4　最尤推定

推定量の構成についてよく用いられる最尤法（さいゆうほう）について，例題を用いて 1 変量の場合を復習しよう．平均 μ（未知），分散 σ^2（既知）の正規母集団からの無作為標本 $X_1, X_2, \ldots, X_n$ の同時確率密度関数 $f(x_1, x_2, \ldots, x_n; \mu)$ は，

$$f(x_1, x_2, \ldots, x_n; \mu) = \prod_{i=1}^{n} f(x_i; \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) \tag{10}$$

である．同時確率密度関数は，変数 $x_1, x_2, \ldots, x_n$ の関数であることに注意する．いま，実現値

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$$

が得られたとする．σ^2 は既知であるため，(10) を変数 $x_1, x_2, \ldots, x_n$ が固定されたパラメータ μ の関数と考えることにする．実現値 $x_1, x_2, \ldots, x_n$ に対して，$f(x_1, x_2, \ldots, x_n; \mu)$ の値は，μ によって大きくなったり小さくなったりする．つまり，得られた実現値は $f(x_1, x_2, \ldots, x_n; \mu)$ の値を大きくする μ をもつ正規母集団から得られたと考える方が自然である．したがって，$f(x_1, x_2, \ldots, x_n; \mu)$ の値を最も大きくする μ を求めれば良さそうであるというのが最尤法の基本的な考え方である．

定義 6. 結合確率 (密度) 関数 $f(x_1, x_2, \ldots, x_n; \theta)$ を $x_1, x_2, \ldots, x_n$ を所与として θ の関数とした $L(\theta; x_1, x_2, \ldots, x_n)$ を**尤度関数**（ゆうどかんすう）（likelihood function）という．つまり，

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \theta)$$

である．また，その対数をとったもの $\ell(\theta; x_1, x_2, \ldots, x_n) = \log L(\theta; x_1, x_2, \ldots, x_n)$ を**対数尤度関数**（たいすうゆうどかんすう）（log-likelihood function）という．

確率密度関数は，確率変数がとる値それぞれの出現しやすさを表したものであった．つまり，確率密度 $f(x; \theta)$ が大きい x は出現しやすく，逆に $f(x; \theta)$ が小さい x は出現しにくい．尤度関数 $L(\theta; x_1, x_2, \ldots, x_n)$ は，$x_1, x_2, \ldots, x_n$ を所与とした θ の関数である．この節の冒頭でも述べたように，θ を動かして尤度関数を最大にする $\hat{\theta}$ を考えるのは自然な発想である．なぜなら，得られたデータ $x_1, x_2, \ldots, x_n$ の出現しやすさを最大にす

る $\hat{\theta}$ を選ぶことができれば，その $\hat{\theta}$ をパラメータとしてもつ母集団からの無作為標本の実現値が $x_1, x_2, \ldots, x_n$ に似た値となる可能性が高くなるからである．

定義 7. 尤度関数（または対数尤度関数）を最大にする値 $\hat{\theta} = \hat{\theta}(x_1, x_2, \ldots, x_n)$ を最尤推定値といい

$$L(\hat{\theta}; x_1, x_2, \ldots, x_n) = \sup_{\theta \in \Theta} L(\theta; x_1, x_2, \ldots, x_n)$$

を満たす．また，$\hat{\theta}(X_1, X_2, \ldots, X_n)$ を θ の**最尤推定量**（maximum likelihood estimator）という．

定義7より，

$$\frac{dL(\theta; x_1, x_2, \ldots, x_n)}{d\theta} = 0 \tag{11}$$

なる方程式の解が θ の最尤推定値の候補になる．ここで，$\log x$ は単調関数であるから，尤度関数 $L(\mu; x_1, x_2, \ldots, x_n)$ の最大化は，対数尤度関数の最大化と同値なので

$$\frac{d\ell(\theta; x_1, x_2, \ldots, x_n)}{d\theta} = 0 \tag{12}$$

を θ について解いても良い．ここに (11)，(12) を**尤度方程式**（likelihood equation）という．

平均 μ (未知)，分散 σ^2 (既知) の正規母集団からの無作為標本 $X_1, X_2, \ldots, X_n$ の同時確率密度関数 $f(x_1, x_2, \ldots, x_n; \mu)$ は，(10) である．対数尤度関数は

$$\ell(\mu; x_1, x_2, \ldots, x_n) = -\frac{n}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

である．対数尤度関数は μ についての関数であり，上に凸であるから尤度方程式

$$\frac{d\ell(\mu; x_1, x_2, \ldots, x_n)}{d\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \tag{13}$$

