單元 34: 雙變量及多變量機率分布
(課本 §5.1, 5.2)

關注二個或多個事件的交集有其必要性, 如

(1) 由 52 張牌中抽二張時, 抽出一張 A (ace) 與一張人頭 (face) 的事件.

(2) 觀察一窩動物中的生存數時, 會考慮事件 A (一窩中的動物數 n) 與事件 B (生存數 y) 的交集.

(3) 觀察一個體的身高與體重, 此乃相當於一對特定身高事件與體重事件的交集.

(4) 對統計學家而言, 重要的事爲: 在取樣過程中的交集, 如 $Y_1, Y_2, \ldots, Y_n$ 表示一實驗的 n 個接續試驗的結果, 它們可代表 n 人的體重 (實驗: 量體重) 或一個人的 n 項具體特徵 (實驗: 抽一人, 觀察其 n 項指定的特徵). 則一特定結果的集合可表成 n 個事件 $(Y_1 = y_1)$, $(Y_2 = y_2)$, $\ldots$, $(Y_n = y_n)$ 的交集, 並記成

$$(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$$

或更簡潔地寫成

$$(y_1,\, y_2,\, \ldots,\, y_n)$$

同時, 對於推論樣本所取自的母體而言, 計算此交集的機率是絕對必要的, 並且是探討多變量機率分布的一主要原因.

多個隨機變數可定義在相同的樣本空間上, 如投擲二個公正的骰子, 則樣本空間

$$\begin{aligned} S &= \{36 \text{ 個樣本點}\} \\ &= \{(y_1, y_2) : y_1 = 1, \ldots, 6,\ y_2 = 1, \ldots, 6\} \end{aligned}$$

則下列的隨機變數均可定義在此樣本空間 S 上:

$$\begin{aligned} Y_1 &: \text{第一個骰子的點數} \\ Y_2 &: \text{第二個骰子的點數} \\ Y_3 &: \text{點數和} \\ Y_4 &: \text{點數乘積} \end{aligned}$$

一個合理的機率定義: 對於 $y_1 = 1, \ldots, 6$, $y_2 = 1, \ldots, 6$,

$$\begin{aligned} P(\text{任一樣本點}) &= P(Y_1 = y_1, Y_2 = y_2) \\ &\overset{\text{表成}}{=} p(y_1, y_2) = \frac{1}{36} \end{aligned}$$

稱作雙變量 (二元) 機率函數 (bivariate probability function), 其 (三維) 相對頻率直方圖如下所示, 且

$$\sum_{y_1=1}^{6}\sum_{y_2=1}^{6} p(y_1,y_2)=1$$

正式定義如下.

定義. 令 Y_1, Y_2 爲二離散隨機變數, 則 Y_1 與 Y_2 的聯合 (二元, 雙變量) 機率質量函數 (joint 或 bivariate probability mass function, 簡稱 joint pmf)

$$p(y_1,y_2) \stackrel{\text{def}}{=} P(Y_1=y_1, Y_2=y_2)$$

其中 $-\infty<y_1<\infty$, $-\infty<y_2<\infty$, 並以

$$(Y_1,Y_2)\sim p(y_1,y_2)$$

表示之.

註. 如同單變量的情形, 雙變量的 joint pmf $p(y_1,y_2)$ 僅在可數個有序對 (y_1,y_2) 上有正的機率值且總和爲 1.

定理 5.1. 若 $(Y_1,Y_2)\sim p(y_1,y_2)$ (亦即, Y_1, Y_2 爲離散隨機變數且其 joint pmf 爲 $p(y_1,y_2)$), 則

(1) 對於所有的 y_1, y_2, $p(y_1, y_2) \geq 0$

(2) 所有正的 $p(y_1, y_2)$ 的總和爲 1, 亦即,

$$\sum_{(y_1,y_2):p(y_1,y_2)>0} p(y_1, y_2) = 1$$

例 1. 設一超商有 3 個收銀台, 且有 2 個顧客在不同的時間任選一收銀台付賬. 令 Y_1 = 選擇收銀台 1 的顧客數, Y_2 = 選擇收銀台 2 的顧客數. 試求 Y_1 與 Y_2 的 joint pmf.

