

第三章

多維隨機變數

3.1 聯合及邊際分佈

到目前為止，我們所討論的隨機變數，都是所謂單變數(univariate)，本章我們要引進多變數(multivariate)，或說多維(或說多變量)隨機變數。

在進行一隨機試驗，較少的時候才只觀測一隨機變數。很少有說只收集一個數據的。例如，做民調，不會只調查一個人；要了解某地區之國民所得，不會只取得一個人的所得。有時候我們要比較兩次考試的成績，則對同一個學生便有兩個數據。身高與體重間之關係，也是兩個變數。這些例子，足以說明多維隨機變數也是我們必須熟悉的。

所謂 **n 維的隨機向量**(n -dimensional random vector)，為一由樣本空間 Ω 映至 n 維歐氏空間 R^n 的函數。有時會說 **n 維(或 n 變數)的隨機變數**。本章頭幾節，我們主要討論 $n = 2$ 的情況， $n > 2$ 的情況至第3.7節再介紹。

假設對 $\forall \omega \in \Omega$ ，有一二維的點 (x, y) 與其對應， $x, y \in R$ ，則便定義出一個二維的隨機向量。

例1.1 投擲一公正骰子兩次，樣本空間 Ω 給在第一章例2.3，共有36個元素。對一樣本點 $\omega = (i, j) \in \Omega$ ，表第一次出現 i ，第二次出現 j 。現定義二

隨機變數

$$X = \text{兩次點數和} = i + j,$$

$$Y = \text{兩次點數差之絕對值} = |i - j|。$$

例如, 若 $(i, j) = (3, 4)$, 則 $X = 7, Y = 1$ 。而對 $(i, j) = (4, 3)$, 亦有 $X = 7, Y = 1$ 。

定義出隨機向量 (X, Y) , 我們便可以求經由 (X, Y) 所定義出的事件之機率。這種事件, 當然都還是 Ω 的一個子集合, 因此才能求出其機率。例如, 我們可以問 $(X, Y) = (7, 1)$, 即 $X = 7$ 且 $Y = 1$, 之機率為何? 即要求 $P((X, Y) = (7, 1))$, 或是 $P(X = 7 \text{ 且 } Y = 1)$ 。爲了簡便, 我們常將後者寫成 $P(X = 7, Y = 1)$, 即以逗號“, ”表“且”。

不難驗證 Ω 中, 使 $X = 7$ 且 $Y = 1$ 的元素, 恰就是 $\omega_1 = (3, 4)$, $\omega_2 = (4, 3)$ 二樣本點。由於每一樣本點之機率皆爲 $1/36$ 。故

$$P(X = 7, Y = 1) = P(\{(3, 4), (4, 3)\}) = \frac{2}{36} = \frac{1}{18}。$$

另外,

$$\begin{aligned} P(X = 7, Y \leq 4) \\ = P(\{(4, 3), (3, 4), (5, 2), (2, 5)\}) = \frac{4}{36} = \frac{1}{9}。 \end{aligned}$$

表1.1 例1.1中 (X, Y) 之聯合p.d.f及邊際p.d.f.

$y \setminus x$	2	3	4	5	6	7	8	9	10	11	12	$f_Y(y)$
0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	0	$\frac{1}{36}$	$\frac{1}{6}$
1	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{5}{18}$
2	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	$\frac{2}{9}$
3	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0	$\frac{1}{6}$
4	0	0	0	0	$\frac{1}{18}$	0	$\frac{1}{18}$	0	0	0	0	$\frac{1}{9}$
5	0	0	0	0	0	$\frac{1}{18}$	0	0	0	0	0	$\frac{1}{18}$
$f_X(x)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$	1

上例中的隨機向量為離散型，因它所有可能的值是可數的(事實上為有限)。對離散型的隨機向量，令

$$f(x, y) = P(X = x, Y = y),$$

稱為 (X, Y) (或省略括號, 只寫 X, Y)之**聯合機率密度函數**(joint probability density function), 或聯合p.d.f.。有時為了強調是 (X, Y) 之聯合p.d.f., 也可寫成 $f_{X,Y}(x, y)$ 。表1.1給出例1.1中 (X, Y) 之聯合p.d.f.。

如同單變數的情況, 由 (X, Y) 之聯合p.d.f., 可完全決定 (X, Y) 之機率分佈。即對 $\forall A \subset R^2$,

$$(1.1) \quad P((x, y) \in A) = \sum_{(x,y) \in A} f(x, y)。$$

由於 (X, Y) 為一離散型的隨機向量, $f(x, y)$ 只對可數個 (x, y) 才不等於0。所以即使 A 包含不可數個點, 如 R^2 上一長方形, (1.1)式的右側, 只會是一可數個 $f(x, y)$ 的和。例如, 取 $A = \{(x, y) | x = 7, y \leq 4\}$ 。這是 R^2 中一半線, 有不可數個點。但由表1.1知, 在 A 中只有 $(x, y) = (7, 1)$ 或 $(7, 3)$ 才會使 $f(x, y) \neq 0$ 。故

$$P(X = 7, Y \leq 4) = f(7, 1) + f(7, 3) = \frac{1}{18} + \frac{1}{18} = \frac{1}{9},$$

如預期的, 與例1.1中所得相同。當求出 (X, Y) 之聯合p.d.f.後, 在求關於 (X, Y) 之機率時, 通常會較回到樣本空間上(如例1.1的作法)去求容易。

對離散型的隨機向量 (X, Y) , 其聯合p.d.f. $f(x, y)$, 要滿足 $f(x, y) \geq 0$, 且存在一可數的集合 $C \subset R^2$, 使得

$$(1.2) \quad \sum_{(x,y) \in C} f(x, y) = 1。$$

反之, 若 $f(x, y)$ 為一由 R^2 至 R 的非負函數, 且只有在一可數的集合 C 才不為0, 並滿足(1.2)式, 必為某一**隨機向量** (X, Y) 之聯合p.d.f.。如同單變數的情況, 這種 (X, Y) 並不唯一。不同的隨機向量, 可以有相同的聯合p.d.f.。

雖然在考慮隨機向量 (X, Y) , 但有時會對其中某一變數有興趣。例如, 我們可能想知道諸如 $X = 3$ 之機率。若 (X, Y) 為一隨機向量, 則 X, Y 分別

為隨機變數。對離散的情況，令 $f_X(x) = P(X = x)$, $f_Y(y) = P(Y = y)$ ，分別稱為 X 及 Y 之邊際(marginal)機率密度函數，或說邊際p.d.f.。下述定理給出由聯合p.d.f.來求邊際p.d.f.的方法。

定理 1.1 設有離散型隨機向量 (X, Y) ，以 $f_{X,Y}(x, y)$ 為其聯合p.d.f.。則

$$(1.3) \quad f_X(x) = \sum_{y \in R} f_{X,Y}(x, y),$$

$$(1.4) \quad f_Y(y) = \sum_{x \in R} f_{X,Y}(x, y)。$$

證明.我們只證關於 $f_X(x)$ 的結果，關於 $f_Y(y)$ 的證明類似。對任意 $x \in R$ ，由於事件 $-\infty < Y < \infty$ 之機率為 1，故

$$\begin{aligned} f_X(x) &= P(X = x) \\ &= P(X = x, -\infty < Y < \infty) \\ &= P((X, Y) \in \{(x, y) | -\infty < y < \infty\}) \\ &= \sum_{y \in R} f_{X,Y}(x, y)。 \end{aligned}$$

證畢。

如果 Y 取非負整數值，則 (1.3) 式成為

$$f_X(x) = \sum_{j=0}^{\infty} f_{X,Y}(x, j),$$

如果 X 取值在正偶數，則

$$f_Y(y) = \sum_{i=1}^{\infty} f_{X,Y}(2i, y)。$$

又 $f_X(x)$, $f_Y(y)$ 須滿足

$$(1.5) \quad \sum_{x \in R} f_X(x) = \sum_{x \in R} \sum_{y \in R} f_{X,Y}(x, y) = \sum_{(x,y) \in R^2} f(x, y) = 1,$$

$$(1.6) \quad \sum_{y \in R} f_Y(y) = \sum_{y \in R} \sum_{x \in R} f_{X,Y}(x, y) = \sum_{(x,y) \in R^2} f(x, y) = 1。$$

因只是對所有可能的值相加, 爲了簡便, 常以

$$\sum_x f_X(x), \quad \sum_y f_Y(y), \quad \sum_{x,y} f(x,y)$$

分別取代

$$\sum_{x \in R} f_X(x), \quad \sum_{y \in R} f_Y(y), \quad \sum_{(x,y) \in R^2} f(x,y)。$$

例1.2 利用定理1.1, 可求出例1.1中 X, Y 之邊際p.d.f.:

$$\begin{aligned} f_X(2) &= \sum_{j=0}^5 f_{X,Y}(2, j) = f_{X,Y}(2, 0) = \frac{1}{36}, \\ f_X(3) &= \sum_{j=0}^5 f_{X,Y}(3, j) = f_{X,Y}(3, 1) = \frac{1}{18}, \\ &\vdots \\ f_Y(0) &= \sum_{i=2}^{12} f_{X,Y}(i, 0) = \frac{1}{6}, \\ &\vdots \end{aligned}$$

我們將 X, Y 之邊際p.d.f. 亦給在表1.1。可驗證

$$\begin{aligned} \sum_x f_X(x) &= \sum_{x=2}^{12} f_X(x) = 1, \\ \sum_y f_Y(y) &= \sum_{y=0}^5 f_Y(y) = 1。 \end{aligned}$$

由 (X, Y) 之聯合p.d.f., 除了知道關於 (X, Y) 之事件的機率, 也可知道僅關於 X , 或僅關於 Y 之事件的機率。

再來我們考慮連續型的隨機向量。一由 R^2 映至 R 的非負函數 $f(x, y)$, 若滿足對 $\forall A \subset R^2$,

$$(1.7) \quad P((X, Y) \in A) = \int \int_A f(x, y) dx dy,$$

便稱為連續型隨機向量 (X, Y) 之聯合p.d.f.。至於 (X, Y) 之邊際p.d.f., 則定義為

$$(1.8) \quad f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy, -\infty < x < \infty,$$

$$(1.9) \quad f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx, -\infty < y < \infty。$$

