4

機　率

4.1 導論

4.2 隨機試驗、樣本空間與事件

4.3 機率測度的方法

4.4 機率的性質

4.5 條件機率與獨立事件

4.6 貝氏定理

本書第1章即已指出統計學的範圍包括敘述統計與推論統計，而第2、3章則專門介紹敘述統計的觀念與方法。從本章開始，我們將逐漸深入推論統計的領域，然而在探討推論統計之前，有一些必備的課題須先加以了解，此即**機率理論**(probability theory)。第4、5、6章我們將專注於機率相關的課題之介紹，包括機率的基本概念、隨機變數與機率分配，以及一些常用的機率分配。

著作權所有 侵害必究

華泰文化
HWA TAI PUBLISHING

4.1 導　論

機率理論是在探討「如何由一已知的母體特徵去推斷一未知的樣本特徵」。例如，假定某校學生（母體）的近視比例為 60%，則我們似乎不可能以隨機抽樣的方式，抽出近視比例剛好為 60%的樣本。但是抽出比例為**可能地** (likely)、**接近於** (close to) 60%的樣本卻是有可能的。機率理論即在於更精確地定義「可能地」與「接近於」的隱含意義，並以具體的數字來衡量此抽象的意義。更具體地說：

> 機率是衡量某一不確定事件可能發生的程度（機會大小），並針對此一不確定事件發生之可能性賦予一量化的數值。

日常生活中，我們常聽到下列的用語：

「本週末可能會下雨，因此郊遊活動很可能要取消」；

「明天某公司的股票可能會漲」；

「今天的棒球賽，甲隊很不可能贏乙隊」。

「很可能」、「可能」、「很不可能」等用語，基本上都表示了不確定性程度的大小，而機率便是利用一數值來說明事件將發生之可能性，其方式是：當一事件極不可能發生時，其機率值接近於 0；相反地，當一事件極可能發生時，其機率值接近於 1。由此可知，機率是一個數值，用以測度某一事件之不確定的程度，其值介於 0 與 1 之間。

綜合言之，機率乃是統計推論的基礎，可藉以尋求更理想的推論方法，並藉機率值的大小及其觀念，以表示推論結果的準確性。此外，機率亦是統計決策理論的重要工具，因為決策往往是在不確定性的情況下判斷，而機率則專門在探討不確定性的問題，因此藉由機率理論，可提供決策抉擇的參考。例如：投資決策、保險公司估計投保人的保費、生產管理人員預測產品的需求量……等等，這些決策都需要根據事件可能發生的機率來決定。

4.2 隨機試驗、樣本空間與事件

機率是針對不確定事件發生之可能性賦予一量化的數值；在此所謂的不確定性是指我們知道某現象會有哪些結果出現，但究竟是哪一種結果出現，則不得而知。因此，我們欲探討機率，必先了解有哪些可能結果，以及這些可能結果之組合的觀念。以下我們介紹幾個名詞的意義，並由此引申出更清晰的機率概念。

著作權所有 侵害必究
華泰文化
HWA TAI PUBLISHING

首先我們說明**隨機試驗**(random experiment)的意義：

觀察一可產生各種可能結果(outcome)的過程，稱為試驗；而若各種可能結果的出現（或發生）具有不確定性，則此一過程便稱為隨機試驗。

試驗一詞的意義極為廣泛，舉凡觀察、實驗、調查、檢驗、抽樣等，皆可謂試驗。例如，擲一骰子、自一副撲克牌抽出一張、對一群顧客進行意見調查以及檢驗一批產品等，皆是試驗的例子。上述這些試驗的例子皆具有一共同的特性，即每次試驗時都可知道有哪些可能結果，但究竟哪一種結果會出現，則具有不確定性。例如，擲一個骰子，則其可能結果為 1、2、3、4、5 或 6，但出現 1 點、2 點或其他則每一次都不確定。此外，拋二枚硬幣、檢驗三件產品、討論明天股票行情……等等，由於各種可能結果的出現皆具有不確定性，因此皆屬於隨機試驗的例子。

