# 單元 50: 有序統計量
(課本 §6.7)

設連續隨機變數

$$Y_1, Y_2, \ldots, Y_n \overset{\text{iid}}{\sim} \text{cdf } F(y) \text{ 以及 pdf } f(y)$$

則排序後的結果

$$Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$$

稱作有序統計量 (order statistics), 其中 $Y_{(i)}$ 表示第 $i$ 個小的隨機變數, 亦即,

$$Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n) \leq Y_{(2)} \leq \cdots \leq Y_{(n-1)} \leq Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$$

**註.** 因爲隨機變數是連續的, 上式中的 “=” 可忽略掉, 而寫成

$$Y_{(1)} < Y_{(2)} < \cdots < Y_{(n-1)} < Y_{(n)}$$

有序統計量的分布如下述.

(1) $Y_{(n)}$ 的 cdf $F_{Y_{(n)}}(y)$ 與 pdf $g_{(n)}(y)$.

因爲

$$Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$$

故對任意的實數 $y$,

$$\left(Y_{(n)} \le y\right) \Leftrightarrow (Y_1 \le y, Y_2 \le y, \ldots, Y_n \le y)$$

因此, 根據 cdf 的定義, 以及上述的等價事件, $Y_{(n)}$ 的 cdf

$$\begin{aligned} F_{Y_{(n)}}(y) &= P\left(Y_{(n)} \le y\right) \\ &= P(Y_1 \le y, Y_2 \le y, \ldots, Y_n \le y) \\ &= P(Y_1 \le y)P(Y_2 \le y)\cdots P(Y_n \le y) \\ &= [F(y)]^n, \ -\infty < y < \infty \end{aligned}$$

其中第三個等號成立乃因爲 $Y_1, Y_2, \ldots, Y_n$ 相互獨立, 第四個等號成立乃因爲它們有共同的 cdf $F(y)$. 接著, 兩邊對 $y$ 微分, 得 $Y_{(n)}$ 的 pdf

$$\begin{aligned} g_{(n)}(y) &= \frac{d}{dy}\left[F_{Y_{(n)}}(y)\right] \\ &= n\,[F(y)]^{n-1} \cdot F'(y) \\ &= n\,[F(y)]^{n-1} f(y), \ -\infty < y < \infty \end{aligned}$$

其中最後一個等號成立乃因爲它們有共同的 pdf $f(y)$.

(2) $Y_{(1)}$ 的 cdf $F_{Y_{(1)}}(y)$ 與 pdf $g_{(1)}(y)$.

因爲

$$Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$$

故對於任意的實數 $y$,

$$\left(Y_{(1)} > y\right) \Leftrightarrow (Y_1 > y, Y_2 > y, \ldots, Y_n > y)$$

所以, 根據 cdf 的定義, 餘集的機率公式, 以及上述的等價事件, $Y_{(1)}$ 的 cdf

$$\begin{aligned} F_{Y_{(1)}}(y) &= P\left(Y_{(1)} \le y\right) \\ &= 1 - P\left(Y_{(1)} > y\right) \\ &= 1 - P(Y_1 > y, Y_2 > y, \ldots, Y_n > y) \\ &= 1 - P(Y_1 > y)P(Y_2 > y)\cdots \\ &\qquad\qquad P(Y_n > y) \\ &= 1 - [1 - F(y)]^n, \ -\infty < y < \infty \end{aligned}$$

其中第四個等號成立乃因爲 $Y_1, Y_2, \ldots, Y_n$ 相互獨立, 第五個等號成立乃因爲它們有共同的 cdf $F(y)$ 且對於 $i = 1, 2, \ldots, n$, $-\infty < y < \infty$,

$$P(Y_i > y) = 1 - P(Y_i \le y) = 1 - F(y)$$

所致. 接著, 兩邊對 $y$ 微分, 得 $Y_{(1)}$ 的 pdf

$$\begin{aligned} g_{(1)}(y) &= \frac{d}{dy}\left[F_{Y_{(1)}}(y)\right] \\ &= -n\,[1 - F(y)]^{n-1} \cdot [-f(y)] \\ &= n\,[1 - F(y)]^{n-1}\, f(y), -\infty < y < \infty \end{aligned}$$