反之, 任一二變數的非負函數 $f(x, y)$, 若滿足

$$(1.10) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1,$$

必為某一連續型的隨機向量 (X, Y) 之聯合p.d.f.。如果(1.10)式中之 $f(x, y)$ 在一有限的區域, 例如長方形 $[a, b] \times [c, d]$, 其中 $a < b, c < d$, 不為0, 則(1.10)式成為

$$\int_c^d \int_a^b f(x, y)dxdy = 1。$$

要注意的是, 有時候不同的聯合p.d.f., 會導致相同的邊際p.d.f.。也就是有可能 (X, Y) 與 (U, V) 分佈不同, 但 $X \stackrel{d}{=} U, Y \stackrel{d}{=} V$ 。習題第10題為一例。

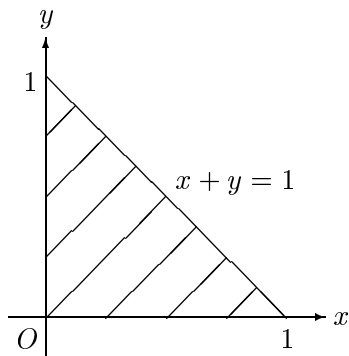
例1.3 設 (X, Y) 之聯合p.d.f.為

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

首先要說明的是, 對上述 f , 有時我們只簡單地寫成

$$f(x, y) = 24xy, 0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1,$$

並隱含著若不在所給的區域, 則 $f(x, y)$ 皆為0。又上述區域可簡化為 $0 \leq x \leq 1, 0 \leq y \leq 1 - x$, 或 $0 \leq y \leq 1, 0 \leq x \leq 1 - y$ 。



我們先驗證 $f(x, y)$ 確為一 p.d.f.:

$$\begin{aligned}
 & \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy \\
 &= \int_0^1 \int_0^{1-y} 24xy dx dy \\
 &= \int_0^1 12x^2 y \Big|_0^{1-y} dy \\
 &= \int_0^1 12y(1-y)^2 dy \\
 &= 12 \left(\frac{1}{2}y^2 - \frac{2}{3}y^3 + \frac{1}{4}y^4 \right) \Big|_0^1 \\
 &= 1 \text{。}
 \end{aligned}$$

其次 X 之邊際 p.d.f. 為

$$\begin{aligned}
 f_X(x) &= \int_0^{1-x} f(x, y) dy \\
 &= \int_0^{1-x} 24xy dy \\
 &= 12x(1-x)^2, 0 \leq x \leq 1 \text{。}
 \end{aligned}$$

故

$$P(X \leq \frac{1}{2}) = \int_0^{\frac{1}{2}} 12x(1-x)^2 dx = \frac{11}{16} \text{。}$$

當然不求邊際 p.d.f., $X \leq 1/2$ 之機率, 亦可由下述積分求得, 答案相同:

$$\int_0^{\frac{1}{2}} \int_0^{1-x} 24xy dy dx = \frac{11}{16} \text{。}$$

同理可求出

$$f_Y(y) = \int_0^{1-y} f(x, y) dx = 12y(1-y)^2, 0 \leq y \leq 1。$$

要注意邊際p.d.f.的範圍。原本 x, y 之範圍是互有影響的($x + y \leq 1$), 但 $f_X(x)$ 之 x 的範圍卻不能有 y , $f_Y(y)$ 之 y 的範圍也不能有 x 。

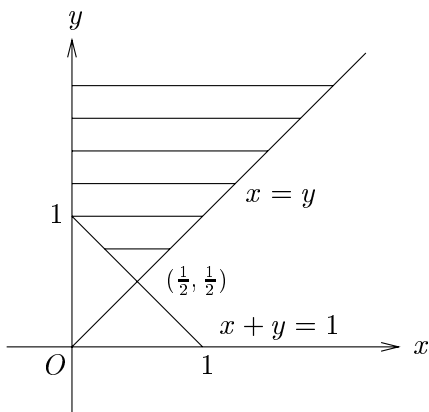
例1.4 設 (X, Y) 之聯合p.d.f.為

$$f(x, y) = e^{-y}, 0 < x < y < \infty。$$

雖然表面上看起來 e^{-y} 中不含 x , 但因 $f(x, y)$ 不為0之處: $0 < x < y < \infty$, 與 x 有關, 所以 $f(x, y)$ 仍與 x 有關。事實上若將 $f(x, y)$ 寫成下述型式就較清楚了:

$$f(x, y) = e^{-y} I_{\{0 < x < y < \infty\}}。$$

底下來求事件 $X + Y \geq 1$ 之機率。由下圖看出 $X + Y \leq 1$ 的機率較好求。又因這是連續型的隨機變數, 故 $X + Y \leq 1$ 與 $X + Y < 1$ 之機率相同。



由上述說明得

$$\begin{aligned} P(X + Y \geq 1) &= 1 - P(X + Y < 1) \\ &= 1 - P(X + Y \leq 1) \end{aligned}$$

$$\begin{aligned}
&= 1 - \int_0^{\frac{1}{2}} \int_x^{1-x} e^{-y} dy dx \\
&= 1 - \int_0^{\frac{1}{2}} (e^{-x} - e^{-(1-x)}) dx \\
&= 2e^{-1/2} - e^{-1}。
\end{aligned}$$

若先對 x 再對 y 積分, 也可求出 $P(X + Y \leq 1)$ 。即

$$P(X + Y \leq 1) = \int_0^{\frac{1}{2}} \int_0^y e^{-y} dx dy + \int_{\frac{1}{2}}^1 \int_0^{1-y} e^{-y} dx dy。$$

會得到相同的答案, 只是此時計算稍繁瑣些, 讀者不妨自行練習。

隨機向量之**聯合分佈函數**的定義為:

$$(1.11) \quad F(x, y) = P(X \leq x, Y \leq y)。$$

對離散型的隨機向量, 聯合分佈函數常沒有簡單的型式, 應用上不是那麼方便。但對連續型的隨機向量, 如同單變數的情況, 若 (X, Y) 之聯合p.d.f.為 $f(x, y)$, 則

$$(1.12) \quad F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du。$$

利用兩變數的微積分基本定理, 在每一 $f(x, y)$ 之連續點 (x, y) , 上式導致

$$(1.13) \quad \frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)。$$

最後對隨機向量 (X, Y) , 令 $g(x, y)$ 為一實值函數, 則 $g(X, Y)$ 仍為一隨機變數。對離散型及連續型, $g(X, Y)$ 之期望值分別為

$$\begin{aligned}
E(g(X, Y)) &= \sum_{x, y} g(x, y) f(x, y), \\
E(g(X, Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy。
\end{aligned}$$

例1.5 承例1.3。我們有

$$\begin{aligned}
 E(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x, y) dx dy \\
 &= \int_0^1 \int_0^{1-y} 24x^2 y^2 dx dy \\
 &= \int_0^1 8y^2 (1-y)^3 dy \\
 &= \left(\frac{8}{3} y^3 - 6y^4 + \frac{24}{5} y^5 - \frac{4}{3} y^6 \right) \Big|_0^1 \\
 &= \frac{8}{3} - 6 + \frac{24}{5} - \frac{4}{3} \\
 &= \frac{2}{15}。
 \end{aligned}$$

另外,

$$E(X) = \int_0^1 \int_0^{1-y} x \cdot f(x, y) dx dy = \int_0^1 \int_0^{1-y} 24x^2 y dx dy = \frac{2}{5}。$$

$E(X)$ 也可利用例1.3中, 所求出的 X 之邊際p.d.f.而得, 即

$$E(X) = \int_0^1 x \cdot f_X(x) dx = \int_0^1 12x^2 (1-x)^2 dx = \frac{2}{5}。$$

答案當然要相同。

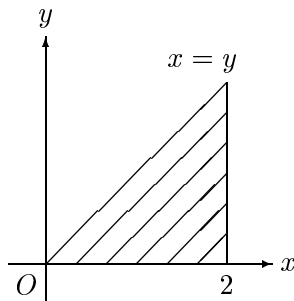
例1.6 設 (X, Y) 之聯合p.d.f.為

$$f(x, y) = \begin{cases} c, & 0 \leq y \leq x \leq 2, \\ 0, & \text{其他。} \end{cases}$$

試求(i) c 之值, (ii) $P(X \geq Y + 1)$, 及(iii) $P(X > x)$, $x \in R$ 。

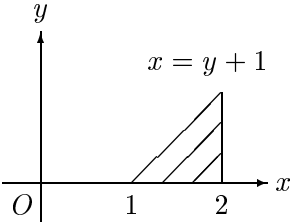
解.(i) 由

$$\begin{aligned}
 1 &= \int_0^2 \int_0^x f(x, y) dy dx \\
 &= \int_0^2 \int_0^x c dy dx \\
 &= \int_0^2 cx dx = 2c。
 \end{aligned}$$



得 $c = 1/2$ 。

(ii)

$$\begin{aligned}
 P(X \geq Y + 1) &= \int_1^2 \int_0^{x-1} \frac{1}{2} dy dx \\
 &= \frac{1}{2} \int_1^2 (x-1) dx \\
 &= \frac{1}{2} \left(\frac{1}{2} x^2 - x \right) \Big|_1^2 = \frac{1}{4}。
 \end{aligned}$$


(iii) 設 $0 < x < 2$ 。則

$$\begin{aligned}
 P(X > x) &= \int_x^2 \int_0^u \frac{1}{2} dy du \\
 &= \int_x^2 \frac{1}{2} u du \\
 &= \frac{1}{4} u^2 \Big|_x^2 = 1 - \frac{1}{4} x^2。
 \end{aligned}$$

