CHAPTER 5

# 機率分配

版權所有

## 章前導讀

第4章所討論的樣本空間，其所包含的元素有些只是定性的描述而非實際的數值。例如，擲一硬幣兩次，可能之結果（樣本點）為 $HH$、$HT$、$TH$ 與 $TT$，皆為一組字母，用以描述所出現的為正面或反面；這些都只是定性的結果，而非一實際之數字。有時候我們所關心的是，將隨機試驗所產生之可能結果，賦予一有意義的數值，如此便可針對數字資料做一些統計分析。基於此一理念，引發我們對隨機試驗之可能結果予以量化，此即隨機變數的觀念。本章將探討隨機變數、機率分配，以及一些相關的課題。基本上，這些都是推論統計之重要的基礎。

華泰文化
HWA TAI PUBLISHING

# 5.1 隨機變數

對一隨機試驗所產生的樣本空間定義一函數，則此樣本空間內的每一元素（樣本點）皆有一函數值與之對應，此一函數即稱為**隨機變數** (random variable)。亦即，隨機變數是將實驗結果以數值來表示。

**隨機變數**

係以樣本空間為定義域的實數值函數。

隨機變數一般皆以大寫字母表示，如 $X$, $Y$, $Z$ 等，而其函數值則以小寫字母表示，如 $x$, $y$, $z$ 等。假定擲一硬幣兩次，則其樣本空間 $S = \{HH, HT, TH, TT\}$；若定義隨機變數 $X$ 代表正面的個數，則 $X$ 之可能值（函數值）為 $x = 0, 1, 2$，其與樣本點之對應關係如圖 5.1 所示：

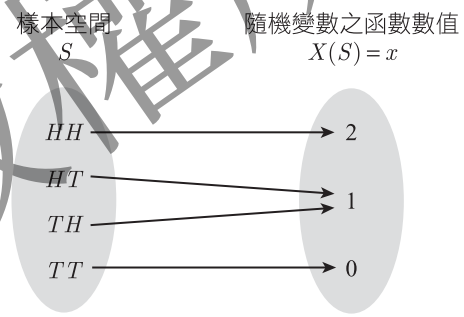



圖 5.1　隨機變數與樣本空間的對應關係

圖 5.1 中 $X(S)$ 乃表示隨機變數 $X$ 是一函數，且以樣本空間 $S$ 為定義域；故 $X(S)$ 即為值域。又，隨機變數之每一可能值實際上皆代表一事件，例如，$x = 2$ 表示事件 $\{HH\}$，$x = 1$ 表示事件 $\{HT\}$ 或 $\{TH\}$，而 $x = 0$ 表示事件 $\{TT\}$。

隨機變數中的「隨機」一詞乃提醒讀者一個事實，即我們事前無法確切地知道隨機試驗會產生哪一種可能結果，而且也不知道與這些結果對應的 $X$ 值。

版權所有
華泰文化
HWA TAI PUBLISHING

## 例題5.1

試寫出下列隨機變數之可能值。

(a) $X$ 代表擲一硬幣 3 次，出現正面之次數。

(b) 一箱子中裝有 10 個球，其中有 4 個紅球，6 個白球。採不放回抽樣，連續抽出 5 個球，令 $Y$ 代表可能出現之紅球個數。

(c) $Z$ 代表某十字路口下個月內發生車禍的次數。

(d) $T$ 代表某一品牌日光燈管的壽命長度。

**解**

(a) $X$ 的可能值為 $x = 0, 1, 2, 3$。

(b) $Y$ 的可能值為 $y = 0, 1, 2, 3, 4$。

(c) $Z$ 的可能值為 $z = 0, 1, 2, \cdots$。

(d) $T$ 的可能值為 $t \geq 0$。

一隨機變數如果其可能值的個數為有限的，或雖無限但為可計數的(countable)，則稱之為**間斷型**或**離散型**(discrete)隨機變數；反之，若一隨機變數其可能值的個數為無限且為不可計數的，亦即係以連續測度者，則稱為**連續型**(continuous)隨機變數。換句話說，若一隨機變數的任二個數值中間做切割後仍有意義，則此變數為連續型，若切割後變得沒有意義，則為間斷型。在例題 5.1 中，(a)、(b)、(c)皆屬於間斷隨機變數，而(d)則為連續隨機變數。在一般的實務應用上，通常連續隨機變數是表示**測量的**(measurable)資料，如身高、體重、溫度與時間等；而間斷隨機變數則表示**計數的**(countable)資料，如不良品的個數、意外事件次數、班級人數等。

