

單元 24: 連續隨機變數的機率分布 (課本 §4.1, 4.2)

非正式定義: 一隨機變數的取值可為一區間中的任一值時, 稱其為連續型隨機變數, 如降雨量 $\in (0, 5)$ 英吋, 機器壽命 $\in (3, 7)$ 年, 並非意謂著: 觀察夠久後至終會觀察到區間中任一值所對應的結果; 而是說區間中沒有一值可被排除在可能值之外. 另外, 連續隨機變數的機率分布需由與離散隨機變數不同的方法發展出.

給出連續隨機變數的正式定義之前, 先定義任意隨機變數的分布函數 (distribution function), 又稱作累積分布函數 (cumulative distribution function).

定義 1. 令 Y 為任一隨機變數, Y 的分布函數 (distribution function, 又稱作累積分布函數, cumulative distribution function, 簡稱作 cdf)

$$F(y) \stackrel{\text{def}}{=} P(Y \leq y), -\infty < y < \infty$$

註. 一隨機變數的分布函數 (cdf) 為一定義在 $(-\infty, \infty)$ 上且值域為 $[0, 1]$ 的函數並且其特性決定出此隨機變數為離散或連續或其它.

例 1. 設 $Y \sim b(2, 1/2)$. 試求 $F(y)$.

<解> 根據 cdf 的定義, 需先求出 Y 的 pmf,

$$\begin{aligned} P(Y = y) &= \binom{2}{y} \left(\frac{1}{2}\right)^y \left(\frac{1}{2}\right)^{2-y} \\ &= \begin{cases} \frac{1}{4}, & y = 0 \\ \frac{1}{2}, & y = 1 \\ \frac{1}{4}, & y = 2 \end{cases} \end{aligned}$$

故,

$$\begin{aligned} F(y) &= P(Y \leq y) \\ &= \begin{cases} P(\emptyset) = 0 & y < 0 \\ P(Y = 0) = \frac{1}{4} & 0 \leq y < 1 \\ P(Y = 0) + P(Y = 1) = \frac{3}{4} & 1 \leq y < 2 \\ P(Y = 0) + P(Y = 1) + P(Y = 2) = 1 & 2 \leq y < \infty \end{cases} \end{aligned}$$

且圖形如下.

稱作階梯函數 (step function): 在可能值之間是水平的 (flat), 在每一個可能值都有一個跳躍 (jump), 其“高度” = “產生此可能值的機率”.

註. 離散隨機變數的累積分布函數 (cdf) 均為階梯函數且僅在可數個可能點上跳躍.

定理 4.1. 分布函數 (cdf) 的性質: 設 $F(y)$ 爲一分布函數, 則

(1) 在負無窮遠的值 (極限)

$$F(-\infty) \equiv \lim_{y \rightarrow -\infty} F(y) = 0$$

因爲 $Y \in (-\infty, \infty)$, 故根據 cdf 的定義,

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

(2) 在正無窮遠的值 (極限)

$$F(\infty) \equiv \lim_{y \rightarrow \infty} F(y) = 1$$

因爲 $Y \in (-\infty, \infty)$, 故根據 cdf 的定義,

$$\begin{aligned} F(\infty) &= P(Y \leq \infty) \\ &= P(S) \text{ (恆發生)} \\ &= 1 \end{aligned}$$

(3) $F(y)$: 非遞減 (nondecreasing), 亦即, 對於任意的 $y_1 < y_2$, 則

$$F(y_1) \leq F(y_2)$$

因爲 $Y \leq y_1$ 恆保證 $Y \leq y_2$, 故

$$(Y \leq y_1) \subseteq (Y \leq y_2)$$

因此,

$$F(y_1) = P(Y \leq y_1) \leq P(Y \leq y_2) = F(y_2)$$

(4) $F(y)$: 右連續 (right continuous), 亦即, 對任一 c ,

$$\lim_{y \rightarrow c^+} F(y) = F(c)$$

證明, 略. 舉例說明, 如例 1, $F(y)$ 在 $y = 0, 1, 2$ 雖不連續但卻是右連續; 在其他點上均連續, 故也是右連續. 因此, 對任一 c , $F(y)$ 在其上都是右連續.

定義 2. 稱隨機變數 Y 爲連續的, 若其分布函數 $F(y)$ 在 $(-\infty, \infty)$ 上是連續的.

註 1. 嚴格的定義: 還要再加上的條件如下, 在任意有限區間內, 除了最多有限個點外, $F'(y)$ 是存在且連續的.

註 2. 設 Y 爲一連續隨機變數, 則對任一實數 y ,

$$P(Y = y) = 0$$

此乃因為若存在 y_0 使得

$$P(Y = y_0) = p_0 > 0$$

則 $F(y)$ 在 y_0 會有一高度為 p_0 的不連續的跳躍 (jump) 而與定義矛盾.

定義 3. 設連續隨機變數 Y 的 cdf 為 $F(y)$, 則

$$f(y) \stackrel{\text{def}}{=} \frac{dF(y)}{dy} = F'(y)$$

當導函數存在時, 並稱其為 Y 的機率密度函數 (probability density function, 簡稱 pdf).

註 1. 由微積分基本定理, cdf $F(y)$ 與 pdf $f(y)$ 之間的關係如下:

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

並如圖所示.

註 2. 機率密度函數 (pdf) 是度量值母體的相對頻率的理論模型.

註 3. 描述離散與連續隨機變數不確定性的 pmf 與 pdf 之間的差異, 如下述.

(1) Y : 離散的,

$$p(y) = P(Y = y)$$

稱作機率質量函數 (pmf), 爲 $Y = y$ 的機率.

