

確率・統計 B

標本分布・推定

若木宏文

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

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

2016.11.25

table of contents

順序統計量

点推定

順序統計量 (1/5)

定義 6.12

X_1, \dots, X_n を大きさ n のランダム標本とする. これを大きさの順に並べて $X_{(1)} \leq \dots \leq X_{(n)}$ とするとき

$$X_{(1)}, \dots, X_{(n)}$$

をこの標本の順序統計量といい, $X_{(i)}$ を第 i 順序統計量とよぶ.

例. $n = 4$, X_1, X_2, X_3, X_4 の実現値が

$$X_1 = 3, X_2 = 4, X_3 = 1, X_4 = 5$$

$$\Rightarrow X_{(1)} = 1, X_{(2)} = 3, X_{(3)} = 4, X_{(4)} = 5$$

順序統計量 (2/5)

ランダム標本の特性量

(1) 標本メディアン (中央値)

$$\tilde{X} = \begin{cases} X_{(k)} & n = 2k - 1 \\ (X_{(k)} + X_{(k+1)})/2 & n = 2k \end{cases}$$

(2) 標本範囲

$$R = X_{(n)} - X_{(1)}$$

(3) 標本中点

$$(X_{(1)} + X_{(n)})/2$$

(4) $100\alpha\%$ 調整平均

$$\frac{1}{n - 2k} \sum_{i=k+1}^{n-k} X_{(i)}, \quad k = [n\alpha]$$

($[a]$ は a を超えない最大の整数, ガウス記号と呼ばれる)

順序統計量 (3/5)

命題 1 (最大値・最小値の分布)

$F(x)$: 母集団分布の分布関数

$f(x)$: 母集団分布の確率密度関数 (連続型の場合)

(1) $X_{(n)} = \max\{X_1, \dots, X_n\}$ の分布関数と確率密度関数

$$G_n(y) = P(X_{(n)} \leq y) = \{F(y)\}^n,$$

$$g_n(y) = \frac{d}{dy}G_n(y) = nf(y)\{F(y)\}^{n-1}$$

(2) $X_{(1)} = \min\{X_1, \dots, X_n\}$ の分布関数と確率密度関数

$$G_1(y) = P(X_{(1)} \leq y) = 1 - \{1 - F(y)\}^n,$$

$$g_1(y) = \frac{d}{dy}G_1(y) = nf(y)\{1 - F(y)\}^{n-1}$$

(証明は板書で)

メモ用紙

メモ用紙

順序統計量 (4/5)

定理 6.11 (順序統計量の分布)

$F(x)$: 母集団分布の分布関数

$f(x)$: 母集団分布の確率密度関数 (連続型の場合)

$$(1) P(X_{(i)} \leq y) = \sum_{k=i}^n \frac{n!}{k!(n-k)!} \{F(y)\}^k \{1 - F(y)\}^{n-k}.$$

(2) $g_i(y)$: $X_{(i)}$ の確率密度関数

$$g_i(y) = \frac{n!}{(i-1)!(n-i)!} \{F(y)\}^{i-1} \{1 - F(y)\}^{n-i} f(y).$$

順序統計量 (5/5)

定理 6.11 (順序統計量の分布 (つづき))

(3) $i < j, y_i \leq y_j$ のとき

$$\begin{aligned}
 & P(X_{(i)} \leq y_i, X_{(j)} \leq y_j) \\
 &= \sum_{k=j}^n \sum_{l=i}^k \frac{n! \{F(y_i)\}^l \{F(y_j) - F(y_i)\}^{k-l} \{1 - F(y_j)\}^{n-k}}{l!(k-l)!(n-k)!}.
 \end{aligned}$$

(4) $g_{ij}(y_i, y_j) : (X_{(i)}, X_{(j)})$ の同時確率密度関数
 $y_i < y_j$ のとき

$$\begin{aligned}
 & g_{ij}(y_i, y_j) \\
 &= \frac{n! F(y_i)^{i-1} \{F(y_j) - F(y_i)\}^{j-i-1} \{1 - F(y_i)\}^{n-j} f(y_i) f(y_j)}{(i-1)!(j-i-1)!(n-j)!}
 \end{aligned}$$

(証明は (1),(2) のみ板書で, (3),(4) はテキストを参照)

メモ用紙

メモ用紙

7. 推定

点推定 (1/7)

$\mathbf{X} = (X_1, \dots, X_n)$: 標本変量 (確率変数)

$\mathcal{P} = \{P_\theta \mid \theta \in \Theta\}$: 統計モデル

$$\mathbf{X} \sim P \in \mathcal{P}$$

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

$$\mathcal{X} = \{\mathbf{x} = (x_1, \dots, x_n \mid f(\mathbf{x}; \theta)) > 0\}$$

パラメータ $\gamma = \gamma(\theta)$ の点推定 :

各 $\mathbf{x} \in \mathcal{X}$ に γ の推定値 $\hat{\gamma} = \hat{\gamma}(\mathbf{x})$ を対応させる方法