## 隨堂練習

1. 假設由 $\{1, 3, 5, 7, 9\}$ 隨機取出二數，令 $X$ 為此二數之和，試完成下列工作。
   (a) 列出全部可能結果，且每一可能結果之 $X$ 值為何？
   (b) 列出全部的 $X$ 值。
2. 在一次隨堂測驗中共有 3 道選擇題：第一題為 3 選 1，第二題為 4 選 1，第三題為 2 選 1。茲有某位學生因考前毫無準備，故決定以隨機方式猜題，令 $X$ 表此學生猜對之題數：
   (a) 列出 $X$ 之全部可能值。
   (b) 求出 $X$ 之機率分配。
   (c) 計算 $P(X \geq 1)$。

版權所有
華泰文化
HWA TAI PUBLISHING

## 5.2 機率分配

隨機變數係將樣本空間所產生的一些基本事件予以數值化的定義，若進一步求得這些事件數值化後的隨機變數之機率，此即**機率分配**(probability distribution)的觀念。也就是說，機率分配乃是針對某些隨機變數之可能值或某一範圍之隨機變數的可能值，求其機率。機率分配本質上亦是一函數，稱為**機率函數**(probability function)，其定義域為隨機變數之可能值，而值域則為相對應的機率值。圖 5.2 繪示了樣本空間、隨機變數與機率分配三者間的關係。

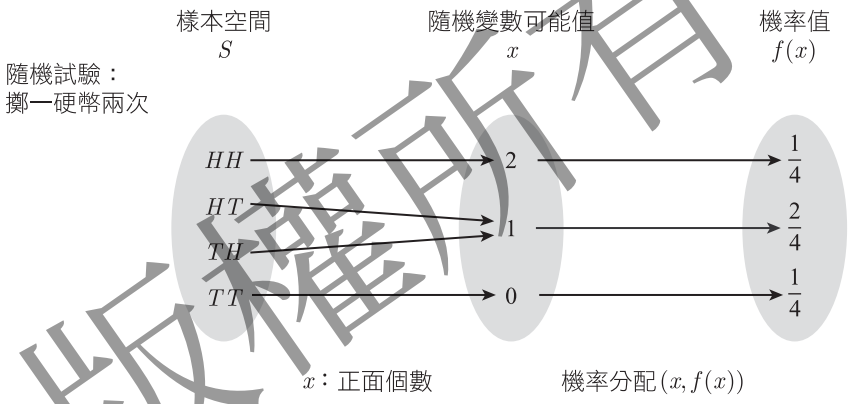



圖 5.2 樣本空間、隨機變數與機率分配三者間的關係

由於隨機變數分為間斷型與連續型，故機率分配亦可分為間斷機率分配與連續機率分配。本節將分別探討此二類型的機率分配，並說明其重要的性質。

### 5.2.1 間斷機率分配

將一間斷隨機變數 $X$ 之所有可能值一一列出，然後計算每一可能值 $x$ 的機率，則可構成間斷機率分配。

**間斷機率分配之意義**

間斷隨機變數 $X$ 之機率分配，指將各個不同的 $x$ 值與其對應之機率值逐一列出的表，有時亦可以公式來取代其詳細的表列。

版權所有
華泰文化
HWA TAI PUBLISHING

## 例題5.2

令 $X$ 代表擲一公正硬幣3次中出現正面之次數，試求 $X$ 之機率分配。

**解**

擲一公正硬幣3次，其樣本空間 $S$ 與隨機變數之可能值 $x$，列示如下表：

| 樣本空間 $S$ | 隨機變數可能值 $x$ | $x$ 之機率值 $f(x)$ |
|---|---|---|
| $HHH$ | 3 | 1/8 |
| $HHT$ | 2 | 1/8 |
| $HTH$ | 2 | 1/8 |
| $THH$ | 2 | 1/8 |
| $HTT$ | 1 | 1/8 |
| $THT$ | 1 | 1/8 |
| $TTH$ | 1 | 1/8 |
| $TTT$ | 0 | 1/8 |