(2) Y : 連續的,

$$P(Y = y) \equiv 0 \neq \text{pdf } f(y)$$

但當 $\Delta y \downarrow 0$,

$$\begin{aligned} & P(y < Y \leq y + \Delta y) \\ &= P(Y \leq y + \Delta y) - P(Y \leq y) \\ &= F(y + \Delta y) - F(y) \\ &= \int_{-\infty}^{y+\Delta y} f(t)dt - \int_{-\infty}^y f(t)dt \quad (\text{根據註 1}) \\ &= \int_y^{y+\Delta y} f(t)dt \\ &\approx \int_y^{y+\Delta y} f(y)dt \\ &= f(y)\Delta y \end{aligned}$$

此乃說明, 當 $\Delta y \downarrow 0$ 時, Y 在 y 附近的機率與 $f(y)$ 有密切關係, 亦即,

$$\approx f(y) \cdot \text{附近長度}$$

定理 4.2. (機率密度函數 pdf 的性質). 設 $f(y)$ 爲一 pdf, 則

(1) 對任意 y 值,

$$f(y) \geq 0$$

因爲 $F(y)$ 爲非遞減, 故

$$F'(y) = f(y) \geq 0$$

(2) 在整個實數線上的積分爲 1, 亦即,

$$\int_{-\infty}^{\infty} f(y) dy = 1$$

因爲根據 cdf 的性質及註 1,

$$1 = F(\infty) = \int_{-\infty}^{\infty} f(y) dy$$

例2. 設隨機變數 Y 的 cdf

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

試求 Y 的 pdf 並繪其圖.

<解> 根據定義, pdf

$$\begin{aligned} f(y) &= F'(y), \text{ 當導函數存在時} \\ &= \begin{cases} \frac{d}{dy}(0) = 0, & y < 0 \\ \frac{d}{dy}(y) = 1, & 0 < y < 1 \\ \frac{d}{dy}(1) = 0, & y > 1 \end{cases} \end{aligned}$$

且 $f(y)$ 在 $y = 0$ 與 1 未定義, 圖形如下.

例 3. 設

$$Y \sim f(y) = \begin{cases} 3y^2, & 0 \leq y \leq 1 \\ 0, & \text{其它 (elsewhere)} \end{cases}$$

試求 $F(y)$ 並繪其圖.

<解> 根據註 1,

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

接著根據 $f(y)$ 的定義, 需要分段積分如下. 首先, 當 $y < 0$,

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

當 $0 \leq y \leq 1$ 時,

$$\begin{aligned} F(y) &= \int_{-\infty}^0 0 dt + \int_0^y 3t^2 dt \\ &= 0 + t^3 \Big|_0^y = y^3 \end{aligned}$$

最後, 當 $y > 1$,

$$\begin{aligned} F(y) &= \int_{-\infty}^0 0dt + \int_0^1 3t^2 dt + \int_1^y 0dt \\ &= 0 + t^3 \Big|_0^1 + 0 = 1 \end{aligned}$$

因此, 得

$$F(y) = \begin{cases} 0, & y < 0 \\ y^3, & 0 \leq y \leq 1 \\ 1, & 1 < y \end{cases}$$

且圖形如下.

定理 4.3. 設 $Y \sim f(y)$ 且 $a < b$, 則

$$P(a \leq Y \leq b) = \int_a^b f(y) dy$$

亦即, 在 $[a, b]$ 上 $f(y)$ 所圍出區域的面積, 如圖所示.

<證> 首先, 根據互斥事件的性質,

$$P(a \leq Y \leq b) = P(Y = a) + P(a < Y \leq b)$$

因為 Y 為連續隨機變數, 故 $P(Y = a) = 0$, 而得

$$P(a \leq Y \leq b) = P(a < Y \leq b) \quad (1)$$

又因為

$$(a < Y \leq b) = (Y \leq b) \setminus (Y \leq a)$$

故, 由 (1) 式, 得

$$\begin{aligned}P(a \leq Y \leq b) &= P(Y \leq b) - P(Y \leq a) \\&= F(b) - F(a) \\&= \int_{-\infty}^b f(y)dy - \int_{-\infty}^a f(y)dy \\&= \int_a^b f(y)dy\end{aligned}$$

註. 若 Y 為連續且 $a < b$, 則

$$P(Y = a) = P(Y = b) = 0$$

故,

$$\begin{aligned}P(a < Y < b) &= P(a \leq Y < b) \\&= P(a < Y \leq b) \\&= P(a \leq Y \leq b) \\&= \int_a^b f(y)dy\end{aligned}$$

注意: 若 Y 為離散, 則, 一般而言, 上式是不成立的.

例 4. 給定

$$f(y) = \begin{cases} cy^2, & 0 \leq y \leq 2 \\ 0, & \text{o.w. (otherwise)} \end{cases}$$

(a) 試求 c 使得 $f(y)$ 爲一合格的 pdf.

(b) 試求 $P(1 \leq Y \leq 2)$ 與 $P(1 < Y < 2)$.

<解> (a) 根據 pdf 的性質 (定理 4.2), c 需要滿足

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(y) dy \\ &= \int_0^2 cy^2 dy \\ &= \left. \frac{c}{3} y^3 \right|_0^2 \\ &= \frac{8}{3} c \end{aligned}$$

所以,

$$c = \frac{3}{8}$$

(b) 因爲 Y 爲連續的, 故

$$\begin{aligned} P(1 < Y < 2) &= P(1 \leq Y \leq 2) \\ &= \int_1^2 \frac{3}{8} y^2 dy \\ &= \left. \frac{1}{8} y^3 \right|_1^2 \\ &= \frac{7}{8} \end{aligned}$$