$\hat{\gamma}(\mathbf{x})$: 推定値, $\hat{\gamma}(\mathbf{X})$: 推定量

点推定 (2/7)

例 7.1

 $X_1, \dots, X_{10} \stackrel{i.i.d.}{\sim} B(1, \theta), \theta \in (0, 1)$

$$(1) \hat{\theta}_1 = (X_1 + \dots + X_{10})/10$$

$$(2) \hat{\theta}_2 = 1/2$$

$$(3) \hat{\theta}_3 = X_1$$

例 7.2

繰り返し測定モデル $X_i = \mu + \varepsilon_i, (i = 1, \dots, n)$

 $\varepsilon_1, \dots, \varepsilon_n \stackrel{i.i.d.}{\sim} \varepsilon, \mathbf{E}(\varepsilon) = 0, \mathbf{Var}(\varepsilon) = \sigma^2$ $\theta = (\mu, \sigma^2)$

$$(1) \gamma(\theta) = \mu : \hat{\mu} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

$$(2) \gamma(\theta) = \sigma^2 : \hat{\sigma}^2 = S_n = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

点推定 (3/7)

良さの基準

- (1) 偏り: $\text{Bias}(\hat{\gamma}) = E(\hat{\gamma}) - \gamma.$
- (2) 分散: $\text{Var}(\hat{\gamma}) = E\{[\hat{\gamma} - E(\hat{\gamma})]^2\}.$
- (3) 平均 2 乗誤差: $\text{MSE}(\hat{\gamma}) = E\{(\hat{\gamma} - \gamma)^2\}.$
- (4) 集中確率: $P(|\hat{\gamma} - \gamma| \leq a), \quad a > 0.$
- (5) 一致性: $n \rightarrow \infty$ のとき, $\hat{\gamma} \xrightarrow{p} \gamma.$

平均 2 乗誤差, 分散, 偏りの 2 乗の間の関係 :

$$\text{MSE}(\hat{\gamma}) = E\{[\hat{\gamma} - E(\hat{\gamma}) + E(\hat{\gamma}) - \gamma]^2\} = \text{Var}(\hat{\gamma}) + \text{Bias}(\hat{\gamma})^2.$$

点推定 (4/7)

定義 7.1

$\hat{\gamma} = \hat{\gamma}(\mathbf{X})$: パラメータ $\gamma = \gamma(\theta)$ の推定量

- (1) $\forall \theta \in \Theta$ に対して, $E(\hat{\gamma}) = \gamma$ となるとき, $\hat{\gamma}$ は**不偏推定量**であるという.
- (2) γ の不偏推定量が存在するとき, γ は**推定可能**であるという.
- (3) γ の不偏推定量 $\hat{\gamma}$ が, 任意の不偏推定量 $\tilde{\gamma}$ に対して

$$\forall \theta \in \Theta \text{ に対して, } \text{Var}_{\theta}(\hat{\gamma}) \leq \text{Var}_{\theta}(\tilde{\gamma})$$

を満たすとき, $\hat{\gamma}$ は γ の**一様最小分散不偏推定量** (UMVUE: Uniformly Minimum Variance Unbiased Estimator) であるという.

点推定 (5/7)

例 7.3 (例 7.1 の続き)

$$(1) E(\hat{\theta}_1) = \theta, \text{Var}(\hat{\theta}_1) = \theta(1 - \theta)/10.$$

$$(2) E(\hat{\theta}_2) = 1/2, \text{Var}(\hat{\theta}_2) = 0.$$

$$(3) E(\hat{\theta}_3) = \theta, \text{Var}(\hat{\theta}_3) = \theta(1 - \theta).$$

点推定 (6/7)

例 7.4

 $X_1, \dots, X_n \stackrel{i.i.d.}{\sim} X, E(X) = \mu, \text{Var}(X) = \sigma^2$ (1) μ の推定

$$\hat{\mu}_{\mathbf{c}} = c_1 X_1 + \dots + c_n X_n$$

$$E(\hat{\mu}_{\mathbf{c}}) = c_1 \mu + \dots + c_n \mu = (c_1 + \dots + c_n) \mu$$

$\hat{\mu}_{\mathbf{c}}$ が不偏であるための必要十分条件は $c_1 + \dots + c_n = 1$

$$\text{Var}(\hat{\mu}_{\mathbf{c}}) = (c_1^2 + \dots + c_n^2) \sigma^2.$$

条件 $c_1 + \dots + c_n = 1$ のもとで $c_1^2 + \dots + c_n^2$ が最小となるのは

$$c_1 = \dots = c_n = \frac{1}{n}$$

メモ用紙

点推定 (7/7)

例 7.4 (続き)

(2) σ^2 の推定

標本分散 $S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ は σ^2 の不偏推定量

$$E \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right\} = \frac{n-1}{n} \sigma^2$$

$n^{-1} \sum_{i=1}^n (X_i - \bar{X})^2$ は σ^2 の不偏推定量ではない。