由於樣本空間之每一樣本點出現的機率皆相同，故每一簡單事件的機率皆為1/8。又 $x = 3$ 出現一次，故其對應的機率值為1/8，而 $x = 2, 1$，各出現3次，故其機率值皆為3/8；同理 $x = 0$ 的機率值為1/8。最後，將 $X$ 之機率分配彙總列於表5.1。

表 5.1 擲一公正硬幣三次，出現正面次數 $X$ 之機率分配

| | $x$ | | | | 總計 |
|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | |
| 機率 | $\frac{1}{8}$ | $\frac{3}{8}$ | $\frac{3}{8}$ | $\frac{1}{8}$ | 1 |

通常以 $x_i$ 代表隨機變數 $X$ 之值，而以 $f(x_i)$ 表示其機率值。例如，上例中 $x_i = 0, 1, 2, 3$，而其對應的機率值為 $f(0), f(1), f(2), f(3)$。$f(x)$ 一般又可以 $f(x) = P(X = x)$ 表示，例如 $f(2) = P(X = 2)$。由於每一 $f(x_i)$ 皆代表機率值，故它們必介於0與1之間，且若將所有可能之 $X$ 值所對應的機率值加總，其和必等於1。

版權所有
華泰文化
HWA TAI PUBLISHING

一間斷隨機變數 $X$ 之機率分配可以函數表示，即：

$$f(x_i) = P(X = x_i)$$

其中 $f(x_i)$ 必須滿足：

(1) 對於 $X$ 中的每一 $x_i$，$0 \le f(x_i) \le 1$。

(2) $\sum_{i=1}^{k} f(x_i) = 1$，$X$ 之可能值為 $x_i, i = 1, 2, \cdots, k$。

機率分配除了可用列表與公式的方式表示外，尚可用圖形來表示，稱為機率分配圖。較為常用的機率分配圖有**線圖** (line chart) 與**機率直方圖** (probability histogram)。圖 5.3 與圖 5.4 乃例題 5.2 之機率分配圖。

線圖是以高度表示機率，所有高度的總和等於 1。而機率直方圖則是以面積表示機率，圖中矩形面積之總和等於 1，這與 2.4.1 節中的相對次數直方圖的觀念相同。

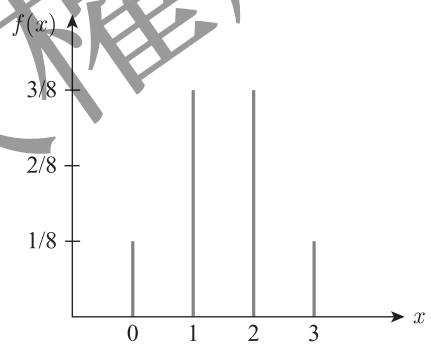



圖 5.3　例題 5.2 之機率分配線圖

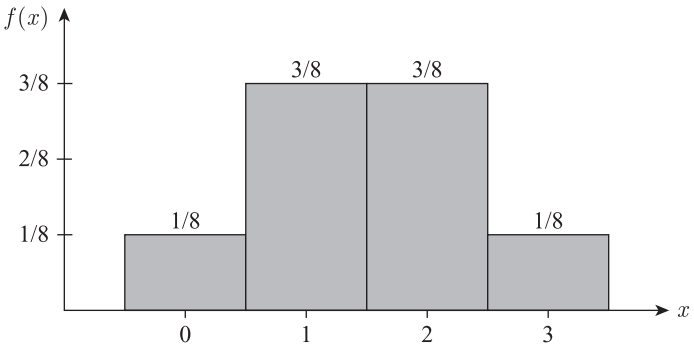



圖 5.4　例題 5.2 之機率直方圖

版權所有
華泰文化
HWA TAI PUBLISHING

## 例題5.3

設有一批產品，產品之不良率為 0.3，現從此批產品隨機抽出 4 個加以檢驗。令 $X$ 表示所檢驗產品中不良品的個數，試求 $X$ 之機率分配，並繪出其圖形。

**解**

由於檢驗每一產品其可能結果為良品 $(G)$ 或不良品 $(D)$，故檢驗 4 個產品之可能情形共有 $2^4 = 16$ 種（可利用樹狀圖一一列出），茲以下表列示各種可能結果及其所對應的 $X$ 值：

| $x=0$ | $x=1$ | $x=2$ | $x=3$ | $x=4$ |
|---|---|---|---|---|
| $GGGG$ | $GGGD$ | $GGDD$ | $GDDD$ | $DDDD$ |
| | $GGDG$ | $GDGD$ | $DGDD$ | |
| | $GDGG$ | $GDDG$ | $DDGD$ | |
| | $DGGG$ | $DGGD$ | $DDDG$ | |
| | | $DGDG$ | | |
| | | $DDGG$ | | |

