定積分的機率意涵

單維彰・2015 年 5 月

我們說 x 為**隨機變數**的意思是，在一個隨機試驗當中，它賦予樣本空間 S 裡的每個樣本點一個實數值。x 之所有可能的值所成的集合，稱為該隨機變數的值域。如果值域是離散的，例如 $\{0, 1\}$ 或者正整數的子集合，則稱 x 為**離散型隨機變數**。若 x 之值域為實數的子集合，則稱 x 為**連續型隨機變數**。當隨機變數用來表示某隨機試驗的整體時，常用大寫字母表示，例如 X；而特定樣本的值，則用對應的小寫字母，例如 x。

一個定義在實數上的函數 f 稱為**機率密度函數** (pdf) 的意思是，它符合以下兩個條件：

(1) $f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

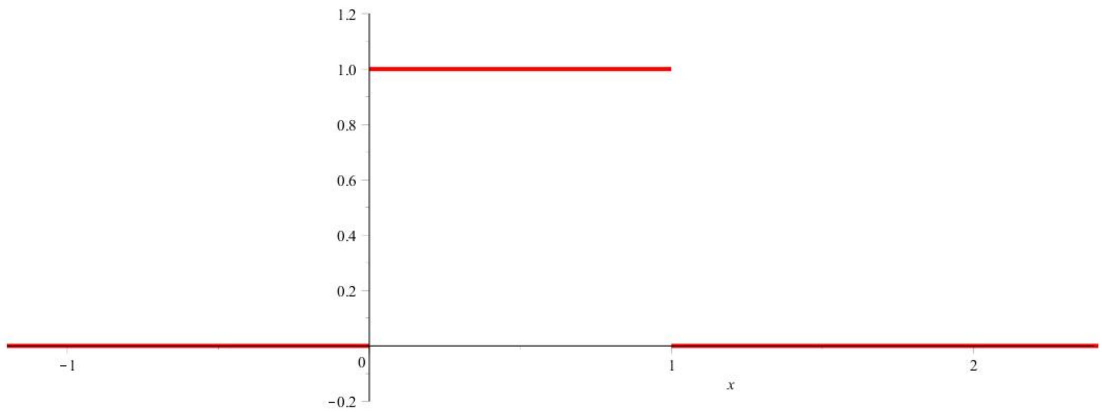
而對於隨機變數 x，f 在 $a \leq x \leq b$ 範圍內的定積分（曲線下面積）是隨機試驗的結果落在 $a \leq x \leq b$ 範圍內的機率。也就是

$$P(a \leq x \leq b) = \int_a^b f(x)\,dx$$

習慣上，機率密度函數的定義域為 $\mathbb{R}$，但也可以只定義在一個區間內，反正只有 $f(x) > 0$ 的部分才對隨機變數的機率發生影響。例如在閉區間 $[0,1]$ 內均勻分佈之 pdf 為

$$f(x) = \begin{cases} 1, & 若\ 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

如下圖。



因為 $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 f(x)\,dx = 1$，很容易檢查 f 確實是一個機率密度函數。

連續型隨機變數相對於離散型的重要差異是：任一個實數的發生機率皆為 0。因為

$$P(x=a)=P(a\le x\le a)=\int_a^a f(x)\,dx=F(a)-F(a)=0$$

所以

$$\text{對任意實數 } a\text{，}\ P(x=a)=0$$

因此，任何一個區間的端點，對連續型隨機變數的機率沒有影響。機率密度函數的定積分$\int_a^b f(x)\,dx$雖然被定義為$P(a\le x\le b)$的意思，但事實上

$$P(a\le x<b)\qquad P(a<x\le b)\qquad P(a<x<b)$$

其實全都一樣。

最後要提醒的是，無上界或無下界範圍的機率，就得做「廣義積分」:

$$P(x\ge a)=\int_a^{\infty} f(x)\,dx \quad \text{與} \quad P(x\le a)=\int_{-\infty}^{a} f(x)\,dx$$

而對單一數值的機率，則要以某個區間來取代。例如$x=3$的意義應該將誤差的觀念帶進來，例如以一般習慣的四捨五入而言，連續型隨機變數 x 的值為 3 的機率是

$$P(2.5\le x<3.5)$$