

第四章

隨機樣本及其他

4.5 順序統計量

在一組隨機樣本中，有時候最大值、最小值或中間值，是我們較有興趣的。例如，過去五十年最大降雨量，幾次跑步最短的時間，台北市房價的中間值等，這些都涉及**順序統計量**(order statistics)。

定義5.1 設 $X_1, \dots, X_n$ 為一組隨機樣本，按照小至大排出，而得 $X_{(1)}, \dots, X_{(n)}$ ，便稱**順序統計量**。

順序統計量滿足 $X_{(1)} \leq \dots \leq X_{(n)}$ 。特別地，

$$X_{(1)} = \min_{1 \leq i \leq n} X_i, \quad X_{(n)} = \max_{1 \leq i \leq n} X_i。$$

另外，

$$R = X_{(n)} - X_{(1)}$$

稱為**樣本全距**(sample range)，或稱**全距統計量**(range statistic)，表最大值與最小值之差距。譬如說班上身高最高與最矮的差距。如果樣本數 n 是

2 第四章 隨機樣本及其他

很清楚的, 則爲了簡便, 也可僅以 $\min X, \max X$ 分別表 $X_{(1)}$ 及 $X_{(n)}$ 。又

$$\text{med}X = \begin{cases} X_{((n+1)/2)} & , \text{若 } n \text{ 是奇數,} \\ (X_{(n/2)} + X_{(n/2+1)})/2 & , \text{若 } n \text{ 是偶數,} \end{cases}$$

表**樣本中位數** (sample median)。當 n 爲偶數時, 有些書對任意介於 $X_{(n/2)}$ 與 $X_{(n/2+1)}$ 間的值, 皆稱爲樣本中位數。

在美國, 職業球隊裡, 例如職棒、職籃, 極少數的球員薪水(或收入)很高, 但大部分的球員薪水是很低的。報紙上, 有時會刊出某位球員與球隊簽下新的合約, 五年薪水一億美元。平均一年兩千萬美元, 尚不包括廣告等收入! 當球員抱怨薪水太低時, 老闆可能會想全隊平均年薪已有百萬美元了, 但球員會想有一半以上的球員年薪少於二十萬美元(職業球員由於球齡短以及會因受傷而結束球員生涯, 所以薪水是較其他行業高)。這兩個觀點都是正確的, 只是一計算平均, 一計算中位數。當討論到收入、價格等含有一些較極端的值(過高或過低), 中位數可能是一較合理的指標值。

若 $X_1, \dots, X_n$ 爲一由連續型的母體所產生之隨機樣本, 則任二隨機變數 $X_i, X_j, i \neq j$, 會相等的機率爲0, 因此 $P(X_{(1)} < \dots < X_{(n)}) = 1$ 。下述定理給出任一順序統計量之分佈。

定理 5.1 設 $X_1, \dots, X_n$ 爲由連續分佈函數 $F(x)$, 且 p.d.f. 爲 $f(x)$, 所產生之隨機樣本, $X_{(1)}, \dots, X_{(n)}$ 表其順序統計量。則 $X_{(j)}$ 之 p.d.f. 爲

$$(5.1) \quad f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} (F(x))^{j-1} (1-F(x))^{n-j} f(x)。$$

證明. 我們先求 $X_{(j)}$ 之分佈函數, 然後微分而得 p.d.f.。事件 $X_{(j)} \leq x$, 等價於 $X_1, \dots, X_n$ 中至少有 j 個小於或等於 x 。因此

$$(5.2) \quad \begin{aligned} F_{X_{(j)}}(x) &= P(X_{(j)} \leq x) \\ &= \sum_{k=j}^n \binom{n}{k} (F(x))^k (1-F(x))^{n-k}。 \end{aligned}$$

將上式對 x 微分, 得

$$\begin{aligned}
 f_{X_{(j)}}(x) &= \sum_{k=j}^n \binom{n}{k} (k(F(x))^{k-1} (1-F(x))^{n-k} \\
 &\quad - (n-k)(F(x))^k (1-F(x))^{n-k-1}) f(x) \\
 &= \binom{n}{j} j(F(x))^{j-1} (1-F(x))^{n-j} f(x) \\
 &\quad + \sum_{k=j+1}^n \binom{n}{k} k(F(x))^{k-1} (1-F(x))^{n-k} f(x) \\
 &\quad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k)(F(x))^k (1-F(x))^{n-k-1} f(x) \circ \\
 &= \frac{n!}{(j-1)!(n-j)!} f(x) (F(x))^{j-1} (1-F(x))^{n-j} \circ
 \end{aligned}$$