註：$x$ 表不良品個數、$G$ 表良品、$D$ 表不良品。

接著計算各個 $X$ 之可能值的機率。例如，$x=0$ 表示 4 個產品皆為良品，而 $P(G) = 0.7$, $P(D) = 0.3$，且檢驗每一產品之結果為獨立事件，因此，

$$f(0) = P(X=0) = P(GGGG) = 0.7 \times 0.7 \times 0.7 \times 0.7 = 0.2401$$

同理，$x=1$ 時，包含 4 個簡單事件，且每一簡單事件皆含有 3 個良品 $(G)$ 與 1 個不良品 $(D)$，故 $P(GGGD) = (0.7)^3 \times 0.3 = 0.1029$，於是 $f(1) = P(X=1) = 4 \times 0.1029 = 0.4116$。其餘 $X$ 值之機率計算如下：

$$\begin{aligned} f(2) &= P(X=2) = 6 \times (0.7)^2 \times (0.3)^2 \\ &= 0.2646 \\ f(3) &= P(X=3) = 4 \times (0.7) \times (0.3)^3 \\ &= 0.0756 \\ f(4) &= P(X=4) = (0.3)^4 \\ &= 0.0081 \end{aligned}$$

將上述所得結果彙總如表 5.2，並繪出其機率直方圖，如圖 5.5 所示。

版權所有
華泰文化
HWA TAI PUBLISHING

表 5.2　例題 5.3 之機率分配

| $x$ | $f(x)$ |
|---|---|
| 0 | 0.2401 |
| 1 | 0.4116 |
| 2 | 0.2646 |
| 3 | 0.0756 |
| 4 | 0.0081 |
| 總計 | 1.0000 |

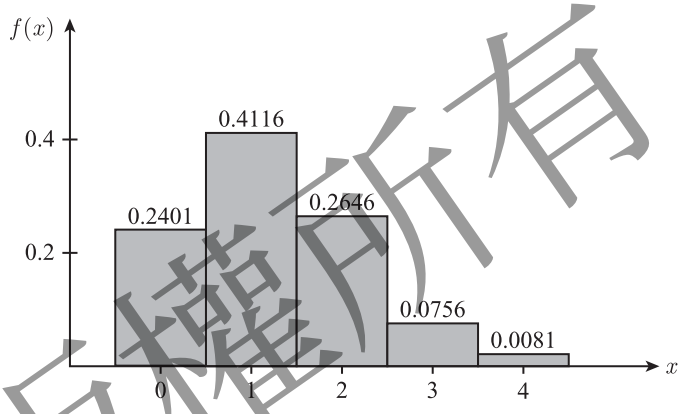



圖 5.5　例題 5.3 之機率直方圖

版權所有

前一章曾述及以相對次數求算機率的觀念，在本章計算機率分配之機率值亦可利用此一方法。此外，相對次數的累積稱為累積相對次數，而機率分配的累積則稱為**累積機率分配**(cumulative probability)，其符號一般以 $F(x)$ 表示。$F(x)$ 的意義乃是自隨機變數 $X$ 之最小的可能值，一直累加至 $x$，亦即：

$$F(x) = \sum_{\substack{i \le x \\ i \in X(S)}}^{x} f(x_i) \tag{5-1}$$

式中 $X(S)$ 表示 $X$ 之可能值的值域。又，若 $a < b$（$a$、$b$為任意實數），則 $X$ 之可能值由 $a$ 至 $b$ 的累積機率可表為 $P(a \le X \le b)$，且

$$\begin{aligned} P(a \le X \le b) &= \sum_{\substack{a \le i \le b \\ i \in X(S)}} f(x_i) \\ &= F(b) - F(a) + f(a) \end{aligned} \tag{5-2}$$

