機率

CH5

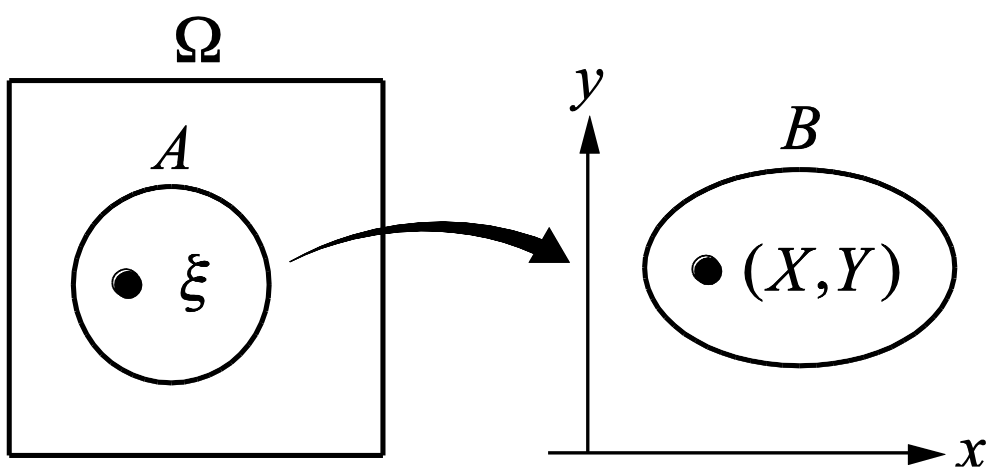
聯合機率分配與隨機樣本

簡介

- 在真實世界中，可能必須面對兩個甚至更多個隨機變數，而這些變數之間可能互相影響。
- 因此，不能只針對個別變數加以分析，而必須適當地刻劃變數間的聯合行為，再據此探討變數的個別行為及變數之間的關係。
- 兩個隨機變數合併起來也稱作雙變量隨機變數(bivariate random variable) 。

多維隨機變數

- 設(Ω, F, P)為一機率空間
- n維的隨機變數為一由樣本空間Ω映到n維歐氏空間R^n的函數
  - 二維隨機變數V = (X, Y)可看成是平面上的隨機點
  - 三維隨機變數V = (X, Y, Z)可看成是空間中的隨機點



5.1 兩隨機變數

「一個隨機變數是一個映射的概念」可以很容易地被一般化來考慮含有 2 個隨機變數的狀況。考慮一個隨機實驗，其樣本空間為 S 且事件類別為 $\mathcal{F}$ 。我們感興趣的是一個函數，它會指派一對實數 $X(\zeta)=(X(\zeta),Y(\zeta))$ 給在 S 中的每一個結果 ζ 。基本上我們正在處理一個向量函數，它會把 S 映射至 R^2 ，也就是從 S 映射至實數平面，如在圖 5.1(a)中所示。我們最感興趣的是包含(X,Y)的事件。

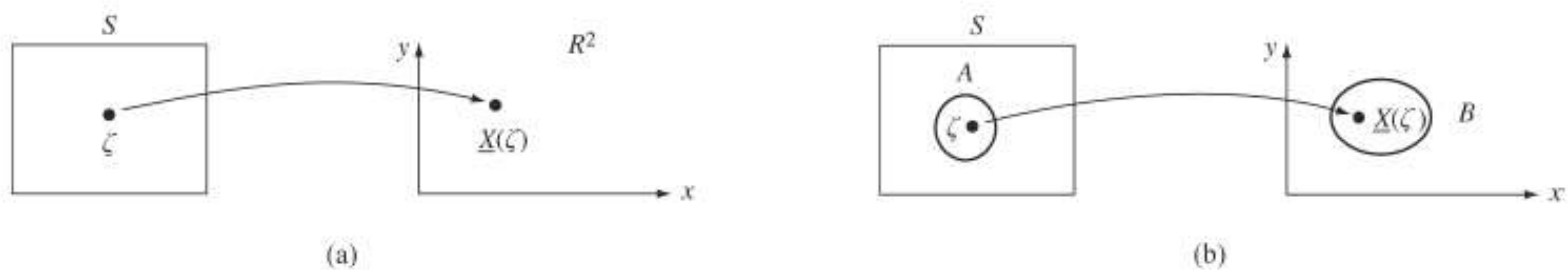


圖 5.1 (a)一個函數指派一對實數給在 S 中的每一個結果。(b)兩隨機變數的等價事件。

請參考範例5.1～5.4，圖5.2～5.4，及公式(5.1)～(5.3)等課本相關內容。

$$B=\left\{\min\left(X,Y\right)\leq 5\right\}$$

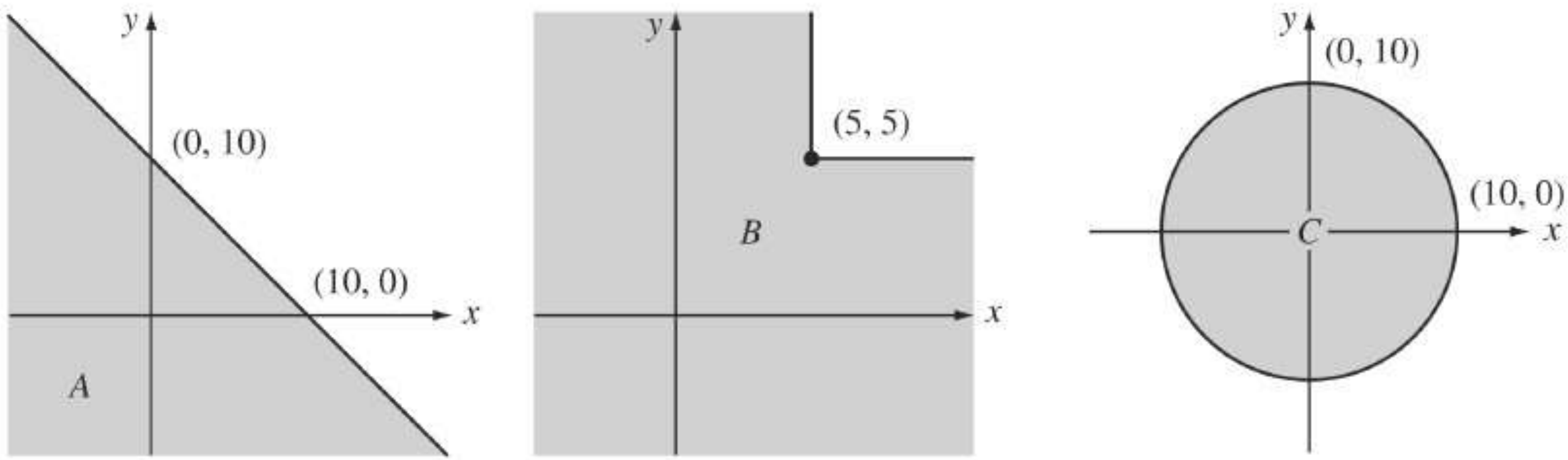


圖 5.2　二維事件的例子

$$A=\left\{X+Y\leq 10\right\}$$

$$C=\left\{X^2+Y^2\leq 100\right\}$$

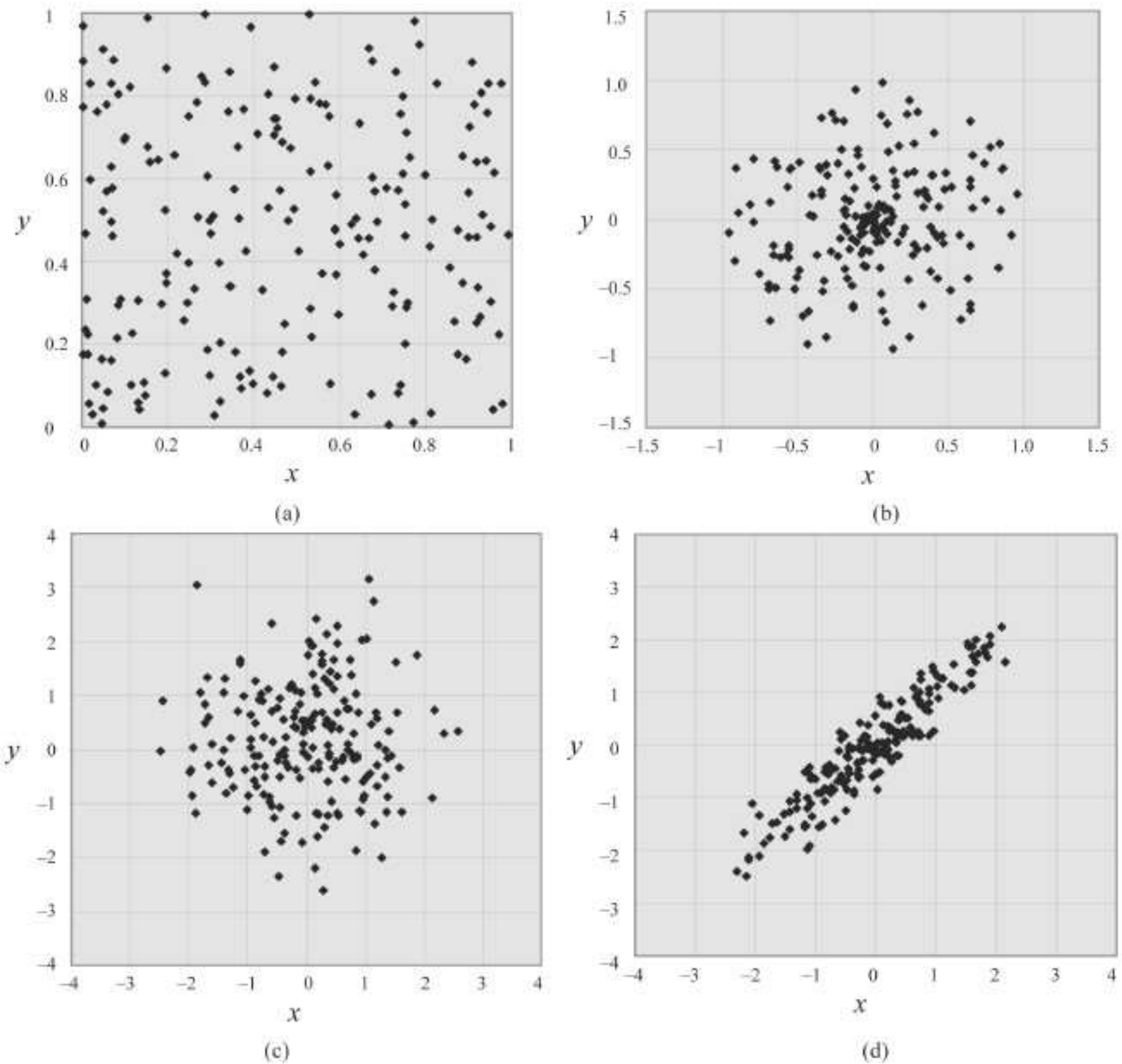
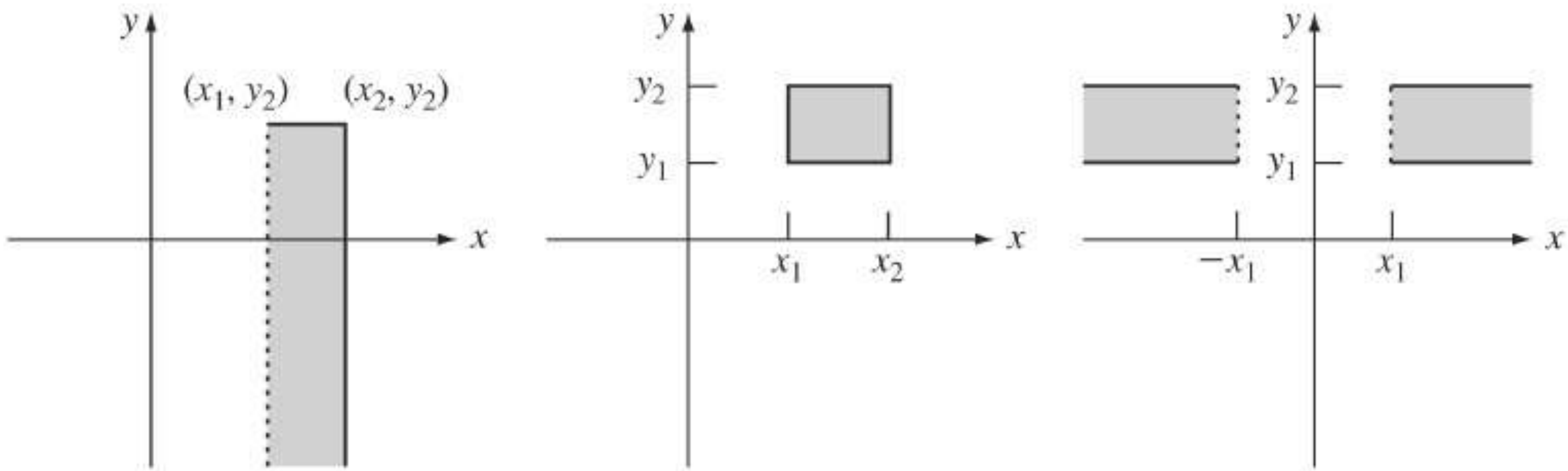


圖 5.3　4 個不同的隨機變數對的分散點狀圖，每一個圖包含 200 個觀察點



$\{x_1 < X \leqslant x_2\} \cap \{Y \leqslant y_2\}$　　$\{x_1 \leqslant X \leqslant x_2\} \cap \{y_1 \leqslant Y \leqslant y_2\}$　　$\{|X| > x_1\} \cap \{y_1 \leqslant Y \leqslant y_2\}$

圖 5.4　某些二維乘積形式的事件

離散隨機變數的聯合分配

- 若 X 和 Y 為離散隨機變數，其實現值分別為 $a_1, a_2, \ldots,$ 和 $b_1, b_2, \ldots$。
- X 和 Y 的**聯合機率密度函數** (joint probability function)
  - 又稱之為**雙變量機率函數** (bivariate probability function),
  - 其函數值為

$$f_{X,Y}(a_i, b_j) = \mathrm{P}(\{X = a_i \text{ 且 } Y = b_j\}), \quad i, j = 1, 2, \ldots$$

離散隨機變數的聯合分配

- EX
  - X = 1,2,3 分別代表所得高低的顧客
  - Y =1,2,3,4 分別代表想投資的標的
  - 則兩者的聯合隨機行為可 用下表的聯合機率加以描述。

	Y=1	Y=2	Y=3	Y=4
X=1	0.1	0	0	0
X=2	0.2	0.1	0.2	0
X=3	0.1	0	0.1	0.2

離散隨機變數對

- 聯合機率質量函數
  - joint probability mass function

$$p_{X,Y}(x, y) = P\left[\underline{\{X = x\} \cap \{Y = y\}}\right]$$
$$\triangleq P[X = x, Y = y] \qquad (x, y) \in R^2 \tag{5.4a}$$

$$p_{X,Y}(x_j, y_k) = P\left[\underline{\{X = x_j\} \cap \{Y = y_k\}}\right]$$
$$\triangleq P[X = x_j, Y = y_k] \quad (x_j, y_k) \in S_{X,Y} \tag{5.4b}$$

$$P[\mathbf{X} \text{ 在 } B \text{ 中}] = \sum_{(x_j, y_k) \text{ 在 } B \text{ 中}} \sum p_{X,Y}(x_j, y_k) \tag{5.5}$$

