac{n_k}{n_j}, \qquad j = 1, 2, \ldots, k-1$$

となる．これを再度 (6.1) に代入すれば，$\widehat{\theta}_k = n_k/n$ となり，

$$\widehat{\theta}_j = \frac{n_j}{n}, \qquad j = 1, 2, \ldots, k$$

となる．ただし，$n = \sum_{\ell=1}^k n_\ell$ とした．つぎに，$\theta_j = n_j/n, j = 1, 2, \ldots, k$ が実際に $l_n(\boldsymbol{\theta})$ を最大にしていることを確認するために，$l_n(\boldsymbol{\theta})$ は $(\theta_1, \theta_2, \ldots, \theta_{k-1})$ 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，

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

となる．

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

定理 6.2 $T(\boldsymbol{X})$ 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 $p(\boldsymbol{x}|\,\theta)$ 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 $\boldsymbol{x}$ に依存する関数 g と $\boldsymbol{x}$ のみに依存する関数 h が存在して，

$$p(\boldsymbol{x}|\,\theta) = h(\boldsymbol{x})g(T(\boldsymbol{x})|\,\theta)$$

と書ける．これより θ に関して $p(\boldsymbol{x}|\,\theta)$ を最大化することは θ に関して $g(T(\boldsymbol{x})|\,\theta)$ を最大化することと同値になる．また，$T(\boldsymbol{x}) = t$ 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 $\boldsymbol{x}$ において $g(T(\boldsymbol{x})|\,\theta)$ 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 6.3 関数 $g: \mathbb{R} \to \mathbb{R}$ が 1 対 1 のとき，$\widehat{\theta}$ が θ の最尤推定量であれば，$g(\widehat{\theta})$ 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

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

と書けるので，

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

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

例 6.9（指数分布族の最尤推定量） $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)'$ は k 母数指数分布族

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

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

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

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

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

となる．系 **??** を用いれば，

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

と

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

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

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

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