確率・統計 B
標本分布・推定

若木宏文

wakaki@math.sci.hiroshima-u.ac.jp

2013.12.18

table of contents

復習

最小 2 乗法と単回帰モデル (1/7)

最小 2 乗法

モデル

$$y_i = g_i(\theta_1, \ldots, \theta_p) + \varepsilon_i, \quad i = 1, \ldots, n$$

g_i : 既知関数
$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$: パラメータ
$\varepsilon_1, \ldots, \epsilon_n$: 誤差, 互いに独立で, 平均 0, 分散 σ^2

$$\text{残差平方和：} \quad \sum_{i=1}^{n} \{y_i - g_i(\theta_1, \ldots, \theta_p)\}^2$$

を最小にするような $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ を求める方法を最小 2 乗法, $\hat{\boldsymbol{\theta}}$ を最小 2 乗推定量という

一様最小分散不偏推定量 (7/10,8/10 より)

定理 7.3 (クラーメルーラオの不等式)

$\boldsymbol{X} = (X_1, \ldots, X_n) \sim \mathsf{P} \in \{\mathsf{P}_{\boldsymbol{\theta}} \,|\, \boldsymbol{\theta} \in \Theta\}, \quad \Theta \subset \mathbb{R}^p$
$f(\boldsymbol{x};\theta)$: P_θ の確率密度関数 (確率関数)
$\gamma(\boldsymbol{\theta})$: 微分可能な関数 (推定したい値)

$\gamma(\boldsymbol{\theta})$ の任意の不偏推定量 $\hat{\gamma}(\boldsymbol{X})$ に対して

$$\mathsf{Var}\{\hat{\gamma}(\boldsymbol{X})\} \geq \dot{\gamma}(\boldsymbol{\theta})' J(\boldsymbol{\theta})^{-1} \dot{\gamma}(\boldsymbol{\theta})$$

が成り立つ. (等号が成立するとき有効推定量という.)
ただし, $\dot{\gamma}(\boldsymbol{\theta}) = \partial\gamma(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $J(\boldsymbol{\theta})$ はフィッシャー情報量行列 :

$$J(\boldsymbol{\theta}) = \mathsf{E}_\theta \left[\left\{ \frac{\partial}{\partial\boldsymbol{\theta}} \log f(\boldsymbol{X}, \boldsymbol{\theta}) \right\} \left\{ \frac{\partial}{\partial\boldsymbol{\theta}} \log f(\boldsymbol{X}, \boldsymbol{\theta}) \right\}' \right]$$

一様最小分散不偏推定量 (9/10)

例 7.5 (単回帰モデル)

$$
\begin{aligned}
&y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \varepsilon_i, \quad i = 1, \ldots, n \\
&\boldsymbol{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \\
&\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_1 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}
\end{aligned}
$$

※. $\boldsymbol{\beta}$ の最小 2 乗推定量は, $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$ と表わさる.

誤差に正規性を仮定すると, $\boldsymbol{y}$ の同時確率密度関数は

$$
f(\boldsymbol{y}; \boldsymbol{\beta}, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - X\boldsymbol{\beta})'(\boldsymbol{y} - X\boldsymbol{\beta})\right\}
$$

一様最小分散不偏推定量 (10/10)

例 7.5 (単回帰モデル (つづき))

$\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$ のフィッシャー情報量行列は

$$J(\boldsymbol{\theta}) = \begin{pmatrix} \frac{1}{\sigma^2} X'X & O \\ O & \frac{n}{2\sigma^4} \end{pmatrix}$$

$\boldsymbol{\ell} = (\ell_1, \ell_2)'$ とするとクラメールーラオの不等式より, ,
$\gamma(\boldsymbol{\beta}) = \boldsymbol{\ell}'\boldsymbol{\beta}$ の任意の不偏推定量に対して

$$\mathrm{Var}\{\hat{\gamma}(\boldsymbol{y})\} \geq \sigma^2 \boldsymbol{\ell}'(X'X)^{-1}\boldsymbol{\ell}$$

一方, $\boldsymbol{\beta}$ の最小 2 乗推定量: $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$ の分散は

$$\mathrm{Var}\{\boldsymbol{\ell}'\hat{\boldsymbol{\beta}}\} = \sigma^2 \boldsymbol{\ell}'(X'X)^{-1}\boldsymbol{\ell}.$$

$\boldsymbol{\ell}'\hat{\boldsymbol{\beta}}$ は, $\boldsymbol{\ell}'\boldsymbol{\beta}$ の有効推定量であり, したがって一様最小分散不偏推定量でもある.