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k) = 1 \tag{5.6}$$

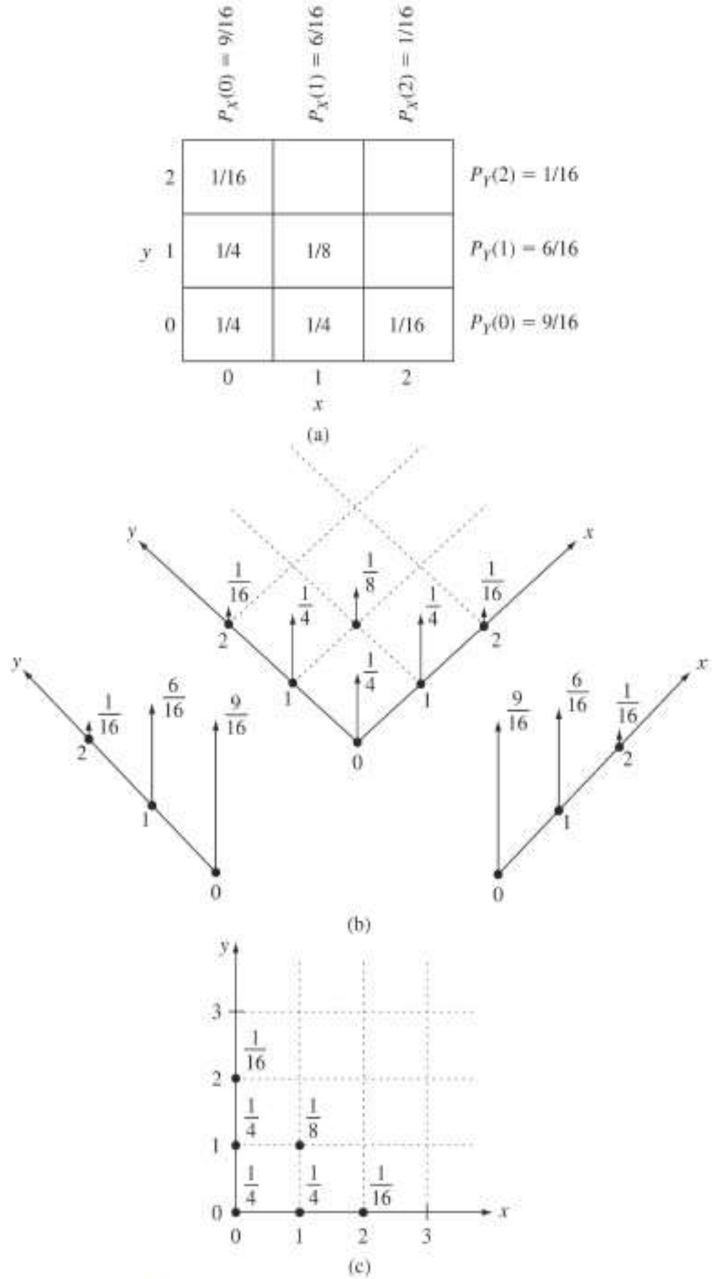


圖 5.5 pmf 的圖形表示：(a)使用表格形式，(b)使用具高度的射線，(c)使用標示有 pmf 值的點

5.1 隨機變數之聯合分配
兩離散型隨機變數

- 定義
  - 令X與Y為定義於一實驗室之樣本空間之兩離散型隨機變數。
  - 每對(x, y)數值之**聯合機率質量函數** (joint probability mass function) $p(x, y)$定義為

$$p(x, y) = P(X = x \text{ 與 } Y = y)$$

其必須滿足 $p(x, y) \geq 0$，與 $\sum_x \sum_y p(x, y) = 1$。

兩離散型隨機變數（續）

- 定義（續）
  - 現令A為 (x, y) 任意配對所形成的集合
    - $A=\{(x, y): x + y=5\}$
    - $A = \{(x, y): \max(x, y) \leq 3\}$。
  - 則 $P[(X, Y) \in A]$ 之機率
    - 可藉由加總 A中配對的聯合機率質量函數求得

$$P[(X, Y) \in A] = \sum_{(x,\,y) \in A} \sum p(x, y)$$

範例 5.6

一個隨機實驗投擲兩個「不公正的」骰子，並注意出現的點數(X,Y)。聯合 pmf $p_{X,Y}(j,k)$的值，其中 $j=1,\ldots,6$ 且 $k=1,\ldots,6$，如在圖 5.6 中之二維圓點圖所示。(j,k)位置的旁邊標示有 $p_{X,Y}(j,k)$ 的值。求出 $P[\min(X,Y)=3]$。

在圖 5.6 中我們有把對應到集合$\{\min(x,y)=3\}$的點圈起來。這個事件的機率爲：

$$\begin{aligned}P[\min(X,Y)=3] &= p_{X,Y}(6,3)+p_{X,Y}(5,3)+p_{X,Y}(4,3)\\&\quad+p_{X,Y}(3,3)+p_{X,Y}(3,4)+p_{X,Y}(3,5)+p_{X,Y}(3,6)\\&=6\left(\frac{1}{42}\right)+\frac{2}{42}=\frac{8}{42}\end{aligned}$$

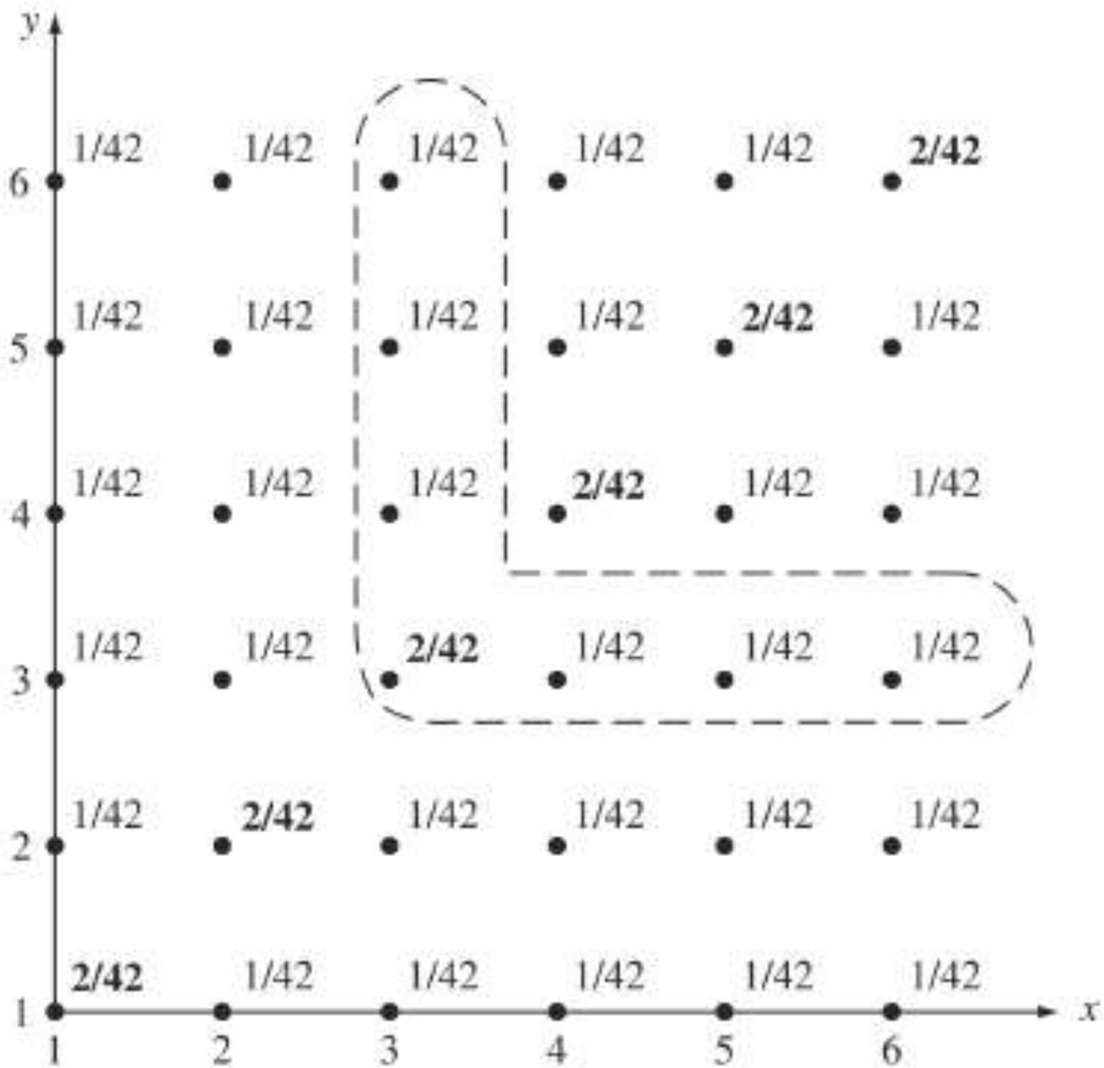


圖 5.6　把包含在 B 中的點圈起來，形成一個區域來展示 pmf。

範例5.1

- 某大型保險仲介業者為一些購買屋險與車險的顧客提供服務。每一種不同的保險皆必須指定扣除條款金額。
- 以車險而言，可選擇＄100 或＄250
- 屋險的選擇則為 0 , $100 與＄200 。

範例5.1（續）

- 若隨機在檔案中抽出一位同時購買兩種保險的顧客。
- 令 X 一車險之扣除條款額，與 Y 一屋險之扣除條款額。
- 可能的（X , Y）配對為（100 , 0) , (100 , 100) , (100 , 200) , (250 , 0) , (250 , 100）與（ 250 , 200)：其聯合機率質量函數說明以上任一配對所對應之機率，任何其他配對機率則為零。

範例5.1（續）

- 假設此聯合機率質量函數表示於以下之聯合機率表（joint probability table）：

$p(x, y)$		y = 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

範例5.1（續）

- 則 $p(100,100) = P(X = 100$ 與 $Y = 100) = P$ (兩類之扣除條款金額皆為 \$100$) = 0.10$。
- $P(Y \geq 100)$ 機率的計算
  - 則是將所有滿足 $y \geq 100$ 之 (x, y) 配對的機率值加總：

Example

- 一台液晶電視的兩個影像處理晶片從有3個晶片的A廠商，2個晶片的B 廠商，和3個晶片的C廠商的盒子內來提供
- 如果 r.v. X表示來自A廠商晶片數目
- 如果 r.v. Y表示來自B 廠商的晶片
- 求其聯合機率函數 $f(x, y)$。

• 邊際機率質量函數

X 的聯合 pmf 提供有關於 X 和 Y 聯合行爲的資訊。我們也關心只有含單一個隨機變數的事件機率。這可以由**邊際機率質量函數(marginal probability mass functions)**來求出：

$$p_X(x_j) = P[X = x_j] = P[X = x_j, Y = \text{任意值}]$$
$$= P[\{X = x_j \text{ 且 } Y = y_1\} \cup \{X = x_j \text{ 且 } Y = y_2\} \cup \ldots]$$
$$= \sum_{k=1}^{\infty} p_{X,Y}(x_j, y_k) \tag{5.7a}$$

類似地，

$$p_Y(y_k) = P[Y = y_k] = \sum_{j=1}^{\infty} p_{X,Y}(x_j, y_k) \tag{5.7b}$$

邊際 pmf 滿足所有一維 pmf 的特性，若要計算僅含單一隨機變數事件的機率，邊際 pmf 提供所需的資訊。

一般而言，知道邊際 pmf 不足以推斷出聯合 pmf。

兩離散型隨機變數（續）

- 定義
  - X 與Y 之**邊際機率質量函數 (marginal probability mass functions)**
  - 分別表示為$p_X(x)$ 與$p_Y(y)$，定義如下：

$$p_X(x) = \sum_{y} p(x, y) \qquad p_Y(y) = \sum_{x} p(x, y)$$

- **邊際分配**則是這兩個隨機變數中任一個變數的個別隨機行為

兩離散型隨機變數（續）

- X 和 Y 具有離散的聯合分配，X 的**邊際機率函數**為

$$f_X(a_i) = P(\{X = a_i\}) = \sum_{\boxed{j}} f_{X,Y}(a_i, b_j), \qquad i = 1,2\ldots$$

而 X 為其他值時，f_X 為 0。

- Y 的**邊際機率函數** f_Y 為

$$f_Y(b_j) = \mathrm{P}(\{Y = b_j\}) = \sum_{\boxed{i}} f_{X,Y}(a_i, b_j), \qquad j = 1,2,\ldots$$

而 Y 為其他值時，f_Y 為 0。

範例5.2(承範例5.1

$p(x, y)$		y = 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- 可能的 X 值為 $x=100$ 與 $x=250$， 所以計算聯合機率表之各列的總和可得

範例5.2（續）(承範例5.1)

- 則 X 之邊際機率質量函數為

範例5.2（續）(承

$p(x, y)$		y = 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- 同樣地，由行總和可得 Y 之邊際機率質量函數為

兩連續型隨機變數

- 定義
  - 令 X 與 Y 為連續型隨機變數。
  - 兩變數之**聯合機率密度函數 (joint probability density function)** $f(x, y)$，為一滿足 $f(x, y) \geq 0$ 與 $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,dx\,dy = 1$ 之函數。

則對任何二維變數集合 A

$$P[(X, Y) \in A] = \iint_A f(x, y)\,dx\,dy$$

兩連續型隨機變數（續）

- 定義（續）
  - 尤其，若A 為二維矩形$\{(x, y): a \le x \le b, c \le y \le d\}$,，則

$$P[(X, Y) \in A] = P(a \le X \le b, c \le Y \le d)$$

$$= \int_a^b \int_c^d f(x, y)\, dy\, dx$$

兩連續型隨機變數（續）

- 我們可將 $f(x, y)$ 視為三維空間中之一曲面，點(x, y) 處之表面高度為 $f(x, y)$。
- 則$P[(x, y) \in A]$ 為此曲面下方及區域 A 上方的體積，與一維情況時曲線下面積相當。
- 說明於圖5.1 。