嚴格來說，隨機試驗必須能滿足下列三個條件：

(1) 試驗可以在相同的條件下重複進行。

(2) 試驗的所有可能結果是明確可知的，且不止一個。

(3) 每次試驗必定僅出現這些可能結果中的一個，但試驗之前無法肯定該次試驗會出現哪一個可能結果。

一隨機試驗之各種可能結果的集合，稱為**樣本空間**(sample space)；樣本空間內的每一元素，稱為**樣本點**(sample point)。

樣本空間是依據隨機試驗可能發生的各種可能結果來界定的，它是以集合的觀念來定義，一般以符號 S 表示。例如：擲一骰子的樣本空間為 $S_1 = \{1, 2, 3, 4, 5, 6\}$；拋二枚硬幣的樣本空間為 $S_2 = \{(H、H), (H、T), (T、H), (T、T)\}$，其中 H 代表正面，T 代表反面。

例題 4.1

自一批產品隨機抽出三件加以檢驗，由於每次檢驗皆僅有二種可能結果，即良品 (G) 與不良品 (D)，且每次的可能結果皆不確定，因此它是隨機試驗。此隨機試驗的樣本空間可用樹狀圖的方式列舉出來，如圖 4.1 所示：

著作權所有 侵害必究
華泰文化
HWA TAI PUBLISHING

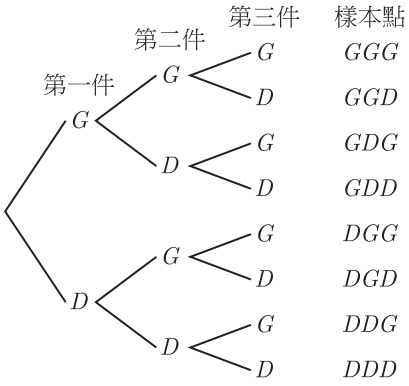


圖 4.1　檢驗三件產品之樣本空間的樹狀圖

圖 4.1 中，樣本點 GGG 代表三件產品皆為良品。樣本空間則為：

$$S = \{GGG, GGD, GDG, GDD, DGG, DGD, DDG, DDD\}$$

樣本空間為一集合，而依據集合所含元素的個數，可定義出**有限樣本空間** (finite sample space) 與**無限樣本空間** (infinite sample space)，亦即：

僅含有限個樣本點的樣本空間，稱為有限樣本空間；含有無限多個樣本點的樣本空間，稱為無限樣本空間。

有限樣本空間與無限樣本空間的例子，參見表 4.1：

表 4.1　有限樣本空間與無限樣本空間之舉例

隨機試驗	樣本空間	有限或無限
1. 擲一骰子	$S_1 = \{1, 2, 3, 4, 5, 6\}$	有限
2. 調查某班級近視人數（設有 45 名學生）	$S_2 = \{0, 1, 2, 3, \cdots, 45\}$	有限
3. 擲一硬幣兩次	$S_3 = \{HH, HT, TH, TT\}$	有限
4. 檢驗一批產品，直到發現有不良品為止，記錄檢查的產品數	$S_4 = \{1, 2, \cdots, n, \cdots\}$	無限
5. 觀察某一燈管之使用壽命	$S_5 = \{t \mid t > 0\}$ t 表壽命時間	無限

在計算隨機試驗可能出現的結果（樣本空間）時，除了可利用例題 4.1 所示的樹狀圖外，亦可利用乘數定理、排列 (permutation) 與組合 (combinations) 來求出樣本空間的個數與組合形式。

著作權所有　侵害必究
華泰文化
HWA TAI PUBLISHING

1. 乘數定理

設一個隨機試驗包含 k 個試驗 $E_1, E_2, \cdots, E_k$，若每一試驗 E_j 有 n_j 種結果，$j = 1, 2, \cdots, k$，則該隨機試驗有 $n_1 \times n_2 \times \cdots \times n_k$ 種可能結果。

2. 排列

自一個含有 m 個元素的集合中，一次抽取 n 個元素（或每抽取一個，抽出不放回，連續抽取 n 個），則共有 P_n^m 個不同排列的樣本組，其排列數為：

$$P_n^m = \frac{m!}{(m-n)!}$$

（其中 $m!$ 視為是 m 階乘，代表： $m! = m(m-1)(m-2)\cdots(1)$）
（例如： $10! = 10 \times 9 \times 8 \times \cdots \times 1$）