至於若 $x \leq 0$, 則 $P(X > x) = 1$, 若 $x \geq 2$, 則 $P(X > x) = 0$ 。

3.2 條件分佈及獨立性

當觀測兩個隨機變數 X 與 Y 時, 兩個變數有時並非不相干。例如, 令 X 表某位學生第一次考試的成績, Y 表第二次考試的成績。如果知道 $X = 100$, 則 $Y = 0$ 的機率應是很低的。又如身高(X)與體重(Y), 如果知道 $X = 180$ (公分), 則 $Y \geq 60$ (公斤)的機率, 應比 $X = 140$ (公分)時, $Y \geq 60$ (公斤)的機率高。上述這些例子顯示, 有時知道 X 的值, 是會提供一些關於 Y 的資訊。

設 (X, Y) 為離散型的隨機向量, 因 $\{X = x\}$ 及 $\{Y = y\}$ 為二事件, 故可以第一章(定義2.1)事件之條件機率的定義, 來定義條件機率 $P(Y = y | X = x)$ 。即只要 $P(X = x) > 0$, 則

$$(2.1) \quad P(Y = y | X = x) = \frac{P(Y = y, X = x)}{P(X = x)}。$$

上式也可以 p.d.f 表示, 我們給一定義如下。

定義2.1 設 (X, Y) 為離散型的隨機向量，以 $f(x, y)$ 為聯合p.d.f., $f_X(x)$, $f_Y(y)$ 為邊際p.d.f.。對任意 x , 只要 $f_X(x) > 0$, 在給定 $X = x$ 之下, Y 之條件p.d.f. $f(y|x)$ 為

$$(2.2) \quad f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)}。$$

對任意 y , 只要 $f_Y(y) > 0$, 在給定 $Y = y$ 之下, X 之條件p.d.f. $f(x|y)$ 為

$$(2.3) \quad f(x|y) = P(X = x|Y = y) = \frac{f(x, y)}{f_Y(y)}。$$

有時為了清楚隨機變數為何, 且避免混淆, 我們會以 $f_{Y|X}(y|x)$ 表在給定 $X = x$ 之下, Y 的條件p.d.f.。也常簡單地以 $Y|X$ 之分佈, 或 $Y|X$ 之p.d.f., 分別表條件分佈及條件p.d.f.。另外, 既然 $f(y|x)$ 被稱為一條件“p.d.f.”, 它必須滿足p.d.f.的要求。這當然沒問題。首先, 因 $f_X(x) > 0, f(x, y) \geq 0$, 故 $f(y|x) \geq 0$ 。其次

$$\sum_y f(y|x) = \sum_y \frac{f(x, y)}{f_X(x)} = \frac{\sum_y f(x, y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1。$$

故 $f(y|x)$ 的確是一p.d.f.。

例2.1 設 (X, Y) 之聯合p.d.f.如下:

$y \setminus x$	1	2	$f_Y(y)$
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{3}{8}$
2	$\frac{9}{16}$	$\frac{1}{16}$	$\frac{5}{8}$
$f_X(x)$	$\frac{11}{16}$	$\frac{5}{16}$	1

我們也求出邊際p.d.f. $f_X(x)$ 及 $f_Y(y)$, 一併填入上表。則

$$\begin{aligned} f_{Y|X}(1|1) &= \frac{f(1, 1)}{f_X(1)} = \frac{1/8}{11/16} = \frac{2}{11}, \\ f_{Y|X}(2|1) &= \frac{f(1, 2)}{f_X(1)} = \frac{9/16}{11/16} = \frac{9}{11}, \\ f_{Y|X}(1|2) &= \frac{f(2, 1)}{f_X(2)} = \frac{1/4}{5/16} = \frac{4}{5}, \\ f_{Y|X}(2|2) &= \frac{f(2, 2)}{f_X(2)} = \frac{1/16}{5/16} = \frac{1}{5}。 \end{aligned}$$

至於 $f_{X|Y}(x|y)$ 留給各位自行求出。可看出

$$\begin{aligned}\sum_y f_{Y|X}(y|1) &= f_{Y|X}(1|1) + f_{Y|X}(2|1) = \frac{2}{11} + \frac{9}{11} = 1, \\ \sum_y f_{Y|X}(y|2) &= f_{Y|X}(1|2) + f_{Y|X}(2|2) = \frac{4}{5} + \frac{1}{5} = 1.\end{aligned}$$

但

$$\begin{aligned}\sum_x f_{Y|X}(1|x) &= f_{Y|X}(1|1) + f_{Y|X}(1|2) = \frac{2}{11} + \frac{4}{5} \neq 1, \\ \sum_x f_{Y|X}(2|x) &= f_{Y|X}(2|1) + f_{Y|X}(2|2) = \frac{9}{11} + \frac{1}{5} \neq 1.\end{aligned}$$

上例顯示，雖 $\sum_y f_{Y|X}(y|x) = 1$ ，但 $\sum_x f_{Y|X}(y|x)$ 卻不一定等於1，其中原因應不難想通。

利用條件p.d.f.可求出各種條件機率。例如，

$$(2.4) \quad P(Y \leq y|X = x) = \sum_{v \leq y} f_{Y|X}(v|x).$$

若 (X, Y) 為連續型的隨機向量，則 $P(X = x) = 0, \forall x \in R$ ，因此無法利用事件的機率來定義給定 $X = x$ 之下， Y 的條件機率，不過經由機率理論的推導，參照離散型的隨機變數的情況，我們給出下述定義。

定義2.2 設 (X, Y) 為連續型的隨機向量，以 $f(x, y)$ 為聯合p.d.f.， $f_X(x)$ ， $f_Y(y)$ 為邊際p.d.f.。對任意 x ，只要 $f_X(x) > 0$ ，在給定 $X = x$ 之下， Y 之條件p.d.f.為

$$(2.5) \quad f(y|x) = \frac{f(x, y)}{f_X(x)}.$$

對任意 y ，只要 $f_Y(y) > 0$ ，在給定 $Y = y$ 之下， X 之條件p.d.f.為

$$(2.6) \quad f(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

如同離散型的情況，亦可證明 $f(y|x)$ 及 $f(x|y)$ 皆確為p.d.f.。

例2.2 承例1.3。可求出

$$\begin{aligned} f(y|x) &= \frac{f(x, y)}{f_X(x)} = \frac{24xy}{12x(1-x)^2} \\ &= \frac{2y}{(1-x)^2}, 0 \leq y \leq 1-x, 0 < x < 1。 \end{aligned}$$

同理

$$f(x|y) = \frac{2x}{(1-y)^2}, 0 \leq x \leq 1-y, 0 < y < 1。$$

注意因 $f_X(0) = f_Y(0) = 0$, 且 $f_X(1) = f_Y(1) = 0$, 故在 $f(y|x)$ 中, $x \neq 0, 1$, 在 $f(x|y)$ 中, $y \neq 0, 1$ 。又可驗證 $f(y|x)$ 及 $f(x|y)$ 都確為 p.d.f.。即

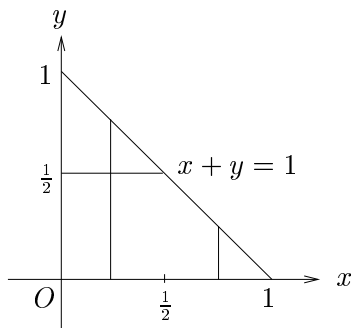
$$\int_0^{1-x} \frac{2y}{(1-x)^2} dy = \int_0^{1-y} \frac{2x}{(1-y)^2} dx = 1。$$

又 $P(Y \leq 1/2|x)$ 是否等於

$$\int_0^{1/2} f(y|x) dy = \int_0^{1/2} \frac{2y}{(1-x)^2} dy = \frac{1}{4(1-x)^2}?$$

要注意的是 y 的範圍要滿足 $0 \leq y \leq 1-x$, 而 $1/2$ 不一定小於 $1-x$ 。正確的答案為(參考附圖)

$$\begin{aligned} P(Y \leq \frac{1}{2}|x) &= \int_0^{\min(1/2, 1-x)} f(y|x) dy \\ &= \int_0^{\min(1/2, 1-x)} \frac{2y}{(1-x)^2} dy \\ &= \frac{(\min(1/2, 1-x))^2}{(1-x)^2} \\ &= \begin{cases} \frac{1}{4(1-x)^2}, & 0 < x \leq \frac{1}{2}, \\ \frac{(1-x)^2}{(1-x)^2} = 1, & \frac{1}{2} < x < 1。 \end{cases} \end{aligned}$$



條件 p.d.f. 亦可用來求 **條件期望值**。對一函數 g , 在給定 $X = x$ 之下, $g(Y)$ 之條件期望值, 以 $E(g(Y)|x)$ 表之, 定義為

$$(2.7) \quad E(g(Y)|x) = \begin{cases} \sum_y g(y) f(y|x) & , \text{離散型,} \\ \int_{-\infty}^{\infty} g(y) f(y|x) dy & , \text{連續型。} \end{cases}$$

條件期望值仍具有第一章定理6.1所列出的一般的期望值之特性。譬如說，由於 $E(Y|x)$ 為一 x 的函數，故當給定 $X = x$ ， $E(Y|x)$ 便非隨機了。因此(利用第一章(6.5)及(6.6)式)

$$(2.8) \quad E(E(Y|x)|x) = E(Y|x),$$

$$(2.9) \quad E(YE(Y|x)|x) = E(Y|x) \cdot E(Y|x) = (E(Y|x))^2。$$

又在誤差的平方之期望值最小下， $E(Y|X)$ 為在知道 X 之下， Y 之最佳預測值。這是第一章定理6.2之推廣，我們留在習題中讓各位自行完成。

例2.3 承例2.1。則

$$\begin{aligned} E(Y|X=1) &= \sum_y yf(y|1) = 1 \cdot \frac{2}{11} + 2 \cdot \frac{9}{11} = \frac{20}{11}, \\ E(Y|X=2) &= \sum_y yf(y|2) = 1 \cdot \frac{4}{5} + 2 \cdot \frac{1}{5} = \frac{6}{5}。 \end{aligned}$$

例2.4 承例2.2。

$$\begin{aligned} E(Y|X=x) &= \int_0^{1-x} yf(y|x)dy \\ &= \int_0^{1-x} \frac{2y^2}{(1-x)^2} dy \\ &= \frac{2(1-x)^3}{3(1-x)^2} \\ &= \frac{2}{3}(1-x), 0 < x < 1。 \end{aligned}$$