十分統計量

十分統計量 (1/8)

定義 7.2

$\boldsymbol{X} = (X_1, \ldots, X_n)' \sim \mathsf{P}$, $\mathsf{P} \in \mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}};\ \boldsymbol{\theta} \in \Theta\}$ とする．統計量 $T = T(\boldsymbol{X})$ は, $T = t$ を与えたときの $\boldsymbol{X}$ の条件付き分布が $\boldsymbol{\theta}$ に無関係であるとき, $\boldsymbol{\theta}$ あるいは, 分布族 $\mathcal{P}$ の十分統計量であるという．

例 7.6

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} B(1, p)$,　$T = \sum_{i=1}^n X_i \sim B(n, p)$

$x_i = 0$ または 1, $t = \sum_{i=1}^n x_i$ とすると

$$
\begin{aligned}
\mathsf{P}(X_1 = x_1, \ldots, X_n = x_n | T = t) &= \frac{\mathsf{P}(X_1 = x_1, \ldots, X_n = x_n, T = t)}{\mathsf{P}(T = t)} \\
&= \frac{p^t (1-p)^{n-t}}{{}_nC_t p^t (1-p)^{n-t}} = \frac{1}{{}_nC_t}
\end{aligned}
$$

したがって, T は p の十分統計量

十分統計量 (2/8)

定理 7.4 (ラオーブラックウエルの定理)

$\boldsymbol{X} = (X_1, \ldots, X_n)' \sim \mathsf{P}$, $\mathsf{P} \in \mathcal{P} = \{\mathsf{P}_{\boldsymbol{\theta}};\ \boldsymbol{\theta} \in \Theta\}$
$T = T(\boldsymbol{X})$: 十分統計量
$\hat{\gamma}(\boldsymbol{X})$: $\gamma(\boldsymbol{\theta})$ の不偏推定量

$\gamma^*(T) = \mathsf{E}\{\hat{\gamma}(\boldsymbol{X})|T\}$ と定義すると

$$\mathsf{E}\{\gamma^*(T)\} = \gamma(\boldsymbol{\theta}) \quad かつ \quad \mathsf{Var}\{\gamma^*(T)\} \leq \mathsf{Var}\{\hat{\gamma}(\boldsymbol{X})\}$$

が成立する. (証明は板書で)

注. $\gamma^*(T)$ は $\boldsymbol{\theta}$ に無関係であるので, 推定量として用いることができる.

十分統計量 (3/8)

定理 7.5 (分解定理)

$f(\boldsymbol{x};\boldsymbol{\theta})$: $\boldsymbol{X} = (X_1, \ldots, X_n)'$ の確率密度関数

$T(\boldsymbol{X})$ が $\boldsymbol{\theta}$ の十分統計量であるための必要十分条件は,

f が非負値関数 g, h を用いて

$$f(\boldsymbol{x};\boldsymbol{\theta}) = g(\boldsymbol{x})h(T(\boldsymbol{x});\boldsymbol{\theta})$$

と書けること. (証明は省略, テキスト p153 参照)

十分統計量 (4/8)

定義 7.3 (完備十分統計量)

T は $\boldsymbol{\theta}$ の十分統計量であるとする.

任意の $\boldsymbol{\theta} \in \Theta$ に対して, $\mathsf{E}_{\boldsymbol{\theta}}\{g(T(\boldsymbol{X}))\} = 0 \Rightarrow$ 確率 *1* で $g(T(\boldsymbol{X})) = 0$

が成り立つとき T は完備であるという.

例 7.6 (続き)

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathsf{B}(1, p), \quad T = \sum_{i=1}^{n} X_i \sim \mathsf{B}(n, p)$

T は p の完備十分統計量である. (証明は板書で)

十分統計量 (5/8)

定理 7.6

$\{f_{\boldsymbol{\theta}};\ \boldsymbol{\theta} \in \Theta\}$ を次で与えられる分布族とする.