3. 組合

自一個含有 m 個元素的集合中，一次抽取 n 個元素，若不考慮 n 個元素被抽中元素的順序，則共有 C_n^m 個不同組合，其組合數為：

$$C_n^m = \frac{m!}{n!(m-n)!}$$

例題 4.2

目前台灣地區盛行的大樂透，乃從 1 號到 49 號任選 6 個號碼，若所下注的 6 個號碼與大樂透開出的號碼完全相同，即中頭彩。(1) 試問簽注大樂透共有幾種可能的號碼組合？(2) 若 6 個號碼簽注的順序，必須和開獎的順序完全一樣，試問有幾種可能的號碼組合？

解

(1) $C_6^{49} = \dfrac{49!}{6!43!} = 13{,}983{,}816$

(2) $P_6^{49} = \dfrac{49!}{43!} = 10{,}068{,}347{,}520$

最後我們來介紹事件的觀念，其定義如下：

事件 (event) 乃樣本空間的部分集合（子集）；每一樣本點皆為樣本空間的子集，故亦皆為事件，稱為簡單事件 (simple event)；而含有兩個以上的樣本點之事件，稱為複合事件 (compound event)。

著作權所有 侵害必究
華泰文化 HWA TAI PUBLISHING

根據上面的定義，空集合 (ϕ) 與樣本空間本身 (S) 乃是二個特殊的子集，故亦為事件，但此二事件有其特別的涵義。空集合所代表的事件，因它不含任何樣本點，故一般稱為**不可能事件**；而事件 S 包含了樣本空間內的所有樣本點，必然會發生，故一般稱為**必然事件**。

例題 4.3

擲一硬幣兩次，其樣本空間為 $\{HH, HT, TH, TT\}$，若分別以 e_1, e_2, e_3, e_4 表示 HH, HT, TH, TT，則 S 亦可寫成：

$$S = \{e_1, e_2, e_3, e_4\}$$

上式 S 中的各元素之排列順序並不重要，我們所關心的只是這些元素（樣本點）所組成的集合。由於 $\{e_1\}, \{e_2\}, \{e_3\}, \{e_4\}$ 皆為 S 的子集，且只包含一個樣本點，故皆為簡單事件。

假如我們令出現一個正面之事件為 E，則事件 E 包含二個樣本點 e_2 與 e_3，亦即 $E = \{e_2, e_3\}$，此時 E 即為複合事件。當某次試驗出現 HH 或 TT 時，則稱事件 E 沒有發生。

例題 4.4

假設一箱子中有 10 個球，其中 7 個為紅球 (R)，3 個為白球 (W)，現在我們自此箱中隨機抽出一球。此時可能出現兩種不同的結果，即所抽出的球不是紅球便是白球，因此其樣本空間為 $S_1 = \{R, W\}$，只含有二個樣本點。

但是從另一個角度來看，我們可將每一次抽出一球視為一簡單事件，而該球的顏色視為其特性。換句話說，我們可以將 7 個紅球附上標籤 $e_1, e_2, \cdots, e_7$，而 3 個白球附上標籤 e_8, e_9, e_{10}。由於分析每一次抽球乃是隨機的，故其樣本空間共有 10 個樣本點，即：

$$S_2 = \{\underbrace{e_1, e_2, \cdots, e_7}_{\text{紅球}}, \underbrace{e_8, e_9, e_{10}}_{\text{白球}}\}$$

如果我們令 A 代表所抽出之球為紅球的事件，則 A 所包含的元素有：

$$A = \{e_1, e_2, \cdots, e_7\}$$

由上述可知，第二種定義樣本空間的方式可能比第一種定義樣本空間的方式來得簡易，特別是當要進一步討論機率之問題時，第二種方式之優點更是多於第一種的定義方式。

著作權所有 侵害必究
華泰文化
HWA TAI PUBLISHING

例 4.4 說明了一個重要的觀念：「就某一特定的問題中，樣本空間之形成並非是唯一的。」因此，當我們考慮如何選擇樣本空間時，應以便於機率問題之處理為原則，而非以簡易地引出樣本空間諸元素為評判的標準。