圖5.1

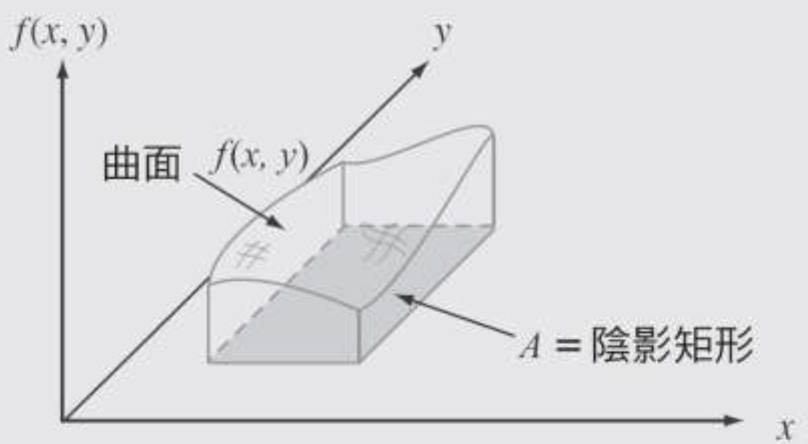


圖 5.1 $P[(X, Y) \in A]$ = 密度函數曲面下及 A 之上的體積

範例5.3

- 銀行經營兩種不同窗口，分別為車行窗口與人行窗口。
- 隨機選擇一天，
  - 令X＝ 車行窗口之使用時間比例
    - 至少一個客戶正在接受服務或正在等候服務
  - Y＝ 人行窗口之使用時間比例
- 則(X, Y) 之可能值所形成的集合為矩形D＝$\{(x, y):0 \le x \le 1, 0 \le y \le 1\}$。

範例5.3（續）

- 假設(X, Y) 之聯合密度函數為

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{其他} \end{cases}$$

範例5.3（續）

- 其他證明此為合法的機率密度函數，必須確定 $f(x, y) \geq 0$，且

範例5.3（續）

- 兩窗口忙碌時間均不超過四分之一的機率為

兩連續型隨機變數（續）

- 定義
  - X 與 Y 的**邊際機率密度函數 (marginal probability density functions)** 分別表示為 $f_X(x)$ 與 $f_Y(y)$，定義如下：

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, \boxed{dy} \qquad -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, \boxed{dx} \qquad -\infty < y < \infty$$

範例5.4(承範例5.3)

- X 的邊際機率密度函數為車行窗口，在不考慮人行窗口的情況下，忙碌時間的機率分配，為

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 \frac{6}{5}(x + y^2)\, dy = \frac{6}{5}x + \frac{2}{5}$$

- 當$0 \le x \le 1$，其他為0

範例5.4（續）(承範例5.3)

- Y 的邊際密度函數為

$$f_Y(y) = \begin{cases} \frac{6}{5}y^2 + \frac{3}{5} & 0 \le y \le 1 \\ 0 & \text{其他} \end{cases}$$

則

範例5.5

- 某堅果公司銷售豪華混合堅果罐頭，內含杏仁、腰果，與花生。
- 假設每個罐頭的淨重正好為1 磅(lb)，但各類堅果之重量分配為隨機的。
- 因為三種堅果的重量總和為1 磅，所以知道其中兩種堅果的聯合機率模型即有第三種堅果重量的所有資訊。

範例5.5（續）

- 令X＝罐頭中之杏仁重量
- Y＝罐頭中之腰果重量
- 則密度為正之區域為
- $D=\{(x, y): 0 \le x \le 1, 0 \le y \le 1, x+y \le 1\}$，圖 5.2 之陰影區域。

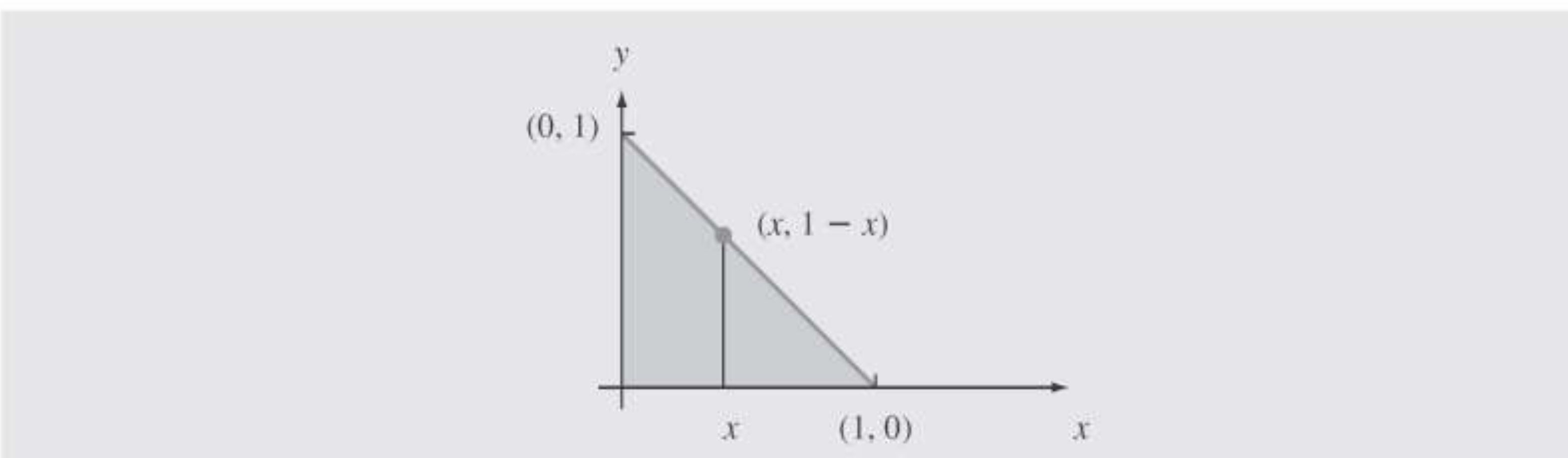


圖 5.2 範例 5.5 中密度為正之區域

範例5.5（續）

- 現在令(X , Y) 之聯合機率密度函數為

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0 & \text{其他} \end{cases}$$

範例5.5（續）

- 任何固定的x， $f(x, y)$ 隨著 y 而增加；任何固定的 y， $f(x, y)$ 隨著 x 而增加。
- 因為“豪華”一詞表示大部分的罐頭中含有較多的杏仁與腰果，而非花生，所以密度函數之較大值應該出現在上邊界處，較小值該接近原點。
- 當(x, y) 沿著各軸向外移動，$f(x, y)$ 所決定曲面之斜率將從0 開始向上攀升。

範例5.5（續）

- 很清楚地，$f(x, y) \geq 0$。
- 要證明聯合機率密度函數的第二個條件，計算二重積分時必須以逐次積分(iterated integral) 方式為之，保持其中一變數不變(如圖5.2 之x)，將另一變數沿著固定變數所定義之直線的積分，最後再對固定變數的可能值進行積分。

範例5.5（續）

- 因此

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dy\, dx = \int_{D}\int f(x, y)\, dy\, dx$$

範例5.5（續）

範例5.5（續）

- 杏仁之邊際機率密度函數可藉由將X 固定於 x 值，沿著通過 x 之垂直線對聯合機率密度函數 $f(x, y)$ 積分：

圖5.3

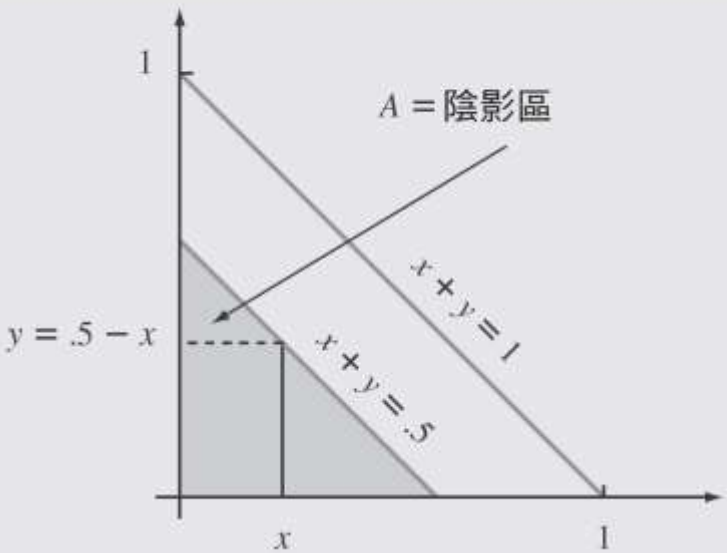


圖 5.3 計算範例 5.5 之 $P[(X, Y) \in A]$

範例5.5（續）

- 藉由$f(x, y)$ 與區域 D 的對稱性， Y 之邊際機率密度函數可透過將$f_X(x)$ 之x 與X 分別以 y 與 Y 替換求得。

X和Y的聯合CDF

包含二維隨機變數事件的基本建構方塊爲半無限矩形，它定義爲$\{(x, y): x \le x_1$ 且 $y \le y_1\}$，如在圖 5.7 中所示。我們使用更爲精簡的表示法$\{x \le x_1, y \le y_1\}$來表示這個區域。**X和 Y的聯合累積分佈函數 (joint cumulative distribution function of X and Y)**被定義爲是事件$\{X \le x_1\} \cap \{Y \le y_1\}$的機率：

$$\boxed{F_{X,Y}(x_1, y_1) = P[X \le x_1, Y \le y_1]} \tag{5.8}$$

$$\boxed{F_{X,Y}(x_1 y_1) = P[X \leqslant x_1, Y \leqslant y_1]}$$

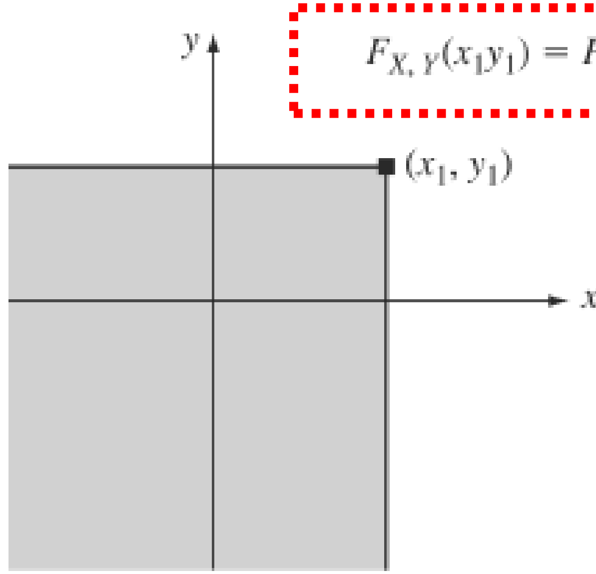


圖 5.7　聯合累積分佈函數被定義為是由點(x_1, y_1)所定義之半無限矩形的機率。

聯合連續隨機變數

- 聯合 pdf & cdf

$$\boxed{F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x} f(\alpha,\beta)\,d\alpha\,d\beta}$$

$$\boxed{f(x,y)=\frac{\partial^2 F(x,y)}{\partial x \partial y}}$$

兩隨機變數的獨立

X和 Y為獨立的隨機變數，假如用 X 定義之任何事件 A_1 獨立於用 Y 定義之任何事件 A_2；也就是說，

$$P[X \text{ 在 } A_1 \text{ 中}, Y \text{ 在 } A_2 \text{ 中}] = P[X \text{ 在 } A_1 \text{ 中}]P[Y \text{ 在 } A_2 \text{ 中}] \tag{5.19}$$

在本節中，我們提出一組簡單的條件來判定何時 X 和 Y 爲獨立的。

假設 X 和 Y 爲一對離散隨機變數，並假設我們對 $A = A_1 \cap A_2$ 這個事件的機率感興趣，其中 A_1 只含有 X，而 A_2 只含有 Y。特別的是，假如 X 和 Y 爲獨立的，那麼 A_1 和 A_2 爲獨立的事件。假如我們令 $A_1 = \{X = x_j\}$ 且 $A_2 = \{Y = y_k\}$，那麼獨立的 X 和 Y 意味著

$$\boxed{\begin{aligned} p_{X,Y}(x_j, y_k) &= P[X = x_j, Y = y_k] \\ &= P[X = x_j]P[Y = y_k] \\ &= p_X(x_j)p_Y(y_k) \quad \text{對所有的 } x_j \text{ 和 } y_k \end{aligned}} \tag{5.20}$$

因此，假如 X 和 Y 爲獨立的離散隨機變數，那麼聯合 pmf 會等於個別邊際 pmf 的乘積。

獨立隨機變數

- 定義
  - 兩隨機變數 X 與 Y 稱為**獨立(independent)**，若對於任意 x 與 y 之配對值皆滿足：

$$\boxed{p(x, y) = p_X(x) \cdot p_Y(y)}$$ 當 X 與 Y 為離散型

或 $$(5.1)$$

$$\boxed{f(x, y) = f_X(x) \cdot f_Y(y)}$$ 當 X 與 Y 為連續型

若非所有之 (x , y) 均滿足 (5.1) 式，則 X 與 Y 稱為**相依 (dependent)**。

範例5.6

$p(x, y)$		y = 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

- 在範例5.1 與5.2 之保險情況,

✓ X 與 Y 獨立需要聯合機率表中每一項機率為其對應之行與列的邊際機率乘積。

範例5.7(承範例5.5)

- 因為$f(x, y)$ 為乘積型式，X 與Y 看似獨立

型式，且密度為正之區域必須為矩形，且邊平行於對應的座標軸。

範例5.8

- 假設兩零件壽命互相獨立，且第一個的壽命，X_1，服從參數為λ_1之指數分配，第二個，X_2，為服從參數為λ_2之指數分配。
- 則聯合機率密度函數為

$$f(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2)$$

$$= \begin{cases} \lambda_1 e^{-\lambda_1 x_1} \cdot \lambda_2 e^{-\lambda_2 x_2} = \lambda_1 \lambda_2 e^{-\lambda_1 x_1 - \lambda_2 x_2} & x_1 > 0, x_2 > 0 \\ 0 & \text{其他} \end{cases}$$

範例5.8（續）

- 令λ_1 ＝1/1000 且λ_2 ＝1/1200， 所以壽命之期望值使用年限分別為1,000 小時與1,200 小時。
- 兩個零件的壽命均至少1,500 小時的機率為

條件分配

- 定義
  - 令 X 與 Y 為兩連續隨機變數，其聯合機率密度函數為 $f(x, y)$，X 之邊際密度函數為$f_X(x)$。
  - 對於任意 X 之值 x,
  - 若 $f_X(x) < 0$，則在**已知$X = x$，Y 的條件機率密度函數(conditional probability density function of Y)** 為

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} \qquad -\infty < y < \infty$$

條件分配（續）

- 定義（續）
  - 若X與Y為離散型，則以機率質量函數取代機率密度函數，
  - 可求得**當$X = x$，Y之條件機率質量函數**。

範例5.12

- 重新考慮範例5.3 與5.4 的狀況， 其牽涉X＝銀行車行窗口之忙碌時間比例與 Y＝人行窗口者之比例。

範例5.12（續）

- 已知$X=0.8$，Y之條件機率密度函數為

範例5.12（續）

- 已知$X=0.8$，人行窗口至多一半時間為忙碌之機率則為

範例5.12（續）

- 利用Y之邊際機率密度函數可得$P(Y \leq 0.5) = 0.350$。同時，$E(Y) = 6$，且在已知$X = 0.8$之情況下，行窗口之期望時間比例(條件期望值)為

兩個隨機變數其函數的期望值

求出兩個或多個隨機變數其函數的期望值的這個問題，類似於求出單一隨機變數其函數的期望值。我們可以證明 $Z = g(X, Y)$ 的期望值可以用以下的表示式求出：

$$E[Z] = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\, dx\, dy & X, Y \text{ 為聯合連續} \\ \sum_i \sum_n g(x_i, y_n) p_{XY}(x_i, y_n) & X, Y \text{ 為離散} \end{cases} \tag{5.25}$$