(3) $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ 的 joint pdf

$$g_{(1)(2)\cdots(n)}(y_1, y_2, \ldots, y_n)$$

直觀的求法: 設 $y_1 < y_2 < \cdots < y_n$. 取 $\Delta y$ 夠小, 令事件

$$\begin{aligned} B &= \Big(y_1 < Y_{(1)} < y_1 + \Delta y, \\ &\qquad y_2 < Y_{(2)} < y_2 + \Delta y, \ldots, \\ &\qquad\qquad y_n < Y_{(n)} < y_n + \Delta y\Big) \end{aligned}$$

對於 $\{1, 2, \ldots, n\}$ 的任一種排列 $(i_1, i_2, \ldots, i_n)$, 令事件

$$\begin{aligned} A_{(i_1, i_2, \ldots, i_n)} &= (y_1 < Y_{i1} < y_1 + \Delta y, \\ &\qquad y_2 < Y_{i2} < y_2 + \Delta y, \ldots, \\ &\qquad\qquad y_n < Y_{in} < y + \Delta y) \end{aligned}$$

共有 $n!$ 種不同的事件. 在 $\Delta y$ 夠小下, 此 $n!$ 個事件爲互斥, 且在 $Y_1, Y_2, \ldots, Y_n$ 同分布的假設下, 此 $n!$ 個事件的機率均等於 $P\left(A_{(1,2,\ldots,n)}\right)$, 因而得出

$$\begin{aligned} P(B) &= P\left(\bigcup_{(i1, i2, \ldots, in)} A_{(i1, i2, \ldots, in)}\right) \\ &= n! P\left(A_{(1,2,\ldots,n)}\right) \end{aligned}$$

又在 $Y_1, Y_2, \ldots, Y_n$ 相互獨立的假設下, 上式相當於

$$\begin{aligned} P(B) &= n! P\left(y_1 < Y_1 < y_1 + \Delta y\right) \\ &\qquad P\left(y_2 < Y_2 < y_2 + \Delta y\right) \cdots \\ &\qquad P\left(y_n < Y_n < y_n + \Delta y\right) \end{aligned}$$

又因爲 $\Delta y$ 夠小, 故對於 $i = 1, 2, \ldots, n$,

$$\begin{aligned} P\left(y_i < Y_i < y_i + \Delta y\right) &= \int_{y_i}^{y_i + \Delta y} f(t)dt \\ &\approx f(y_i)\Delta y \end{aligned}$$

因而由上式導出

$$\begin{aligned} P(B) &\approx n! \left[f(y_1)\Delta y\right]\left[f(y_2)\Delta y\right] \cdots \left[f(y_n)\Delta y\right] \\ &= n! f(y_1) f(y_2) \cdots f(y_n) (\Delta y)^n \qquad (1) \end{aligned}$$

接著, 以 $Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)}$ 的 joint pdf 的角度, 在 $\Delta y$ 夠小下,

$$\begin{aligned} P(B) &= \int_{y_1}^{y_1 + \Delta y} \int_{y_2}^{y_2 + \Delta y} \cdots \int_{y_n}^{y_n + \Delta y} \\ &\qquad g_{(1)(2)\cdots(n)}(t_1, t_2, \ldots, t_n) \\ &\qquad dt_n \cdots dt_2 dt_1 \\ &\approx g_{(1)(2)\cdots(n)}(y_1, y_2, \ldots, y_n)(\Delta y)^n \quad (2) \end{aligned}$$

最後, 比較 (1) 式及 (2) 式, 得

$$\begin{aligned} & g_{(1)(2)\cdots(n)}(y_1, y_2, \ldots, y_n) \\ &= \begin{cases} n! f(y_1) f(y_2) \cdots f(y_n), & y_1 \le y_2 \le \cdots \le y_n \\ 0, & \text{其它} \end{cases} \end{aligned}$$

其中等號右邊大括號的第二式成立乃因爲

$$Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$$

故當 $y_1 \le y_2 \le \cdots \le y_n$ 不成立時, $B = \emptyset$ 所致.