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) = \exp\left\{\sum_{i=1}^{p} \eta_i(\boldsymbol{\theta}) T_i(\boldsymbol{x}) + D(\boldsymbol{\theta}) + S(\boldsymbol{x})\right\}.$$

$\boldsymbol{x} \in \mathbb{R}^q$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)' \in \Theta$, Θ は $\mathbb{R}^p$ の開区間,
$T_1, \ldots, T_p, S : \mathbb{R}^q$ 上で定義された関数
D は Θ 上で定義された関数

$p \leq q$ であり, $\boldsymbol{\eta}(\boldsymbol{\theta}) = (\eta_1(\boldsymbol{\theta}), \ldots, \eta_p(\boldsymbol{\theta}))'$ による Θ の像 $\boldsymbol{\eta}(\Theta)$ が, $\mathbb{R}^p$ の開集合を含むとすると

$$\boldsymbol{T} = (T_1(\boldsymbol{X}), \ldots, T_p(\boldsymbol{X}))'$$

は $\boldsymbol{\theta}$ の完備十分統計量.

十分統計量 (6/8)

定理 7.7 (完備十分統計量と一様最小分散不偏推定量)

$\boldsymbol{X} = (X_1, \ldots, X_n)' \sim \mathrm{P}$, $\mathrm{P} \in \mathcal{P} = \{\mathrm{P}_{\boldsymbol{\theta}};\ \boldsymbol{\theta} \in \Theta\}$
$\boldsymbol{T} = \boldsymbol{T}(\boldsymbol{X})$: $\boldsymbol{\theta}$ の完備十分統計量

$\hat{\gamma}(\boldsymbol{T})$ が $\gamma(\boldsymbol{\theta})$ の不偏推定量
$\Rightarrow$ $\hat{\gamma}(\boldsymbol{T})$ は一様最小分散不偏推定量

(証明は板書で)

例 7.7

$X_1, \ldots, X_n$ を正規母集団 $N(\mu, \sigma^2)$ からのランダム標本とする.
$X_1, \ldots, X_n$ の同時確率密度関数は

$$f(x_1, \ldots, x_n; \mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 + n\bar{x}_n^2\right) + \frac{n\mu}{\sigma^2}\bar{x}_n - \frac{n\mu^2}{2\sigma^2}\right\}$$

と表されるので

十分統計量 (7/8)

例 7.7 (続き)

$$\boldsymbol{T} = (T_1, T_2)' = \left(\sum_{i=1}^{n} (X_i - \bar{X}_n)^2 + n\bar{X}_n^2,\ \bar{X}_n \right)'$$

は完備十分統計量である．定理 6.8 より標本分散

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 = \frac{1}{n-1} (T_1 - nT_2)$$

は, σ^2 の不偏推定量であり,
定理 7.7 より, 一様最小分散不偏推定量

十分統計量 (8/8)

例 7.7 (続き)

$N(\mu, \sigma^2)$ のフィッシャー情報量行列は

$$\begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^4} \end{pmatrix}$$

となるので, 系 7.1 より, σ^2 の不偏推定量 $\hat{\sigma}^2$ に関する情報量不等式は $\mathrm{Var}\{\hat{\sigma}^2\} \geq \frac{\sigma^4}{2n}$ となる. 定理 6.8 と χ^2_{n-1} の分散が $2(n-1)$ となることから

$$\mathrm{Var}(S_n^2) = \frac{2\sigma^4}{n-1}$$

となり, 情報量不等式の下限より大きくなる. したがって, σ^2 の有効推定量は存在しない.

最尤法

最尤法 (1/6)

定義 7.4

$f(\boldsymbol{x})$: $\boldsymbol{X} = (X_1, \ldots, X_n)'$ の確率密度関数
確率密度関数 $g(\boldsymbol{x})$ に対して

$$\mathrm{KL}(f; g) = -\mathsf{E}\left\{\log \frac{g(\boldsymbol{X})}{f(\boldsymbol{X})}\right\}$$

を g の f からのカルバックーライブラーの擬距離とよぶ.

補題 7.1

任意の g に対して, $\mathrm{KL}(f; g) \geq 0$ が成り立ち, 等号は確率 1 で $f(\boldsymbol{X}) = g(\boldsymbol{X})$ のときに限る.
(証明は板書で)

最尤法 (2/6)

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} f(x; \boldsymbol{\theta}_0)$, $\boldsymbol{\theta}_0 \in \Theta \subset \mathbb{R}^p$

(仮定 A0) $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$ ならば $\mathrm{P}(f(X; \boldsymbol{\theta}) \neq f(X; \boldsymbol{\theta}')) > 0$.

