

No.11 点推定

点推定

- 母集団分布を特徴づける定数を**母数**という。例: 平均 (母平均), 分散 (母分散)

例

- 母平均 μ (未知) の母集団から大きさ 10 の無作為標本を抽出したところ次のようになった:

62.3 71.8 48.4 33.4 $\cdots$ 45.8 45.4

この値から母平均 μ の値を推定する方法を考える。

- 大きさ n の無作為標本は n 個の互いに独立で同じ分布 (母集団分布) に従う確率変数 $X_1, X_2, \dots, X_n$ とする。
- 抽出した後, 観測された値を**実現値**といい, $x_1, x_2, \dots, x_n$ で現す。
- 上の 62.3, 71.8, $\dots$, 45.4 は大きさ 10 の標本から定まる確率変数 $X_1, X_2, \dots, X_{10}$ の実現値である。
- $\bar{x} = \frac{x_1 + \dots + x_{10}}{10} = 52.5$ は標本平均 $\bar{X}$ の実現値

母平均の点推定 (推測)

母平均 μ , 母分散 σ^2 の母集団から大きさ n の無作為標本を抽出したとき

- $\bar{X}$ の平均と分散の公式より $E[\bar{X}] = \mu, V[\bar{X}] = \frac{\sigma^2}{n}$
- 標本抽出するごとに $\bar{X}$ はいろんな値を取りうるそれら全ての平均は μ と一致し, $V[\bar{X}] = \sigma^2/n \rightarrow 0$ ($n \rightarrow \infty$) より $\bar{X}$ の値は n が大きいとき μ のまわりに多く分布する。よって, 母平均 μ を推定するには標本平均 $\bar{X}$ の実現値 $\bar{x}$ を用いるのが自然
- 母数を統計量の 1 つの実現値で推定することを**点推定**という。
- μ を $\bar{X}$ で推定するとき, 確率変数である統計量 $\bar{X}$ は母平均 μ の**推定量**, その実現値を母平均 μ の**推定値**という。
- 点推定には**よい性質**をもった推定量が必要である。よい性質とは**一致性**, **不偏性**と呼ばれる性質である。

一致性

- 推定量 $T = T(X_1, X_2, \dots, X_n)$ を推定量 (確率変数の関数) が母数 θ の**一致推定量**であるとは, $k > 0$ に対して,

$$\lim_{n \rightarrow \infty} P(|T - \theta| < k) = 1$$

を満たすことである (データが増えると, 推定量 T はほとんど θ のまわりにある)。

定理 (チェビシェフの不等式)

確率変数 X が平均 $\mu = E[X]$, 分散 $\sigma^2 = V[X]$ であるとき, 定数 k に対して, 次が成り立つ:

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

証明 X が連続型の場合に証明する.

$$\begin{aligned}\sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx = \int_{|x - \mu| \geq k} (x - \mu)^2 f(x) dx + \int_{|x - \mu| < k} (x - \mu)^2 f(x) dx \\ &\geq \int_{|x - \mu| \geq k} (x - \mu)^2 f(x) dx \\ &\geq k^2 \int_{|x - \mu| \geq k} f(x) dx = k^2 P(|X - \mu| \geq k)\end{aligned}$$

ゆえに, $P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$ を得る. $\square$

標本平均 $\bar{X}$ は母平均 μ の一致推定量 (大数の法則)

$$0 \leq P(|\bar{X} - \mu| \geq k) \leq \lim_{n \rightarrow \infty} \frac{V[\bar{X}]}{k^2} \leq \frac{\sigma^2}{nk^2} \rightarrow 0 \quad (n \rightarrow \infty)$$

であるから, 余事象を考えると,

$$\lim_{n \rightarrow \infty} P(|\bar{X} - \mu| < k) = \lim_{n \rightarrow \infty} (1 - P(|\bar{X} - \mu| \geq k)) = 1$$

を得る.

不偏性 推定量 $T = T(X_1, X_2, \dots, X_n)$ が母数 θ の**不偏推定量**であるとは,

$$E[T] = \theta$$

を満たすことである.

例 $E[\bar{X}] = \mu$ なので標本平均は母平均の不偏推定量である.

注 $n = 5$ のとき, $X^* = \frac{1}{10}X_1 + \frac{1}{5}X_2 + \frac{2}{5}X_3 + \frac{1}{5}X_4 + \frac{1}{10}X_5$ とおくと

$$E[X^*] = \mu, V[X^*] = \frac{13}{50}\sigma^2$$

を満たす. $V[\bar{X}] = \frac{1}{5}\sigma^2 < E[X^*]$ より, $\bar{X}$ のほうが, X^* よりよい不偏推定量.

分散の不偏推定量

- 母平均 μ が既知 $\rightarrow Y_i = (X_i - \mu)^2$ とおくと, $E(\frac{1}{n} \sum_i Y_i) = \frac{1}{n} \sum_i E(Y_i) = \sigma^2$ が成り立つ.

- 母平均 μ が未知 $\rightarrow U^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2 : \sigma^2$ の不偏推定量である.

証明 $E[U^2] = \frac{1}{n-1} \sum_{i=1}^n E[(X_i - \bar{X})^2] = \sigma^2$ を示せばよい.

$$\begin{aligned} E[(X_i - \bar{X})^2] &= E[\{(X_i - \mu) - (\bar{X} - \mu)\}^2] \\ &= E[(X_i - \mu)^2] - 2E[(X_i - \mu)(\bar{X} - \mu)] + E[(\bar{X} - \mu)^2] \end{aligned}$$

ここで, $E[(X_i - \mu)^2] = \sigma^2$ であり, $E[(\bar{X} - \mu)^2] = V[\bar{X}] = \frac{\sigma^2}{n}$ である. さらに,

$$\begin{aligned} E[(X_i - \mu)(\bar{X} - \mu)] &= E[(X_i - \mu)\left(\frac{1}{n} \sum_j X_j - \frac{n}{n} \mu\right)] = E[(X_i - \mu)\left(\frac{1}{n} \sum_j (X_j - \mu)\right)] \\ &= \frac{1}{n} \sum_{j=1}^n E[(X_i - \mu)(X_j - \mu)] \\ &= \frac{1}{n} \left\{ E[(X_i - \mu)^2] + \sum_{j \neq i} E[(X_i - \mu)(X_j - \mu)] \right\} \\ &= \frac{1}{n} \left\{ E[(X_i - \mu)^2] + \sum_{j \neq i} E[(X_i - \mu)] E[(X_j - \mu)] \right\} \\ &= \frac{\sigma^2}{n} + 0 = \frac{\sigma^2}{n} \end{aligned}$$

よって, $E[(X_i - \bar{X})^2] = \sigma^2 - \frac{2\sigma^2}{n} + \frac{\sigma^2}{n} = \frac{n-1}{n} \sigma^2$ より,

$$E[V] = \frac{1}{n-1} \cdot n \cdot \frac{n-1}{n} \sigma^2 = \sigma^2.$$