此處用到

$$\begin{aligned}
 &\sum_{k=j+1}^n \binom{n}{k} k(F(x))^{k-1} (1-F(x))^{n-k} f(x) \\
 &= \sum_{k=j}^{n-1} \binom{n}{k+1} (k+1)(F(x))^k (1-F(x))^{n-k-1} f(x) \\
 &= \sum_{k=j}^{n-1} \frac{n!(k+1)}{(k+1)!(n-k-1)!} (F(x))^k (1-F(x))^{n-k-1} f(x) \\
 &= \sum_{k=j}^{n-1} \frac{n!(n-k)}{k!(n-k)!} (F(x))^k (1-F(x))^{n-k-1} f(x) \\
 &= \sum_{k=j}^{n-1} \binom{n}{k} (n-k)(F(x))^k (1-F(x))^{n-k-1} f(x) \circ
 \end{aligned}$$

證畢。

例5.1 設 $X_1, \dots, X_n$ 為由 $\mathcal{U}(0, 1)$ 分佈所產生隨機樣本。則 $F_X(x) = x, x \in (0, 1)$ 。利用定理5.1, 得

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!} x^{j-1} (1-x)^{n-j}$$

$$= \frac{\Gamma(n+1)}{\Gamma(j)\Gamma(n-j+1)} x^{j-1}(1-x)^{n-j}, x \in (0, 1)。$$

因此 $X_{(j)}$ 有 $Be(j, n-j+1)$ 分佈。

例5.2 設 $X_1, \dots, X_n$ 為由分佈函數 F 所產生之隨機樣本。則

$$\begin{aligned} P(X_{(n)} \leq x) &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\ &= \prod_{i=1}^n P(X_i \leq x) = (F(x))^n, x \in R。 \end{aligned}$$

當然在(5.2)式中令 $j = n$, 也會得到上式。若 F 之 p.d.f. 為 f , 則經由對 x 微分, 得

$$(5.3) \quad f_{X_{(n)}}(x) = n(F(x))^{n-1}f(x), x \in R。$$

與(5.1)式中令 $j = n$ 所得一致。

另外,

$$\begin{aligned} P(X_{(1)} \leq x) &= 1 - P(X_{(1)} > x) \\ &= 1 - P(X_1 > x, X_2 > x, \dots, X_n > x) \\ &= 1 - \prod_{i=1}^n P(X_i > x) \\ &= 1 - (1 - F(x))^n, x \in R。 \end{aligned}$$

經由對 x 微分, 得

$$(5.4) \quad f_{X_{(1)}}(x) = n(1 - F(x))^{n-1}f(x), x \in R,$$

仍與在(5.1)式中令 $j = 1$ 所得一致。

例如, 若 $X_1, \dots, X_n$ 之共同分佈為 $\mathcal{E}(\lambda)$, 則由(5.4)式, 即得 $X_{(1)} \sim \mathcal{E}(n\lambda)$, 仍有指數分佈, 只是參數改為 $n\lambda$ 。此為一有趣的結果。

定理5.2 設 $X_1, \dots, X_n$ 為由連續分佈函數 $F(x)$, 且 p.d.f. 為 $f(x)$, 所產生之隨機樣本, $X_{(1)}, \dots, X_{(n)}$ 表其順序統計量。則 $X_{(i)}, X_{(j)}, 1 \leq i < j \leq n$, 之

聯合p.d.f. 爲

$$(5.5) \quad f_{X_{(i)}, X_{(j)}}(y_i, y_j) = \frac{n!}{(i-1)!(j-i-1)(n-j)!} (F(y_i))^{i-1} (F(y_j) - F(y_i))^{j-i-1} \cdot (1 - F(y_j))^{n-j} f(y_i) f(y_j), -\infty < y_i < y_j < \infty.$$

上定理的證明我們略去了。三個或更多個順序統計量的聯合p.d.f. 亦可求出。如設整數 $1 \leq k_1 < \cdots < k_r \leq n$, 則 $X_{(k_1)}, \cdots, X_{(k_r)}$ 之聯合p.d.f. 爲

$$(5.6) \quad f_{X_{(k_1)}, \dots, X_{(k_r)}}(y_{k_1}, \dots, y_{k_r}) = \frac{n!}{(k_1-1)!(k_2-k_1-1)! \cdots (n-k_r)!} (F(y_{k_1}))^{k_1-1} \cdot (F(y_{k_2}) - F(y_{k_1}))^{k_2-k_1-1} \cdots (1 - F(y_{k_r}))^{n-k_r} f(y_{k_1}) \cdots f(y_{k_r}), \\ -\infty < y_{k_1} < \cdots < y_{k_r} < \infty.$$

又 $X_{(1)}, \cdots, X_{(n)}$ 之聯合p.d.f. 爲

$$(5.7) \quad f_{X_{(1)}, \dots, X_{(n)}}(y_1, \dots, y_n) = n! f(y_1) \cdots f(y_n), \\ -\infty < y_1 < \cdots < y_n < \infty.$$

