

# 確率・統計 B 推定

若木宏文

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

<http://home.hiroshima-u.ac.jp/wakaki/lecture/probstatB16/index.shtml>

2017.1.6

# table of contents

最小 2 乗法と単回帰モデル

一様最小分散不偏推定量

十分統計量

# 最小 2 乗法と単回帰モデル (1/7)

## 最小 2 乗法 モデル

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

$g_i$  : 既知関数

$\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)'$  : パラメータ

$\varepsilon_1, \dots, \varepsilon_n$  : 誤差, 互いに独立で, 平均 0, 分散  $\sigma^2$

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

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

# 最小 2 乗法と単回帰モデル (2/7)

## 単回帰モデル

$(y_1, x_1), \dots, (y_n, x_n)$  : データ

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = 1, \dots, n$$

を単回帰モデルと呼ぶ.

仮定

- (1)  $(\alpha, \beta)$  : 未知パラメータ
- (2)  $x_1, \dots, x_n$  : 既知定数
- (3)  $\varepsilon_1, \dots, \varepsilon_n$  : 互いに独立な確率変数
- (4)  $E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2, i = 1, \dots, n$

$y_i$  を目的変数 (応答変数),  $x_i$  を説明変数,  $\beta$  を回帰係数 と呼ぶ

## 最小 2 乗法と単回帰モデル (3/7)

## 残差平方和

$$Q(\alpha, \beta) = \sum_{i=1}^n \{y_i - \alpha - \beta x_i\}^2$$

$(\alpha, \beta)$  の最小 2 乗推定量は

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}, \quad \hat{\beta} = \frac{s_{xy}}{s_x^2},$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

$$s_{xy} = \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}), \quad s_x^2 = s_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2,$$

(証明は板書で)

# メモ用紙

## 最小 2 乗法と単回帰モデル (4/7)

## 定理 7.1

$$(1) E(\hat{\alpha}) = \alpha, E(\hat{\beta}) = \beta$$

$$(2) \text{Var}(\hat{\alpha}) = \sigma^2 \left( \frac{1}{n} + \frac{\bar{x}^2}{s_{xx}} \right), \quad \text{Var}(\hat{\beta}) = \frac{\sigma^2}{s_{xx}}, \quad \text{Cov}(\hat{\alpha}, \hat{\beta}) = -\frac{\sigma^2 \bar{x}}{s_{xx}}$$

## 補題 1

確率行列  $\mathbf{X} = (X_{ij})$  ( $m \times n$ ) の平均は, 次のように定義される.

$$E(\mathbf{X}) = \begin{pmatrix} E(X_{11}) & \cdots & E(X_{1n}) \\ \vdots & & \vdots \\ E(X_{m1}) & \cdots & E(X_{mn}) \end{pmatrix}$$

$\mathbf{A}$  ( $p \times m$ ),  $\mathbf{B}$  ( $n \times q$ ),  $\mathbf{C}$  ( $p \times q$ ) を定数行列とすると,  
 $E(\mathbf{AXB} + \mathbf{C}) = \mathbf{AE}(\mathbf{X})\mathbf{B} + \mathbf{C}$  が成り立つ.

(証明は板書で)

# メモ用紙



# メモ用紙

## メモ用紙

## 最小 2 乗法と単回帰モデル (5/7)

### 定理 7.2 (ガウス–マルコフの定理)

単回帰モデル  $y_i = \alpha + \beta x_i + \varepsilon_i$  ( $i = 1, \dots, n$ )  
において

$$E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2, \quad \text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad (i \neq j = 1, \dots, n)$$

とする. 定数  $l_1, l_2$  に対して

- (1)  $\hat{\theta} = l_1 \hat{\alpha} + l_2 \hat{\beta}$  は,  $\theta = \theta(\alpha, \beta) = l_1 \alpha + l_2 \beta$  の線形不偏推定量
- (2) 任意の  $\theta$  の線形不偏推定量  $\tilde{\theta}$  に対して

$$\text{Var}(\hat{\theta}) \leq \text{Var}(\tilde{\theta})$$

注. ここでの線形不偏推定量とは,  $y_1, \dots, y_n$  の一次結合  $a_1 y_1 + \dots + a_n y_n$  で,  $\theta$  の不偏推定量となるもののこと.

