確率・統計 B
一様最小分散不偏推定量

若木宏文

wakaki@hiroshima-u.ac.jp
http://home.hiroshima-u.ac.jp/ wakaki/lecture/probstatB17/index.shtml

2017.12.01

table of contents

一様最小分散不偏推定量

十分統計量

一様最小分散不偏推定量 (1/11)

定理 1 (クラメール–ラオの不等式 (1 次元パラメーターの場合))

$\boldsymbol{X} = (X_1, \ldots, X_n) \sim \mathsf{P} \in \{\mathsf{P}_\theta \,|\, \theta \in \Theta\}, \quad \Theta \subset \mathbb{R}$
$f(\boldsymbol{x}; \theta)$: P_θ の確率密度関数 (確率関数)

(仮定) 各 $\theta_0 \in \Theta$ に対して θ_0 の近傍 U と 関数 $M(\boldsymbol{x})$ が存在して, $\theta \in U$ ならば

$$\left| \frac{\partial f(\boldsymbol{x}; \theta)}{\partial \theta} \right| \leq M(\boldsymbol{x}), \text{ かつ } \mathsf{E}_\theta \left[\left\{ \frac{M(\boldsymbol{X})}{f(\boldsymbol{X}; \theta)} \right\}^2 \right] < \infty$$

θ の任意の不偏推定量 $\hat{\theta}$ に対して

$$\mathsf{Var}(\hat{\theta}) \geq \frac{1}{J(\theta)}$$

一様最小分散不偏推定量 (2/11)

定理 1 (クラメール－ラオの不等式 (つづき))

ただし,

$$\mathsf{E}_\theta[g(\boldsymbol{X})] = \begin{cases} \displaystyle\iint g(\boldsymbol{x})f(\boldsymbol{x};\theta)dx_1\cdots dx_n & (\text{連続型}) \\ \displaystyle\sum_{j=1}^{\infty} g(\boldsymbol{x}_j)f(\boldsymbol{x}_j;\theta) & (\text{離散型}) \end{cases},$$

$$J(\theta) = \mathsf{E}_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f(\boldsymbol{X};\theta)\right\}^2\right]$$

であり, $J(\theta) > 0$ と仮定する.

- 定理 1 の $J(\theta)$ をフィッシャー情報量と呼ぶ.
- $\mathrm{Var}(\hat{\theta}) = \frac{1}{J(\theta)}$ となる不偏推定量が存在するとき, それを有効推定量と呼ぶ

一様最小分散不偏推定量 (3/11)

命題 1 (シュワルツの不等式)

可測関数 g, h に対して

$$\{\mathsf{E}[g(\boldsymbol{X})h(\boldsymbol{X})]\}^2 \leq \mathsf{E}[\{g(\boldsymbol{X})\}^2]\mathsf{E}[\{h(\boldsymbol{X})\}^2]$$

(証明のヒント)

任意の実数 t に対して

$$\begin{aligned}&\mathsf{E}[\{tg(\boldsymbol{X}) + h(\boldsymbol{X})\}^2] \\ &\quad = t^2\mathsf{E}[\{g(\boldsymbol{X})\}^2] + 2t\mathsf{E}[g(\boldsymbol{X})h(\boldsymbol{X})] + \mathsf{E}[\{h(\boldsymbol{X})\}^2] \geq 0\end{aligned}$$

であることから, 2 次方程式の判別式を考える.

一様最小分散不偏推定量 (4/11)

補題 1

定理 1 の仮定の下, $\mathsf{E}_\theta[\{g(\boldsymbol{X})\}^2] < \infty$ ならば

$$\frac{d}{d\theta}\mathsf{E}_\theta[g(\boldsymbol{X})] = \mathsf{E}_\theta\left[g(\boldsymbol{X})\frac{\partial \log f(\boldsymbol{X};\theta)}{\partial\theta}\right]$$

証明 (連続型の場合) $\int\int \sim dx_1 \cdots dx_n = \int \sim d\boldsymbol{x}$ と書く.

$$\left|g(\boldsymbol{x})\frac{\partial f(\boldsymbol{x};\theta)}{\partial\theta}\right| \le |g(\boldsymbol{x})|\left\{\frac{M(\boldsymbol{x})}{f(\boldsymbol{x};\theta)}\right\}f(\boldsymbol{x};\theta),$$

$$\int |g(\boldsymbol{x})|\left\{\frac{M(\boldsymbol{x})}{f(\boldsymbol{x};\theta)}\right\}f(\boldsymbol{x};\theta)d\boldsymbol{x} = \mathsf{E}_\theta\left[|g(\boldsymbol{X})|\left\{\frac{M(\boldsymbol{X})}{f(\boldsymbol{X};\theta)}\right\}\right]$$

$$\le \left\{\mathsf{E}_\theta[\{g(\boldsymbol{X})\}^2]\mathsf{E}_\theta\left[\left\{\frac{M(\boldsymbol{X})}{f(\boldsymbol{X};\theta)}\right\}^2\right]\right\}^{1/2} < \infty$$

(シュワルツの不等式)

一様最小分散不偏推定量 (5/11)

優収束定理より

$$
\begin{aligned}
\frac{d}{d\theta}\mathsf{E}_\theta[g(\boldsymbol{X})] &= \frac{d}{d\theta}\int g(\boldsymbol{x})f(\boldsymbol{x};\theta)d\boldsymbol{x} = \int g(\boldsymbol{x})\frac{\partial f(\boldsymbol{x};\theta)}{\partial\theta}d\boldsymbol{x} \\
&= \int g(\boldsymbol{x})\frac{\partial \log f(\boldsymbol{x};\theta)}{\partial\theta}f(\boldsymbol{x};\theta)d\boldsymbol{x} = \mathsf{E}_\theta\left[g(\boldsymbol{X})\frac{\partial \log f(\boldsymbol{X};\theta)}{\partial\theta}\right]
\end{aligned}
$$

離散型の場合には, 積分を級数で置き換えればよい.

