確率・統計 B
標本分布・推定

若木宏文

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

2014.12.10

table of contents

正規母集団からの統計量の分布 (1/10)

母集団分布が正規分布の場合
　標本平均と標本分散の関数の分布を導出する.
　平均や分散に関する「区間推定」や「検定」で必要となる.

定理 6.8
$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ とする. このとき, 次が成り立つ.
(1) $\bar{X}_n \sim N(\mu, \sigma^2/n)$.
(2) $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2$.
(3) $\bar{X}_n$ と S_n は互いに独立である.

(証明には, 定義といくつかの定理が必要)

正規母集団からの統計量の分布 (2/10)

定義 6.1

カイ 2 乗分布　確率変数 X の確率密度関数が

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(\frac{n}{2})2^{n/2}} e^{-x/2} x^{n/2-1}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

で与えられるとき, X は自由度 n のカイ **2** 乗分布に従うといい, $X \sim \chi_n^2$ とかく.

正規母集団からの統計量の分布 (3/10)

$\boldsymbol{X} = (X_1, \ldots, X_n)'$ の確率密度関数が

$$f(x_1, \ldots, x_n) = (2\pi)^{-n/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{x} - \boldsymbol{\mu})' \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \right\}$$

と表せるとき，n 次元正規分布に従うといい，
$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$ と書く．ここで，

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}.$$

パラメータ $\boldsymbol{\mu}$ は任意の定数，すなわち，$-\infty < \mu_i < \infty$
$(i = 1, \ldots, n)$ で，Σ は正定値行列である．

正規母集団からの統計量の分布 (4/10)

命題 1

$\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$, $Z_1, \ldots, Z_n \overset{i.i.d.}{\sim} N(0,1)$ とする.

(1) $\boldsymbol{Z} \sim N_n(\boldsymbol{0}, I_n)$
ただし, I_n は n 次単位行列.

(2) A : n 次正則行列, $\boldsymbol{b}$: n 次元ベクトル
$A\boldsymbol{Z} + \boldsymbol{b} \sim N_n(\boldsymbol{b}, AA')$

(3) $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$
$\Rightarrow \quad \mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu}, \mathrm{Var}(\boldsymbol{X}) = \Sigma,$
$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) := \mathrm{E}(e^{i\boldsymbol{t}'\boldsymbol{X}}) = \exp(i\boldsymbol{t}'\boldsymbol{\mu} - \boldsymbol{t}'\Sigma\boldsymbol{t}/2)$

(4) B : $m \times n$ 行列, $\mathrm{rank} B = m$, $\boldsymbol{b}$: m 次元ベクトル,
$\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$, $\quad \boldsymbol{Y} = B\boldsymbol{X} + \boldsymbol{b}$
$\Rightarrow \quad \boldsymbol{Y} \sim N_m(B\boldsymbol{\mu} + \boldsymbol{b}, B\Sigma B')$

((3), (4) はテキストの定理 6.5, 証明は板書で)

正規母集団からの統計量の分布 (5/10)

定理 6.7

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(0,1)$ とし, $V = X_1^2 + \cdots + X_n^2$ とする. このとき, $V \sim \chi_n^2$ である.
(証明は板書)

定理 6.8

$X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ とする. このとき, 次が成り立つ.

(1) $\bar{X}_n \sim N(\mu, \sigma^2/n)$.

(2) $(n-1)S_n^2/\sigma^2 \sim \chi_{n-1}^2$.

(3) $\bar{X}_n$ と S_n は互いに独立である.

