2015年後期
応用確率統計

⑨ 平均と分散の不偏推定

河野 行雄

kawano@pe.titech.ac.jp

2015年12月17日

復習

■ 推定量と推定値

推定量: $\hat{\theta}(X_1, \cdots, X_n)$　　推定値: $\hat{\theta}(x_1, \cdots, x_n)$

■ 不偏性

$$E(\hat{\theta} \mid \theta) = E^{X_1, \cdots, X_n}\left(\hat{\theta}(X_1, \cdots, X_n) \mid \theta\right) = \theta$$

■ 標本平均と幾何平均

$$\hat{\theta} = \frac{X_1 + \cdots + X_n}{n}$$ ⇨ 不偏

$$\hat{\theta} = \left(\prod_{i=1}^{n} X_i\right)^{1/n}$$ ⇨ 不偏ではない

平均値の不偏推定

■ 観測データ

例： $X_i = \theta + \varepsilon_i \qquad (i = 1, \cdots, n)$

確率変数　　未知母数　　誤差

誤差の分布： $f(\varepsilon_i)$　　平均： $E(\varepsilon_i) = 0$　　分散： $V(\varepsilon_i) = \sigma^2$

■ 最適な推定量

不偏： $E(\hat{\theta} \mid \theta) = E^{X_1, \cdots, X_n}\left(\hat{\theta}(X_1, \cdots, X_n) \mid \theta\right) = \theta$

分散最小： $\hat{\theta}^* = \arg\min_{\hat{\theta}} V(\hat{\theta} \mid \theta)$

誤差の分布：
- 未知： $\hat{\theta}^* =$ 標本平均
- 既知： $\hat{\theta}^* =$ ピットマン推定量

標本平均

■ 標本平均

誤差の分布未知のときの最小分散推定量

$$\hat{\theta}^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

■ 不偏性

$$E\left(\hat{\theta}^*\right) = \frac{1}{n}\sum_{i=1}^{n} E\left(X_i\right) = \theta$$

■ 最小分散

分散がクラーメル・ラオの下限に一致： $V\left(\hat{\theta}^*\right) = {}^{1}\!/_{nI}$

クラメール・ラオの不等式

$$V\left(\hat{\theta}\right) \geq {}^{1}\!/_{nI} \qquad I = E\left(\left(\frac{\partial}{\partial\theta}\log f\left(X,\theta\right)\right)^2\right)$$

標本平均の分散

■ 標本平均

$$\hat{\theta}^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

■ 標本平均の分散

$$V\left(\hat{\theta}^*\right) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = E\left(\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n}\theta\right)^2\right)$$

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

$$= \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right)^2 = \frac{\sigma^2}{n}$$

クラメール・ラオの下限

■ クラメール・ラオの不等式

$$V\left(\hat{\theta}\right) \geq {}^{1}\!/_{nI} \qquad I = E\left(\left(\frac{\partial}{\partial\theta}\log f\left(X,\theta\right)\right)^2\right)$$

■ 誤差が正規分布のとき

$$f\left(x,\theta\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{\left(x-\theta\right)^2}{2\sigma^2}\right)$$

$$\log f\left(x,\theta\right) = -\frac{1}{2}\log 2\pi\sigma^2 - \frac{\left(x-\theta\right)^2}{2\sigma^2} \quad \Rightarrow \quad \frac{1}{nI} = \frac{\sigma^2}{n} = V\left(\hat{\theta}^*\right)$$

$$E\left(\left(\frac{\partial}{\partial\theta}\log f\left(X,\theta\right)\right)^2\right) = E\left(\frac{\left(X-\theta\right)^2}{\sigma^4}\right) = \frac{1}{\sigma^2}$$

一致性

■ 一致推定量

観測数を増やすと推定量が未知母数に一致（確率収束）

$$\lim_{n\to\infty} P\left(\left|\hat{\theta}^* - \theta\right| \geq \varepsilon\right) = 0$$

■ 標本平均

$$\hat{\theta}^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E\left(\hat{\theta}^*\right) = \theta \qquad \lim_{n\to\infty} V\left(\hat{\theta}^*\right) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$