一様最小分散不偏推定量 (6/11)

定理 1 の証明

補題 2 で $g(\boldsymbol{x}) \equiv 1$ ととると

$$0 = \mathsf{E}_\theta\left[\frac{\partial \log f(\boldsymbol{X};\theta)}{\partial\theta}\right] \tag{1}$$

$\hat{\theta} = \hat{\theta}(\boldsymbol{X})$ を θ の不偏推定量として, $g(\boldsymbol{x}) = \hat{\theta}(\boldsymbol{x})$ ととると

$$1 = \mathsf{E}_\theta\left[\hat{\theta}\frac{\partial \log f(\boldsymbol{X};\theta)}{\partial\theta}\right] \tag{2}$$

$(2) - (1) \times \theta$

$$1 = \mathsf{E}_\theta\left[(\hat{\theta} - \theta)\frac{\partial \log f(\boldsymbol{X};\theta)}{\partial\theta}\right]$$

$$1^2 \leq \mathsf{E}_\theta[(\hat{\theta} - \theta)^2]\mathsf{E}_\theta\left[\left\{\frac{\partial \log f(\boldsymbol{X};\theta)}{\partial\theta}\right\}^2\right] = \mathsf{Var}(\hat{\theta})J(\theta)$$

一様最小分散不偏推定量 (7/11)

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

(仮定) 各 $\boldsymbol{\theta}_0$ に対して, $\boldsymbol{\theta}_0$ の近傍 U と関数 $M(\boldsymbol{x})$ が存在して, $\boldsymbol{\theta} \in U$ ならば

$$\left\| \frac{\partial f(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq M(\boldsymbol{x}), \text{ かつ, } \mathsf{E}_{\boldsymbol{\theta}}\left[\left\{ \frac{M(\boldsymbol{X})}{f(\boldsymbol{X}, \boldsymbol{\theta})} \right\}^2 \right] < \infty$$

$\gamma(\boldsymbol{\theta})$ 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.

一様最小分散不偏推定量 (8/11)

定理 7.3 (クラメール–ラオの不等式 (つづき))

ただし, $\dot{\gamma}(\boldsymbol{\theta}) = \partial\gamma(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$,

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

であり, $J(\boldsymbol{\theta})$ は正則であると仮定する.

$J(\boldsymbol{\theta})$ をフィッシャー情報量行列と呼ぶ. (証明は省略, テキスト 7.3 節参照)

一様最小分散不偏推定量 (9/11)

系 7.3

$X_1, X_2, \ldots, X_n$: P からのランダム標本

$\mathrm{P} \in \{\mathrm{P}_{\boldsymbol{\theta}} \,|\, \boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)' \in \Theta\}, \quad \Theta \subset \mathbb{R}^p$

$J(\boldsymbol{\theta})$: フィッシャー情報量行列

$\Rightarrow$

(i) $\boldsymbol{X} = (X_1, \ldots, X_n)'$ のフィッシャー情報量行列は $nJ(\boldsymbol{\theta})$.

(ii) $J(\boldsymbol{\theta})^{-1}$ の (k, k) 成分を $J^{(kk)}(\boldsymbol{\theta})$ とおくと θ_k の任意の不偏推定量 $\hat{\theta}_k$ に対して

$$\mathrm{Var}(\hat{\theta}_k) \geq \frac{J^{(kk)}(\boldsymbol{\theta})}{n}.$$

一様最小分散不偏推定量 (10/11)

例 7.5 (単回帰モデル)

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

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 (11/11)

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, $\gamma(\boldsymbol{\beta}) = \boldsymbol{\ell}'\boldsymbol{\beta}$ の任意の不偏推定量に対して

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

一方, $\boldsymbol{\beta}$ の最小 2 乗推定量は, $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\boldsymbol{y}$ と表わされ,

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

$\boldsymbol{\ell}'\hat{\boldsymbol{\beta}}$ 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 (1/9)

定義 7.2

$\boldsymbol{X} = (X_1, \ldots, X_n)' \sim \mathrm{P}$, $\mathrm{P} \in \mathcal{P} = \{\mathrm{P}_{\boldsymbol{\theta}};\ \boldsymbol{\theta} \in \Theta\}$ 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

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

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

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

したがって, T は p の十分統計量

十分統計量 (1/9)

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

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

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

$$\mathsf{P}(X_1 = x_1, \ldots, X_n = x_n | T = t) = \frac{\mathsf{P}(X_1 = x_1, \ldots, X_n = x_n, T = t)}{\mathsf{P}(T = t)}$$

$$= \frac{p^t (1-p)^{n-t}}{{}_nC_t p^t (1-p)^{n-t}} = \frac{1}{{}_nC_t}$$

したがって, T は p の十分統計量

十分統計量 (2/9)

定理 7.4 (ラオーブラックウエルの定理)

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

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

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

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 (3/9)

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 (4/9)

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, $\mathrm{E}_{\boldsymbol{\theta}}\{g(T(\boldsymbol{X}))\} = 0 \Rightarrow$ 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)

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

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 (4/9)

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 (5/9)

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 (6/9)

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

(証明は板書で)

十分統計量 (7/9)

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 (8/9)

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 (9/9)

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 7.3 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.