を解くと，尤度関数の最大化となる．式 (13) を解くと

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

となり，平均 μ の最尤推定量は

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

である．この例の場合，最尤法によって構成された推定量が有効推定量でもあることに注意する．一般に，最尤推定量が良い性質を持つことが知られている．しかし，不偏性や有効性のような推定の誤差に関する記述は定義の中にはない．したがって，最尤推定量の評価は，それぞれの問題設定において確認しなければならない．

次に，パラメータ θ が 1 つではなく，2 つある場合を扱う．平均 μ（未知），分散 σ^2（未知）の正規母集団からの無作為標本 $X_1, X_2, \ldots, X_n$ の同時確率密度関数 $f(x_1, x_2, \ldots, x_n; \mu, \sigma^2)$ は，既知の分散を未知として扱った (10) である．したがって，対数尤度関数は

$$\ell(\mu, \sigma^2; x_1, x_2, \ldots, x_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

である．尤度方程式は

$$\begin{aligned}
\frac{\partial \ell(\mu, \sigma^2; x_1, x_2, \ldots, x_n)}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \\
\frac{\partial \ell(\mu, \sigma^2; x_1, x_2, \ldots, x_n)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0
\end{aligned}$$

である．この連立方程式を μ と σ^2 について解くと，

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

を得る．以上から，平均 μ と分散 σ^2 の最尤推定量は，それぞれ標本平均 $\overline{X}$ と標本分散 S^2 である．ただし，

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

ここから，p 変量の場合を考えよう．平均ベクトル $\boldsymbol{\mu}$（未知），共分散行列 $\boldsymbol{\Sigma}$（未知）の p 変量正規母集団からの無作為標本 $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ の同時確率密度関数 $f(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ は，

$$\prod_{i=1}^{n} f(\boldsymbol{x}_i; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{pn/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu})\right]$$

である．ただし，

$$\boldsymbol{X}_i = (X_{1i}, \ldots, X_{pi})', \quad \boldsymbol{x}_i = (x_{1i}, \ldots, x_{pi})'$$

であり，$n > p$ とする．したがって，対数尤度関数は

$$\ell = -\frac{pn}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}) \tag{14}$$

である．ここで，1 変量の自然な拡張として，標本平均 $\overline{X}$ に対応して，

$$\overline{\boldsymbol{X}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{X}_i = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_{1i} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} X_{pi} \end{pmatrix} = \begin{pmatrix} \overline{X}_1 \\ \vdots \\ \overline{X}_p \end{pmatrix}$$

n 倍した標本分散 nS^2 に対応して，

$$\begin{aligned}
\boldsymbol{A} &= \sum_{i=1}^{n}(\boldsymbol{X}_i - \overline{\boldsymbol{X}})(\boldsymbol{X}_i - \overline{\boldsymbol{X}})' \\
&= \sum_{i=1}^{n}\begin{pmatrix} X_{1i} - \overline{X}_1 \\ \vdots \\ X_{pi} - \overline{X}_p \end{pmatrix}(X_{1i} - \overline{X}_1, \ldots, X_{pi} - \overline{X}_p) \\
&= \sum_{i=1}^{n}\begin{pmatrix} (X_{1i} - \overline{X}_1)^2 & \cdots & (X_{1i} - \overline{X}_1)(X_{pi} - \overline{X}_p) \\ \vdots & \ddots & \vdots \\ (X_{pi} - \overline{X}_p)(X_{1i} - \overline{X}_1) & \cdots & (X_{pi} - \overline{X}_p)^2 \end{pmatrix} \\
&= \left(\sum_{i=1}^{n}(X_{si} - \overline{X}_s)(X_{ti} - \overline{X}_t)\right)
\end{aligned}$$

を導入する．

補題 1. 任意のベクトル $\boldsymbol{b}$ に対して，

$$\sum_{i=1}^{n}(\boldsymbol{X}_i-\boldsymbol{b})(\boldsymbol{X}_i-\boldsymbol{b})' = \sum_{i=1}^{n}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})(\boldsymbol{X}_i-\overline{\boldsymbol{X}})' + n(\overline{\boldsymbol{X}}-\boldsymbol{b})(\overline{\boldsymbol{X}}-\boldsymbol{b})'$$

証明.