範例 5.25　獨立的隨機變數其函數的乘積

假設 X 和 Y 爲獨立的隨機變數，並令 $g(X, Y) = g_1(X) g_2(Y)$。求出 $E[g(X, Y)] = E[g_1(X) g_2(Y)]$。

$$
\begin{aligned}
E[g_1(X) g_2(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1(x') g_2(y') f_X(x') f_Y(y') \, dx' \, dy' \\
&= \left\{ \int_{-\infty}^{\infty} g_1(x') f_X(x') \, dx' \right\} \left\{ \int_{-\infty}^{\infty} g_2(y') f_Y(y') \, dy' \right\} \\
&= E[g_1(X)] E[g_2(Y)]
\end{aligned}
$$

離散隨機變數的動差

- 若隨機變數 X 和 Y 具有離散的聯合分配，其實現值分別為 $a_1, a_2, \ldots$和 $b_1, b_2 \ldots,$
- 則 X 的期望值 (平均數、第一階動差) 為

$$
\begin{aligned}
\mathrm{E}(X) &= \sum_i \sum_j a_i f_{X,Y}(a_i, b_j) \\
&= \sum_i a_i \sum_j f_{X,Y}(a_i, b_j) \\
&= \sum_i a_i f_X(a_i)
\end{aligned}
$$

離散隨機變數的動差

- 令X 和 Y 代表兩個隨機變數。
- 則對任意實數 a 和 b,
  - $E(aX+bY) = aE(X) + bE(Y)$。
- X 和 Y 的變異數 (第二階中央動差) 為
  - $var(X) = E(X^2) – [E(X)]^2$
  - $var(Y) = E(Y^2) – [E(Y)]^2$

離散隨機變數的邊際分配 -- 實例

- 根據下例 中的資料，可以求算其邊際機率如下：

	Y=1	Y=2	Y=3	Y=4	X的邊際機率	X的累積機率
X=1	0.1	0	0	0	0.1	0.1
X=2	0.2	0.1	0.2	0	0.5	0.6
X=3	0.1	0	0.1	0.2	0.4	1
Y的邊際機率	0.4	0.1	0.3	0.2	1	-
Y的累積機率	0.4	0.5	0.8	1	-	-

除了在特殊情況下，聯合機率通常不能由邊際機率決定。

離散隨機變數的動差

- 動差

5.2 期望值，共變異數與相關性

- 提要
  - 令X與 Y 為聯合分配之隨機變數，其機率質量函數為 $p(x, y)$ 抑或機率密度函數為$f(x, y)$，根據這些變數為離散型或連續型而決定。則函數$h(X, Y)$之期望值，表示為$E[h(X, Y)]$ 或$\mu_{h(x,y)}$，為

$$E[h(X, Y)] = \begin{cases} \sum_x \sum_y h(x, y) \cdot p(x, y) & \text{若 } X \text{ 與 } Y \text{ 離散型} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x, y) \cdot f(x, y)\, dx\, dy & \text{若 } X \text{ 與 } Y \text{ 連續型} \end{cases}$$

範例5.13

- 五個朋友皆購買某場演唱會的門票。
- 若門票的座位為某列1-5 號，且將票隨機分配給五個人，則這五位朋友中任意兩位間隔座位數的期望值為何？

範例5.13（續）

- 令X與 Y分別表示第一個人及第二個人的座位號碼。可能的(X, Y) 配對為$\{(1, 2), (1, 3), ..., (5, 4)\}$，且$(X, Y)$ 的聯合機率質量函數為

$$p(x, y) = \begin{cases} \dfrac{1}{20} & x = 1, \ldots, 5;\ y = 1, \ldots, 5;\ x \neq y \\ 0 & \text{其他} \end{cases}$$

範例5.13（續）

- 分開兩個座位的號碼為$h(X, Y) = |X - Y| - 1$。以下表格為每個可能配對(x, y) 之$h(x, y)$ 值。

$h(x, y)$		x=1	2	3	4	5
y	1	—	0	1	2	3
	2	0	—	0	1	2
	3	1	0	—	0	1
	4	2	1	0	—	0
	5	3	2	1	0	—

範例5.13（續）

- 因此

範例5.14

- 在範例5.5，一個一磅罐頭中的杏仁數 X 與腰果數 Y 之聯合機率密度函數為

$$f(x, y) = \begin{cases} 24xy & 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1 \\ 0 & \text{其他} \end{cases}$$

範例5.14（續）

- 若一磅杏仁的成本為$1.00， 一磅腰果的成本為$1.50， 且一磅花生的成本為$0.50， 則罐頭內含物之總成本為

範例5.14（續）

- 總成本之期望值為

共變異數

- 對於離散隨機變數 X 和 Y，E(XY) 稱為 X 和 Y 的**交叉動差** (cross moment):

$$\mathrm{E}(XY) = \sum_i \sum_j a_i b_j f_{X,Y}(a_i, b_j)$$

交叉動差會受隨機變數位置改變的影響。

- 共變異數是利用隨機變數與平均數之間的差距來計算交差動差，一般以 cov(X,Y) 表示，其公式如下：

$$\mathrm{cov}(X,Y) = \mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))]$$

- 共變異數描述的是兩隨機變數和其均數之差距的共同變動。
  - **當共變異數為正值**，這些差距平均而言呈同方向變動。
  - **當共變異數為負值**，這些差距平均而言呈反方向變動。

共變異數

- 定義
  - 兩隨機變數 X 與 Y 間之**共變異數 (covariance)** 為

$$\operatorname{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$
$$= \begin{cases} \displaystyle\sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y) & X, Y \text{ 為離散型} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y)f(x, y)\, dx\, dy & X, Y \text{ 為連續型} \end{cases}$$

共變異數（續）

- **因為**$X-\mu_X$ **與**$Y-\mu_Y$ **為兩變數對其個別平均數之偏差**值，故共變異數為偏差乘積的期望值。
- **注意** $\mathrm{Cov}\,(X, X) = E[(X-\mu_X)^2] = V(X)$。

共變異數（續）

- 此定義之基本原理如下 ：
  - 假設 X 與 Y 間為強烈正相關，意味當 X 之值大，Y 之值有大的趨勢；X 小，Y 亦小。
  - 則大部分之機率質量或密度將使 $(x-\mu_X)$ 與 $(y-\mu_Y)$ 同時為正(X 與 Y 皆高出其個別的平均數) 或同時為負，故 $(x-\mu_X)(y-\mu_Y)$ 之乘積值趨向正值。

共變異數（續）

- 因此強烈正相關的情況，Cov (X, Y) 應非常正。
- 強烈負相關的情況，($x-\mu_X$) 與($y-\mu_Y$) 的符號有相反趨勢，乘積將為負。
- 因此強烈負相關的情況，Cov(X, Y) 應為非常負。
- 若X 與Y 並非強烈相關，則正與負之乘積有相抵銷的趨勢，共變異數將會趨近於 0。

圖5.4

- 圖5.4 說明不同的可能性。

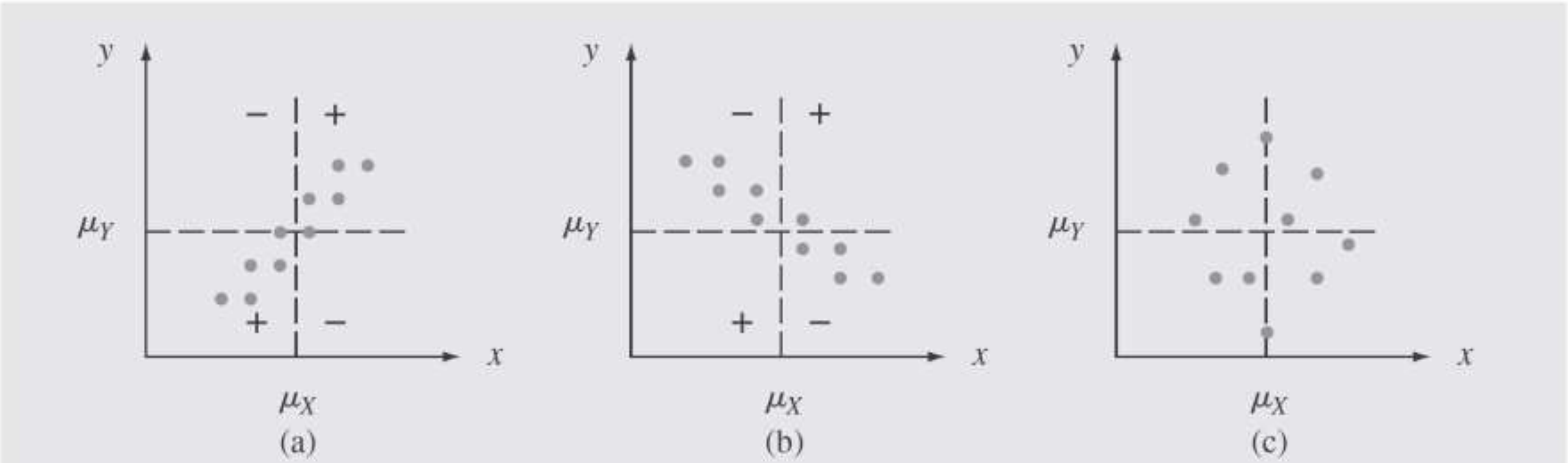


圖 5.4 對應於圖中點之十配對資料，每對之 $p(x, y) = 1/10$；(a) 正共變異數；(b) 負共變異數；(c) 共變異數趨近零

範例5.15

- 範例5.1 之 X＝汽車保險扣除條款金額， 與Y＝房屋保險扣除條款金額之聯合及邊際機率質量函數為

$p(x, y)$		y: 0	100	200
x	100	0.20	0.10	0.20
	250	0.05	0.15	0.30

x	100	250
$p_X(x)$	0.5	0.5

y	0	100	200
$p_Y(y)$	0.25	0.25	0.5

範例5.15（續）

- 可得 $\mu_X = \sum x p_X(x) = 175$ 與 $\mu_Y = 125$，因此

共變異數（續）

- 提要

$$\mathrm{Cov}(X, Y) = E(XY) - \mu_X \cdot \mu_Y$$

- 此式之證明需將$(X-\mu_X)(Y-\mu_Y)$展開，然後計算各項之期望值。

範例5.16(承範例5.5)

- X＝罐頭中之杏仁數及Y＝罐頭中之腰果數的聯合與邊際機率密度函數為

$$f(x, y) = \begin{cases} 24xy & 0 \le x \le 1, 0 \le y \le 1, x + y \le 1 \\ 0 & \text{其他} \end{cases}$$

$$f_X(x) = \begin{cases} 12x(1-x)^2 & 0 \le x \le 1 \\ 0 & \text{其他} \end{cases}$$

範例5.16（續）(承範例5.5)

- 將$f_X(x)$ 中以 y 取代 x 即可得$f_Y(y)$。可以很簡單地證明 $\mu_X = \mu_Y = \frac{2}{5}$, 且

範例5.16（續）(承範例5.5)

- 因此，$\text{Cov}(X, Y) = \frac{2}{15} - \left(\frac{2}{5}\right)\left(\frac{2}{5}\right) = \frac{2}{15} - \frac{4}{25} = -\frac{2}{75}$。變異數為負是合理的，因為杏仁愈多意味腰果愈少。

相關性

- 定義
  - X 與 Y 之**相關係數(correlation coefficient)**，以 $\text{Cov}(X, Y)$，$\rho_{X,Y}$，或僅以 ρ 表示，定義為

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \cdot \sigma_Y}$$

範例5.17

- 簡單證明範例 5.15 保險情境中，

相關性（續）

- 提要

1. 若a 與c 同時為正或負，則

$$\mathrm{Corr}\,(aX+b, cY+d)=\mathrm{Corr}\,(X, Y)$$

2. 對任意兩個隨機變數 X 與 Y,

$-1 \leq \mathrm{Corr}(X, Y) \leq 1$。

相關性（續）

- 提要1 說明相關係數不受度量單位所導致的線性轉換影響。
- 根據提要2，最強烈可能的正相關為$\rho = +1$，且最強烈可能的負相關為$\rho = -1$。
- 兩變數之關係於$|\rho| \geq 0.8$ 時被稱為強烈相關，$0.5 < |\rho| < 0.8$ 為溫和相關，$|\rho| \leq 0.5$ 為弱相關。

相關性（續）

- 提要

  1. 若 X 與 Y 為獨立，則 $\rho=0$，但 $\rho=0$ 並不意味獨立。

  2. 在某些數值 a 或 b，$\rho=1$或-1

  若且唯若 $Y=aX+b$，其中 $a\neq 0$。

範例5.18

- 令 X 與 Y 為離散型隨機變數，且聯合機率質量函數為

$$p(x, y) = \begin{cases} \dfrac{1}{4} & (x, y) = (-4, 1), (4, -1), (2, 2), (-2, -2) \\ 0 & \text{其他} \end{cases}$$

機率質量為正之點標示於圖5.5 的(x, y) 座標系統中。

範例5.18（續）

- 從圖中可以很明顯地看出 X 之值可以完全由 Y 之值來決定，且反之亦然，因此兩變數為完全相依。但是，因為

圖5.5

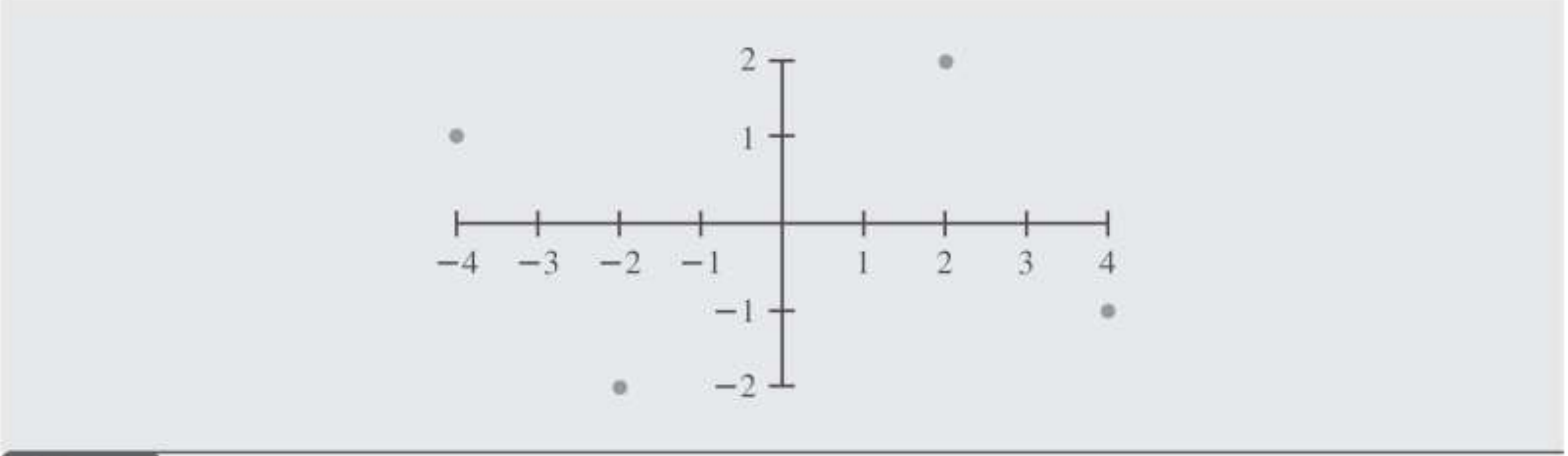


圖 5.5 範例 5.18 之配對母體

多隨機變數

- 定義
  - 若$X_1, X_2, \ldots, X_n$ 皆為離散型隨機變數，則這些變數之聯合機率質量函數為

$$p(x_1, x_2, \ldots, x_n)$$
$$= P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