(証明は省略. テキスト 9 章を参照)

## 一様最小分散不偏推定量 (1/10)

定理 1 (クラメールーラオの不等式 (1次元パラメーターの場合))

$\mathbf{X} = (X_1, \dots, X_n) \sim P \in \{P_\theta \mid \theta \in \Theta\}$ ,  $\Theta \subset \mathbb{R}$

$f(\mathbf{x}; \theta)$ :  $P_\theta$  の確率密度関数 (確率関数)

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

$$\left| \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta} \right| \leq M(\mathbf{x}), \text{ かつ } E_{\theta_0} \left[ \left\{ \frac{M(\mathbf{X})}{f(\mathbf{X}; \theta_0)} \right\}^2 \right] < \infty$$

$\theta$  の任意の不偏推定量  $\hat{\theta}$  に対して

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

## 一様最小分散不偏推定量 (2/10)

## 定理 1 (クラメル-ラオの不等式 (つづき))

ただし,

$$E_{\theta}[g(\mathbf{X})] = \begin{cases} \iint g(\mathbf{x}) f(\mathbf{x}; \theta) dx_1 \cdots dx_n & (\text{連続型}) \\ \sum_{j=1}^{\infty} g(\mathbf{x}_j) f(\mathbf{x}_j; \theta) & (\text{離散型}) \end{cases},$$
$$J(\theta) = E_{\theta} \left[ \left\{ \frac{\partial}{\partial \theta} \log f(\mathbf{X}; \theta) \right\}^2 \right]$$

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.
- $\text{Var}(\hat{\theta}) = \frac{1}{J(\theta)}$  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 (3/10)

### 命題 1 (シュワルツの不等式)

可測関数  $g, h$  に対して

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

### (証明のヒント)

任意の実数  $t$  に対して

$$\begin{aligned} E[\{tg(\mathbf{X}) + h(\mathbf{X})\}^2] \\ = t^2E[\{g(\mathbf{X})\}^2] + 2tE[g(\mathbf{X})h(\mathbf{X})] + E[\{h(\mathbf{X})\}^2] \geq 0 \end{aligned}$$

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 (4/10)

## 補題 2

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

$$\frac{d}{d\theta} E_{\theta}[g(\mathbf{X})] = E_{\theta}\left[g(\mathbf{X}) \frac{\partial \log f(\mathbf{X}; \theta)}{\partial \theta}\right]$$

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

$$\begin{aligned} \left|g(\mathbf{x}) \frac{\partial f(\mathbf{x}; \theta)}{\partial \theta}\right| &\leq |g(\mathbf{x})| \left\{\frac{M(\mathbf{x})}{f(\mathbf{x}; \theta_0)}\right\} f(\mathbf{x}; \theta_0), \\ \int |g(\mathbf{x})| \left\{\frac{M(\mathbf{x})}{f(\mathbf{x}; \theta_0)}\right\} f(\mathbf{x}; \theta_0) d\mathbf{x} &= E_{\theta_0} \left[|g(\mathbf{X})| \left\{\frac{M(\mathbf{X})}{f(\mathbf{X}; \theta_0)}\right\}\right] \\ &\leq \left\{E_{\theta}[\{g(\mathbf{X})\}^2] E_{\theta_0} \left[\left\{\frac{M(\mathbf{X})}{f(\mathbf{X}; \theta_0)}\right\}^2]\right\}^{1/2} < \infty \end{aligned}$$

(シュワルツの不等式)