又

$$\begin{aligned} E(Y^2|X=x) &= \int_0^{1-x} y^2 \frac{2y}{(1-x)^2} dy \\ &= \frac{1}{(1-x)^2} \cdot \frac{1}{2}(1-x)^4 = \frac{1}{2}(1-x)^2, 0 < x < 1。 \end{aligned}$$

因此

$$\begin{aligned}
& E(2Y^2 + 3Y|X = x) \\
&= 2E(Y^2|X = x) + 3E(Y|X = x) \\
&= (1-x)^2 + 2(1-x) = (1-x)(3-x), 0 < x < 1。
\end{aligned}$$

在給定 $X = x$ 之下, 除了可求條件期望值 $E(Y|X = x)$, 亦可求**條件變異數** $\text{Var}(Y|X = x)$, 或以 $\text{Var}(Y|x)$ 表之, 其定義為

$$(2.10) \quad \text{Var}(Y|x) = E((Y - E(Y|x))^2|x)。$$

以上式來定義條件變異數是有道理的。變異數是量測一變數(此處為 Y)與其期望值(此處為 $E(Y|x)$)之離差平方(此處為 $(Y - E(Y|x))^2$)的期望值(此處為給定 $X = x$ 之下的期望值), 故得(2.10)式。由(2.10)式立即導致(證明留在習題第3題)

$$(2.11) \quad \text{Var}(Y|x) = E(Y^2|x) - (E(Y|x))^2。$$

例2.5 承例1.3, 2.2, 及2.4。可得

$$\begin{aligned}
\text{Var}(Y|x) &= E(Y^2|x) - (E(Y|x))^2 \\
&= \frac{1}{2}(1-x)^2 - \frac{4}{9}(1-x)^2 \\
&= \frac{1}{18}(1-x)^2, 0 < x < 1。
\end{aligned}$$

變異數與 x 有關。當 $x \rightarrow 1$ 時, 變異數 $\rightarrow 0$, 這是合理的, 由例1.3中的圖形可看出, 當 $x \rightarrow 1$ 時, $y \rightarrow 0$, 變異愈來愈小, 變異數自然也愈來愈小。甚至因 x 愈大時, y 的活動範圍愈小($0 \leq y \leq 1-x$), 變異也就愈小。事實上, 我們所求出的 $\text{Var}(Y|x)$ 在 $0 < x < 1$ 間, 的確為 x 的漸減函數。

例2.6 設 X, Y 之聯合p.d.f.為

$$f(x, y) = \frac{1}{64}(x+y), 0 \leq x, y \leq 4。$$

試求(i) $f_X(x)$, (ii) $f(y|x)$, (iii) $E(Y|x)$, 及(iv) $\text{Var}(Y|x)$, $0 \leq x \leq 4$ 。

解.(i) 依定義, 即得

$$f_X(x) = \int_0^4 \frac{1}{64}(x+y)dy = \frac{1}{16}x + \frac{1}{8}, 0 \leq x \leq 4。$$

(ii) 利用(i)得

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{(x+y)/64}{x/16 + 1/8} = \frac{x+y}{4x+8}, 0 \leq y \leq 4, 0 \leq x \leq 4。$$

(iii) 利用(ii)得

$$\begin{aligned} E(Y|x) &= \int_0^4 y \cdot f(y|x)dy \\ &= \int_0^4 y \cdot \frac{x+y}{4x+8}dy \\ &= \frac{1}{4x+8} \int_0^4 (xy + y^2)dy \\ &= \frac{8x + 64/3}{4x+8} = \frac{6x+16}{3x+6}, 0 \leq x \leq 4。 \end{aligned}$$

(iv) 先求出

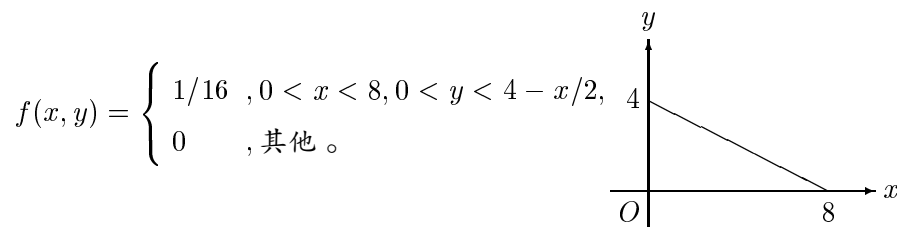
$$\begin{aligned} E(Y^2|x) &= \int_0^4 y^2 \cdot f(y|x)dy \\ &= \int_0^4 y^2 \cdot \frac{x+y}{4x+8}dy \\ &= \frac{1}{4x+8} \int_0^4 (xy^2 + y^3)dy \\ &= \frac{64x/3 + 64}{4x+8} = \frac{16x+48}{3x+6}, 0 \leq x \leq 4, \end{aligned}$$

因此

$$\begin{aligned} \text{Var}(Y|x) &= E(Y^2|x) - (E(Y|x))^2 \\ &= \frac{16x+48}{3x+6} - \left(\frac{6x+16}{3x+6}\right)^2 \\ &= \frac{12x^2 + 48x + 32}{(3x+6)^2}, 0 \leq x \leq 4。 \end{aligned}$$

例2.7 設 (X, Y) 在 xy 平面上, 以 $(0, 0)$, $(8, 0)$ 及 $(0, 4)$ 為三頂點之三角形內(不含三邊)均勻分佈。試求 $f(y|x)$, $E(Y|x)$, 及 $\text{Var}(Y|x)$ 。

解. 首先 (X, Y) 之聯合p.d.f.為



由此得

$$f_X(x) = \int_0^{4-x/2} \frac{1}{16} dy = \frac{4-x/2}{16} = \frac{8-x}{32}, 0 < x < 8。$$

則

$$f(y|x) = \frac{1/16}{(8-x)/32} = \frac{2}{8-x}, 0 < y < 4 - x/2, 0 < x < 8。$$

可看出給定 $X = x$, Y 在區間 $(0, 4 - x/2)$ 均勻分佈。最後, 利用對 $\forall a > 0$, 有 $\mathcal{U}(0, a)$ 分佈之隨機變數的期望值為 $a/2$, 變異數為 $a^2/12$, 即得

$$\begin{aligned} E(Y|x) &= \frac{1}{2} \left(4 - \frac{x}{2}\right) = 2 - \frac{x}{4}, 0 < x < 8, \\ \text{Var}(Y|x) &= \frac{1}{12} \left(4 - \frac{x}{2}\right)^2 = \frac{(8-x)^2}{48}, 0 < x < 8。 \end{aligned}$$

讀者亦可自行求出 $X|Y = y$ 有 $\mathcal{U}(0, 8 - 2y)$ 分佈, 且 $E(X|y) = 4 - y$, $\text{Var}(X|y) = (4 - y)^2/3, 0 < y < 4。$

註2.1 設 A 為平面上一區域, 且設 A 的面積(以 $S(A)$ 表之)存在且為正。則 (X, Y) 在 A 上均勻分佈, 表 (X, Y) 之聯合 p.d.f. 為

$$f(x, y) = \frac{1}{S(A)}, (x, y) \in A。$$

例2.8 設 (X, Y) 之聯合 p.d.f. 為

$$f(x, y) = \frac{1}{2} x^3 e^{-x(1+y)}, x, y > 0。$$

試求 $f(y|x)$, $E(Y|x)$ 及 $\text{Var}(Y|x)$ 。

解. 首先

$$f_X(x) = \int_0^\infty \frac{1}{2} x^3 e^{-x(1+y)} dy = \frac{1}{2} x^2 e^{-x}, x > 0。$$

因此

$$f(y|x) = \frac{x^3 e^{-x(1+y)}/2}{x^2 e^{-x}/2} = x e^{-xy}, y > 0, x > 0。$$

即 $Y|X = x$ 有 $\mathcal{E}(x)$ 分佈。因此

$$E(Y|x) = \frac{1}{x}, \quad \text{Var}(Y|x) = \frac{1}{x^2}, x > 0。$$

讀者亦可自行求出 $f_Y(y) = 3(1+y)^{-4}$, $y > 0$ 。由此可得對 $\forall y > 0$, $X|Y = y$ 有 $\Gamma(4, (1+y)^{-1})$ 分佈, 且 $E(X|y) = 4/(1+y)$, $\text{Var}(X|y) = 4/(1+y)^2$, $y > 0$ 。

例2.9 設有一函數

$$f(x, y) = \begin{cases} 4xy - 2x - 2y + 2, & 0 < x, y < 1, \\ 0, & \text{其他。} \end{cases}$$

(i) 試證 f 爲一二維之 p.d.f.;

(ii) 假設 (X, Y) 以 f 爲聯合 p.d.f., 試求 $f(y|x)$, $E(Y|x)$ 及 $\text{Var}(Y|x)$ 。

解. (i) 對 $0 < x, y < 1$,

$$4xy - 2x - 2y + 2 = 2xy + 2(1-x)(1-y) > 0,$$

即 f 爲非負。又

$$\begin{aligned} \int_0^1 \int_0^1 f(x, y) dy dx &= \int_0^1 \int_0^1 (4xy - 2x - 2y + 2) dy dx \\ &= 1 - 1 - 1 + 2 = 1, \end{aligned}$$

故 f 爲一二維之 p.d.f.。

(ii) 首先

$$\begin{aligned} f_X(x) &= \int_0^1 (4xy - 2x - 2y + 2) dy \\ &= 2x - 2x - 1 + 2 = 1, 0 < x < 1。 \end{aligned}$$

因此

$$f(y|x) = 4xy - 2x - 2y + 2, 0 < y < 1, 0 < x < 1。$$

利用上述條件p.d.f., 即得

$$\begin{aligned} E(Y|x) &= \int_0^1 y(4xy - 2x - 2y + 2)dy = \frac{1}{3}(1+x), 0 < x < 1, \\ \text{Var}(Y|x) &= \int_0^1 y^2(4xy - 2x - 2y + 2)dy - \frac{1}{9}(1+x)^2 \\ &= \frac{1}{6}(1+2x) - \frac{1}{9}(1+x)^2 \\ &= \frac{1}{18}(1+2x-2x^2), 0 < x < 1. \end{aligned}$$

讀者可否看出 X, Y 之邊際分佈皆為 $\mathcal{U}(0, 1)$, 但 (X, Y) 並無二維的均勻分佈。又可否立即看出, 對本例, $f(x|y)$, $E(X|y)$ 及 $\text{Var}(X|y)$ 分別為何?