多隨機變數（續）

- 定義（續）
  - 若隨機變數為連續，則$X_1, X_2, ..., X_n$ 之聯合機率密度函數$f(x_1, x_2, ..., x_n)$ 將使任意 n 區間$[a_1, b_1], ..., [a_n, b_n]$ 均能滿足

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n)$$

$$= \int_{a_1}^{b_1} \ldots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

多隨機變數（續）

- 在二項實驗中，每個試驗只有兩種可能的結果。現在考慮一由 n 次獨立且相同試驗所構成的實驗，其中每次試驗可能得到的結果是 r 種之一。令 $p_1 = P$(試驗產生第i 種結果)，且定義隨機變數 $X_i =$ 得到第i 種結果之次數($i = 1, ..., r$)。我們稱此種實驗為**多項實驗**(multinomial experiment)，且$X_1, ..., X_r$ 之聯合機率函數則稱為**多項分配**(multinomial distribution)。

多隨機變數（續）

- 可證明$X_1, ..., X_r$之聯合機率質量函數為

$$p(x_1, \ldots, x_r) = \begin{cases} \dfrac{n!}{(x_1!)(x_2!)\cdot\cdots\cdot(x_r!)} p_1^{x_1}\cdot\cdots\cdot p_r^{x_r} & x_i = 0, 1, 2, \ldots, \text{且 } x_1 + \cdots + x_r = n \\ 0 & \text{其他} \end{cases}$$

  在$r＝2$的情況下其為二項分配，且 $X_1＝$成功次數，$X_2＝n－X_1＝$失敗次數。

範例5.9

- 若測得十個獨立取得之豌豆切片的對偶基因(allele)，且$p_1 = P(\text{AA})$, $p_2 = P(\text{Aa})$, $p_3 = P(\text{aa})$, $X_1 = \text{AA}$ 發生次數, $X_2 = \text{Aa}$ 發生次數, 及$X_3 = \text{aa}$ 發生次數, 則這些X_i 之多項機率質量函數為

$$p(x_1, x_2, x_3) = \frac{10!}{(x_1!)(x_2!)(x_3!)} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$$x_i = 0, 1, \ldots \quad \text{且 } x_1 + x_2 + x_3 = 10$$

範例5.9（續）

- 若$p_1 = p_3 = 0.25$, $p_2 = 0.5$,

$$P(X_1 = 2, X_2 = 5, X_3 = 3)$$
$$= p(2, 5, 3)$$
$$= \frac{10!}{2!\,5!\,3!}(0.25)^2(0.5)^5(0.25)^3 = 0.0769$$

範例5.10

- 當利用某種方法蒐集某地區之固定體積的石頭樣本，會得到四種石頭種類。令X_1，X_2，與X_3 表示隨機抽樣所抽出的第1，2，與3 種石頭的體積比例(第四種石頭的體積比例為$1-X_1-X_2-X_3$，故變數X_4實為多餘)。

範例5.10（續）

- 若X_1，X_2，X_3 之聯合機率密度函數為

$$f(x_1, x_2, x_3) = \begin{cases} kx_1x_2(1-x_3) & 0 \le x_1 \le 1, 0 \le x_2 \le 1, 0 \le x_3 \le 1, x_1+x_2+x_3 \le 1 \\ 0 & \text{其他} \end{cases}$$

則k 可以下式決定

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1, x_2, x_3)\, dx_3\, dx_2\, dx_1$$

$$= \int_0^1 \left\{ \int_0^{1-x_1} \left[\int_0^{1-x_1-x_2} kx_1x_2(1-x_3)\, dx_3 \right] dx_2 \right\} dx_1$$

範例5.10（續）

- 此重積分之值為$k/144$， 故$k=144$ 。第1，2 種石頭共佔有體積比例最多為50%的機率為

$$P(X_1+X_2\leq 0.5)=\iiint\limits_{\left\{\substack{0\leq x_i\leq 1 \quad i=1,2,3 \\ x_1+x_2+x_3\leq 1,\ x_1+x_2\leq 0.5}\right\}} f(x_1,x_2,x_3)\,dx_3\,dx_2\,dx_1$$

$$=\int_0^{0.5}\left\{\int_0^{0.5-x_1}\left[\int_0^{1-x_1-x_2}144x_1x_2(1-x_3)\,dx_3\right]dx_2\right\}$$

$$=0.6066$$

多隨機變數（續）

- **獨立**(independent)
- 定義
  - 若其每一子集合 $X_{1_1}, X_{1_2}, \ldots, X_{i_k}$ (每兩個, 每三個等等) 之聯合機率質量函數或聯合機率密度函數均等於其邊際機率質量函數或邊際機率密度函數之乘積，則隨機變數$X_1, X_2, \ldots, X_n$ 稱為**獨立**(independent) 。

範例5.11

- 若$X_1, ..., X_n$ 表示n 個零件的使用壽命，這些零件互相獨立運作，且每個零件的使用壽命服從參數為λ 之指數分配，則

$$
\begin{aligned}
& f(x_1, x_2, \ldots, x_n) \\
&= (\lambda e^{-\lambda x_1}) \cdot (\lambda e^{-\lambda x_2}) \cdot \cdots \cdot (\lambda e^{-\lambda x_n}) \\
&= \begin{cases} \lambda^n e^{-\lambda \Sigma x_i} & x_1 \geq 0, x_2 \geq 0, \ldots, x_n \geq 0 \\ 0 & \text{其他} \end{cases}
\end{aligned}
$$

範例5.11（續）

- 若這些 n 個零件組成一個系統，其中一個零件故障，此系統即停止運作，則此系統可維持運作超過t 時間的機率為

$$
\begin{aligned}
P(X_1 > t, \ldots, X_n > t) &= \int_t^{\infty} \cdots \int_t^{\infty} f(x_1, \ldots, x_n)\, dx_1 \ldots dx_n \\
&= \left(\int_t^{\infty} \lambda e^{-\lambda x_1}\, dx_1\right) \cdots \left(\int_t^{\infty} \lambda e^{-\lambda x_n}\, dx_n\right) \\
&= (e^{-\lambda t})^n = e^{-n\lambda t}
\end{aligned}
$$

範例5.11（續）

- 因此,

$$P\,(\text{系統壽命} \leq t) = 1 - e^{-n\lambda t} \quad t \geq 0$$

此表示系統壽命為服從參數$n\lambda$ 之指數分配；系統壽命之期望值為$1/n\lambda$。

5.3 統計量與其分配

- **視每一觀察**值為隨機變數，並以X_1, $X_2, \ldots, X_n$ **表示樣本**。
- **觀察**值的變異進而意味任何樣本觀察值**函數之**值也隨樣本的不同而不同。
- $\overline{x}$ **之**值，s **之**值等等，同樣具有不確定性。

範例5.19

- 假設隨機抽樣某種原料之強度服從參數為$\alpha=2$(形狀)，與$\beta=5$(比例) 之韋伯分配(Weibull distribution) 。其對應之密度曲線如圖5.6 所示。

範例5.19（續）

- 由4.5 節之公式可得

  $\mu = E(X) = 4.4311$

  $\tilde{\mu} = 4.1628$

  $\sigma^2 = V(X) = 5.365$

  $\sigma = 2.316$

  因為此分配呈現右偏，平均數超過中位數。

範例5.19（續）

- 我們利用MINITAB 產生此分配下六組不同樣本，每組之$n=10$ (各有10 個材料強度之六個不同群組)。結果顯示於表5.1，表中包含每個樣本的樣本平均數，樣本中位數，與樣本標準差。

範例5.19（續）

- 首先請注意在任一組樣本中之10 個觀察值與其他樣本者完全不同。其次，六個樣本之樣本平均數，樣本中位數，與樣本標準差皆彼此相異。10% 的截尾平均數，四分散佈值等亦然。

圖5.6

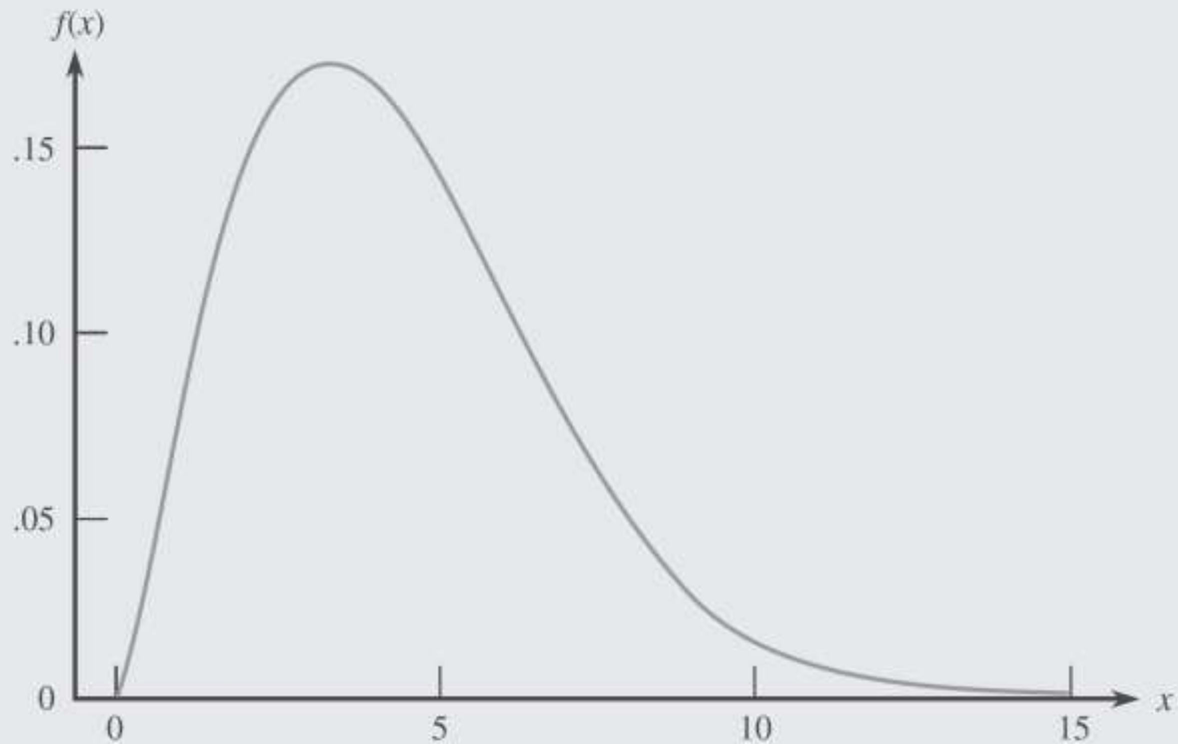


圖 5.6 範例 5.19 之韋伯分配密度曲線

表5.1

表 5.1 範例 5.19 中具韋伯分配之樣本

樣本	1	2	3	4	5	6
1	6.1171	5.07611	3.46710	1.55601	3.12372	8.93795
2	4.1600	6.79279	2.71938	4.56941	6.09685	3.92487
3	3.1950	4.43259	5.88129	4.79870	3.41181	8.76202
4	0.6694	8.55752	5.14915	2.49759	1.65409	7.05569
5	1.8552	6.82487	4.99635	2.33267	2.29512	2.30932
6	5.2316	7.39958	5.86887	4.01295	2.12583	5.94195
7	2.7609	2.14755	6.05918	9.08845	3.20938	6.74166
8	10.2185	8.50628	1.80119	3.25728	3.23209	1.75468
9	5.2438	5.49510	4.21994	3.70132	6.84426	4.91827
10	4.5590	4.04525	2.12934	5.50134	4.20694	7.26081
$\bar{x}$	4.401	5.928	4.229	4.132	3.620	5.761
$\tilde{x}$	4.360	6.144	4.608	3.857	3.221	6.342
s	2.642	2.062	1.611	2.124	1.678	2.496

範例5.19（續）

- 此外，任一特定樣本之樣本平均數皆可視為母體平均數μ ，其值已知為4.4311，之點估計值(“點”因其為單一數字，相當於數字線中之單一點) 。六組樣本所計算出的估計值與被估計值均不相同。第二個及第六個樣本之估計值皆過大，第五組樣本卻明顯低估。

範例5.19（續）

- 同樣地，樣本標準差提供母體標準差之點估計值。六組樣本所估計之值皆會與實際值多少有些微誤差。

範例5.19（續）

- 總之，個別樣本之觀察值將因樣本不同而有所不同，故通常任何計算自樣本資料之值與用於估計母體特徵值的樣本特徵值，實際上將不會與被估計之值完全吻合。

5.3 統計量與其分配（續）

- 定義
  - **統計量 (statistic)** 為任何由樣本資料計算所得之值。在取得資料之前，將產生的統計量之值為不確定的。因此，統計量為一隨機變數且以大寫字母表示；小寫字母則用於表示計算或觀察所得之統計量之值。

5.3 統計量與其分配（續）

- 因此樣本平均數，視為一統計量 (在樣本抽樣前或實驗完成前)，以 $\overline{X}$ 表示；此統計量之計算值為 $\overline{x}$。同樣地，S 表示樣本標準差，被視為一統計量，其計算值為 s。

5.3 統計量與其分配（續）

- 任何統計量，為一隨機變數，皆具有機率分配。
- 統計量之機率分配經常被稱為**樣本分配** (sampling distribution)。

隨機樣本

- 定義
  - 隨機變數 $X_1, X_2, ..., X_n$ 被稱為構成一組 (簡單) 樣本數 (或大小) 為 n 之**隨機樣本 (random samples)**，若

    1. X_i 為獨立隨機變數。

    2. 每個 X_i 皆有相同的機率分配。

樣本分配之推演

- 假設統計量為“非常簡單”之X_i 函數，則可利機率原理獲取統計量的分配。

範例5.20

- 一大型汽車服務中心對四，六，與八汽缸汽車之定期保養收費分別為$40， $45，與$50 。若20% 定期保養之汽車為四汽缸，30% 為六汽缸，與50% 為八汽缸，則隨機抽取一定期保養汽車，其收益的機率分配為

x	40	45	50
$p(x)$	0.2	0.3	0.5

其 $\mu = 46.5, \sigma^2 = 15.25$ (5.2)

