確率・統計 B
標本分布

若木宏文

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

2014.11.26

table of contents

母集団分布と統計モデル (2/15)

重要な語句

- 母集団 (population)：実験や調査の対象
  - 有限母集団：対象の個数が有限 (例. 日本人全員)
  - 無限母集団：対象の個数が無限 (例. 生産システムでの不良品率の調査)
- 標本抽出 (sampling)：対象の一部を取り出すこと
- 標本 (sample)：取り出された対象
- 標本数 (sample size)：取り出された対象の数 (n で表すことがほとんど)
- 実現値：観測された値 $x_1, \ldots, x_n$
- 標本値：抽出された対象の特性値 (ここでは $\hat{\mu} = \bar{x}$)

母集団分布と統計モデル (4/15)

定義 6.1 (無作為抽出)

母集団に含まれるどの対象も同じ確率で選ばれるような方法で標本を取り出すこと

1. 復元抽出：母集団に含まれる対象をひとつづつ選んでゆくとき, 前に選ばれた対象も何度も選べる抽出法（標本は独立）
2. 非復元抽出：どの対象も一度しか選べない抽出法（標本は独立でないが, 標本数に比べ母集団が大きいときは独立と見なしてもよい）

母集団分布と統計モデル (5/15, 8/15)

定義 6.2 (母集団分布)

母集団分布：母集団から対象を 1 つ無作為抽出して, 特性 X の値を測定するときの, 確率変数 X で分布

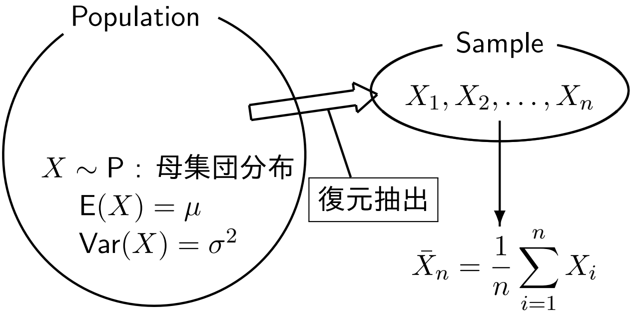
実現値：実際に測定することによって得られた X の値 x

定義 6.3 (標本と母集団分布の数学的な定義)

確率変数 $X_1, \ldots, X_n$ が互いに独立で, 同一の確率分布 P に従うとき, $X_1, \ldots, X_n$ は, 分布 P からの大きさ $\boldsymbol{n}$ のランダム標本, あるいは無作為標本といい, P を母集団分布という.

$$X_1, \ldots, X_n \sim \text{ i.i.d. P}$$

母集団分布と統計モデル (9/15)



$$X_1, X_2, \ldots, X_n \sim \text{ i.i.d. P}, \quad \mathrm{E}(X_i) = \mu,\ \mathrm{Var}(X_i) = \sigma^2$$

$$\bar{X}_n \xrightarrow{p} \mu \ (n \to \infty),$$

$$\mathrm{P}(|\bar{X}_n - \mu| \le \delta) \approx \mathrm{P}(|Z| \le \frac{\sqrt{n}\delta}{\sqrt{\sigma^2}}), \quad Z \sim N(0,1)$$

$$\left(\frac{\sqrt{n}\delta}{\sqrt{\sigma^2}} = 1.96 \Rightarrow \mathrm{P}(|\bar{X}_n - \mu| \le \delta) \approx 0.95 \right)$$

母集団分布と統計モデル (10/15)

定義 6.4

(統計モデル) 母集団分布が, ある確率分布の集まり

$$\mathcal{P} = \{\mathsf{P}_\theta;\, \theta \in \Theta\}, \quad \Theta \subset \mathbb{R}^p \ (p \text{ はある自然数})$$

に属すると想定されるとき, $\mathcal{P}$ を母集団分布の統計モデルという.
このとき, θ を母数, Θ を母数空間とよぶ.