正規母集団からの統計量の分布 (6/10)

(証明のポイント)
$\boldsymbol{X} = (X_1, \dots, X_n)' \Rightarrow \boldsymbol{X} \sim N_n(\mu \mathbf{1}_n, \sigma^2 I_n) \quad (\mathbf{1}_n = (1, \dots, 1)')$
H : 第 1 行が $(1/\sqrt{n}, \dots, 1/\sqrt{n})$ である n 次の直交行列

$$\boldsymbol{Y} = (Y_1, \dots, Y_n)' = H\boldsymbol{X}$$

と定めると

$$\boldsymbol{Y} \sim N_n(\mu H \mathbf{1}_n, \sigma^2 I_n)$$

と定めると

$$\mathrm{E}(\boldsymbol{Y}) = \mu \begin{pmatrix} 1/\sqrt{n} & \cdots & 1/\sqrt{n} \\ * & \cdots & * \\ \vdots & & \vdots \\ * & \cdots & * \end{pmatrix} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \sqrt{n}\mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$Y_1, \dots, Y_n$ は独立, $Y_1 \sim N(\sqrt{n}\mu, \sigma^2)$, $Y_2, \dots, Y_n \overset{i.i.d.}{\sim} N(0, \sigma^2)$

正規母集団からの統計量の分布 (7/10)

定義 6.10

(t-分布) 確率変数 T の確率密度関数が

$$f_T(t) = \frac{\Gamma\left(\frac{1}{2}(n+1)\right)}{\sqrt{n\pi}\Gamma\left(\frac{1}{2}n\right)}\left(1+\frac{1}{n}t^2\right)^{-(n+1)/2}$$

で与えられるとき, X は自由度 n の t-分布に従うといい, $T \sim t_n$ とかく.

定理 6.9

$X \sim N(0,1)$, $Y \sim \chi_n^2$, X と Y は独立とする. このとき

$$T = \frac{X}{\sqrt{Y/n}} \sim t_n$$

(証明は省略, テキスト 125 ページ参照)

正規母集団からの統計量の分布 (8/10)

注 6.5

正規母集団 $N(\mu, \sigma^2)$ からの大きさ n の標本 $X_1, \ldots, X_n$ に基づく標本平均, 標本分散を $\bar{X}_n$, S_n^2 とする. このとき

$$T = t(X_1, \ldots, X_n) = \frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \sim t_{n-1}^2$$

正規母集団からの統計量の分布 (9/10)

定義 6.11

F-分布確率変数 V の確率密度関数が

$$f_V(v) = \begin{cases} \dfrac{\Gamma\left(\frac{1}{2}(m+n)\right)}{\Gamma\left(\frac{1}{2}m\right)\Gamma\left(\frac{1}{2}n\right)} \left(\dfrac{m}{n}\right)^{m/2} v^{m/2-1} \left(1+\dfrac{m}{n}v\right)^{-(m+n)/2}, & \\ & v > 0 \\ 0, & v \leq 0 \end{cases}$$

で与えられるとき, V は自由度 m, n の F-分布に従うといい, $V \sim F_{m,n}$ とかく.

正規母集団からの統計量の分布 (10/10)

定理 6.10

$X \sim \chi^2_m$, $Y \sim \chi^2_n$, X と Y は独立とする．このとき

$$V = \frac{X/m}{Y/n}$$

は自由度 m, n の F-分布に従う．

(証明は省略, テキスト 127 ページ参照)

順序統計量 (1/5)

定義 6.12

$X_1, \ldots, X_n$ 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 $X_{(1)} \le \cdots \le X_{(n)}$ とするとき

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

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
$X_1 = 3, X_2 = 4, X_3 = 1, X_4 = 5$
$\Rightarrow \quad X_{(1)} = 1, X_{(2)} = 3, X_{(3)} = 4, X_4 = 5$

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) 100α%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 2 (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) $X_{(n)} = \max\{X_1, \ldots, X_n\}$ の分布関数と確率密度関数

$$G_n(y) = \mathsf{P}(X_{(n)} \leq y) = \{F(y)\}^n,$$
$$g_n(y) = \frac{d}{dy} G_n(y) = nf(y)\{F(y)\}^{n-1}$$

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

$$G_1(y) = \mathsf{P}(X_{(1)} \leq y) = 1 - \{1 - F(y)\}^n,$$
$$g_1(y) = \frac{d}{dy} G_n(y) = nf(y)\{1 - F(y)\}^{n-1}$$

(証明は板書で)

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

(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

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

(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

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