(4) 對於 $1 \le k \le n$, 第 $k$ 個有序統計量 $Y_{(k)}$ 的 pdf $g_{(k)}(y)$.

直觀的求法: 對於 $-\infty < y_k < \infty$, 當 $\Delta y$ 夠小時, 如圖示,

$$\begin{aligned} & P(y_k < Y_{(k)} < y_k + \Delta y) \\ & \quad \approx \ P(Y_1, Y_2, \ldots, Y_n \text{ 中有 } (k-1) \text{ 個小於 } y_k, \text{一} \\ & \qquad \text{個介於 } y_k \text{ 與 } y_k + \Delta y \text{ 間}, (n-k) \text{ 個大於 } y_k) \end{aligned}$$

接著, 因爲 $Y_1, Y_2, \ldots, Y_n$ 爲獨立同分布, 故上式中近似符號右邊的機率爲三項分布, 其中落在第一類的機率爲

$$F(y_k)$$

有 $(k-1)$ 個; 落在第二類的機率近似於

$$f(y_k)\Delta y$$

有一個; 落在第三類的機率爲

$$1 - F(y_k)$$

有 $(n-k)$ 個, 因而得出

$$\begin{aligned} & P(y_k < Y_{(k)} < y_k + \Delta y) \\ & \quad \approx \binom{n}{k-1,1,n-k} [F(y_k)]^{k-1} \cdot \\ & \qquad\qquad f(y_k)\Delta y \cdot [1-F(y_k)]^{n-k} \\ & \quad = \frac{n!}{(k-1)!(n-k)!} [F(y_k)]^{k-1} f(y_k) \cdot \\ & \qquad\qquad [1-F(y_k)]^{n-k} \Delta y \quad (3) \end{aligned}$$

又由 $Y_{(k)}$ 的 marginal pdf 的角度, 在 $\Delta y$ 夠小下,

$$P(y_k < Y_{(k)} < y_k + \Delta y) \approx g_{(k)}(y_k)\Delta y \qquad (4)$$

因此, 比較 (3) 式與 (4) 式, 得

$$\begin{aligned} g_{(k)}(y_k) &= \frac{n!}{(k-1)!(n-k)!} [F(y_k)]^{k-1} \cdot \\ & \qquad [1-F(y_k)]^{n-k} f(y_k) \end{aligned}$$

其中 $-\infty < y_k < \infty$.

(5) 對任意的 $1 \le j < k \le n$, $Y_{(j)}$ 與 $Y_{(k)}$ 的 joint pdf $g_{(j)(k)}(y_j, y_k)$.

直觀的求法: 同 (4), 如圖示, 當 $-\infty < y_j < y_k < \infty$ 時, $Y_1, Y_2, \ldots, Y_n$ 中, 需要有 $(j-1)$ 個小於 $y_j$, 個別的機率爲

$$F(y_j)$$

一個介於 $y_j$ 與 $y_j + \Delta y$ 間, 其機率近似於

$$f(y_j)\Delta y$$

$(k-1-j)$ 個介於 $y_j$ 與 $y_k$ 間, 個別的機率爲

$$F(y_k) - F(y_j)$$

一個介於 $y_k$ 與 $y_k + \Delta y$ 間, 其機率近似於

$$f(y_k)\Delta y$$

以及 $(n-k)$ 個大於 $y_k$, 個別的機率爲

$$1 - F(y_k)$$

因而 $Y_{(j)}$ 與 $Y_{(k)}$ 的 joint pdf

$$\begin{aligned} & g_{(j)(k)}(y_j, y_k) \\ = \; & \frac{n!}{(j-1)!(k-1-j)!(n-k)!}\left[F(y_j)\right]^{j-1} \cdot \\ & \left[F(y_k) - F(y_j)\right]^{k-1-j}\left[1 - F(y_k)\right]^{n-k} \cdot \\ & f(y_j)f(y_k),\ -\infty < y_j < y_k < \infty \end{aligned}$$

例 1. 設某種電子零件的壽命

$$Y \sim f(y) = \begin{cases} \frac{1}{100}e^{-y/100}, & y > 0 \\ 0, & \text{其它} \end{cases}$$

其中的單位爲小時. 若某系統由二個獨立運作的此種零件以串聯方式 (in series) 組成, 亦即, 只要其中一個零件壞了, 此系統就壞了. 令 $X$ 爲此系統的壽命. 試求 $X$ 的 pdf.