由例2.1及2.2知, 給定 $X = x$ 之下, Y 的分佈可能會隨 x 而變。所以在求條件分佈時, 我們得到一分佈族: 對每一 x 有一分佈。若我們想描述整個分佈族, 可以“ $Y|X$ ”(Y給定 X , 或說給定 X 之下 Y 之分佈)來表示。

例如, 若 X 為一取正整數值之隨機變數, 且在給定 $X = x$ 之下, Y 有 $\mathcal{B}(x, p)$ 分佈, 則可以說 $Y|X$ 有 $\mathcal{B}(X, p)$ 分佈。採用符號“ $Y|X$ ”, 或是分佈中有一參數為隨機變數, 則我們便描述出一條件分佈族。同樣地, $E(g(Y)|x)$ 亦為一 x 的函數, 而 $E(g(Y)|X)$ 則為一隨機變數, 其值依 X 之值而定。若 $X = x$, 則隨機變數 $E(g(Y)|X)$ 之值便為 $E(g(Y)|x)$ 。例如, 在例2.4中, 我們可以寫

$$E(Y|x) = \frac{2}{3}(1-x), \quad E(Y^2|x) = \frac{1}{2}(1-x)^2.$$

在前述幾個例子中, 給定 $X = x$, Y 之分佈均隨 x 而變。但有些時候, 給定 $X = x$, Y 之分佈並未改變, 亦即知道 X 之值, 對 Y 並未提供任何資訊。此時我們說 X 與 Y 獨立。

定義2.3 設二隨機變數 X, Y 之聯合p.d.f.為 $f(x, y)$, 邊際p.d.f.分別為 $f_X(x)$, $f_Y(y)$ 。若

$$(2.12) \quad f(x, y) = f_X(x)f_Y(y), \forall x, y \in R,$$

便稱 X 與 Y 獨立。

由定義2.1及2.2知, 對離散型或連續型的隨機向量 (X, Y) ,

$$(2.13) \quad f(y|x) = \frac{f(x, y)}{f_X(x)}, \text{ 只要 } f_X(x) > 0。$$

因此

$$(2.14) \quad f(x, y) = f(y|x)f_X(x), \text{ 只要 } f_X(x) > 0。$$

若 X 與 Y 獨立, 則(2.12)式成立, 因此

$$(2.15) \quad f(y|x) = f_Y(y), \text{ 只要 } f_X(x) > 0。$$

由上式又得對 $\forall A \subset R$, 只要 $f_X(x) > 0$, 則

$$P(Y \in A|x) = \int_A f(y|x)dy = \int_A f_Y(y)dy = P(Y \in A),$$

即知道 $X = x$ 之值, 對求 Y 的機率並無影響。

例2.10 在例2.1中, 因

$$f_X(1)f_Y(1) = \frac{11}{16} \cdot \frac{3}{8} \neq \frac{1}{8} = f(1, 1),$$

故 X 與 Y 不獨立。當然由 $f(y|x)$ 之值與 x 有關亦可判定 X 與 Y 不獨立。

例2.11 設 (X, Y) 有如例2.8之聯合p.d.f., 即

$$f(x, y) = \frac{1}{2}x^3e^{-x(1+y)}, x, y > 0。$$

試判斷 X 與 Y 是否獨立。

解. 首先由例2.8, X, Y 之邊際p.d.f.分別為

$$f_X(x) = \frac{1}{2}x^2e^{-x}, x > 0,$$

$$f_Y(y) = \frac{3}{(1+y)^4}, y > 0。$$

顯然

$$f(1, 1) = \frac{1}{2}e^{-2} \neq \frac{1}{2}e^{-1} \cdot \frac{3}{(1+1)^4} = f_X(1) \cdot f_Y(1)。$$

故 X 與 Y 不獨立。

要利用(2.12)式來決定 X 與 Y 是否獨立, 便須知 X 與 Y 之邊際p.d.f. $f_X(x)$ 及 $f_Y(y)$ 。利用下述定理, 有時可讓檢驗工作容易些。

定理2.1 設隨機向量 (X, Y) 以 $f(x, y)$ 為聯合p.d.f.。則 X 與 Y 獨立, 若且唯若存在函數 $g(x)$ 及 $h(y)$, 使得對 $\forall x, y \in R$,

$$(2.16) \quad f(x, y) = g(x)h(y)。$$

證明.若 X 與 Y 獨立, 依定義(2.12)式便成立。取 $g(x) = f_X(x)$, $h(y) = f_Y(y)$, 則(2.16)式成立。這便證明了“唯若”的部分。對“若”的部分, 我們先證連續型隨機變數的情況。

假設存在 $g(x)$ 及 $h(y)$, 使得(2.16)式成立。令

$$a = \int_{-\infty}^{\infty} g(x)dx, \quad b = \int_{-\infty}^{\infty} h(y)dy,$$

則

$$(2.17) \quad \begin{aligned} ab &= \int_{-\infty}^{\infty} g(x)dx \cdot \int_{-\infty}^{\infty} h(y)dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)dxdy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy \\ &= 1, \end{aligned}$$

最後一等式成立, 是因 $f(x, y)$ 為一聯合p.d.f.。

另外, (2.16)式導致

$$(2.18) \quad f_X(x) = \int_{-\infty}^{\infty} g(x)h(y)dy = g(x)b,$$

$$(2.19) \quad f_Y(y) = \int_{-\infty}^{\infty} g(x)h(y)dx = h(y)a。$$

利用(2.17)、(2.18)及(2.19)式得

$$f(x, y) = g(x)h(y) = g(x)h(y)ab = f_X(x)f_Y(y)。$$

得證 X 與 Y 獨立。

至於離散型的情況, 只要將上述積分皆改為和即可。

下述系理為定理2.1之一立即的推論。

系理2.1 設隨機向量 (X, Y) 以 $F(x, y)$ 為其聯合分佈函數, 則 X 與 Y 獨立, 若且唯若

$$(2.20) \quad F(x, y) = F_X(x)F_Y(y), x, y \in R,$$

若且唯若存在函數 $G(x)$ 及 $H(y)$, 使得對 $\forall x, y \in R$,

$$(2.21) \quad F(x, y) = G(x)H(y)。$$

利用定理2.1, 立即可判定例2.11中的 X 與 Y 不獨立。底下再給一例。

例2.12 設 (X, Y) 之聯合p.d.f. 為

$$f(x, y) = \frac{1}{121,500} x^3 y^2 e^{-x/3-y/5}, x, y > 0。$$

則由定理2.1知, X 與 Y 獨立。又讀者宜練習由 $f(x, y)$ 看出 X, Y 分別有 $\Gamma(4, 3)$ 及 $\Gamma(3, 5)$ 分佈。

在利用獨立的定義, 或定理2.1, 以檢驗二隨機變數是否獨立, 要留意的是使 $f(x, y) > 0$ 之 (x, y) 的集合, 必須是 $\{(x, y) | x \in A, y \in B\}$ 的型式, 其中 $A = \{x | f_X(x) > 0\}$, $B = \{y | f_Y(y) > 0\}$, 即為 R^2 上一有“乘積”型式的區域, 通常以 $A \times B$ 表之。換句話說, x, y 的範圍要彼此無關。在例1.3中, 雖然 $f(x, y) = 24xy$, 表面上看起來, 可寫成 $g(x)h(y)$ 的型式, 但使 $f(x, y) > 0$ 之 x, y 的集合為 $\{0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1\}$, 為平面上—三角形, 並無 $A \times B$ 的型式。範圍如果是 $x^2 + y^2 \leq 4$ 當然也不行。更明確地說, 例1.3中之 $f(x, y)$ 應寫成

$$f(x, y) = 24xy I_{\{0 \leq x \leq 1, 0 \leq y \leq 1, x+y \leq 1\}}。$$

因其中指示函數的部分, 無法寫成一個 x 的函數乘上一個 y 的函數, 所以 X 與 Y 不獨立。

當 X 與 Y 獨立時, 有關 X 與 Y 的機率或期望值都會較好求。我們給一定理。

定理2.2 設 X, Y 為二獨立的隨機變數。則

(i) 對 $\forall A, B \subset R$,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B),$$

即事件 $\{X \in A\}$ 與 $\{Y \in B\}$ 相互獨立。

(ii) 設有函數 $g(x)$ 及 $h(y)$, 滿足 $E(g(X))$ 及 $E(h(Y))$ 皆存在。則

$$E(g(X)h(Y)) = E(g(X)) \cdot E(h(Y))。$$

證明.我們仍只證連續型隨機變數的情況。先證(ii)。由於 X 與 Y 獨立, $f(x, y) = f_X(x)f_Y(y)$, 故

$$\begin{aligned} E(g(X)h(Y)) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x, y)dxdy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{\infty} h(y)f_Y(y)\left(\int_{-\infty}^{\infty} g(x)f_X(x)dx\right)dy \\ &= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy \\ &= E(g(X)) \cdot E(h(Y))。 \end{aligned}$$

得證(ii)。

其次對於(i)的部分, 可仿(ii)的證明, 或利用(ii)來證明。事實上(ii)是比(i)更一般的結果。

取

$$g(x) = I_A(x), \quad h(y) = I_B(y),$$

分別表 A, B 之指示函數。則 $g(x)h(y)$ 為 R^2 上之集合 C 之指示函數, 其中 $C = \{(x, y) | x \in A, y \in B\}$ 。又

$$E(g(X)) = P(X \in A), \quad E(h(Y)) = P(Y \in B),$$

$$E(g(X)h(Y)) = P((X, Y) \in C)。$$

故得

$$\begin{aligned}
 P(X \in A, Y \in B) &= P((X, Y) \in C) \\
 &= E(g(X)h(Y)) \\
 &= E(g(X))E(h(Y)) \\
 &= P(X \in A)P(Y \in B) \text{。}
 \end{aligned}$$