大数の強法則から

$$\frac{1}{n} \sum_{i=1}^{n} \log \frac{f(X_i; \boldsymbol{\theta})}{f(X_i; \boldsymbol{\theta}_0)} \overset{a.s.}{\to} -\mathrm{KL}(f(*; \boldsymbol{\theta}_0); f(*; \boldsymbol{\theta})) \quad (n \to \infty)$$

である.

補題 7.1 より 右辺は $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ で最大値をとる.

左辺を最大とする $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}(X_1, \ldots, X_n)$ は,
$n \to \infty$ のとき θ_0 に収束 (A0 以外にも条件が必要)

最尤法 (3/6)

尤度関数

標本変量 $\boldsymbol{X} = (X_1, \ldots X_n)'$ に対する統計モデルを

$$\mathrm{P}(\boldsymbol{X} \in A) = \begin{cases} \displaystyle\sum_{\boldsymbol{x} \in A} f(\boldsymbol{x}; \boldsymbol{\theta}), & \boldsymbol{X} \text{ が離散型の場合} \\ \displaystyle\int_{\boldsymbol{x} \in A} f(\boldsymbol{x}; \boldsymbol{\theta}) d\boldsymbol{x}, & \boldsymbol{X} \text{ が連続型の場合} \end{cases}$$

とする．ここに, $\boldsymbol{\theta} \in \Theta$. 確率密度関数 $f(\boldsymbol{x}; \boldsymbol{\theta})$ を ($\boldsymbol{x}$ を固定して) $\boldsymbol{\theta}$ の関数とみなすとき

$$L(\boldsymbol{\theta}; \boldsymbol{x}) \quad (= f(\boldsymbol{x}; \boldsymbol{\theta})) \tag{1}$$

と表し, 尤度関数とよぶ.

最尤法 (4/6)

定義 7.5 (最尤推定量)

尤度関数 $L(\boldsymbol{\theta};\boldsymbol{x})$ の最大を実現する $\boldsymbol{\theta}$ を $\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ と表し, $\boldsymbol{\theta}$ の最尤推定値という.

$$L(\hat{\boldsymbol{\theta}};\boldsymbol{x})=\sup_{\boldsymbol{\theta}\in\Theta}L(\boldsymbol{\theta};\boldsymbol{x}).$$

$\hat{\boldsymbol{\theta}}(\boldsymbol{X})$ を $\boldsymbol{\theta}$ の最尤推定量という.

対数尤度 $\log L(\boldsymbol{\theta};\boldsymbol{x})$ を $\ell(\boldsymbol{\theta};\boldsymbol{x})$ と表す. 多くの場合, 最尤推定値は尤度方程式

$$\frac{\partial}{\partial\boldsymbol{\theta}}\ell(\boldsymbol{\theta};\boldsymbol{x})=\mathbf{0}$$

の解として与えられる.

最尤法 (5/6)

例 7.7

(1) $X_1, \ldots, X_n \overset{i.i.d.}{\sim} B(1, \theta)$, $\theta \in (0, 1)$

尤度関数: $L(\theta; \boldsymbol{x}) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$

最尤推定値: $\hat{\theta} = \bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$

(2) $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, $\boldsymbol{\theta} = (\mu, \sigma^2)'$

尤度関数: $L(\boldsymbol{\theta}; \boldsymbol{x}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right\}$

最尤推定値: $\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = s^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x})^2$

(3) $X_1, \ldots, X_n \overset{i.i.d.}{\sim} U(0, \theta)$

尤度関数: $L(\theta, x_1, \ldots, x_n) = \frac{1}{\theta^n},\ 0 < x_i \leq \theta,\ \theta \in \Theta = (0, \infty)$

最尤推定値:$\hat{\theta} = \max_{i=1,\ldots,n} x_i$

最尤法 (6/6)

最尤推定量の性質

(i) 不偏性

$\mathrm{E}(\hat{\theta}) = \theta$

(ii) 一致性

$\hat{\theta} \xrightarrow{p} \theta \quad (n \to \infty)$

(iii) 漸近正規性

標準化すると $n \to \infty$ のとき正規分布に分布収束する性質

例 7.8

	不偏性	一致性	漸近正規性
例 7.7(1)	○	○	○
例 7.7(2)$\hat{\mu}$	○	○	○
例 7.7(2)$\hat{\sigma}^2$	○	○	○
例 7.7(3)	×	○	×