$\longrightarrow$ チェビシェフの不等式（大数の法則）から一致推定量

標本分散

■ 標本分散

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

■ 不偏性

$$E\left(\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right) = E\left(\sum_{i=1}^{n}\left(\left(X_i - \theta\right) - \left(\overline{X} - \theta\right)\right)^2\right)$$

$$= E\left(\sum_{i=1}^{n}\left(X_i - \theta\right)^2 - 2\boxed{\sum_{i=1}^{n}\left(X_i - \theta\right)}\left(\overline{X} - \theta\right) + \sum_{i=1}^{n}\left(\overline{X} - \theta\right)^2\right)$$

$$\boxed{n\left(\overline{X} - \theta\right)}$$

$$= E\left(\sum_{i=1}^{n}\left(X_i - \theta\right)^2 - n\left(\overline{X} - \theta\right)^2\right) = nV(X) - nV\left(\overline{X}\right)$$

$$= n\sigma^2 - \sigma^2 = (n-1)\sigma^2$$

→ 不偏ではない

不偏分散

■ 標本分散

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

■ 不偏分散

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

$$E\left(\hat{\sigma}^2\right)=\frac{n-1}{n-1}\sigma^2=\sigma^2$$

← 観測値一つ分に相当する情報を平均に使ったため

位置共変推定量

■ 位置共変推定量

$$\hat{\theta}(X_1 + c, \cdots, X_n + c) = \hat{\theta}(X_1, \cdots, X_n) + c$$

⟶ 例えば標本平均

■ 変数変換

例:　$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad Y_1 = X_2 - X_1, \cdots, Y_{n-1} = X_n - X_1$

■ 標本平均の補正

$$\hat{\theta}^* = \bar{X} - E_0\left(\bar{X} \mid Y_1, Y_2, \cdots, Y_{n-1}\right)$$

⟶ $V\left(\hat{\theta}^*\right) \le V\left(\hat{\theta}\right)$ 誤差の分布が既知であれば分散を削減可能

ピットマン推定量

■ 標本平均

誤差の分布既知のときの最小分散位置共変推定量

$$\hat{\theta}^* = \overline{X} - E_0\left(\overline{X} \mid Y_1, Y_2, \cdots, Y_n\right)$$

$$= \int \theta g\left(\theta; X_1, X_2, \cdots X_n\right) d\theta$$

■ 尤度関数

観測値の同時確率密度関数

$$g\left(\theta; X_1, X_2, \cdots X_n\right) = \frac{\prod_{i=1}^{n} f\left(X_i - \theta\right)}{\int \prod_{i=1}^{n} f\left(X_i - \theta\right) d\theta}$$

ピットマン推定量の例

■ 観測データ

$$X = \theta + \varepsilon \qquad f(\varepsilon) = \begin{cases} 1, & |\varepsilon| \le 0.5 \\ 0, & \text{otherwise} \end{cases}$$

■ 尤度関数

$$\prod_{i=1}^{n} f(X_i - \theta) = \begin{cases} 1, & \alpha \le \theta \le \beta \\ 0, & \text{otherwise} \end{cases} \qquad \begin{aligned} \alpha &= \max_i X_i - 0.5 \\ \beta &= \min_i X_i + 0.5 \end{aligned}$$

■ ピットマン推定量

$$\hat{\theta}^* = \frac{\int_\alpha^\beta \theta d\theta}{\int_\alpha^\beta d\theta} = \frac{1}{2}(\beta + \alpha) = \frac{1}{2}(\max X_i + \min X_i)$$

ピットマン推定量の例

■ 標本平均

$$\hat{\theta} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$V\left(\hat{\theta}\right) = \frac{\sigma^2}{n} = \frac{1}{12n} \to \frac{1}{12}, \frac{1}{24}, \frac{1}{36}, \frac{1}{48} \quad \longleftarrow \quad \sigma^2 = \int_{-0.5}^{0.5} x^2 dx = \frac{1}{12}$$

■ ピットマン推定量

$$\hat{\theta}^* = \frac{1}{2}\left(\max X_i + \min X_i\right) = \frac{1}{2}(a+b)$$

$$a = \min_i X_i$$

$$b = \max_i X_i$$

$$V\left(\hat{\theta}^*\right) = \int_{-0.5}^{0.5}\int_{-0.5}^{b}\left(\frac{a+b}{2}\right)^2 \boxed{n(n-1)(b-a)^{n-2}} dadb$$

$$= \frac{1}{2(n+2)(n+1)} \to \frac{1}{12}, \frac{1}{24}, \frac{1}{40}, \frac{1}{60}$$

(n-2)個の観測値が区間[a,b]にある確率

まとめ

- 誤差分布未知のときの最小分散推定量

$$\hat{\theta}^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- 誤差分布既知のときの最小分散位置共変推定量

$$\hat{\theta}^* = \frac{\int \theta \prod_{i=1}^{n} f\left(X_i - \theta\right) d\theta}{\int \prod_{i=1}^{n} f\left(X_i - \theta\right) d\theta}$$

- 不偏分散

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