<解> 對於 $i = 1, 2, 3$, $j = 1, 2, 3$, 令樣本點 $\{i, j\}$ = 第 1 個顧客選擇收銀台 i, 且第 2 個顧客選擇收銀台 j. 根據 mn 律, 共有 $3 \times 3 = 9$ 個樣本點, 組成樣本空間

$$\begin{aligned} S = & [\{1,1\},\{1,2\},\{1,3\},\{2,1\},\{2,2\},\{2,3\}, \\ & \{3,1\},\{3,2\},\{3,3\}] \end{aligned}$$

因爲任選, 每一樣本點都是相同可能地發生, 故機率均爲 $\frac{1}{9}$. 又

$$\{1,1\} \Leftrightarrow (Y_1 = 2, Y_2 = 0)$$

由此導出

$$p(2,0)=\frac{1}{9}$$

同理,

$$\{1,2\},\{2,1\}\Leftrightarrow(Y_1=1,Y_2=1)$$

故,

$$p(1,1)=\frac{2}{9}$$

接著,

$$\{1,3\},\{3,1\}\Leftrightarrow(Y_1=1,Y_2=0)$$

由此得

$$p(1,0)=\frac{2}{9}$$

又因爲

$$\{2,2\}\Leftrightarrow(Y_1=0,Y_2=2)$$

故可導出

$$p(0,2)=\frac{1}{9}$$

最後,

$$\{2,3\},\{3,2\}\Leftrightarrow(Y_1=0,Y_2=1)$$

故得

$$p(0,1) = \frac{2}{9}$$

以及由

$$\{3,3\} \Leftrightarrow (Y_1 = 0, Y_2 = 0)$$

可導出

$$p(0,0) = \frac{1}{9}$$

所以, joint pmf $p(y_1, y_2)$ 可列成下表:

y_2 \ y_1	0	1	2
0	1/9	2/9	1/9
1	2/9	2/9	0
2	1/9	0	0

同單變量隨機變數, 聯合離散與聯合連續隨機變數的差別可由聯合分布表現出, 其正式定義如下.

定義. 對任意的隨機變數 Y_1 與 Y_2, 它們的聯合 (雙變量) 分布函數 (joint (bivariate) distribution function, 簡稱 joint cdf)

$$F(y_1, y_2) \overset{\text{def}}{=} P(Y_1 \le y_1, Y_2 \le y_2), \quad -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

註 1. 若 Y_1, Y_2 爲離散的, 則 joint cdf

$$F(y_1, y_2) = \sum_{t_1=-\infty}^{y_1} \sum_{t_2=-\infty}^{y_2} p(t_1, t_2)$$

例 2. 試求例 1 中, Y_1, Y_2 的 $F(-1, 2)$, $F(1.5, 2)$ 與 $F(5, 7)$.

<解> 根據 joint cdf 的定義, 首先,

$$F(-1, 2) = P(Y_1 \le -1, Y_2 \le 2) = P(\emptyset) = 0$$

其中第二個等號成立此乃因爲 $Y_1 \ge 0$ 所致. 接著, 再根據上述有關離散隨機變數的註 1,

$$\begin{aligned} F(1.5, 2) &= P(Y_1 \le 1.5, Y_2 \le 2) \\ &= p(0,0) + p(0,1) + p(0,2) + \\ &\qquad p(1,0) + p(1,1) + p(1,2) \\ &= \frac{1}{9} + \frac{2}{9} + \frac{1}{9} + \frac{2}{9} + \frac{2}{9} + 0 = \frac{8}{9} \end{aligned}$$