<解> 因爲串聯, 故

$$X = \min(Y_1, Y_2)$$

其中 $Y_i =$ 第 $i$ 個零件的壽命, 它們獨立同分布且共同的 cdf

$$F(y) = \begin{cases} 1 - e^{-y/100}, & y > 0 \\ 0, & \text{其它} \end{cases}$$

所以, 由公式或圖示, $X$ 的 pdf

$$f_X(y) = g_{(1)}(y) = 2\,[1 - F(y)]\,f(y) \tag{5}$$

再根據 pdf $f(y)$ 與 cdf $F(y)$ 在 $y > 0$ 時, 分別取值

$$\frac{1}{100}e^{-y/100} \text{ 與 } 1 - e^{-y/100}$$

其它地方取值 0, 由 (5) 式得 $X$ 的 pdf

$$\begin{aligned} f_X(y) &= \begin{cases} 2 \cdot e^{-y/100} \cdot \frac{1}{100}e^{-y/100}, & y > 0 \\ 0, & \text{其它} \end{cases} \\ &= \begin{cases} \frac{1}{50}e^{-y/50}, & y > 0 \\ 0, & \text{其它} \end{cases} \end{aligned} \tag{6}$$

**註.** 由 (6) 式知, 串聯形成的系統壽命

$$X = \min(Y_1, Y_2) \sim \exp(50)$$

由此導出

$$E(X) = 50 = \frac{100}{2}$$

然而對於 $i = 1, 2$, 組成系統的零件 $i$ 的壽命

$$Y_i \sim \exp(100)$$

以及零件平均壽命

$$E(Y_i) = 100$$

卻爲系統平均壽命的二倍.

**例 2.** 承接例 1. 若系統由並聯 (in parallel) 組成, 亦即, 二個零件全壞了, 系統才壞掉. 試求系統壽命 $X$ 的 pdf.

<解> 因爲並聯, 故

$$X = \max(Y_1, Y_2)$$

因此, 由公式或圖示, $X$ 的 pdf

$$f_X(y) = g_{(2)}(y) = 2F(y)f(y) \tag{7}$$

再根據 pdf $f(y)$ 與 cdf $F(y)$ 在 $y > 0$ 時, 分別取值

$$\frac{1}{100}e^{-y/100} \text{ 與 } 1 - e^{-y/100}$$

其它地方取值 0, 由 (7) 式得 $X$ 的 pdf

$$
\begin{aligned}
& f_X(y) \\
&= \begin{cases} 2 \cdot \left(1 - e^{-y/100}\right) \cdot \frac{1}{100}e^{-y/100}, & y > 0 \\ 0, & \text{其它} \end{cases} \\
&= \begin{cases} \frac{1}{50}\left(e^{-y/100} - e^{-y/50}\right), & y > 0 \\ 0, & \text{其它} \end{cases} \qquad (8)
\end{aligned}
$$

**註.** 由 (8) 知, 由並聯形成的系統壽命

$$X = \max(Y_1, Y_2)$$

不是指數隨機變數.

**例 3.** 設 $Y_1, Y_2, \ldots, Y_5$ 爲一由 unif$(0,1)$ 分布所選出的隨機樣本. 試求 (a) $Y_{(2)}$ 的 pdf, 以及 (b) $Y_{(2)}$ 與 $Y_{(4)}$ 的 joint pdf.