在(5.7)式中, $n!$ 的出現是很顯然的: 對任一組 $y_1 < \cdots < y_n$, 有 $n!$ 組 $x_1, \cdots, x_n$, 其順序統計量均對應 $y_1, \cdots, y_n$ 。

底下給幾個例子。

例5.3 設 X_1, X_2, X_3 爲由 p.d.f. $f(x) = 2x$, $0 < x < 1$, 所產生之隨機樣本。則 $X_{(1)}, X_{(2)}, X_{(3)}$ 之聯合p.d.f. 爲

$$f_{X_{(1)}, X_{(2)}, X_{(3)}}(y_1, y_2, y_3) = 3! \cdot (2y_1)(2y_2)(2y_3) \\ = 48y_1y_2y_3, 0 < y_1 < y_2 < y_3.$$

由此可得 $X_{(1)}$ 之邊際p.d.f.

$$(5.8) \quad f_{X_{(1)}}(y_1) = \int_{y_1}^1 \int_{y_2}^1 48y_1y_2y_3 dy_3 dy_2$$

$$= 6y_1(1 - y_1^2)^2, 0 < y_1 < 1。$$

又 $X_{(1)}, X_{(3)}$ 之聯合 p.d.f. 爲

$$\begin{aligned} (5.9) \quad f_{X_{(1)}, X_{(3)}}(y_1, y_3) &= \int_{y_1}^{y_3} 48y_1y_2y_3dy_2 \\ &= 24y_1y_3(y_3^2 - y_1^2), 0 < y_1 < y_3 < 1。 \end{aligned}$$

可驗證(5.8)及(5.9)式, 與利用(5.1)式及(5.5)式所得相同。

例5.4 設 $X_1, \dots, X_n$ 爲由 $\mathcal{U}(0, 1)$ 分佈所產生之隨機樣本。令全距 $R = X_{(n)} - X_{(1)}$, 半全距(midrange) $V = (X_{(1)} + X_{(n)})/2$ 。試求 R, V 之聯合 p.d.f., 邊際 p.d.f., $E(R)$ 及 $E(V)$ 。

解. 首先

$$f_{X_{(1)}, X_{(n)}}(y_1, y_n) = n(n-1)(y_n - y_1)^{n-2}, 0 < y_1 < y_n < 1。$$

利用變數代換, 得

$$f_{R,V}(r, v) = n(n-1)r^{n-2}, 0 < r < 1, r/2 < v < 1 - r/2。$$

由此又得 R 之邊際 p.d.f. 爲

$$\begin{aligned} f_R(r) &= \int_{r/2}^{1-r/2} n(n-1)r^{n-2}dv \\ &= n(n-1)(1-r)r^{n-2}, 0 < r < 1。 \end{aligned}$$

可看出 R 有 $Be(n-1, 2)$ 分佈。又可求出

$$f_V(v) = \begin{cases} \int_0^{2v} n(n-1)r^{n-2}dr \\ \quad = n(2v)^{n-1}, 0 < v \leq 1/2, \\ \int_0^{2(1-v)} n(n-1)r^{n-2}dr \\ \quad = n(2(1-v))^{n-1}, 1/2 < v \leq 1。 \end{cases}$$

有了 R 及 V 之邊際 p.d.f., 便可求出 $E(R)$ 及 $E(V)$ 。如

$$\begin{aligned} E(R) &= \int_0^1 n(n-1)(1-r)r^{n-2} \cdot r dr \\ &= n(n-1) \left(\int_0^1 r^{n-1} dr - \int_0^1 r^n dr \right) \\ &= n(n-1) \left(\frac{1}{n} - \frac{1}{n+1} \right) \\ &= \frac{n-1}{n+1}。 \end{aligned}$$

但 $E(R)$ 亦可如下得到:

$$\begin{aligned} E(R) &= E(X_{(n)} - X_{(1)}) = E(X_{(n)}) - E(X_{(1)}) \\ &= \frac{n}{n+1} - \frac{1}{n+1} = \frac{n-1}{n+1}, \end{aligned}$$

$$\begin{aligned} E(V) &= E\left(\frac{X_{(1)} + X_{(2)}}{2}\right) = \frac{1}{2}(E(X_{(1)}) + E(X_{(2)})) \\ &= \frac{1}{2}\left(\frac{1}{n+1} + \frac{n}{n+1}\right) = \frac{1}{2}。 \end{aligned}$$

此處用到 $X_{(1)} \sim Be(1, n)$, $X_{(n)} \sim Be(n, 1)$ (見例5.1)。

例5.5 某廠牌燈管宣稱可使用10,000小時, 某辦公室最近安裝40支該廠牌的燈管, 才使用1個月便壞了一支。這是否合理呢?