最後, 因爲 Y_1 與 Y_2 均小於或等於 2, 故事件

$$(Y_1 \le 5, Y_2 \le 7)$$

恆發生, 因而

$$F(5, 7) = P(Y_1 \le 5, Y_2 \le 7) = P(S) = 1$$

註 2. 因爲 $Y_1 \leq 2$ 且 $Y_2 \leq 2$, 故, 若

$$\min(y_1, y_2) \geq 2$$

則

$$F(y_1, y_2) = P(S) = 1$$

又因爲 $Y_1 \geq 0$ 且 $Y_2 \geq 0$, 故, 若

$$\min(y_1, y_2) < 0$$

則

$$F(y_1, y_2) = P(\emptyset) = 0$$

註 3. 稱隨機變數 Y_1, Y_2 爲聯合連續的, 若其聯合分布函數 $F(y_1, y_2)$ 在整個平面 (亦即, $-\infty < y_1 < \infty$, $-\infty < y_2 < \infty$) 上連續. 嚴格定義如下述.

定義. 設連續隨機變數 Y_1 與 Y_2 的 joint cdf 爲 $F(y_1, y_2)$. 若存在一非負函數 $f(y_1, y_2)$ 使得對於所有的

$$-\infty < y_1 < \infty, -\infty < y_2 < \infty$$

聯合累積分布函數

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2) dt_2 dt_1$$

則稱 Y_1, Y_2 爲聯合連續隨機變數, 且稱 $f(y_1, y_2)$ 爲其聯合機率密度函數 (joint probability density function, 簡稱 joint pdf), 並以

$$(Y_1, Y_2) \sim f(y_1, y_2)$$

表示之.

定理 5.2. (joint cdf 的性質). 設

$$(Y_1, Y_2) \sim F(y_1, y_2)$$

亦即, Y_1, Y_2 的 joint cdf 爲 $F(y_1, y_2)$, 則

(1) 對任意的 y_1, y_2, 在負無窮遠的值 (極限)

$$F(-\infty, -\infty) = F(-\infty, y_2) = F(y_1, -\infty) = 0$$

因爲 $Y_1 \in (-\infty, \infty)$ 且 $Y_2 \in (-\infty, \infty)$, 故根據 joint cdf 的定義,

$$\begin{aligned} F(-\infty, -\infty) &= P(Y_1 \leq -\infty, Y_2 \leq -\infty) \\ &= P(\emptyset) \text{ (恆不發生)} \\ &= 0 \end{aligned}$$

同理,

$$\begin{aligned} F(-\infty, y_2) &= P(Y_1 \leq -\infty, Y_2 \leq y_2) \\ &= P(\emptyset) \text{ (恆不發生)} \\ &= 0 \end{aligned}$$

以及

$$
\begin{aligned}
F(y_1, -\infty) &= P(Y_1 \le y_1, Y_2 \le -\infty) \\
&= P(\emptyset) \text{ (恆不發生)} \\
&= 0
\end{aligned}
$$

(2) 在正無窮遠的值 (極限)

$$F(\infty, \infty) = 1$$

因爲 Y_1 與 Y_2 均小於 ∞, 故根據 joint cdf 的定義,

$$
\begin{aligned}
F(\infty, \infty) &= P(Y_1 \le \infty, Y_2 \le \infty) \\
&= P(S) \text{ (恆發生)} \\
&= 1
\end{aligned}
$$

(3) 對任意的 $y_1 \le y_1^*$ 以及 $y_2 \le y_2^*$, 則

$$
\begin{aligned}
&F(y_1^*, y_2^*) - F(y_1^*, y_2) - F(y_1, y_2^*) \\
&\qquad\qquad + F(y_1, y_2) \ge 0
\end{aligned}
$$

根據圖示, 上式不等號的左邊等於

$$P(y_1 < Y_1 \le y_1^*,\ y_2 < Y_2 \le y_2^*)$$

故大於或等於 0.