範例5.20（續）

- 假設某天只有兩個服務與定期保養有關。令X_1＝第一個定期保養服務的收益，且X_2＝第二個的收益。假設X_1與X_2為獨立，其機率分配顯示於(5.2) 式[故X_1與X_2構成(5.2) 式分配之一隨機樣本]。表5.2列出可能的(x_1, x_2)配對，每個的發生機率[假設獨立，以(5.2) 式計算]，及$\bar{x}$與s^2之值。

範例5.20（續）

- 現在為獲得X，每次保養之樣本平均收益的機率分配，我們必須思考每個可能之x值以求算其機率。

表5.2

表 5.2 範例 5.20 之結果、機率，及 $\bar{x}$ 與 s^2 之值

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
40	40	0.04	40	0
40	45	0.06	42.5	12.5
40	50	0.10	45	50
45	40	0.06	42.5	12.5
45	45	0.09	45	0
45	50	0.15	47.5	12.5
50	40	0.10	45	50
50	45	0.15	47.5	12.5
50	50	0.25	50	0

範例5.20（續）

- 例如，$x=45$ ，表中出現三次，其機率為0.10，0.09 與0.10 ，故

$$p_{\overline{X}}(45) = P(\overline{X} = 45) = 0.10 + 0.09 + 0.10 = 0.29$$

同樣地，

$$\begin{aligned} p_{S^2}(50) &= P(S^2 = 50) \\ &= P(X_1 = 40, X_2 = 50 \quad \text{或} \quad X_1 = 50, X_2 = 40) \\ &= 0.10 + 0.10 = 0.20 \end{aligned}$$

範例5.20（續）

- (5.3) 與(5.4) 式為$\overline{X}$ 與S^2 完整的樣本分配。

$\bar{x}$	40	42.5	45	47.5	50
$p_{\overline{X}}(\bar{x})$	0.04	0.12	0.29	0.30	0.25

(5.3)

s^2	0	12.5	50
$p_{S^2}(s^2)$	0.38	0.42	0.20

(5.4)

範例5.20（續）

- 圖5.7 為(5.2) 式的原分配與(5.3) 式之$\overline{X}$分配的機率直方圖。此圖首先透露$\overline{X}$分配之平均數(期望值) 等於原分配的平均數46.5 ，因為兩者之直方圖的中心點位置相同。

圖5.7

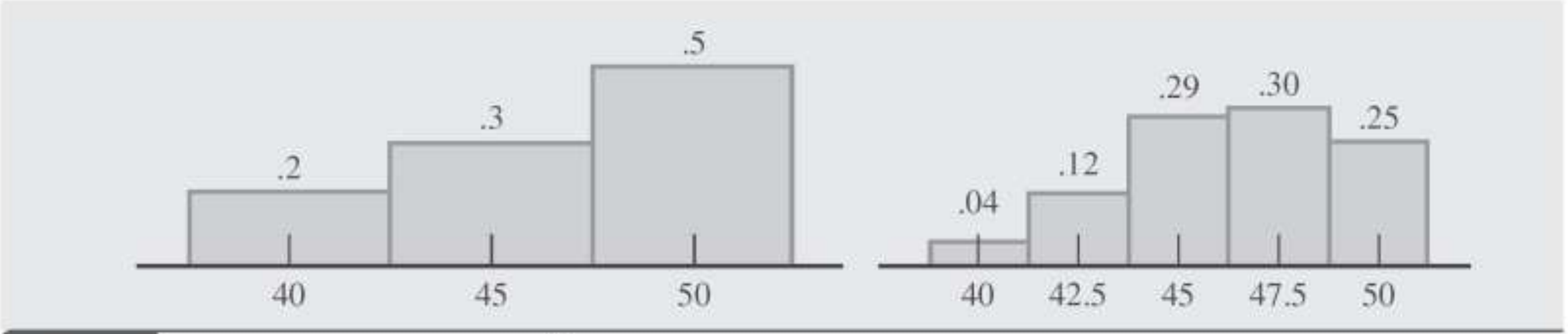


圖 5.7 範例 5.20 原分配與 $\bar{X}$ 分配之機率直方圖

範例5.20（續）

- 由(5.3) 式可知

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \overline{x} p_{\overline{X}}(\overline{x})$$
$$= (40)(0.04) + \cdots + (50)(0.25)$$
$$= 46.5 = \mu$$

範例5.20（續）

- 其次，$\overline{X}$ 分配之分散程度(變異程度) 較原分配為小，因為機率質量向平均數移動。再一次由(5.3) 式可知

$$\sigma_{\overline{X}}^2 = V(\overline{X}) = \sum \overline{x}^2 \cdot p_{\overline{X}}(\overline{x}) - \mu_{\overline{X}}^2$$
$$= (40)^2(0.04) + \cdots + (50)^2(0.25) - (46.5)^2$$
$$= 7.625 = \frac{15.25}{2} = \frac{\sigma^2}{2}$$

$\overline{X}$ 之變異數恰為原變異數的一半(因為n＝2)。

範例5.20（續）

- S^2 之平均數為

$$\mu_{S^2} = E(S^2) = \sum s^2 \cdot p_{S^2}(s^2)$$
$$= (0)(0.38) + (12.5)(0.42) + (50)(0.20)$$
$$= 15.25 = \sigma^2$$

亦即，X 之樣本分配以母體平均數μ 為中心，且S^2 之樣本分配以母體變異數σ^2 為中心。

範例5.20（續）

- 若某天有四個定期保養，則樣本平均數$\overline{X}$將依據四個X_i 的隨機樣本，每個皆服從 (5.2) 式的分配。經過較多的計算最後產生如下$n=4$ 之$\overline{X}$ 的機率質量函數

$\bar{x}$	40	41.25	42.5	43.75	45	46.25	47.5	48.75	50
$p_{\overline{X}}(\bar{x})$	0.0016	0.0096	0.0376	0.0936	0.1761	0.2340	0.2350	0.1500	0.0625

由此可得，$\mu_{\overline{X}}=46.50=\mu$與
$\sigma^2_{\overline{X}} = 3.8125 = \sigma^2/4$ 。

範例5.20（續）

- 圖5.8 為此機率質量函數之機率直方圖。

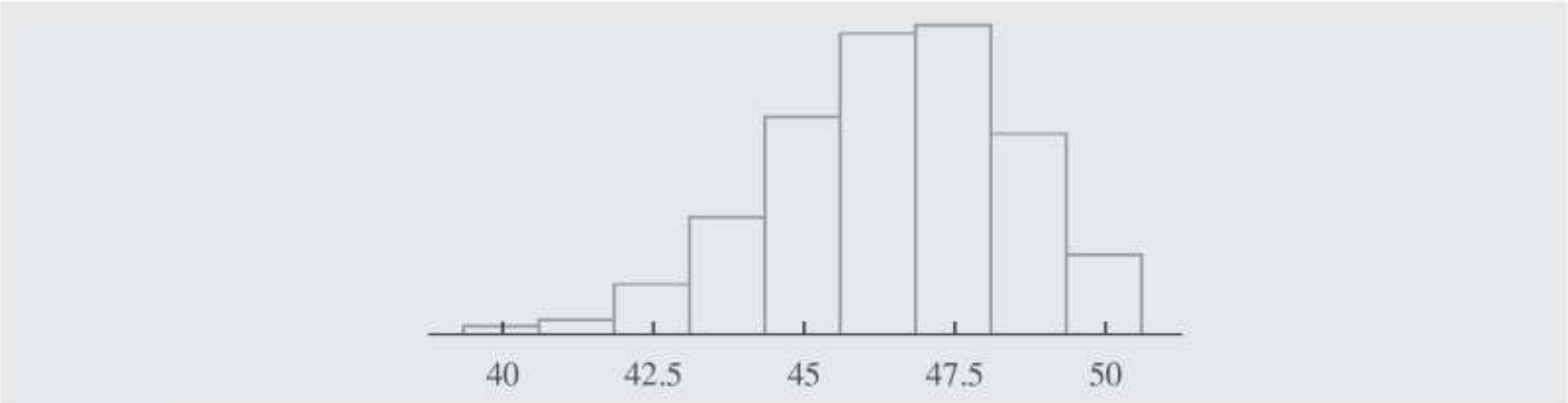


圖 5.8 範例 5.20，在 $n=4$ 時，$\overline{X}$ 之機率直方圖

範例5.21

- 銀行某類型交易的服務時間為一隨機變數，其服從參數為λ之指數分配。假設X_1與X_2為兩名不同客戶的服務時間，且彼此獨立。考慮這兩名顧客的總服務時間$T_o = X_1 + X_2$，亦為一統計量。

範例5.21（續）

- T_o 之累積分配函數，於$t \geq 0$ 時，

$$F_{T_o}(t) = P(X_1 + X_2 \leq t) = \iint_{\{(x_1, x_2): x_1 + x_2 \leq t\}} f(x_1, x_2)\, dx_1\, dx_2$$

$$= \int_0^t \int_0^{t-x_1} \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2}\, dx_2\, dx_1 = \int_0^t [\lambda e^{-\lambda x_1} - \lambda e^{-\lambda t}]\, dx_1$$

$$= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}$$

積分範圍顯示於圖5.9 。

圖5.9

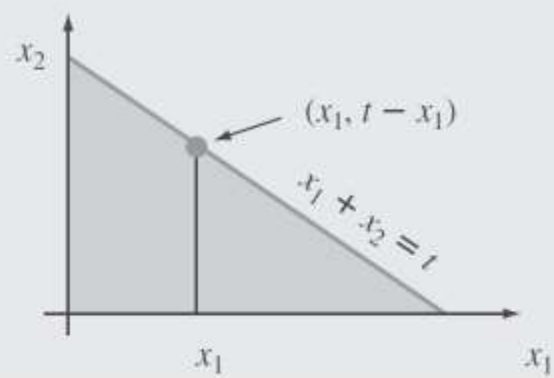


圖 5.9 範例 5.21 求取 T_o 累積分配函數的積分範圍

範例5.21（續）

- 將$F_{T_o}(t)$ 微分可獲得T_o 之機率密度函數。

$$f_{T_o}(t) = \begin{cases} \lambda^2 t e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \qquad (5.5)$$

此為伽瑪機率密度函數($\alpha=2$ 且$\beta=1/\lambda$)。

範例5.21（續）

- $\overline{X} = T_o/2$ 之機率密度函數，可藉由$\{\overline{X} \leq \overline{x}\}$若且唯若$\{T_o \leq 2\overline{x}\}$ 的關係求得，如

$$f_{\overline{X}}(\overline{x}) = \begin{cases} 4\lambda^2 \overline{x} e^{-2\lambda \overline{x}} & \overline{x} \geq 0 \\ 0 & \overline{x} < 0 \end{cases} \qquad (5.6)$$

原指數分配之平均數與變異數為$m = 1/\lambda$與$\sigma^2 = 1/\lambda^2$。

範例5.21（續）

- 由(5.5) 與(5.6) 式，可證明$E(\overline{X}) = 1/\lambda$，$V(\overline{X}) = 1/(2\lambda^2)$，$E(T_o) = 2/\lambda$ 與$V(T_o) = 2/\lambda^2$。此一結果再度透露X，T_o 之平均數及變異數與具原分配者有著某種普遍關係。

模擬實驗

- 第二個獲得有關統計量之樣本分配的方法為進行模擬實驗。此種方法通常用於當以機率原理推演過於困難或複雜時。這種實驗實際上皆需由電腦輔助執行。必須說明實驗的特性如下：
  1. 感興趣之統計量($\overline{X}$，S^2，某特定之截尾平均數等)
  2. 母體分配(μ＝100 與σ＝15 之常態，低限A＝5 與高限B＝10 之均勻分配等)
  3. 樣本數 n (如，n＝10 或n＝50)
  4. 重複次數 k (如，k＝500)

模擬實驗（續）

- 則利用電腦取得 k 個不同的隨機樣本，每個均來自指定的母體分配，樣本數皆為 n。對於每個樣本，計算其統計量，並繪製 k 個計算值之直方圖。此直方圖提供統計量的近似樣本分配。k 之值愈大，將會愈接近實際分配($k \to \infty$ 時，將會得到實際的樣本分配)。
- 若統計量"相當簡單"，通常k＝500 或 1,000 就足夠。

範例5.22

- 我們第一個模擬實驗的母體服從$\mu = 8.25$與$\sigma = 0.75$之常態分配，如同圖5.10。

圖5.10

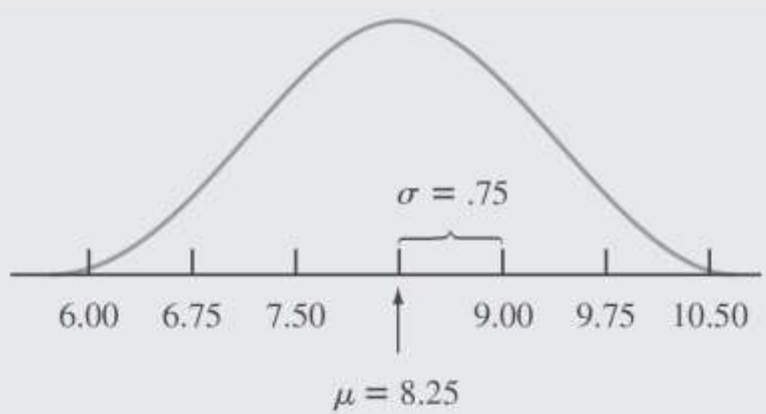


圖 5.10 常態分配，$\mu=8.25$ 與 $\sigma=0.75$

範例5.22（續）

- 我們實際上執行了四個不同的實驗，且每個實驗重複進行500 次。在第一次實驗中，利用MINITAB 軟體產生500 個$n=5$之觀察值樣本，其他三個樣本數分別為$n=10$，$n=20$ 與$n=30$ 。每個樣本皆計算樣本平均數，其$\overline{x}$ 之直方圖示如圖5.11 。

範例5.22（續）

- 首先要注意的是直方圖的形狀。就合理近似性而言，以上四個圖形皆類似常態曲線。若每個直方圖由超出500 甚多的$\bar{x}$值所組成，則圖形會更明顯地接近常態。其次，每個直方圖的中心點皆近似8.25，即被抽樣之母體平均數。若直方圖建立於無限多之$\bar{x}$ 值序列上，則其中心會正好為母體平均數，8.25 。

範例5.22（續）

- 直方圖要注意的最後一點為它們彼此間的分散程度。n 值愈小，則樣本分配相對於平均數的分散範圍就愈大。這就是為什麼相較於另外兩個較小的樣本數，$n=20$ 與$n=30$ 之直方圖建立在較窄的組距上。對於較大之樣本數，大部分的x 值皆非常接近8.25 。此即為平均的效果。

範例5.22（續）

- 當n 為小時，一特殊的x 值將導致值距離中心較遠。當樣本數較大時，任何特殊的x 值，在與其他樣本值平均下，仍舊會得到一接近μ 之$\bar{x}$ 值。結合以上入微觀察所獲得的結果，應該與你的直覺相呼應：**建立在較大 n 之$\overline{X}$ 比建立在小 n 之$\overline{X}$ 有距離 μ 較近的趨勢**。

