確率・統計 B
標本分布

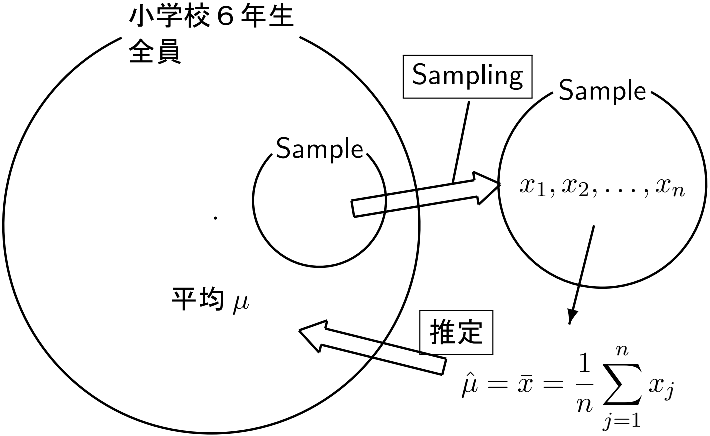
若木宏文

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

2013.10.30

母集団分布と統計モデル (1/15)

例 1 (日本の小学 6 年生の平均身長が知りたい)



小学 6 年生全員の身長を調べるのは大変！

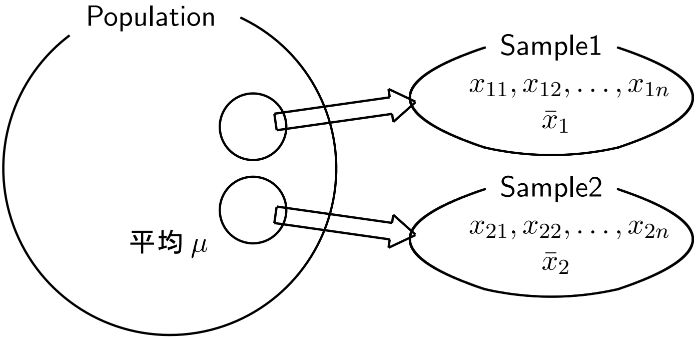
母集団分布と統計モデル (2/15)

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 (3/15)

疑問点



- $\hat{\mu} = \bar{x}$ は母集団全体での平均 μ に近い値をとるのか？
- 標本を取り直すと $\bar{x}$ の値も変わるが, どのくらい値が変わるのか？

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)

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

例 2 (不良品の調査)

母集団: ある製品の在庫. 不良品が $100 \times p$ パーセント含まれているとする

母集団から無作為に製品を 1 個取り出して,

不良品なら $X = 1$, 良品なら $X = 0$ とする

X の母集団分布は, 繰り返し数 1, 成功確率 p の二項分布

母集団分布と統計モデル (6/15)

例 2 (不良品の調査（つづき）)

復元抽出で無作為に n 個標本を取り出し, 不良品かどうかを表す変数を

$$X_1, \ldots, X_n$$

とすると, $X_1, \ldots, X_n$ は独立に, 母集団分布と同じ分布 ($B(1, p)$) に従がう.

非復元抽出で無作為に n 個標本を取り出す場合
母集団に含まれる製品の個数を N 個とし, その内, m 個が不良品であるとすると母集団分布は $B(1, \frac{m}{N})$

$X_1 = \cdots = X_{n-1} = 1$ のときの X_n の条件付分布：$B(1, \frac{m-n+1}{N-n+1})$
$X_1 = \cdots = X_{n-1} = 0$ のときの X_n の条件付分布：$B(1, \frac{m}{N-n+1})$

母集団分布と統計モデル (7/15)

母集団分布の分布関数 (有限母集団の場合)

母集団に属する対象の個数を N 個とする
$x_1, \ldots, x_N$: 各対象の特性 X の値

X の母集団分布の分布関数は

$$F(x) = \mathrm{P}(X \leq x) = \frac{x_i \leq x \text{ である } x_i \text{ の個数}}{N}$$

$\Rightarrow$ テキスト 113 ページ, 例 6.2

母集団分布と統計モデル (8/15)

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 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.

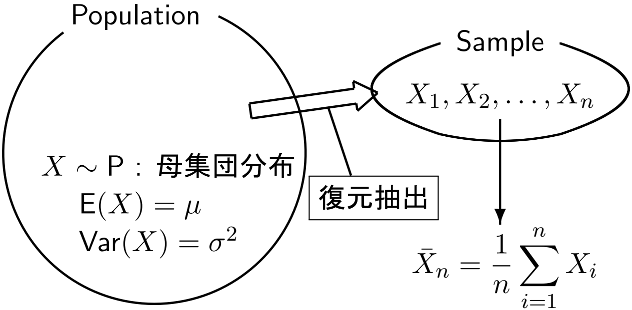
注 6.1 (i.i.d.)

$X_1, \ldots, X_n$ が独立 (independently) に, 同一の (identically) 分布 P に従う (distributed) とき,

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

と書く

母集団分布と統計モデル (9/15)



$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

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

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

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

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

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.

母集団分布と統計モデル (11/15)

例 6.3 (くり返し測定モデル)

$x_1, \ldots, x_n$: 未知の量 μ をくり返して測定して得られる観測値

母集団 : 測定対象を無限回測定して得られるすべての値

$\Rightarrow$ $x_1, \ldots, x_n$ は, 無限母集団からの大きさ n の標本 $X_1, \ldots, X_n$ の実現値

くり返し測定に対してよく想定される統計モデル :

$$X_i = \mu + \varepsilon_i, \quad i = 1, \ldots, n$$

母集団分布と統計モデル (12/15)

例 6.3 (くり返し測定モデル (つづき))

$$X_i = \mu + \varepsilon_i, \quad i = 1, \ldots, n$$

繰り返しモデルに関する仮定

(1) $\varepsilon_1, \ldots, \varepsilon_n$ は互い独立である.

(2) $\varepsilon_1, \ldots, \varepsilon_n$ は, それぞれ, 同一分布に従う.

(3) $\mathrm{E}\{\varepsilon_i\} = 0$, $i = 1, \ldots, n$.

(4) $\mathrm{Var}[\varepsilon_i] = \sigma^2$, $i = 1, \ldots, n$.

ここに, $-\infty < \mu < \infty$, $\sigma^2 > 0$.

(5) $\varepsilon_1, \ldots, \varepsilon_n$ は正規分布に従う.

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

母集団分布と統計モデル (13/15)

定義 6.5 (経験分布関数)

$X_1, \dots, X_n$: ある分布 P からの大きさ n のランダム標本
$x_1, \dots, x_n$: その実現値
$x_1, \dots, x_n$ 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 $X_1, \dots, X_n$ 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) $\sqrt{n}\{F_n(x) - F(x)\} \xrightarrow{d} N(0, F(x)(1 - F(x)))$.

(証明は板書で)