4.3 機率測度的方法

事件是機率測度的對象。在一項隨機試驗中，如果其出現的結果是事件 E 所包含的樣本點之一，則稱事件 E 發生；否則，事件 E 沒有發生。至於如何賦予一事件某機率值呢？一般而言，賦予機率值的方式有三種，以下分別說明之。

4.3.1 古典方法

使用古典方法測度事件的機率，乃基於下列限制條件與假設：

1. 限制條件：樣本空間必須是有限的樣本空間。
2. 基本假設：假設樣本空間內的每一樣本點之出現的機會皆相同。

若符合上述的假設條件，則可依古典方法測度事件的機率：

古典方法的機率測度

在一有限的樣本空間 S 中，某一事件 E 的機率 $P(E)$ 定義為：

$$P(E)=\frac{n(E)}{n(S)} \tag{4-1}$$

式中 $n(S)$ 與 $n(E)$ 分別代表樣本空間與事件所包含的樣本點個數。

例題 4.5

擲一公正硬幣兩次，試求出現一個正面的機率。

解 正如例 4.3 所示，其樣本空間包含 4 個樣本點，即 $S=\{HH, HT, TH, TT\}$。又所謂公正硬幣乃表示此 4 個樣本點中，每一樣本點出現之機會相等，因此符合古典方法的假設條件。令事件 E 代表出現一個正面，即 $E=\{HT, TH\}$，於是：

$$P(E)=\frac{n(E)}{n(S)}=\frac{2}{4}=\frac{1}{2}$$

著作權所有 侵者必究
華泰文化
HWA TAI PUBLISHING

例題 4.6

在例 4.4 中，如果我們抽樣的方式是隨機自箱中抽出一個球，則所抽出球為紅球的機率為何？

解 由於係採隨機抽球的方式，因此箱中的 10 個球每一球被抽中的機會皆相等；又樣本空間 $S_2 = \{e_1, e_2, \cdots, e_7, e_8, e_9, e_{10}\}$（$e_1, \cdots, e_7$：紅球；$e_8, e_9, e_{10}$：白球）為一有限的樣本空間，故符合古典方法的假設條件。令 E 代表紅球的事件，則 $E = \{e_1, e_2, \cdots, e_7\}$，於是：

$$P(E) = \frac{n(E)}{n(S)} = \frac{7}{10}$$

本例題亦可從另一角度來推理：由於每一次所抽出的球不是紅球 R，就是白球 W，因此樣本空間 $S_1 = \{R, W\}$。又由於紅球與白球個數之比例為 7：3，故其機率亦為此一相同比例，亦即：

$$P(E) = \frac{7}{7+3} = \frac{7}{10}$$

4.3.2 相對次數方法

古典方法的使用，必須符合兩個假設條件，故其應用性受到限制。實際上有許多的情況是：也許樣本空間內的每一樣本點發生之機會相等，但卻無法明確的建立其樣本空間；另外，如果將一骰子截去一角，則我們便沒有理由再認為每一面出現的機會相等；再者，當我們欲求一個人將在 50 歲死亡的機率時，我們必須去調查每年中 50 歲死亡占該年中全部死亡人數之比例。以上所述的諸種情況，在在顯示了我們並無堅強的理由支持我們必須採用古典方法以求事件的機率。

當每一樣本點發生之機會相等的假設不再適用時，我們將如何測度某事件發生之機率呢？此時唯一的思考路線是，重複進行此一試驗許多次，並觀察該事件出現次數的比例。這種方式即是相對次數方法的觀念。具體言之，**相對次數方法**的意義如下：

相對次數方法之機率測度

一隨機試驗重複進行 N 次，若事件 E 出現 n 次，則其機率 $P(E)$ 約為：

$$P(E) \approx \frac{n}{N} \tag{4-2}$$

著作權所有 侵害必究
華泰文化 HWA TAI PUBLISHING

(4-2)式中，$\frac{n}{N}$ 實際上即代表相對次數，這也就是相對次數方法名稱的由來。

例題 4.7

令 E 代表擲一骰子而得到6點的事件，如果擲一骰子100次，而出現6點共23次，則 E 的相對次數為 $\frac{23}{100}=0.23$，此即事件 E 之機率的估計值。同樣的，若再擲此骰子100次，其中6點出現18次，則此時的 $P(E)\approx\frac{18}{100}=0.18$。如果將此前後兩次的實驗結合之，則 $N=200$，其中6點出現之相對次數為 $\frac{23+18}{200}=\frac{41}{200}=0.205$，亦可將之視為事件 E 之機率的估計值。