定理 5.3. (joint pdf 的性質). 設

$$(Y_1, Y_2) \sim \text{joint pdf } f(y_1, y_2)$$

則

(1) 對所有的 y_1, y_2,

$$f(y_1, y_2) \geq 0$$

(2) 在整個平面上的積分爲 1, 亦即,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 = 1$$

因爲根據 joint cdf 的性質以及 joint pdf 的定義,

$$1 = F(\infty, \infty) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2$$

註 4. 對任意的 $a_1 \leq a_2$, $b_1 \leq b_2$, 根據 joint pdf $f(y_1, y_2)$ 與 joint cdf $F(y_1, y_2)$ 的定義,

$$\begin{aligned} & P(a_1 \leq Y_1 \leq a_2, b_1 \leq Y_2 \leq b_2) \\ = \;& f(y_1, y_1) \text{ 在區域 } [a_1, a_2] \times [b_1, b_2] \text{ 上的體積} \\ = \;& \int_{b_1}^{b_2}\int_{a_1}^{a_2} f(y_1, y_2) dy_1 dy_2 \end{aligned}$$

亦即, 聯合連續隨機變數 (Y_1, Y_2) 落在平面上任一區域的機率等於其 joint pdf $f(y_1, y_2)$ 在此區域上的積分.

例 3. 設一放射性粒子會隨機地落在一邊長爲 1 的方塊內, 亦即, 粒子落在面積相同的區域內的機會是一樣的, 如圖示. 令 Y_1, Y_2 分別爲粒子位置的橫坐標與縱坐標, 則合理的 joint pdf

$$f(y_1, y_2) = \begin{cases} 1, & 0 \le y_1 \le 1, 0 \le y_2 \le 1 \\ 0, & \text{其它} \end{cases}$$

此乃一二變量版的均勻分布. 試求 (a) joint pdf $f(y_1, y_2)$ 的圖形, (b) $F(.2, .4)$, 以及 (c) $P(.1 \le Y_1 \le .3, 0 \le Y_2 \le .5)$.

<解> (a) joint pdf $f(y_1, y_2)$ 爲一在單位正方形上高度爲 1 的水平面以及其餘地方高度爲 0 的水平面的聯集, 如圖示.

(b) 根據 joint cdf 的定義, 以及 $f(y_1, y_2)$ 的公式,

$$\begin{aligned} F(.2, .4) &= \int_{-\infty}^{.2} \int_{-\infty}^{.4} f(y_1, y_2) dy_2 dy_1 \\ &= \int_0^{.2} \int_0^{.4} 1 dy_2 dy_1 \\ &= f(y_1, y_2) \equiv 1 \text{ 在 } [0, .2] \times [0, .4] \\ &\quad \text{上的體積} \\ &= 1 \cdot ([0, .2] \times [0, .4] \text{ 的面積}) = .08 \end{aligned}$$

或直接積分, 得

$$
\begin{aligned}
F(.2,.4) &= \int_0^{.2}\int_0^{.4} 1 dy_2 dy_1 \\
&= \int_0^{.2}\left(y_2\Big|_0^{.4}\right)dy_1 \\
&= \int_0^{.2}(.4)dy_1 \\
&= (.4)y_1\Big|_0^{.2} \\
&= (.4)(.2) = .08
\end{aligned}
$$

(c) 根據註 4, 亦即, (Y_1, Y_2) 落在一區域內的機率等於其 joint pdf $f(y_1, y_2)$ 在此區域上的積分,

$$
\begin{aligned}
&P(.1 \le Y_1 \le .3,\ 0 \le Y_2 \le .5) \\
&= \int_{.1}^{.3}\int_0^{.5} f(y_1, y_2) dy_2 dy_1 \\
&= \int_{.1}^{.3}\int_0^{.5} 1 dy_2 dy_1 \\
&= 1\cdot([.1,.3]\times[0,.5] \text{ 的面積}) \\
&= 1\cdot(.3-.1)(.5) = .1
\end{aligned}
$$

例 4. 令 $Y_1 =$ 庫存石油的比率, $Y_2 =$ 銷售石油的比率 (均針對石油容器而言), 則 $0 \le Y_2 \le Y_1 \le 1$. 假設

$$
(Y_1, Y_2) \sim f(y_1, y_2) = \begin{cases} 3y_1, & 0 \le y_2 \le y_1 \le 1 \\ 0, & \text{其它} \end{cases}
$$

亦即, 如圖示, $f(y_1, y_2) \geq 0$, 且

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 \\
&= \int\int_R f(y_1, y_2) dy_1 dy_2
\end{aligned}$$