華泰文化
HWA TAI PUBLISHING

請注意，對間斷機率分配而言，$P(X < a)$ 並不等於 $P(X \leq a)$；前者不包含機率值 $f(a)$，亦即：

$$F(a) = P(X \leq a) = P(X < a) + f(a)$$

**例題5.4**

令 $X$ 表示某學校400名教師出版教科書冊數，假設其次數分配如下表所示：

| 出版教科書冊數 $x$ | 人數 | 相對人數 |
|---|---|---|
| 0 | 61 | 0.15 |
| 1 | 153 | 0.38 |
| 2 | 106 | 0.27 |
| 3 | 56 | 0.14 |
| 4 | 24 | 0.06 |
| 總計 | 400 | 1.00 |

試求 $X$ 之機率分配，並計算 $P(X > 1)$ 與 $P(X \leq 2)$。

**解**

將表中的相對人數視為機率之估計值，則可求得 $X$ 之近似機率分配。事實上，當總人數愈多時，則此機率估計值愈接近真正的機率。茲將 $X$ 之機率分配列示如下：

| | $x$ | | | | | 總計 |
|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | |
| $f(x)$ | 0.15 | 0.38 | 0.27 | 0.14 | 0.06 | 1.00 |

接著，計算 $P(X > 1)$與$P(X \leq 2)$ 如下：

$$\begin{aligned} P(X > 1) &= f(2) + f(3) + f(4) \\ &= 1 - [f(0) + f(1)] \\ &= 1 - (0.15 + 0.38) = 0.47 \\ P(X \leq 2) &= f(0) + f(1) + f(2) \\ &= 0.15 + 0.38 + 0.27 = 0.80 \end{aligned}$$

版權所有
華泰文化
HWA TAI PUBLISHING

## 5.2.2 連續機率分配

根據第 4 章，機率有時可視為長期相對次數的量度，而連續機率分配之觀念則亦可由一組包含很多觀測值之相對次數的直方圖所形成。在相對次數直方圖中，當資料之觀測值個數增加，且直方圖之組距間隔縮小，則在間隔無限縮小的情況下，即可構成連續機率分配的直方圖。茲舉一例說明之：令 $X$ 表初生嬰兒之體重，則 $X$ 為連續隨機變數，此時如何建立 $X$ 之機率分配？首先，假定有 100 個初生嬰兒體重的資料，而此資料乃以 1 磅為組距，其相對次數直方圖如圖 5.6(a) 所示。根據 2.4 節的說明，相對次數直方圖具有下列的特性：

(1) 直方圖之面積總和為 1。
(2) 就某些組的邊界點 $a$ 與 $b$ 而言，介於 $a$ 與 $b$ 之相對次數為該區間內，直方圖所表示的面積。

譬如，圖 5.6(a) 中，7.5～9.5 磅之區間內，包含有 53 個初生嬰兒，則其相對次數（即機率）為 $53/100 = 0.53$。

其次，假設我們調查之初生嬰兒增為 500 個，且組距縮小為 0.25 磅，其相對次數直方圖繪示如圖 5.6(b)，此可視為圖 5.6(a) 之精製圖形（請注意：如果只縮短組距而不增加觀測值，則整個分配之形狀將會改變）。圖 5.6(b) 仍具有上述直方圖之特性 (1) 與 (2)。

依此方式，進一步將圖 5.6(b) 再予精密細分，則其相對次數直方圖為觀測值個數非常多，且各組組距更為縮小。在這種不斷精密細分直方圖的過程中，兩兩相隔的長方形（直方）將逐漸消失，而原先直方圖之上端將變成一平滑曲線 (smoothed curve)，如圖 5.6(c) 所示。由於機率可解釋為長期之相對次數，故此曲線乃代表相對次數直方圖之極限形式，其中總機率為 1，並分配至每一隨機變數 $X$ 之值。圖 5.6(c) 中的曲線稱為連續隨機變數 $X$ 之**機率密度曲線** (probability density curve)，其數學式 $f(x)$ 稱為連續隨機變數 $X$ 之**機率密度函數** (probability density function, *pdf*)。

版權所有
華泰文化
HWA TAI PUBLISHING