單元 8: 一些常見的不偏估計量
(課本 §8.3)

正式求得目標參數的點估計量的方法詳述於第九章, 此處僅以直觀法呈現某些估計量如下.

1. 設 $Y_1, \ldots, Y_n$ 爲一大小爲 n 的隨機樣本且共同的期望值

$$E(Y_i) = \mu$$

以及變異數

$$\mathrm{Var}(Y_i) = \sigma^2$$

其中 $i = 1, \ldots, n$, 則一個母體期望值 μ 的估計量可選爲

$$\widehat{\mu} = \overline{Y} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} Y_i$$

且

$$E(\widehat{\mu}) = \mu$$

故 $\widehat{\mu} = \overline{Y}$ 是不偏的 (unbiased). 又估計量的變異數

$$\sigma_{\widehat{\mu}}^2 \stackrel{\text{def}}{=} \mathrm{Var}(\widehat{\mu}) = \frac{\sigma^2}{n}$$

得 n 愈大, 估計量變異數 $\sigma_{\widehat{\mu}}^2$ 愈小.

2. 設 $X_1, \ldots, X_n$ 爲 n 次 Bernoulli 試驗且成功機率爲 p, 則

$$Y \stackrel{\text{def}}{=} \sum_{i=1}^{n} X_i \sim \text{binomial}(n, p)$$

且一個母體比率 (proportion) p 的估計量可選爲

$$\widehat{p} \stackrel{\text{def}}{=} \frac{Y}{n} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

由此得

$$E(\widehat{p}) = p$$

故 $\widehat{p} = \frac{Y}{n}$ 是不偏的 (unbiased). 又估計量 $\widehat{p}$ 的變異數

$$\sigma_{\widehat{p}}^2 = \frac{pq}{n}$$

亦是 n 愈大, $\sigma_{\widehat{p}}^2$ 愈小.

3. 設 $Y_{11}, \ldots, Y_{1n_1}$ 爲一大小是 n_1 的隨機樣本且共同期望值

$$E(Y_{1i}) = \mu_1$$

以及變異數

$$\text{Var}(Y_{1i}) = \sigma_1^2$$

其中 $i = 1, \ldots, n_1$. 又 $Y_{21}, \ldots, Y_{2n_2}$ 爲大小是 n_2 的隨機樣本且共同期望值

$$E(Y_{2i}) = \mu_2$$

以及變異數

$$\mathrm{Var}(Y_{2i}) = \sigma_2^2$$

其中 $i = 1, \ldots, n_2$. 此外, 此二樣本亦相互獨立, 則一個母體期望值的差

$$\theta \stackrel{\text{def}}{=} \mu_1 - \mu_2$$

的估計量可選爲

$$\widehat{\theta} \stackrel{\text{def}}{=} \overline{Y}_1 - \overline{Y}_2$$

其中 $\overline{Y}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_{1i}$ 且 $\overline{Y}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_{2i}$. 由此得

$$E(\widehat{\theta}) = \mu_1 - \mu_2 = \theta$$

故 $\widehat{\theta} = \overline{Y}_1 - \overline{Y}_2$ 是不偏的. 又估計量 $\widehat{\theta}$ 的變異數

$$\sigma_{\widehat{\theta}}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

亦得 n_1 與 n_2 愈大, $\sigma_{\widehat{\theta}}^2$ 愈小.

4. 設 $X_{11}, \ldots, X_{1n_1}$ 爲 n_1 次的 Bernoulli 試驗, 其成功機率爲 p_1. 又 $X_{21}, \ldots, X_{2n_2}$ 爲另一 n_2 次的

Bernoulli 試驗, 其成功機率爲 p_2. 此外, 它們之間相互獨立, 則

$$Y_1 \stackrel{\text{def}}{=} \sum_{i=1}^{n_1} X_{1i} \sim \text{binomial}(n_1, p_1)$$

且

$$Y_2 \stackrel{\text{def}}{=} \sum_{i=1}^{n_2} X_{2i} \sim \text{binomial}(n_2, p_2)$$

又一母體比率的差

$$\theta \stackrel{\text{def}}{=} p_1 - p_2$$

的估計量可選爲

$$\widehat{\theta} \stackrel{\text{def}}{=} \frac{Y_1}{n_1} - \frac{Y_2}{n_2} = \widehat{p}_1 - \widehat{p}_2$$

則

$$E(\widehat{\theta}) = p_1 - p_2 = \theta$$

故 $\widehat{\theta} = \widehat{p}_1 - \widehat{p}_2$ 是不偏的, 且估計量的變異數

$$\sigma^2_{\widehat{\theta}} = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

亦是 n_1 與 n_2 愈大, $\sigma^2_{\widehat{\theta}}$ 愈小.

註. 上述估計量 $\widehat{\theta}$ 的 $E(\widehat{\theta})$ 與 $\text{Var}(\widehat{\theta})$ 均成立, 無論母體機率密度函數的型式爲何.

5. 設 $Y_1, \dots, Y_n$ 獨立同分布且

$$E(Y_i) = \mu$$

以及

$$\mathsf{Var}(Y_i) = \sigma^2$$

其中 $i = 1, \dots, n$. 一個母體變異數

$$\sigma^2 = E[(Y - E(Y))^2]$$

的估計量可選爲

$$\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \stackrel{\text{def}}{=} S'^2$$

則

$$E(\widehat{\sigma^2}) = E(S'^2) = \frac{n-1}{n}\sigma^2$$

故 $\widehat{\sigma^2} = S'^2$ 是偏差的. 但若將估計量選爲

$$\widehat{\sigma^2} = S^2 \stackrel{\text{def}}{=} \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2$$

則

$$E(\widehat{\sigma^2}) = E(S^2) = \sigma^2$$

而得 $\widehat{\sigma^2} = S^2$ 是不偏的.

註 1. 前四個估計量在大樣本下均可被常態分布近似.

註 2. 多大才算大? (1) 對大多數的母體, 當小到 $n = 5$ 時, $\overline{Y}$ 的分布呈現丘狀 (mound-shaped, 土墩型, 投手板型) 且當 $n = 30$ 或更大時, 則會迅速地呈現出常態型, 如圖示.

(2) 對於二項分布, 大小 n 因 p 而定, 因爲當 p 愈靠近 0 或 1, $\hat{p}$ 的分布愈不對稱 (asymmetric), 如圖示. 因此, 一般的原則是選擇 n 使得

$$p \pm 3\sqrt{pq/n} \in (0, 1)$$

則 $\hat{p}$ 的分布將爲丘狀且趨近於常態分布, 如圖示.