假設例4.7之實驗重複進行，並記錄及調整事件 E 之相對次數，則可能得出如圖4.2所示的分布圖。由圖4.2可發現，當 N 變動時，E 的相對次數亦隨之上、下波動，唯當 N 愈來愈大時，其相對次數將漸趨於穩定。當然，不同的人分別進行相同 N 次之試驗，其所得的相對次數之趨勢線不盡相同。然而，就很大的 N 而言，其相對次數之穩定的趨勢值應該相同。

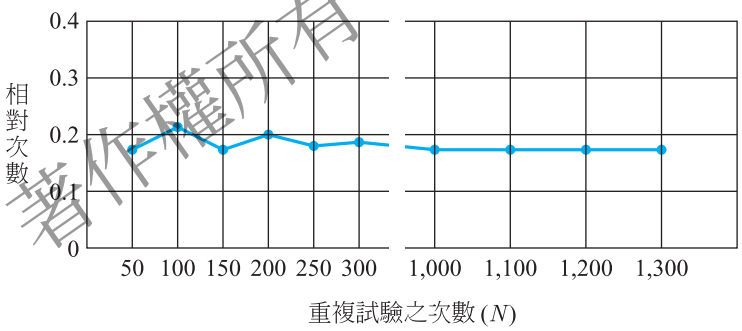


圖 4.2　相對次數之穩定性

相對次數方法的重要觀念即在於，當試驗次數 N 趨於無窮大時，則相對次數極限值（穩定的趨勢值）可作為事件機率的測度，亦即：

$$P(E)=\lim_{N\to\infty}\frac{n}{N}$$

當然，我們不可能觀察到無限多次的試驗，但這只是理論上的說法，亦即：當 N 愈來愈大時，則相對次數愈趨於穩定，此現象稱為統計的規律性。由此可知，以相對次數方法測度機率時，事件的機率意指在一長期實驗中，發生次數相對於總試驗次數的比例。欲使求得事件機率之估計值愈準確，則理論告訴我們「重複愈多次試驗愈佳」。

著作權所有 侵害必究
華泰文化 HWA TAI PUBLISHING

根據相對次數方法的觀念，人口學者往往對於生育資料進行無數的研究，藉以探討某特定年齡組死亡的相對次數，並以之作為某一年齡組死亡機率的估計值。

例題 4.8

假定根據某人口資料顯示，最近 10,000 個人死亡的年齡如表 4.2 所示。試求出一新生嬰兒至少活到 70 歲的機率。

表 4.2　10,000 個人之死亡年齡記錄

年齡（歲）	死亡人數	相對次數
未滿 10	323	0.0323
10～19	65	0.0065
20～29	121	0.0121
30～39	184	0.0184
40～49	431	0.0431
50～59	969	0.0969
60～69	1,821	0.1821
70～79	2,728	0.2728
80 以上	3,358	0.3358
總計	10,000	1.0

解 根據相對次數方法的觀念，我們以相對次數作為機率之估計值。因此，至少活到 70 歲的相對次數為：

$$\frac{2,728 + 3,358}{10,000} = 0.6086$$

以此作為該新生嬰兒能活到 70 歲以上的機率之估計值。當然，由表 4.2，我們亦可估計出某新生嬰兒可活到多少歲之機率。

相對次數方法測度事件機率的應用範圍很廣，只要可進行重複試驗，或者存有過去某事件發生的相對次數之資料，皆可加以運用。

4.3.3 主觀方法

當隨機試驗缺乏明顯的形式，亦不可能重複進行多次的試驗，或者亦缺乏過去類似事件的資料時，該如何估計某一事件的機率？換句話說，無法利用前述兩種測度機率的方法時，通常會憑個人的經驗或直覺，以判斷事件發生的機率，此即所謂的**主觀機率** (subjective probability) 測度方法。

著作權所有 侵害必究
華泰文化 HWA TAI PUBLISHING