圖5.11

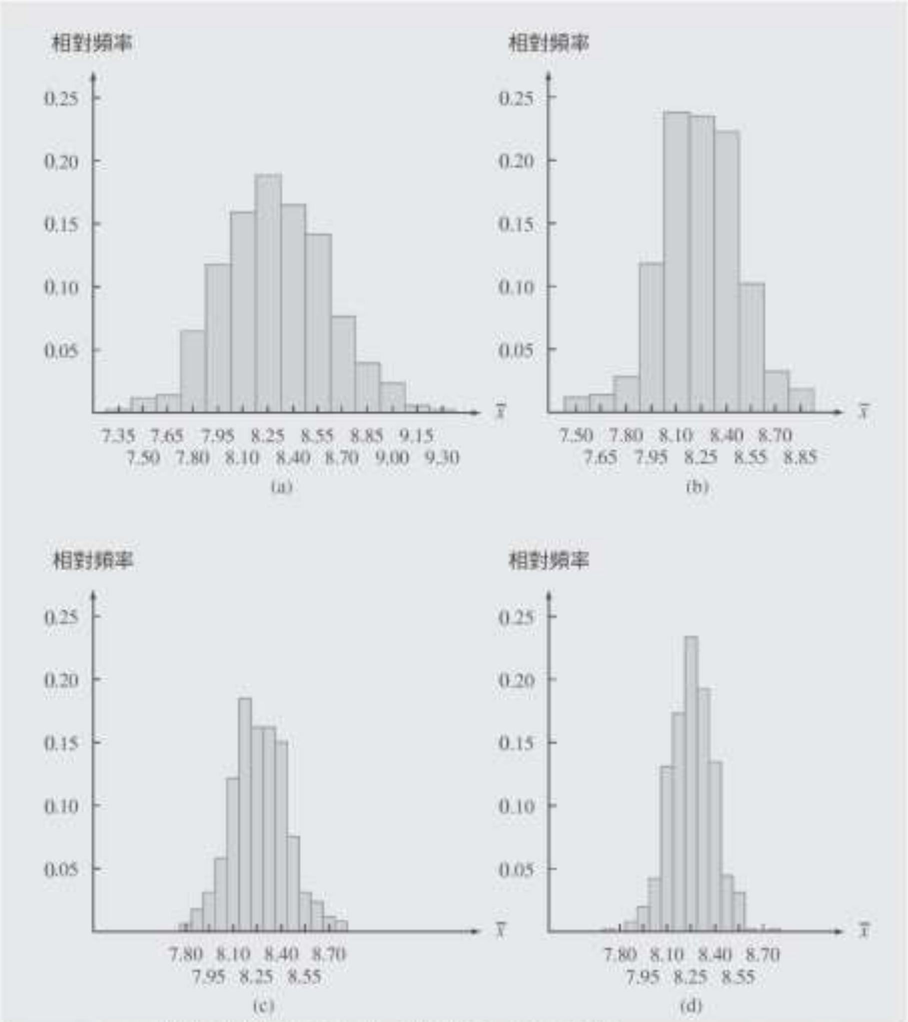


圖 5.11 500 個樣本的樣本直方圖，每個皆含 n 個觀察值：(a) $n=5$；(b) $n=10$；(c) $n=20$；(d) $n=30$

範例5.23

- 考慮一母體分配非常偏斜之模擬實驗。圖 5.12 為某型電子控制器壽命的密度曲線 [此實際上為一對數常態分配，其 $E(\ln(X)) = 3$ 與 $V(\ln(X)) = 1.6$]。感興趣的統計量仍是樣本平均數 $\bar{X}$。

範例5.23（續）

- 此一實驗重複 500 次，且如同範例 5.22，考慮四種樣本數。所產生之直方圖與由 MINITAB 根據 500 個 $n = 30$ 上之 $\bar{x}$ 值所繪製之常態機率圖，顯示於圖 5.13。

圖5.12

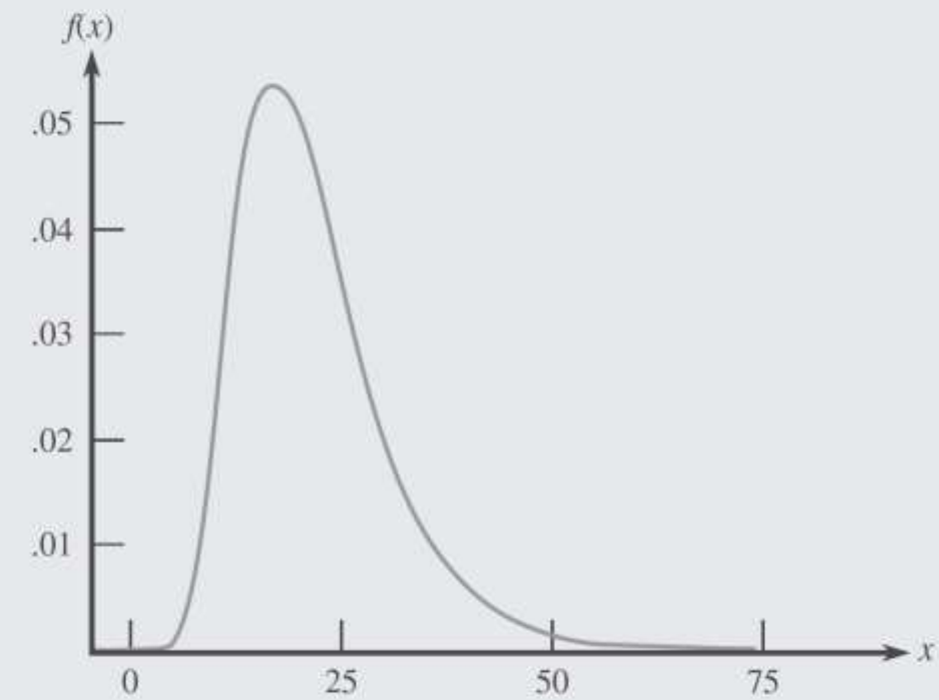


圖 5.12 範例 5.23 模擬實驗的密度曲線 [$E(X)=2.7584$，$V(X)=82.1449$]

範例5.23（續）

- 不同於常態情況，這些直方圖的形狀皆不同。尤其，隨著樣本數n 的增加，它們的偏斜程度漸緩。四個不同樣本數之500個$\bar{x}$ 值的平均數皆非常靠近母體分配之平均數。若每個直方圖皆建立在無限的$\bar{x}$ 值序列上，而非只有500 ，則四個樣本之中心點將恰為21.7584 。因此不同的n 值會使形狀改變，但樣本分配之中心點$\bar{X}$ 則不變。

範例5.23（續）

- 比較圖5.13 之四個直方圖亦顯示當n 增加時，直方圖之分散程度減少。增加n 值將導致更高度集中於母體平均數，並使直方圖更接近常態曲線。圖5.13(d) 之直方圖與圖5.13(e) 之常態機率圖提供有效的證據證明樣本數$n=30$ 足以克服母體分配之偏斜，並使$\overline{X}$ 的樣本分配近似常態。

圖5.13

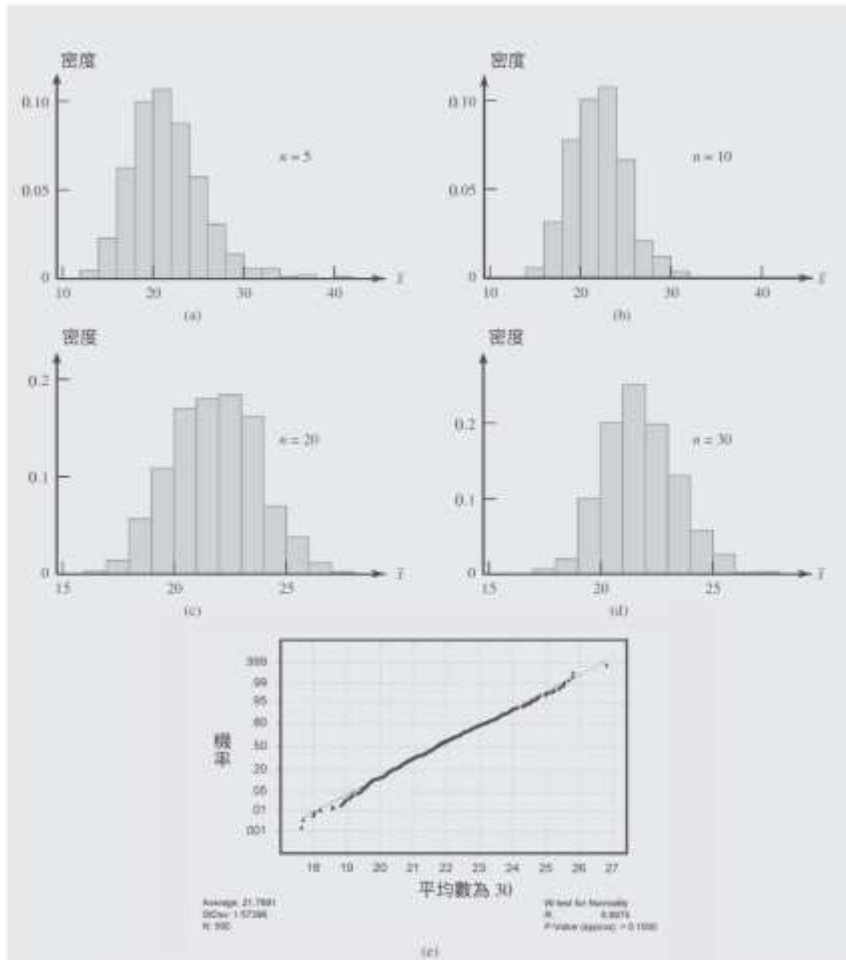


圖 5.13 範例 5.23 之模擬實驗結果：(a) $n=5$ 之 $\bar{x}$ 直方圖；(b) $n=10$ 之 $\bar{x}$ 直方圖；(c) $n=20$ 之 $\bar{x}$ 直方圖；(d) $n=30$ 之 $\bar{x}$ 直方圖；(e) $n=30$ 之常態機率圖 (輸出自 MINITAB)

5.4 樣本平均數之分配

- 提要
  - 令$X_1, X_2, \ldots, X_n$ 為取自平均數μ 與標準差σ 之分配的隨機樣本。則

    1. $E(\overline{X}) = \mu_{\overline{X}} = \mu$

    2. $V(\overline{X}) = \sigma^2_{\overline{X}} = \sigma^2/n$ 且 $\sigma_{\overline{X}} = \sigma/\sqrt{n}$

    此外，令$T_o = X_1 + \ldots + X_n$ (樣本總和)，則$E(T_o) = n\mu$,，與 $\sigma_{T_o} = \sqrt{n}\sigma$。

5.4 樣本平均數之分配（續）

- 結果1，$\overline{X}$ 之樣本(即機率) 分配的中心點正是取樣母體的平均數。
- 結果2 ，當樣本數n 增加，$\overline{X}$ 分配將更向μ 集中。

範例5.24

- 於一鈦缺口樣品之張力疲勞測試 (notched tensile fatigue test) 中，第一次聲頻發射 (用以指示裂縫形成) 之週期數的期望值為 μ ＝28,000 且週期數的標準差為 σ＝5,000。令 $X_1, X_2, ..., X_{25}$ 為一大小為 25 的隨機樣本，其中每個 X_i 皆表示不同隨機選取樣品的週期數。

範例5.24（續）

- 則發生第一次聲頻發射前之週期數樣本平均數之期望值為 $E(\overline{X}) = \mu = 28{,}000$，且 25 個樣品之總週期數的期望值為 $E(T_o) = n\mu = 25(28{,}000) = 700{,}000$。$X$ 與 T_o 之標準差為

$$\sigma_{\overline{X}} = \sigma/\sqrt{n} = \frac{5{,}000}{\sqrt{25}} = 1{,}000$$

$$\sigma_{T_o} = \sqrt{n}\sigma = \sqrt{25}(5{,}000) = 25{,}000$$

範例5.24（續）

- 若樣本數增加到$n=100$，$E(\overline{X})$ 不變，但 $\sigma_{\overline{X}}=500$為前值的一半(樣本數必須為原來的四倍才能使$\overline{X}$之標準差變為原來的一半)。

常態母體分配情況

- 提要
  - 令$X_1, X_2, \ldots, X_n$ 為取自平均數μ與標準差σ之常態分配的隨機樣本。則對任何n，X 皆為常態分配(平均數μ與標準差$\sqrt{n}\sigma$)。

常態母體分配情況（續）

- 當母體分配為常態時，我們可得知$\overline{X}$與T_o分配的所有資訊。尤其，如$P(a \le \overline{X} \le b)$與$P(c \le T_0 \le d)$ 等的機率可藉由標準化求得。圖5.14 說明此提要。

圖5.14

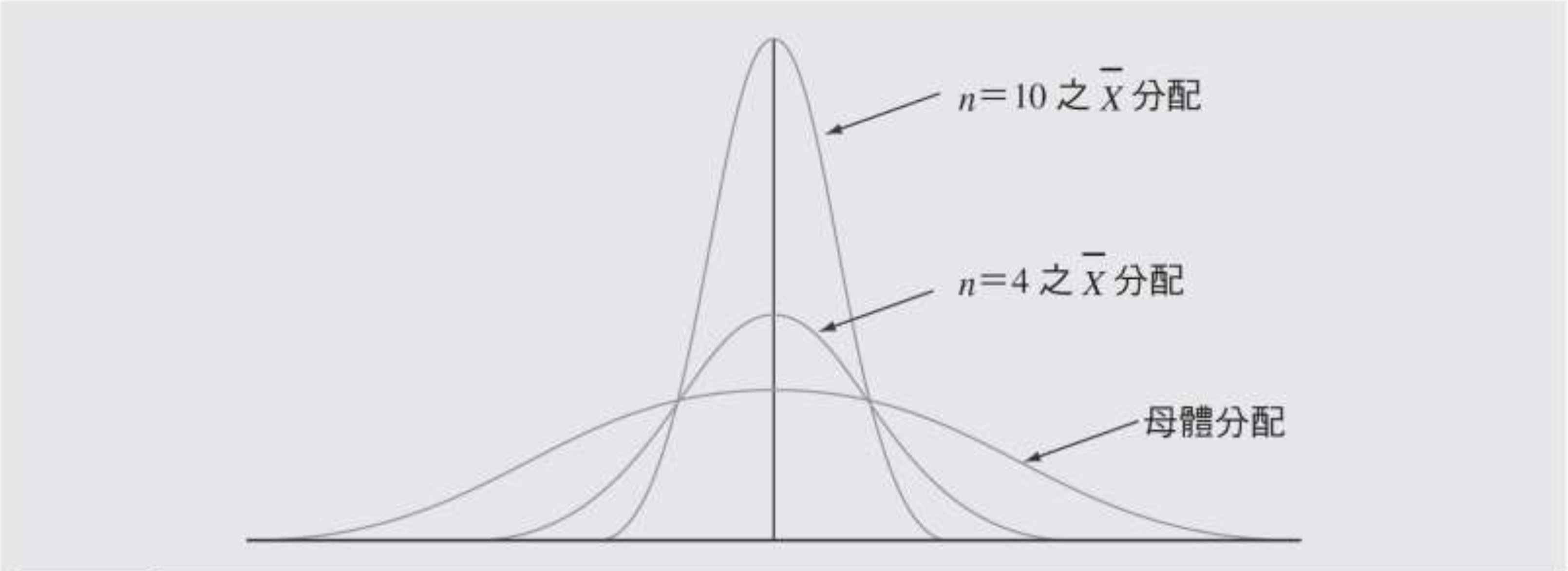


圖 5.14 常態母體分配與 $\overline{X}$ 之樣本分配

範例5.25

- 隨機選取一隻某亞種的老鼠，其通過迷宮的時間為一服從μ ＝1.5 分鐘，且σ＝0.35 分鐘之常態分配隨機變數。假設隨機抽樣五隻老鼠。令$X_1, \ldots, X_5$ 表示牠們停留在迷宮的時間。假設X_i 為此常態分配的隨機樣本，五隻老鼠的時間總和T_o ＝X_1＋⋯＋X_5 介於6 到8 分鐘之間的機率為何？