$$\begin{aligned}\sum_{i=1}^{n}(\boldsymbol{X}_i-\boldsymbol{b})(\boldsymbol{X}_i-\boldsymbol{b})' &= \sum_{i=1}^{n}(\boldsymbol{X}_i-\overline{\boldsymbol{X}}+\overline{\boldsymbol{X}}-\boldsymbol{b})(\boldsymbol{X}_i-\overline{\boldsymbol{X}}+\overline{\boldsymbol{X}}-\boldsymbol{b})' \\ &= \boxed{\text{読者の演習}} \\ &= \sum_{i=1}^{n}(\boldsymbol{X}_i-\overline{\boldsymbol{X}})(\boldsymbol{X}_i-\overline{\boldsymbol{X}})' + n(\overline{\boldsymbol{X}}-\boldsymbol{b})(\overline{\boldsymbol{X}}-\boldsymbol{b})'\end{aligned}$$

最後の等号は，

$$\sum_{i=1}^{n}(\boldsymbol{X}_i-\overline{\boldsymbol{X}}) = \sum_{i=1}^{n}\boldsymbol{X}_i - n\overline{\boldsymbol{X}} = n\overline{\boldsymbol{X}} - n\overline{\boldsymbol{X}} = \boldsymbol{0}$$

による． □

補題 1 から，

$$\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})' = \sum_{i=1}^{n}(\boldsymbol{x}_i-\overline{\boldsymbol{x}})(\boldsymbol{x}_i-\overline{\boldsymbol{x}})' + n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})(\overline{\boldsymbol{x}}-\boldsymbol{\mu})'$$

である．ただし，

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}x_{1i} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n}x_{pi} \end{pmatrix} = \begin{pmatrix} \overline{x}_1 \\ \vdots \\ \overline{x}_p \end{pmatrix}$$

である．また，Tr を行列の対角和（トレース）とし，

$$\begin{aligned}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) &= \mathrm{Tr}\left(\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right) \\ &= \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\sum_{i=1}^{n}(\boldsymbol{x}_i-\boldsymbol{\mu})(\boldsymbol{x}_i-\boldsymbol{\mu})'\right) \\ &= \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}\left(\hat{\boldsymbol{A}}+n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})(\overline{\boldsymbol{x}}-\boldsymbol{\mu})'\right)\right) \\ &= \mathrm{Tr}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{A}} + \mathrm{Tr}\left(\boldsymbol{\Sigma}^{-1}n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})(\overline{\boldsymbol{x}}-\boldsymbol{\mu})'\right) \\ &= \mathrm{Tr}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{A}} + n(\overline{\boldsymbol{x}}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}}-\boldsymbol{\mu})\end{aligned}$$

と表すことができる．ただし，

$$\hat{\boldsymbol{A}} = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'$$

である．したがって，対数尤度 (14) は

$$\ell = -\frac{pn}{2}\log(2\pi) - \frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{Tr}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{A}} - \frac{n}{2}(\overline{\boldsymbol{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\overline{\boldsymbol{x}} - \boldsymbol{\mu}) \tag{15}$$

となる．

補題 2. $\boldsymbol{D}$ が次数 p の正定値行列ならば，正定値行列 $\boldsymbol{G}$ に関して，

$$f(\boldsymbol{G}) = -n\log|\boldsymbol{G}| - \mathrm{Tr}\boldsymbol{G}^{-1}\boldsymbol{D}$$

の最大値が存在し，$\boldsymbol{G} = (1/n)\boldsymbol{D}$ のとき最大値

$$f((1/n)\boldsymbol{D}) = pn\log n - n\log|\boldsymbol{D}| - pn$$

をとる．

以上から，$\boldsymbol{\Sigma}$ が正定値行列より $\boldsymbol{\Sigma}^{-1}$ も正定値行列であるから，式 (15) の右辺第 4 項の二次形式が 0 のとき対数尤度は最大になるから，

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{x}}$$

を得る．次に，式 (15) の右辺第 1 項は定数であるから無視して，第 2 項と第 3 項に着目すると

$$-\frac{n}{2}\log|\boldsymbol{\Sigma}| - \frac{1}{2}\mathrm{Tr}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{A}} = \frac{1}{2}\left(-n\log|\boldsymbol{\Sigma}| - \mathrm{Tr}\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{A}}\right)$$

であり，右辺の () の中を最大化すれば良いことがわかる．ここで，補題 2 を用いれば，$\boldsymbol{\Sigma} = (1/n)\hat{\boldsymbol{A}}$ のとき最大になるから

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\hat{\boldsymbol{A}}$$

を得る．したがって，次の定理を得る．

定理 9. $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ が平均ベクトル $\boldsymbol{\mu}$，共分散行列 $\boldsymbol{\Sigma}$ の p 変量正規母集団からの無作為標本ならば，$\boldsymbol{\mu}$ と $\boldsymbol{\Sigma}$ の最尤推定量はそれぞれ

$$\hat{\boldsymbol{\mu}} = \overline{\boldsymbol{X}}, \quad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{A}$$

である．