

單元 44: 隨機變數的函數簡介 (課本 §6.1)

統計的目標乃是根據取自於母體的樣本所含的資訊，對母體做推論並配以優質程度的度量 (measure of goodness)，如估計母體平均值 μ 的直觀作法為

(1) 由母體中抽出 n 個觀察值的隨機樣本

$$y_1, y_2, \dots, y_n$$

(2) 以樣本平均值

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

估計 μ .

問. 此種估計法有多好?

答. (1) 因為估計會隨著抽出的隨機樣本而變，所以估計法可寫成

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

其中 Y_i 為第 i 個可能的觀察值 (為一隨機變數), 並稱

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$$

(乃一 $Y_1, Y_2, \dots, Y_n$ 的函數, 故也是一隨機變數) 為 μ 的估計量 (estimator), 在某一特定選出的 n 個觀察值

$$y_1, y_2, \dots, y_n$$

下, $\bar{Y}$ 所給出的值

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

稱作 μ 的估計值.

(2) 一種優質程度的度量為估計誤差 (error of estimation), 定義為

$$|\bar{y} - \mu|$$

也是會隨著觀察值而變, 故無法確知它是否會小於某一特定的 B 值, 但若確定了估計量 $\bar{Y}$ 的分布後, 就可得到

$$|\bar{Y} - \mu| \leq B$$

的機率, 並以此作為一種優質程度的度量. 因此, 關鍵乃在於決定 $\bar{Y}$ (隨機變數 $Y_1, Y_2, \dots, Y_n$ 的函數) 的機率分布.

(3) 決定 $\bar{Y}$ (或隨機變數的函數) 的機率分布, 又必須要先求出隨機變數 $Y_1, Y_2, \dots, Y_n$ 的聯合機率分布. 一個常見的假設是母體大小遠大於樣本大小, 故可視 $Y_1, Y_2, \dots, Y_n$ 為獨立同分布, 亦即, 相互獨立且有共同的 pdf $f(y)$, 或共同的 pmf $p(y)$, 並敘述成 “ $Y_1, Y_2, \dots, Y_n$ 為取自於, 以 $f(y)$ 為其 pdf (或 $p(y)$ 為其 pmf) 的母體的隨機樣本”, 且以

$$Y_1, Y_2, \dots, Y_n \stackrel{\text{iid}}{\sim} f(y) \text{ (或 } p(y))$$

表示之. 因此, 隨機變數

$$Y_1, Y_2, \dots, Y_n$$

的 joint pmf (離散情況)

$$p(y_1, y_2, \dots, y_n) = p(y_1)p(y_2) \cdots p(y_n)$$

或 joint pdf (連續情況)

$$f(y_1, y_2, \dots, y_n) = f(y_1)f(y_2) \cdots f(y_n)$$

註. iid 為 independent, and identically distributed 的縮寫.