例. (繰り返し測定モデル)

$$X_1, \ldots, X_n \sim \text{ i.i.d. } N(\mu, \sigma^2).$$

母集団分布と統計モデル (13/15)

定義 6.5 (経験分布関数)

$X_1, \ldots, X_n$：ある分布 P からの大きさ n のランダム標本
$x_1, \ldots, x_n$：その実現値
$x_1, \ldots, x_n$ から定義される関数

$$F_n(x) = \frac{x_i \leq x \text{ である } x_i \text{ の個数}}{n}$$

を経験分布関数という.
n 個のランダム標本 $X_1, \ldots, X_n$ の経験分布関数は形式上, 全く同じ記号を用いて

$$F_n(x) = \frac{X_i \leq x \text{ である } X_i \text{ の個数}}{n}$$

と定義される.

母集団分布と統計モデル (14/15)

確率変数 X^* ： 離散型確率変数, $\mathsf{P}(X^* = x_i) = 1/n,\ i = 1, \ldots, n$
$\Rightarrow$ X^* の分布関数は $F_n(x)$

経験分布関数によって定まる離散型分布, すなわち X^* の分布を経験分布とよぶ.

母集団分布と統計モデル (15/15)

定理 6.1 (経験分布関数の性質)

$X_1, \ldots, X_n$：ある分布 P からの大きさ n のランダム標本

$F(x)$：分布 P の分布関数

F_n：ランダム標本から定義される経験分布関数

$\Rightarrow$

(1) $nF_n(x) \sim B(n, F(x))$.

(2) $F_n(x) \xrightarrow{p} F(x)$.

(3) $\sqrt{n}\{F_n(x) - F(x)\} \xrightarrow{d} N(0, F(x)(1 - F(x)))$.

統計量の分布 (1/8)

定義 6.6

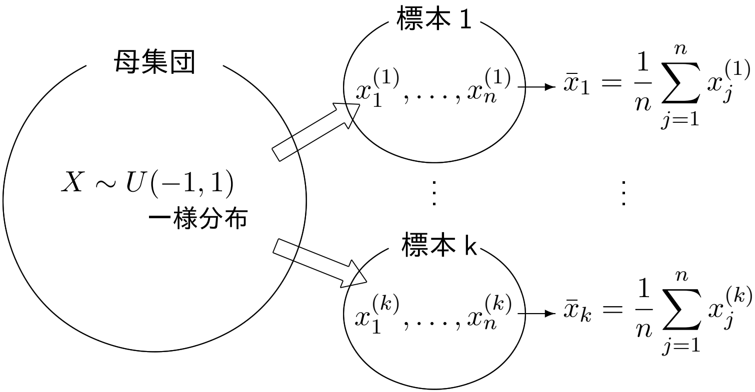
統計量

$t(x_1, \ldots, x_n)$: n 変数のボレル可測関数,
未知パラメータは含まれないものとする.
$X_1, \ldots, X_n$: ランダム標本
$\Rightarrow T = t(X_1, \ldots, X_n)$ を統計量とよぶ.

統計量の分布や, その特性値が知りたい.

統計量の分布 (2/8)

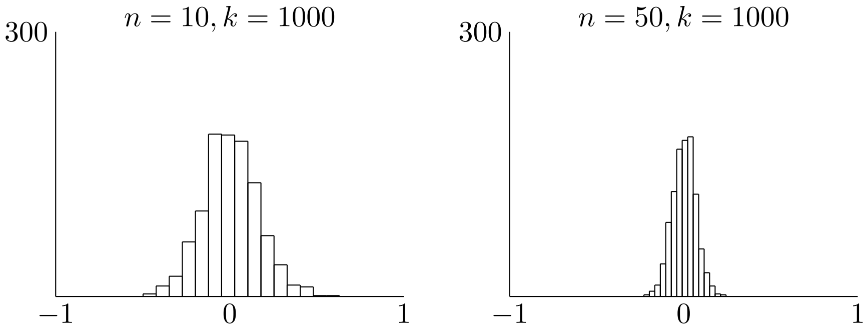
例. 統計量の分布 (数値実験)