<解> 首先, 由題意, $Y_1, Y_2, \ldots, Y_5$ 爲獨立同分布且共同的 pdf

$$f(y) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{其它} \end{cases}$$

以及共同的 cdf

$$F(y) = \begin{cases} 0, & y < 0 \\ y, & 0 \le y \le 1 \\ 1, & y > 1 \end{cases}$$

此乃因爲由 cdf 的定義以及 pdf $f(y)$ 的取值, 當 $y < 0$ 時,

$$F(y) = \int_{-\infty}^{y} f(t)dt = \int_{-\infty}^{y} 0dt = 0$$

當 $0 \le y \le 1$ 時,

$$F(y) = \int_{-\infty}^{0} 0dt + \int_{0}^{y} 1dt = y$$

以及當 $y > 1$ 時,

$$F(y) = \int_{-\infty}^{0} 0dt + \int_{0}^{1} 1dt + \int_{1}^{y} 0dt = 1$$

所致.

(a) 由公式或圖示, 對於 $-\infty < y_2 < \infty$, $Y_{(2)}$ 的 pdf

$$g_{(2)}(y_2) = \frac{5!}{1!1!3!} F(y_2) \left[1 - F(y_2)\right]^3 f(y_2)$$

再根據 pdf $f(y)$ 與 cdf $F(y)$ 的取值情況, 當 $y_2 < 0$ 時, 由上式得

$$g_{(2)}(y_2) = 20(0)(1-0)^3(0) = 0$$

當 $0 \le y_2 \le 1$ 時,

$$\begin{aligned} g_{(2)}(y_2) &= 20(y_2)(1-y_2)^3(1) \\ &= 20y_2(1-y_2)^3 \end{aligned}$$

當 $y_2 > 1$ 時,

$$g_{(2)}(y_2) = 20(1)(1-1)^3(0) = 0$$

因此, 綜合上述, 得 $Y_{(2)}$ 的 pdf

$$g_{(2)}(y_2) = \begin{cases} 20y_2(1-y_2)^3, & 0 \le y_2 \le 1 \\ 0, & \text{其它} \end{cases}$$

註. 由 (a) 的結論,

$$Y_{(2)} \sim \text{Beta}(2,4)$$

一般而言, 若

$$Y_1, Y_2, \ldots, Y_n \overset{\text{iid}}{\sim} \text{unif}(0,1)$$

則第 $k$ 個有序統計量 ($k$th order statistics)

$$Y_{(k)} \sim \text{Beta}(k, n-k+1)$$

爲何如此? 根據公式或圖示, 對於 $-\infty < y_k < \infty$, $Y_{(k)}$ 的 pdf

$$\begin{aligned} g_{(k)}(y_k) &= \frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1} \cdot \\ &\qquad [1-F(y_k)]^{n-k} f(y_k) \end{aligned}$$

再根據 pdf $f(y)$ 與 cdf $F(y)$ 的取值情況, 如同 (a) 的推導過程, 由上式得 $Y_{(k)}$ 的 pdf

$$g_{(k)}(y_k) = \begin{cases} \frac{\Gamma(n+1)}{\Gamma(k)\Gamma(n-k+1)} y_k^{k-1} \cdot (1-y_k)^{(n-k+1)-1}, & 0 \le y_k \le 1 \\ 0, & \text{其它} \end{cases}$$

剛好就是 Beta$(k, n-k+1)$ 的 pdf. 因此, 對於 $1 \le k \le n$, 第 $k$ 個有序統計量

$$Y_{(k)} \sim \text{Beta}(k, n-k+1)$$

(b) 根據公式或圖示, 對於 $-\infty < y_2 < y_4 < \infty$, $Y_{(2)}$ 與 $Y_{(4)}$ 的 joint pdf

$$g_{(2)(4)}(y_2, y_4) = \frac{5!}{1!1!1!1!1!} F(y_2) \cdot [F(y_4) - F(y_2)] \cdot [1 - F(y_4)] f(y_2) f(y_4)$$

再根據 pdf $f(y)$ 與 cdf $F(y)$ 的取值情況, 如同 (a) 的推導過程, 由上式得 $Y_{(2)}$ 與 $Y_{(4)}$ 的 joint pdf

$$g_{(2)(4)}(y_2, y_4) = \begin{cases} 5! y_2 (y_4 - y_2)(1 - y_4), & 0 \le y_2 < y_4 \le 1 \\ 0, & \text{其它} \end{cases}$$