證畢。

例2.13 設 X, Y 為二獨立的隨機變數, X 有 $\mathcal{E}(2)$ 分佈, Y 有 $\mathcal{U}[0, 3]$ 分佈。則

$$\begin{aligned}
 P(X \geq 3, Y \geq 2) &= P(X \geq 3)P(Y \geq 2) \\
 &= \int_3^\infty 2e^{-2x} dx \cdot \int_2^3 \frac{1}{3} dy \\
 &= \frac{1}{3}e^{-6} \text{。}
 \end{aligned}$$

又

$$\begin{aligned}
 E(XY^2) &= E(X) \cdot E(Y^2) = \frac{1}{2} \cdot \int_0^3 \frac{1}{3} y^2 dy \\
 &= \frac{1}{2} \cdot 3 = \frac{3}{2} \text{。}
 \end{aligned}$$

例2.14 設 U_1, U_2 為二獨立的隨機變數, 且皆有 $\mathcal{U}(0, 1]$ 分佈。又設平面上有一隨機的矩形 A , 四個頂點之座標分別為 $(0, 0)$, $(U_1, 0)$, (U_1, U_2) 及 $(0, U_2)$ 。則 A 之面積的期望值與中位數何者為大?

解. 首先面積 U_1U_2 之期望值為

$$E(U_1U_2) = E(U_1)E(U_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \text{。}$$

次求面積 U_1U_2 之中位數。對 $\forall 0 < y \leq 1$,

$$\begin{aligned}
 P(U_1U_2 \geq y) &= \int_y^1 \left(\int_{y/u_1}^1 du_2 \right) du_1 \\
 &= \int_y^1 \left(1 - \frac{y}{u_1} \right) du_1
 \end{aligned}$$

$$\begin{aligned}
&= (u_1 - y \log u_1) \Big|_y^1 \\
&= 1 - y(1 - \log y) \text{。}
\end{aligned}$$

而若 $y > 1$, 則 $P(U_1 U_2 \geq y) = 0$, 若 $y \leq 0$, 則 $P(U_1 U_2 \geq y) = 1$ 。由此即得

$$P(U_1 U_2 \geq \frac{1}{4}) = 1 - \frac{1}{4}(1 + \log 4) \doteq 0.403 \text{。}$$

故 $U_1 U_2$ 之中位數小於 $1/4$ 。即 $U_1 U_2$ 之中位數比期望值小。

例2.15 自區間 $(0, 1)$ 隨機地取兩數, 以 X 及 Y 表之。求最接近 X/Y 的整數恰為一偶數之機率。

解. 具有對稱性概念的人, 看到又是均勻分佈, 又是奇數、偶數, 會以為此機率應為 $1/2$ 。底下我們來看對不對。首先由題意知, X, Y 為二獨立的隨機變數, 且皆有 $\mathcal{U}(0, 1)$ 分佈。故 (X, Y) 之聯合 p.d.f. 為

$$f(x, y) = 1, 0 < x, y < 1 \text{。}$$

對任意 X 及 Y , 最接近 X/Y 的整數恰為偶數, 若且唯若 $X/Y < 0.5$, 或存在一正整數 n , 使得 $2n - 0.5 < X/Y < 2n + 0.5$ 。現因

$$P(X/Y < 0.5) = P(2X < Y) = \int_0^{0.5} \int_{2x}^1 dy dx = \frac{1}{4},$$

又

$$\begin{aligned}
&P(2n - 0.5 < X/Y < 2n + 0.5) \\
&= P\left(\frac{2X}{4n+1} < Y < \frac{2X}{4n-1}\right) \\
&= \int_0^1 \left(\frac{2x}{4n-1} - \frac{2x}{4n+1}\right) dx \\
&= \frac{1}{4n-1} - \frac{1}{4n+1},
\end{aligned}$$

故

$$P(\text{最接近 } X/Y \text{ 的整數恰為一偶數})$$

$$\begin{aligned}
&= \frac{1}{4} + \sum_{n=1}^{\infty} \left(\frac{1}{4n-1} - \frac{1}{4n+1} \right) \\
&= \frac{1}{4} + \left(\frac{1}{3} - \frac{1}{5} + \frac{1}{7} - \frac{1}{9} + \cdots \right) \\
&= \frac{1}{4} + 1 - \left(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \cdots \right) \\
&= \frac{1}{4} + 1 - \arctan 1 \\
&= \frac{1}{4} + 1 - \frac{\pi}{4} = \frac{5-\pi}{4} < \frac{1}{2}。
\end{aligned}$$

事實上我們所求即為 $X/Y \in S$ 之機率, 其中

$$S = (0, 0.5) \cup (1.5, 2] \cup (2, 2.5) \cup (3.5, 4] \cup (4, 4.5) \cup \cdots,$$

包含每一整數區間 $(i, i+1], i \geq 0$, 的一半部分。但此機率卻小於 $1/2$, 即使 X, Y 原來均有獨立的均勻分佈。

註2.2 此處用到

$$\arctan x = x - \frac{1}{3}x^3 + \frac{1}{5}x^5 - \cdots + \frac{(-1)^n}{2n+1}x^{2n+1} + \cdots, |x| \leq 1。$$

底下檢驗二隨機變數獨立的定理, 亦為定理2.1之一推論。我們略去其證明。

定理2.3 設有二隨機變數 X, Y 。則 X 與 Y 獨立, 若且唯若

$$(2.22) \quad \psi(s, t) = E(e^{isX+itY}) = E(e^{isX})E(e^{itY}), s, t \in R。$$

上式左側稱為 X, Y 之聯合特徵函數。定理2.3即是說 X 與 Y 獨立, 若且唯若其聯合特徵函數等於 X, Y 各自特徵函數之積。定理2.3也有母函數(針對取非負整數值之隨機變數), 拉普拉斯轉換(針對取非負整數值之隨機變數)及動差母函數之版本。只要這些轉換至少在一二維的開區間存在(母函數及拉普拉斯轉換沒問題, 動差母函數 $M(s, t) = E(e^{sX+tY})$ 有時只有在 $s = t = 0$ 才存在)。

定理2.3有下述一重要的推論。

系理2.2 設 X, Y 為二獨立的隨機變數。則

$$(2.23) \quad E(e^{it(X+Y)}) = E(e^{itX})E(e^{itY}), t \in R。$$

同樣地,系理2.2亦有關於其他三個轉換的版本。

例2.16 設 X, Y 為二獨立的隨機變數, 各有 $\mathcal{P}(\lambda)$ 及 $\mathcal{P}(\mu)$ 的分佈。則由系理2.2, $X + Y$ 之拉普拉斯轉換為

$$\begin{aligned} \phi(s, t) &= E(e^{-s(X+Y)}) = E(e^{-sX})E(e^{-sY}) \\ &= e^{-\lambda(1-e^{-s})}e^{-\mu(1-e^{-s})} \\ &= e^{-(\lambda+\mu)(1-e^{-s})}, s \geq 0。 \end{aligned}$$

可看出 $e^{-(\lambda+\mu)(1-e^{-s})}$ 對應 $\mathcal{P}(\lambda+\mu)$ 分佈之拉普拉斯轉換, 故 $X+Y$ 有 $\mathcal{P}(\lambda+\mu)$ 分佈。

利用適當的轉換, 可求出一些具有常見分佈(參數可能不同)之隨機變數和的分佈。我們列出如下, 證明則留在習題。

定理2.4 設 X_1, X_2 為二獨立的隨機變數。令 $S = X_1 + X_2$ 。則對 $i = 1, 2$,

- (i) 若 X_i 有 $\mathcal{P}(\lambda_i)$ 分佈, 則 S 有 $\mathcal{P}(\lambda_1 + \lambda_2)$ 分佈;
- (ii) 若 X_i 有 $\mathcal{B}(n_i, p)$ 分佈, 則 S 有 $\mathcal{B}(n_1 + n_2, p)$ 分佈;
- (iii) 若 X_i 有 $\Gamma(\alpha_i, \beta)$ 分佈, 則 S 有 $\Gamma(\alpha_1 + \alpha_2, \beta)$ 分佈;
- (iv) 若 X_i 有 $\chi_{n_i}^2$ 分佈, 則 S 有 $\chi_{n_1+n_2}^2$ 分佈;
- (v) 若 X_i 有 $\mathcal{N}(\mu_i, \sigma_i^2)$ 分佈, 則 S 有 $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ 分佈;
- (vi) 若 X_i 有自0開始之 $Ge(p)$ 分佈, 則 S 有 $\mathcal{NB}(2, p)$ 分佈;
- (vii) 若 X_i 有 $\mathcal{NB}(r_i, p)$ 分佈, 則 S 有 $\mathcal{NB}(r_1 + r_2, p)$ 分佈。

並非所有獨立的同一類分佈之隨機變數的和仍有同一類分佈(當然如果將幾何分佈視為負二項分佈, 則(vi)為(vii)之一特例)。定理2.4(vi)為一例。另外, 設 X_1, X_2 獨立。則當 X_1, X_2 皆有 $\mathcal{U}(0, 1)$ 分佈時, $X_1 + X_2$ 並無均勻分佈。讀者也可練習再找一些例子。