統計量の分布 (3/8)

例. 統計量の分布 (数値実験)(つづき)

標本平均の分布

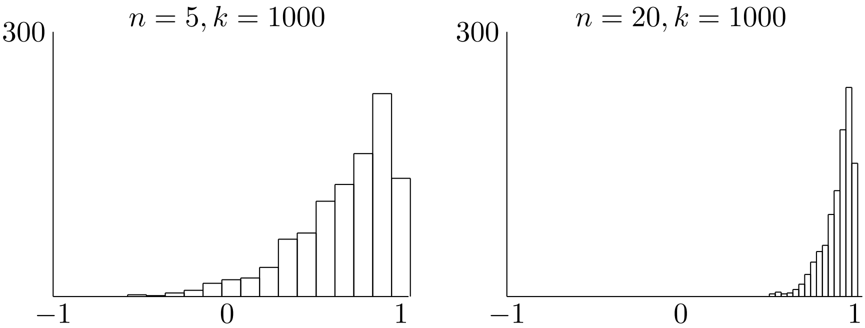


$\bar{x} = 0$ を中心にほぼ左右対称に分布している
標本数 (n) が大きくなると, 分布の「幅」が狭くなる

統計量の分布 (4/8)

例. 統計量の分布 (数値実験)(つづき)

最大値の分布 ($t = \max\{x_1, \ldots, x_n\}$)



右側に偏った分布
標本数 (n) が大きくなると, 分布の「幅」が狭くなる

統計量の分布 (5/8)

定義 6.7

標本平均, 標本分散

$X_1, \ldots, X_n$: ランダム標本

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

をそれぞれ, 標本平均, 標本分散とよぶ.

統計量の分布 (6/8)

P_n：ランダム標本 $X_1, \ldots, X_n$ の経験分布
X^*：P_n に従う確率変数

$\Rightarrow$

$$\mathsf{E}(X^*) = \sum_{i=1}^{n} X_i \mathsf{P}(X^* = X_i) = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$

$$\mathsf{Var}(X^*) = \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \mathsf{P}(X^* = X_i) = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

標本平均は, 経験分布の平均

標本分散は, 経験分布の分散の $\dfrac{n}{n-1}$ 倍

統計量の分布 (7/8)

定理 6.2

$X_1, \ldots, X_n$ をランダム標本とし

$$\mathrm{E}(X_i) = \mu, \quad \mathrm{Var}(X_i) = \sigma^2, \quad i = 1, \ldots, n$$

とする．このとき次が成り立つ．

(1) $\mathrm{E}(\bar{X}_n) = \mu$.

(2) $\mathrm{Var}(\bar{X}_n) = \frac{1}{n}\sigma^2$.

(3) $\mathrm{E}(S_n^2) = \sigma^2$.

(証明 は (3) のみ板書で)

統計量の分布 (8/8)

定理 6.3

定理 6.2 と同じ仮定のもとで次が成り立つ.

(1) $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$.

(2) $S_n^2 \xrightarrow{p} \sigma^2$.

(3) $\dfrac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{d} N(0,1)$.

(証明は板書で)

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

(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) $\boldsymbol{Z} \sim N_n(\mathbf{0}, I_n)$
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) $\boldsymbol{X} \sim N_n(\boldsymbol{\mu}, \Sigma)$
$\Rightarrow$ $\mathsf{E}(\boldsymbol{X}) = \boldsymbol{\mu}, \mathsf{Var}(\boldsymbol{X}) = \Sigma$,
$\varphi_{\boldsymbol{X}}(\boldsymbol{t}) := \mathsf{E}(e^{i\boldsymbol{t}'\boldsymbol{X}}) = \exp(i\boldsymbol{t}'\boldsymbol{\mu} - \boldsymbol{t}'\Sigma\boldsymbol{t}/2)$

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

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

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

と定めると

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

と定めると

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

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

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

$X \sim N(0,1)$, $Y \sim \chi^2_n$, 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

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

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)