解. 爲了便於計算, 我們假設燈管的壽命有指數分佈。即設燈管壽命 $X_1, \dots, X_{40}$ 爲 i.i.d. 之 $\mathcal{E}(\lambda)$ 分佈, 期望值 $\lambda^{-1} = 10,000$ (小時)。又設每月上班25天, 每天開燈10小時。

對一特定的燈管, 使用1個月(250小時)內會壞的機率爲

$$P(X \leq 250) = 1 - e^{-250/10,000} = 1 - e^{-0.025} \doteq 1 - 0.9753 = 0.0247,$$

的確不大。但40支燈管1個月內至少壞一支的機率, 相當於最小順序統計量 $X_{(1)} \leq 250$ 之機率。此機率可利用例2.5, $X_{(1)}$ 有 $\mathcal{E}(40\lambda) = \mathcal{E}(1/250)$ 分佈, 或直接求:

$$P(X_{(1)} \leq 250) = 1 - P(X_{(1)} > 250)$$

$$\begin{aligned}
&= 1 - P(X_1 > 250, \dots, X_{40} > 250) \\
&= 1 - (P(X > 250))^{40} \\
&= (e^{-250/10,000})^{40} \\
&= 1 - e^{-1} \doteq 1 - 0.3679 = 0.6321。
\end{aligned}$$

便不算太小了。

至於5個月(1,250小時)內至少壞一支燈管的機率為

$$\begin{aligned}
P(X_{(1)} \leq 1,250) &= 1 - (P(X > 1,250))^{40} \\
&= 1 - (e^{-1,250/10,000})^{40} \\
&= 1 - e^{-5} \doteq 0.99326,
\end{aligned}$$

非常接近1。但一特定的燈管, 5個月內壞的機率依然不大:

$$P(X \leq 1,250) = 1 - e^{1,250/10,000} = 1 - e^{-0.125} \doteq 1 - 0.8825 = 0.1175。$$

最後, 雖平均壽命為10,000小時, 但一燈管10年(30,000小時)後仍可使用的可能性有多大呢? 即要求最大順序統計量 $X_{(40)}$ 要大於30,000之機率:

$$\begin{aligned}
P(X_{(40)} > 30,000) &= 1 - (P(X \leq 30,000))^{40} \\
&= 1 - (1 - e^{30,000/10,000})^{40} \\
&= 1 - (1 - e^{-3})^{40} \doteq 1 - 0.1297 = 0.8703。
\end{aligned}$$

此機率不小。雖一特定的燈管使用10年以上的機率為 $e^{-3} \doteq 0.04979$ 很小。上述這些似令人驚訝的結果, 都是因燈管數較多所造成的。

底下我們來看一些關於順序統計量之極限結果。

例5.6 設 $X_1, \dots, X_n$ 為由 $\mathcal{U}[0, \theta]$ 分佈所產生之隨機樣本, $\theta > 0$ 。令 $Y_n = X_{(n)}$, $G_n(y) = P(Y_n \leq y)$ 。則對 $\forall 0 < y < \theta$,

$$G_n(y) = (P(X \leq y))^n = y^n,$$

且 $G_n(y) = 0$, 若 $y \leq 0$, $G_n(y) = 1$, 若 $y \geq \theta$ 。故得

$$\lim_{n \rightarrow \infty} G_n(y) = G(y) = \begin{cases} 0, & 0 \leq y < \theta, \\ 1, & y \geq \theta. \end{cases}$$

$G(y)$ 對應常數隨機變數 $Y \equiv \theta$ 之分佈函數。即證出

$$Y_n \xrightarrow[n \rightarrow \infty]{d} \theta。$$

又由 (2.22) 式, 得

$$(5.10) \quad Y_n \xrightarrow[n \rightarrow \infty]{p} \theta。$$

也可依定義 2.1, 直接證明 (5.10) 式。

例5.7 設 $X_1, \dots, X_n$ 為由 p.d.f. $f(x) = (1+x)^{-2}$, $x > 0$, 所產生之隨機樣本, $F(x) = 1 - (1+x)^{-1}$, $x > 0$ 。令 $Y_n = nX_{(1)}$, $G_n(y) = P(Y_n \leq y)$ 。則對 $\forall y > 0$,

$$\begin{aligned} G_n(y) &= P(nX_{(1)} \leq y) \\ &= 1 - P(X_{(1)} > y/n) \\ &= 1 - (1 + y/n)^{-n} \\ &\xrightarrow{n \rightarrow \infty} 1 - e^{-y}。 \end{aligned}$$

上述極限見第二章註2.3。又 $G_n(y) = 0$, $\forall y \leq 0$ 。故得

$$Y_n \xrightarrow[n \rightarrow \infty]{d} \mathcal{E}(1)。$$

次令 $U_n = X_{(n)}/n$ 。則可證明 $U_n \xrightarrow{d} U$, 其中 $P(U \leq u) = e^{-1/u}$, $u > 0$ (習題第18題)。