我們再給一有趣的例子。

例2.17 對一正整數 i ，假設任選一正整數是 i 的倍數之機率為 $1/i$ 。這個假設是合理的，原因則讓大家自行想想。利用

$$(2.24) \quad \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}。$$

試求任選二正整數互質的機率 p 。

解.任選的二正整數以 X, Y 表之，依題意 X 與 Y 應獨立。以 (X, Y) 表 X 與 Y 之最大公因數，互質就是 $(X, Y) = 1$ 。 $(X, Y) = i$ ，表 $X = iU, Y = iV$ ，其中 $(U, V) = 1$ 。 X, Y 皆要屬於 i 的倍數之集合 $\{i \cdot 1, i \cdot 2, \dots\}$ ，此機率為 $(1/i) \cdot (1/i) = 1/i^2$ ，又 $P((U, V) = 1) = p$ 。故 $P((X, Y) = i) = (1/i^2) \cdot p = p/i^2$ 。因此

$$1 = \sum_{i=1}^{\infty} P((X, Y) = i) = \sum_{i=1}^{\infty} \frac{p}{i^2}。$$

故得

$$p = \left(\sum_{i=1}^{\infty} \frac{1}{i^2} \right)^{-1} = \frac{6}{\pi^2} \doteq 0.6079。$$

即任選二正整數互質比不互質較可能發生。

我們再給兩個補充說明。首先，若(2.12)(或(2.16))式在一集合 A 不成立，但 $\int \int_A dx dy = 0$ ，則 X 與 Y 仍為獨立的隨機變數。例如，若

$$f(x, y) = \begin{cases} e^{-\lambda x - \mu y} & , x, y > 0, x \neq y, \\ e^{-\lambda x y} & , x, y > 0, x = y。 \end{cases}$$

令 $A = \{(x, y) | x, y > 0, x = y\}$ 。由於

$$\int \int_A dx dy = \int_0^{\infty} \int_y^y 1 dx dy = \int_0^{\infty} 0 dy = 0,$$

故 X 與 Y 獨立。

其次,若 X 與 Y 獨立,由(2.15)式,可得

$$E(Y|X) = E(Y),$$

條件期望值與給定的 X 無關。反之若上式成立,是否導致 X 與 Y 獨立呢?此答案我們留在第3.4節的習題第5題。

本節最後我們來介紹**偏斜常態分佈**(skew normal distribution)。此分佈族包含常態分佈,且與常態分佈有一些相同或相似的性質。但此分佈由於有一**偏斜參數**(skewness parameter),適合當做有**偏斜數據**(即p.d.f.並非左右對稱,期望值也不等於中位數)的機率模型。

一隨機變數 Z ,若其p.d.f.為

$$(2.25) \quad f(z) = 2\phi(z)\Phi(\lambda z), -\infty < z < \infty,$$

其中 $\lambda \in R$,且 $\phi(z)$, $\Phi(z)$ 分別為 $\mathcal{N}(0, 1)$ 分佈之p.d.f.及分佈函數,便稱為有參數 λ 之**偏斜常態分佈**,以 $\mathcal{SN}(\lambda)$ 表之。

圖2.1給出兩個 $\mathcal{SN}(\lambda)$ 分佈p.d.f.的圖形。可看出圖形是有偏斜的,不像常態分佈的圖形為左右對稱。

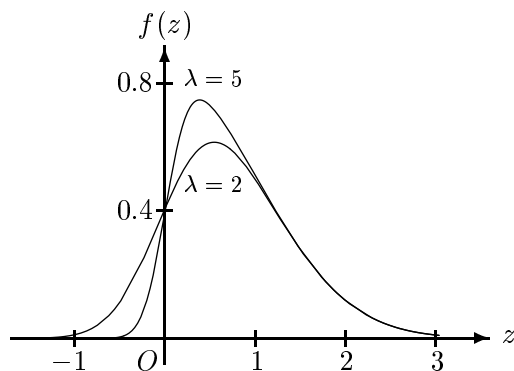


圖2.1 $\mathcal{SN}(2)$ 及 $\mathcal{SN}(5)$ p.d.f.之圖形

我們列出一些偏斜常態分佈簡單的性質如下。

- (i) $\mathcal{SN}(0) = \mathcal{N}(0, 1)$;
- (ii) 若 $Z \sim \mathcal{SN}(\lambda)$, 則 $-Z \sim \mathcal{SN}(-\lambda)$;
- (iii) 若 $Z \sim \mathcal{SN}(\lambda)$, 則 $Z^2 \sim \chi_1^2$;

(iv) 若 $Z \sim \mathcal{SN}(\lambda)$, 則 Z 之分佈函數為

$$\Phi(z; \lambda) = P(Z \leq z) = 2 \int_{-\infty}^z \int_{-\infty}^{\lambda t} \phi(t)\phi(u)du dt;$$

(v) $1 - \Phi(-z; \lambda) = \Phi(z; -\lambda)$;

(vi) $\Phi(z; 1) = (\Phi(z))^2$ 。

上述這些性質的證明並不困難, 因此留在習題中, 其他性質可參考 Azzalini(1985)。

由表面的形式, 一時可能不易看出(2.23)式的確定義出一p.d.f.。即使由性質(iv), 也不易看出

$$\lim_{z \rightarrow \infty} \Phi(z; \lambda) = 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda t} \phi(t)\phi(u)du dt = 1。$$

事實上, 設 X, Y 為二獨立的隨機變數, 且皆有 $\mathcal{N}(0, 1)$ 分佈, 因 $X \stackrel{d}{=} (-X)$, $Y \stackrel{d}{=} (-Y)$, 故

$$-(X - \lambda Y) = (-X) - \lambda(-Y) \stackrel{d}{=} X - \lambda Y。$$

因此

$$P(X - \lambda Y < 0) = \frac{1}{2}。$$

又

$$\begin{aligned} \frac{1}{2} &= P(X - \lambda Y < 0) = P(X < \lambda Y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda y} \phi(x)\phi(y)dx dy \\ &= \int_{-\infty}^{\infty} \phi(y)\Phi(\lambda y)dy。 \end{aligned}$$

由於 $2\phi(z)\Phi(\lambda z) \geq 0$, 且

$$\int_{-\infty}^{\infty} 2\phi(z)\Phi(\lambda z)dz = 1,$$

故 $2\phi(z)\Phi(\lambda z)$ 為一p.d.f.。習題裡我們會給出較一般的結果。底下先給一例。

例2.18 設 X, Y 為二獨立的隨機變數, 且皆有 $\mathcal{N}(0, 1)$ 分佈。令

$$Z = \begin{cases} X & , \text{若 } \lambda X > Y, \\ -X & , \text{若 } \lambda X \leq Y, \end{cases}$$

試證對 $\forall \lambda \in R$, Z 有 $\mathcal{SN}(\lambda)$ 分佈。

證明.我們只證 $\lambda > 0$ 的情況, $\lambda \leq 0$ 的情況可類似地證出。底下來求 Z 之分佈函數。

$$\begin{aligned} (2.26) \quad F(z) &= P(Z \leq z) \\ &= P(Z \leq z, \lambda X > Y) + P(Z \leq z, \lambda X \leq Y) \\ &= P(X \leq z, X > Y/\lambda) + P(-X \leq z, \lambda X \leq Y) \\ &= P(Y/\lambda < X \leq z, Y \leq \lambda z) + P(-z \leq X \leq Y/\lambda, Y \geq -\lambda z) \\ &= \int_{-\infty}^{\lambda z} \int_{y/\lambda}^z \phi(x)\phi(y)dx dy + \int_{-\lambda z}^{\infty} \int_{-z}^{y/\lambda} \phi(x)\phi(y)dx dy \\ &= \int_{-\infty}^{\lambda z} (\Phi(z) - \Phi(y/\lambda))\phi(y)dy + \int_{-\lambda z}^{\infty} (\Phi(y/\lambda) - \Phi(-z))\phi(y)dy \\ &= \Phi(z)\Phi(\lambda z) - \int_{-\infty}^{\lambda z} \Phi(y/\lambda)\phi(y)dy \\ &\quad + \int_{-\lambda z}^{\infty} \Phi(y/\lambda)\phi(y)dy - \Phi(-z)(1 - \Phi(-\lambda z))。 \end{aligned}$$

此處用到

$$P(X \leq z, X > Y/\lambda, Y > \lambda z) = 0,$$

$$P(-X \leq z, \lambda X \leq Y, Y < -\lambda z) = 0,$$

$$P(X \leq z, X > Y/\lambda, Y \leq \lambda z) = P(Y/\lambda < X \leq z, Y \leq \lambda z),$$

且

$$P(-X \leq z, \lambda X \leq Y, Y \geq -\lambda z) = P(-z \leq X \leq Y/\lambda, Y \geq -\lambda z)。$$

現利用

$$\phi(z) = \phi(-z), \quad \Phi(-z) = 1 - \Phi(z),$$

及微積分基本定理, 由(2.26)式即得

$$(2.27) \quad \frac{dF(z)}{dz} = 2\phi(z)\Phi(\lambda z), -\infty < z < \infty.$$

得證 Z 有 $SN(\lambda)$ 分佈。

3.3 變數代換

在1.5節我們曾介紹單變數的變數代換。對隨機向量 (X, Y) ，有時我們會對 X, Y 的二函數，如

$$(3.1) \quad U = g_1(X, Y),$$

$$(3.2) \quad V = g_2(X, Y),$$

有興趣。如果令 $A = \{(x, y) | f(x, y) > 0\}$ ，其中 $f(x, y)$ 為 X, Y 之聯合 p.d.f., $B = \{(u, v) | u = g_1(x, y), v = g_2(x, y), (x, y) \in A\}$ 。則 (U, V) 便是由 A 映至 B 之一變數代換。有時也可能只對一個函數 $W = g(X, Y)$ 有興趣。這兩種情況都屬於兩個變數的變數代換。

例如，令 X 表某位學生第一次考試的成績， Y 表第二次考試的成績。我們可能想對其進步成績 $U = Y - X$ ，與總分 $V = X + Y$ 有興趣，也可能只想對其進步的比率 $W = (Y - X)/X$ 有興趣。 (U, V) 或是單獨的 U, V ， W 皆是將原來的隨機向量 (X, Y) 變換(或說轉換)至新的變數。

設由 (X, Y) 經由(3.1)及(3.2)式變換至 (U, V) 。若由(3.1)及(3.2)式可解出唯一的 $X = h_1(U, V)$, $Y = h_2(U, V)$ ，則我們說由 (X, Y) 至 (U, V) 的變換，為 $1-1$ 。此時有

$$(3.3) \quad u = g_1(x, y), v = g_2(x, y) \Leftrightarrow x = h_1(u, v), y = h_2(u, v).$$