範例5.25（續）

- 從提要中得知，T_o 服從常態分配，其μ_{T_o} ＝$n\mu$ ＝5(1.5)＝7.5 且變異數為$\sigma^2_{T_o}$ ＝$n\sigma^2$ ＝5(0.1225)＝0.6125，故標準差為σ_{T_o} ＝ 0.783 。

範例5.25（續）

- 為將T_o標準化，將其減去μ_{T_o}，再除以σ_{T_o}：

$$P(6 \le T_o \le 8) = P\left(\frac{6-7.5}{0.783} \le Z \le \frac{8-7.5}{0.783}\right)$$

$$= P(-1.92 \le Z \le 0.64)$$

$$= \Phi(0.64) - \Phi(-1.92) = 0.7115$$

範例5.25（續）

- 求算樣本平均時間$\overline{X}$(常態分配變數) 至多為2.0 分鐘的機率，需要與 $\mu_{\overline{X}}=\mu=1.5$ 及 $\sigma_{\overline{X}}=\sigma/\sqrt{n}=0.35/\sqrt{5}=0.1565$。則

$$P(\overline{X}\le 2.0)=P\left(Z\le\frac{2.0-1.5}{0.1565}\right)$$
$$=P(Z\le 3.19)$$
$$=\Phi(3.19)=0.9993$$

中央極限定理

- 定理　**中央極限定理(CLT)**
  - 令$X_1, X_2, \ldots, X_n$為取自平均數μ，變異數σ^2之常態分配的隨機樣本。若n夠大，則$\overline{X}$近似服從常態分配，其$\mu_{\overline{X}} = \mu$與$\sigma^2_{\overline{X}} = \sigma^2/n$，且$T_o$亦近似服從常態分配，其$\mu_{T_o} = n\mu$與$\sigma^2_{T_o} = n\sigma^2$之值愈大，近似性愈佳。

中央極限定理（續）

- 圖5.15 為中央極限定理的敘述。根據中央極限定理，當n 夠大，且我們欲計算如$P(a \le \overline{X} \le b)$ 之機率，只要“假裝”$\overline{X}$ 為常態，將其標準化，並利用常態表。所得的答案將接近正確。

圖5.15

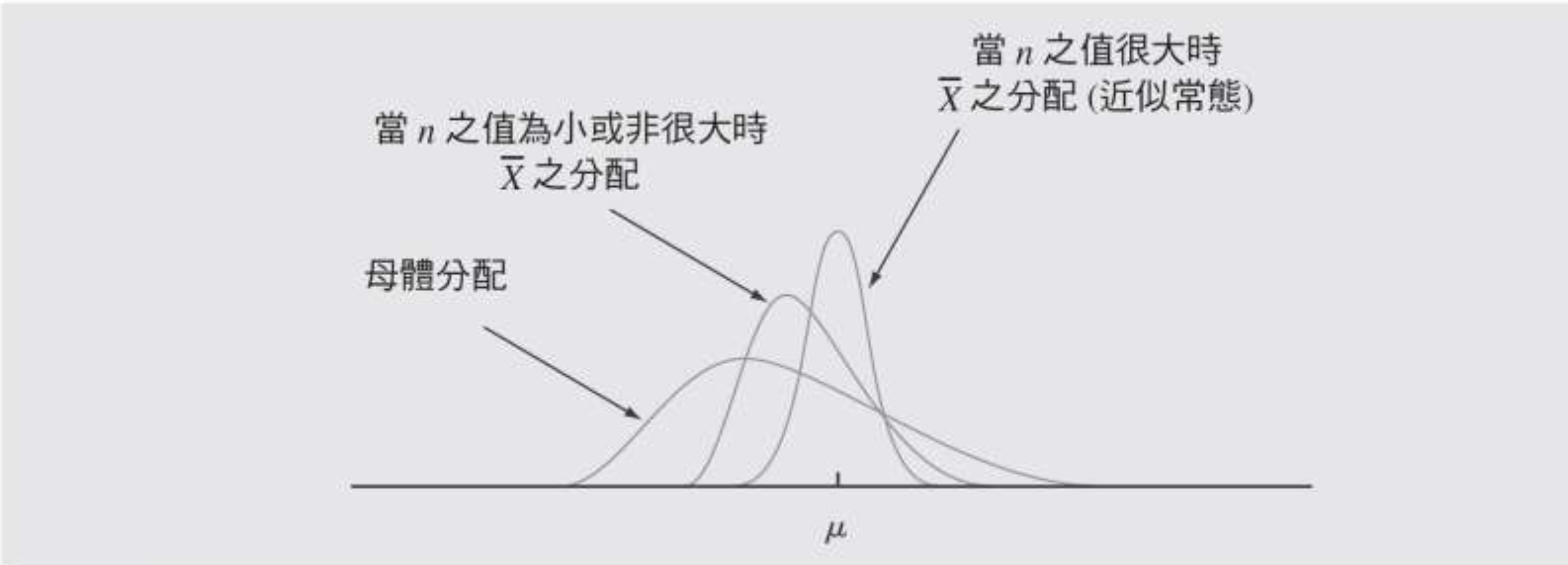


圖 5.15 中央極限定理的敘述

範例5.26

- 一批化學製品中驗出某種特別雜質含量為一隨機變數，其平均數為4.0 克，標準差為1.5 克。若籌備50 批獨立的化學製品，雜質含量之樣本平均數 X 介於3.5 到3.8 克的(近似) 機率為何？

範例5.26（續）

- 依據稍後將陳述的經驗法則，$n = 50$ 已經大到可以使用中央極限定理。因此，$\bar{X}$ 近似服從 $\mu_{\bar{X}} = 4.0$ 與 $\sigma_{\bar{X}} = 1.5/\sqrt{50} = 0.2121$ 之常態分配，故

$$P(3.5 \leq \bar{X} \leq 3.8) \approx P\left(\frac{3.5-4.0}{0.2121} \leq Z \leq \frac{3.8-4.0}{0.2121}\right)$$
$$= \Phi(-0.94) - \Phi(-2.36) = 0.1645$$

範例5.27

- 某消費者組織例行性報告其測試每部新車的重大瑕疵數量。假設某一模型的新車，其重大瑕疵數量為一隨機變數，平均數為3.2 ，標準差為2.4 。於此模型新車的100 個隨機取樣之車子中，重大瑕疵數量之樣本平均數超過4 的可能性為何？

範例5.27（續）

- 令X_i 表示隨機樣本第i 部車的重大瑕疵數量。注意X_i 為一離散型隨機變數，但無論變數為離散型或連續型，中央極限定理皆適用。同時，相較於平均數，雖然此一非負變數的標準差算是非常大，透露此分配為右偏，但在大樣本數情況下，$\overline{X}$ 仍近似常態分配。

範例5.27（續）

- 利用 $\mu_{\overline{X}}=3.2$ 與 $\sigma_{\overline{X}}=2.4$，

$$P(\overline{X} > 4) \approx P\left(Z > \frac{4-3.2}{0.24}\right)$$

$$= 1 - \Phi(3.33) = 0.0004$$

中央極限定理（續）

- 經驗法則
  - 若 $n > 30$，適用中央極限定理。

中央極限定理之其他應用

- 中央極限定理可佐證第四章介紹之以常態近似法估計二項分配機率的正當性。二項分配之變數 X ，為二項實驗中的成功次數，實驗包含 n 個獨立成功／失敗的試驗，且任何試驗之 $p=P(S)$。現在定義新的隨機變數 $X_1, X_2, \ldots, X_n$ 為

$$X_i = \begin{cases} 1 & \text{若第 } i \text{ 次試驗為成功} \\ 0 & \text{若第 } i \text{ 次試驗為失敗} \end{cases} \qquad (i = 1, \ldots, n)$$

中央極限定理之其他應用（續）

- 因為試驗間為獨立，且每次試行之 $P(S)$ 固定，故 Xi 為 iid (一取自伯努利分配之隨機樣本)。中央極限定理表示若 n 夠大，則 X_i 之總和與平均數皆近似於常態分配。
- 當 X_i 全部加總，每個發生 S 的皆加 1，且每個 F 的皆加 0，所以 $X_1+X_2+\cdots+X_n=X$。X_i 之樣本平均數為 X/n，即成功的樣本比例。

中央極限定理之其他應用（續）

- 當 n 夠大時，X 和 X/n 皆近似常態。近似常態所需的樣本數決定於 p 值。當 p靠近 0.5 時，每個 X_i 的分配皆相當對稱 (見圖 5.17)，當 p 接近 0 或 1 時，分配則相當偏斜。欲利用常態近似法，唯有當 $np \geq 10$ 與 $n(1-p) \geq 10$ 才適合，其可保證 n 夠大到足以克服伯努利分配的任何偏斜。

圖5.17

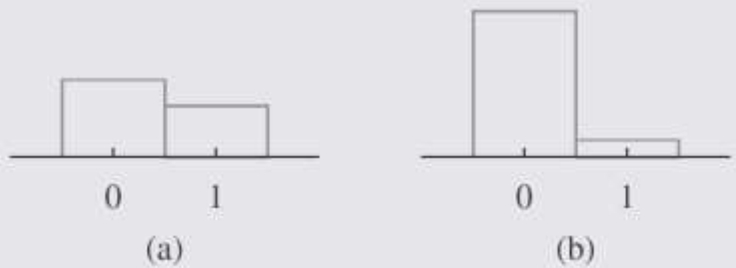


圖 5.17 兩個伯努利分配：(a) $p5-4$ (相當對稱)；(b) $p5-1$ (十分偏斜)

中央極限定理之其他應用（續）

- 提要
  - 令 $X_1, X_2, \ldots, X_n$為來自僅可能出現正值分配 $[P(Xi\cdot 0)=1]$ 之隨機樣本。則若 n 夠大，乘積 $Y= X_1X_2\cdots\cdot X_n$ 將近似對數常態分配。

5.5 線性組合之分配

- 定義
  - 已知n 個隨機變數$X_1, \ldots, X_n$ 與n 個常數$a_1,\ldots, a_n$，則隨機變數

$$Y = a_1X_1 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i \qquad (5.7)$$

稱為X_i 之**線性組合(linear combination)** 。

5.5 線性組合之分配（續）

- 提要
  - 令$X_1, X_2, \ldots, X_n$ 之平均數分別為$\mu_1, \mu_2, \ldots, \mu_n$，變異數分別為$\sigma^2{}_1, \sigma^2{}_2, \ldots, \sigma^2{}_n$。

    1.無論Xi 是否獨立，

$$
\begin{aligned}
&E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\
&= a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n) \\
&= a_1\mu_1 + \cdots + a_n\mu_n \qquad (5.8)
\end{aligned}
$$

5.5 線性組合之分配（續）

- 提要（續）

2. 若$X_1, X_2, ..., X_n$為獨立，則

$$
\begin{aligned}
&V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) \\
&= a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) \\
&= a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 \qquad (5.9)
\end{aligned}
$$

且

$$\sigma_{a_1X_1+\cdots+a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2} \quad (5.10)$$

5.5 線性組合之分配（續）

- 提要（續）

  3. 對於任何$X_1, X_2, ..., X_n$ ，

$$V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j \mathrm{Cov}(X_i, X_j) \tag{5.11}$$

範例5.28

- 加油站販售三種等級的汽油；普通，高級，與超級。定價分別為每加侖\$21.20，\$21.35 ，與\$21.50。令X_1，X_2，與X_3表示這些汽油某日的銷售量(加侖)。

範例5.28（續）

- 假設X_i 為獨立，且$\mu_1=1{,}000$, $\mu_2=500$, $\mu_3=300$, $\sigma_1=100$, $\sigma_2=80$ 與$\sigma_3=50$。其販售的收益為$Y=21.2X_1+21.35X_2+21.5X_3$，且

$$E(Y) = 21.2\mu_1 + 21.35\mu_2 + 21.5\mu_3 = \$4{,}125$$

$$V(Y) = (21.2)^2\sigma_1^2 + (21.35)^2\sigma_2^2 + (21.5)^2\sigma_3^2 = 104{,}025$$

$$\sigma_Y = \sqrt{104{,}025} = \$322.53$$

兩隨機變數之差異

- 推論
  - $E(X_1-X_2)=E(X_1)-E(X_2)$，且若 X_1 與 X_2 獨立，則 $V(X_1-X_2)=V(X_1)+V(X_2)$。

範例5.29

- 某汽車製造商之某型汽車裝置有六汽缸引擎，或四汽缸引擎。令X_1與X_2分別為獨立且隨機選取之六汽缸引擎與四汽缸引擎汽車的燃油效率。

範例5.29（續）

- 已知$\mu_1 = 22$，$\mu_2 = 26$，$\sigma_1 = 1.2$，且$\sigma_2 = 1.5$，

$$E(X_1 - X_2) = \mu_1 - \mu_2 = 22 - 26 = -4$$

$$V(X_1 - X_2) = \sigma_1^2 + \sigma_2^2 = (1.2)^2 + (1.5)^2 = 3.69$$

$$\sigma_{X_1 - X_2} = \sqrt{3.69} = 1.92$$

範例5.29（續）

- 若我們將變數變更標籤，故X_1為四汽缸引擎者，則$E(X_1 - X_2) = 4$，但差之變異數仍為3.69 。

常態隨機變數情況

- 提要
  - 若$X_1, X_2, \ldots, X_n$為互相獨立的常態隨機變數(可能具不同的平均數與／或變異數)，則任何Xi 的線性組合亦為常態分配。尤其，兩獨立常態分配變數差$X_1 - X_2$ 為常態分配。

範例5.30(承範例5.28)

- 某日販售三種等級汽油所得之總收益為$Y = 21.2X_1 + 21.35X_2 + 21.5X_3$，經計算(承範例$\mu_Y = 4{,}125$ 與(假設為獨立) $\sigma_Y = 322.53$。若X_i 為常態分配，收益超過4,500 的5.28) 機率為：

$$P(Y > 4{,}500) = P\left(Z > \frac{4{,}500 - 4{,}125}{322.53}\right)$$

$$= P(Z > 1.16) = 1 - \Phi(1.16) = 0.1230$$