## 一様最小分散不偏推定量 (5/10)

優収束定理より

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

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 (6/10)

## 定理 1 の証明

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

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

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

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

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

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

## 一様最小分散不偏推定量 (7/10)

定理 7.3 (クラメル-ラオの不等式)

 $\mathbf{X} = (X_1, \dots, X_n) \sim P \in \{P_{\boldsymbol{\theta}} \mid \boldsymbol{\theta} \in \Theta\}, \quad \Theta \subset \mathbb{R}^p$  $f(\mathbf{x}; \boldsymbol{\theta})$ :  $P_{\boldsymbol{\theta}}$  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  $M(\mathbf{x})$  が存在して,  
 $\boldsymbol{\theta} \in U$  ならば

$$\left\| \frac{\partial f(\mathbf{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq M(\mathbf{x}), \quad \text{かつ, } E_{\boldsymbol{\theta}_0} \left[ \left\{ \frac{M(\mathbf{X})}{f(\mathbf{X}, \boldsymbol{\theta}_0)} \right\}^2 \right] < \infty$$

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

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

が成り立つ.

## 一様最小分散不偏推定量 (8/10)

定理 7.3 (クラメールーラオの不等式 (つづき))

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

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

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 (9/10)

## 例 7.5 (単回帰モデル)

$$y_i = \beta_0 + \beta_1(x_i - \bar{x}) + \varepsilon_i, \quad i = 1, \dots, n$$

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_1 - \bar{x} \\ \vdots & \vdots \\ 1 & x_n - \bar{x} \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

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

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

## 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))

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

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

$\ell = (\ell_1, \ell_2)'$  とすると,  $\gamma(\beta) = \ell'\beta$  の任意の不偏推定量に対して

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

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

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

$\ell'\hat{\beta}$  は,  $\ell'\beta$  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

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

### 例 7.6

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

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

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

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)

$$\mathbf{X} = (X_1, \dots, X_n)' \sim P, P \in \mathcal{P} = \{P_{\theta}; \theta \in \Theta\}$$

$T = T(\mathbf{X})$  : 十分統計量

$\hat{\gamma}(\mathbf{X})$  :  $\gamma(\theta)$  の不偏推定量

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

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

が成立する. (証明は板書で)

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

# メモ用紙



# メモ用紙

## 十分統計量 (3/9)

### 定理 7.5 (分解定理)

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

$T(\mathbf{X})$  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

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

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  $\theta$  の十分統計量であるとする.

任意の  $\theta \in \Theta$  に対して,  $E_{\theta}\{g(T(\mathbf{X}))\} = 0 \Rightarrow$  確率 1 で  
 $g(T(\mathbf{X})) = 0$

が成り立つとき  $T$  は完備であるという.

### 例 7.6 (続き)

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

$T$  は  $p$  の完備十分統計量である. (証明は板書で)

## メモ用紙

## 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.

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

$\mathbf{x} \in \mathbb{R}^q$ ,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)' \in \Theta$ ,  $\Theta$  は  $\mathbb{R}^p$  の開区間,

$T_1, \dots, T_p, S : \mathbb{R}^q$  上で定義された関数

$D$  は  $\Theta$  上で定義された関数

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

$$\mathbf{T} = (T_1(\mathbf{X}), \dots, T_p(\mathbf{X}))'$$

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)

$$\mathbf{X} = (X_1, \dots, X_n)' \sim P, P \in \mathcal{P} = \{P_\theta; \theta \in \Theta\}$$

$\mathbf{T} = \mathbf{T}(\mathbf{X})$  :  $\theta$  の完備十分統計量

$\hat{\gamma}(\mathbf{T})$  が  $\gamma(\theta)$  の不偏推定量

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

(証明は板書で)

## メモ用紙

# メモ用紙



## 十分統計量 (7/9)

### 例 7.7

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

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

と表されるので

## 十分統計量 (8/9)

## 例 7.7 (続き)

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

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

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

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