其中區域

$$R : \begin{cases} 0 \leq y_1 \leq 1 \\ 0 \leq y_2 \leq y_1 \end{cases}$$

接著, 根據 Fubini 定理以及區域 R 的表示法, 上式的二元 (雙變數) 積分等於如下的重積分, 亦即,

$$\begin{aligned}
&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(y_1, y_2) dy_1 dy_2 \\
&= \int_0^1 \int_0^{y_1} 3y_1 dy_2 dy_1 \\
&= \int_0^1 \left[3y_1 y_2 \Big|_{y_2=0}^{y_1} \right] dy_1 \\
&= \int_0^1 3y_1^2 dy_1 \\
&= y_1^3 \Big|_0^1 = 1
\end{aligned}$$

故 $f(y_1, y_2)$ 確實爲一個 joint pdf. 試求庫存小於容器的一半且銷售量超過容器 1/4 的機率.

<解> 根據題意,

$$所求 = P(Y_1 < 1/2,\ 1/4 < Y_2)$$

再根據 (Y_1, Y_2) 落在一區域內的機率等於其 joint pdf 在此區域上的積分以及 Fubini 定理, 上式相當於

$$\text{所求} = \int_{-\infty}^{1/2}\int_{1/4}^{\infty} f(y_1, y_2)dy_2dy_1$$

又根據圖示, $f(y_1, y_2)$ 僅在區域

$$R: \begin{cases} 1/4 < y_1 < 1/2 \\ 1/4 < y_2 < y_1 \end{cases}$$

上爲正且取值 $3y_1$, 故得

$$\begin{aligned}
\text{所求} &= \int_{1/4}^{1/2}\int_{1/4}^{y_1} 3y_1dy_2dy_1 \\
&= \int_{1/4}^{1/2}\left[3y_1y_2\Big|_{y_2=1/4}^{y_1}\right]dy_1 \\
&= \int_{1/4}^{1/2}\left(3y_1^2 - \frac{3}{4}y_1\right)dy_1 \\
&= y_1^3 - \frac{3}{8}y_1^2\Big|_{1/4}^{1/2} \\
&= \left(\frac{1}{8} - \frac{3}{32}\right) - \left(\frac{1}{64} - \frac{3}{128}\right) \\
&= \frac{16 - 12 - 2 + 3}{128} = \frac{5}{128}
\end{aligned}$$

註 5. 推廣至 n 變量.

(1) 離散情況: 聯合質量函數 (joint pmf)

$$\begin{aligned} & p(y_1, y_2, \ldots, y_n) \\ = \ & P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n) \end{aligned}$$

聯合累積分布函數 (joint cdf)

$$\begin{aligned} & F(y_1, y_2, \ldots, y_n) \\ = \ & P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n) \\ = \ & \sum_{t_1=-\infty}^{y_1} \sum_{t_2=-\infty}^{y_2} \cdots \sum_{t_n=-\infty}^{y_n} p(t_1, t_2, \ldots, t_n) \end{aligned}$$

並以

$$(Y_1, Y_2, \ldots, Y_n) \sim p(y_1, y_2, \ldots, y_n)$$

或

$$(Y_1, Y_2, \ldots, Y_n) \sim F(y_1, y_2, \ldots, y_n)$$

表示之.

(2) 連續情況: 以聯合密度函數 $f(y_1, y_2, \ldots, y_n)$ 描述 $(Y_1, Y_2, \ldots, Y_n)$ 在點 $(y_1, y_2, \ldots, y_n)$ 附近的機率, 並且滿足

$$(Y_1, Y_2, \ldots, Y_n) \sim f(y_1, y_2, \ldots, y_n)$$

若且爲若聯合累積分布函數 (joint cdf)

$$
\begin{aligned}
& F(y_1, y_2, \ldots, y_n) \\
&\quad = P(Y_1 \le y_1, Y_2 \le y_2, \ldots, Y_n \le y_n) \\
&\quad = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f(t_1, t_2, \ldots, t_n) \\
&\qquad\qquad dt_n \cdots dt_2 dt_1
\end{aligned}
$$