例如，設

$$U = X + Y, \quad V = X - Y,$$

則

$$X = \frac{1}{2}(U + V) = h_1(U, V), \quad Y = \frac{1}{2}(U - V) = h_2(U, V).$$

故 (X, Y) 至 (U, V) 為 $1-1$ 變換。另外，設 X, Y 為二正的隨機變數，且令

$$U = XY, \quad V = X/Y,$$

則

$$X = (UV)^{1/2} = h_1(U, V), \quad Y = (U/V)^{1/2} = h_2(U, V),$$

因此 (X, Y) 至 (U, V) 仍為1-1變換。

若 (X, Y) 為離散型的隨機向量，以 $f(x, y)$ 為聯合p.d.f.。則由(3.3)式， (U, V) 之聯合p.d.f.為

$$(3.4) \quad f_{U,V}(u, v) = f(h_1(u, v), h_2(u, v))。$$

此因

$$\begin{aligned} f_{U,V}(u, v) &= P(g_1(X, Y) = u, g_2(X, Y) = v) \\ &= P(X = h_1(u, v), Y = h_2(u, v)) \\ &= f(h_1(u, v), h_2(u, v))。 \end{aligned}$$

做變數代換，務必要將新的支撐(即前述 B)找出來。見下例。

例3.1 設 (X, Y) 之聯合p.d.f.為

$$f(x, y) = \frac{n!}{x!y!(n-x-y)!} p_1^x p_2^y (1-p_1-p_2)^{n-x-y},$$

其中 $p_1, p_2 \geq 0$, $p_1 + p_2 \leq 1$, x, y 為非負整數，且 $x + y \leq n$ 。令

$$U = X + Y, \quad V = X/(X + Y)。$$

其中當 $X = Y = 0$ 時，令 $V = 0$ 。則

$$X = UV = h_1(U, V), \quad Y = U(1 - V) = h_2(U, V),$$

故 (X, Y) 至 (U, V) 為1-1的變換。因此 (U, V) 之聯合p.d.f.為

$$\begin{aligned} f_{U,V}(u, v) &= f(uv, u(1-v)) \\ &= \frac{n!}{(uv)!(u(1-v))!(n-u)!} p_1^{uv} p_2^{u(1-v)} (1-p_1-p_2)^{n-u}, \end{aligned}$$

其中 $uv, u(1-v)$ 為非負整數，且 $0 \leq u \leq n$ 。

在做變數代換時，要小心求出新的變數之範圍。底下亦為一例。

例3.2 設 X, Y 為二獨立的隨機變數，分別有 $\mathcal{P}(\lambda)$ 及 $\mathcal{P}(\mu)$ 分佈。則 (X, Y) 之聯合p.d.f.為

$$f(x, y) = \frac{e^{-\lambda} \lambda^x}{x!} \frac{e^{-\mu} \mu^y}{y!}, x, y = 0, 1, 2, \dots。$$

我們想求 $U = X + Y$ 之分佈。

雖然 (X, Y) 至 U 並非1-1的變換，但若引進 $V = Y$ ，則 (X, Y) 至 (U, V) 為1-1的變換，且

$$X = U - V, \quad Y = V。$$

故

$$\begin{aligned} f_{U,V}(u, v) &= f(u - v, v) \\ &= \frac{e^{-\lambda} \lambda^{u-v}}{(u-v)!} \frac{e^{-\mu} \mu^v}{v!}, v = 0, 1, 2, \dots, u = v, v+1, v+2, \dots。 \end{aligned}$$

有了聯合p.d.f.，便可求出 U 的邊際p.d.f.:

$$\begin{aligned} f_U(u) &= \sum_{v=0}^u \frac{e^{-\lambda} \lambda^{u-v}}{(u-v)!} \frac{e^{-\mu} \mu^v}{v!} \\ &= e^{-(\lambda+\mu)} \sum_{v=0}^u \frac{\lambda^{u-v} \mu^v}{(u-v)! v!} \\ &= \frac{e^{-(\lambda+\mu)}}{u!} \sum_{v=0}^u \frac{u!}{(u-v)! v!} \lambda^{u-v} \mu^v \\ &= \frac{e^{-(\lambda+\mu)} (\lambda + \mu)^u}{u!}, \end{aligned}$$

其中倒數第二個等號右側的和等於 $(\lambda + \mu)^u$ 是用到二項式定理。故得 U 有 $\mathcal{P}(\lambda + \mu)$ 分佈，與例2.16利用拉普拉斯轉換所得相同。

讀者可利用例3.2中的方法，對二獨立的隨機變數 X, Y ，且分別有 $\mathcal{B}(m, p)$ 及 $\mathcal{B}(n, p)$ 分佈，導出 $X + Y$ 有 $\mathcal{B}(m + n, p)$ 分佈，仍與利用拉普拉斯轉換所得相同。

有時我們只對 X, Y 之一函數 $U = g(X, Y)$ 有興趣, 例3.2便提供一典型的作法: 先引進另一隨機變數 $V = h(X, Y)$, 然後求 (U, V) 之聯合分佈, 再求 U 之邊際分佈。

若 (X, Y) 為連續型的隨機向量, 仍以 $f(x, y)$ 為其聯合p.d.f., 則 (U, V) 之聯合p.d.f., 可如單變數的情況(見第一章(5.5)式)得到。即

$$(3.5) \quad f_{U,V}(u, v) = |J(u, v)|f(h_1(u, v), h_2(u, v)),$$

其中

$$(3.6) \quad J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial y}{\partial u} \frac{\partial x}{\partial v},$$

而

$$\begin{aligned} \frac{\partial x}{\partial u} &= \frac{\partial h_1(u, v)}{\partial u}, & \frac{\partial x}{\partial v} &= \frac{\partial h_1(u, v)}{\partial v}, \\ \frac{\partial y}{\partial u} &= \frac{\partial h_2(u, v)}{\partial u}, & \frac{\partial y}{\partial v} &= \frac{\partial h_2(u, v)}{\partial v}. \end{aligned}$$

J 稱為變換 $x = h_1(u, v)$, $y = h_2(u, v)$ 之**雅可彬**(Jacobian)。令

$$B = \{(u, v) | u = g_1(x, y), v = g_2(x, y)\}。$$

我們假設 J 在 B 中不恆為0。(3.5)式的由來是利用重積分之變數代換。即

$$\begin{aligned} F_{U,V}(u, v) &= P(U \leq u, V \leq v) \\ &= P(g_1(X, Y) \leq u, g_2(X, Y) \leq v) \\ &= \int \int_A f(x, y) dx dy \\ &= \int_{-\infty}^v \int_{-\infty}^u |J(s, t)| f(h_1(s, t), h_2(s, t)) ds dt, \end{aligned}$$

其中 $A = \{(x, y) | g_1(x, y) \leq u, g_2(x, y) \leq v\}$, 且用到重積分之變數代換, 令 $s = g_1(x, y)$, $t = g_2(x, y)$ 。再利用

$$f_{U,V}(u, v) = \frac{\partial^2}{\partial u \partial v} F_{U,V}(u, v),$$

及微積分基本定理, 即得(3.5)式。

定理3.1 設 X, Y 為二獨立的隨機變數, 分別有 $\Gamma(p, \lambda)$ 及 $\Gamma(q, \lambda)$ 分佈。令 $U = X/(X + Y), V = X + Y$ 。則 U 與 V 獨立, 且 U 有 $\mathcal{B}e(p, q)$ 分佈, V 有 $\Gamma(p + q, \lambda)$ 分佈。

證明. 首先 X, Y 之聯合p.d.f.為

$$f(x, y) = \frac{x^{p-1}e^{-x/\lambda}}{\Gamma(p)\lambda^p} \frac{y^{q-1}e^{-y/\lambda}}{\Gamma(q)\lambda^q}, x, y > 0。$$

由

$$u = x/(x + y), \quad v = x + y, \quad 0 < u < 1, \quad v > 0。$$

解出

$$x = h_1(u, v) = uv, y = h_2(u, v) = v(1 - u)。$$

則

$$J = \begin{vmatrix} v & u \\ -v & 1 - u \end{vmatrix} = v。$$

故得

$$\begin{aligned} f_{U,V}(u, v) &= |J|f(h_1(u, v), h_2(u, v)) \\ &= v \frac{(uv)^{p-1}e^{-uv/\lambda}(v(1-u))^{q-1}e^{-v(1-u)/\lambda}}{\Gamma(p)\Gamma(q)\lambda^{p+q}}, \\ &0 < u < \infty, 0 < v < 1。 \end{aligned}$$

經化簡後得

$$f_{U,V}(u, v) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1} \frac{v^{p+q-1}e^{-v/\lambda}}{\Gamma(p+q)\lambda^{p+q}}。$$

由定義2.3知, U 與 V 獨立, 且分別有 $\mathcal{B}e(p, q)$ 及 $\Gamma(p + q, \lambda)$ 分佈。

我們說明一下定理3.1。首先此定理涵蓋定理2.4之(iii)。另外, 此定理為數理統計中一重要且特殊的結果。由表面的型式看來, U, V 不

像會獨立: $U = X/(X + Y)$, 而 $V = X + Y$ 剛好是 U 的分母。但對 X, Y 皆為 gamma 分佈, 且有同樣的尺度參數, 則 U 與 V 獨立。事實上此為 gamma 分佈之一特性。即若已知 X 與 Y 為取正值的隨機變數, 且知 $V = X + Y$ 與 $U = X/(X + Y)$ 獨立, 則 X 與 Y 皆為 gamma 分佈, 且有相同的尺度參數, 這是 Lukacs(1955) 所證出, 現已有許多推廣。

底下結果為定理 3.1 之一立即的推論。

系理 3.1 設 U 與 V 獨立, 且 U 有 $Be(p, q)$ 分佈, V 有 $\Gamma(p + q, \lambda)$ 分佈。則 UV 有 $\Gamma(p, \lambda)$ 分佈。

系理 3.1 指出, 當參數有某特殊的關係, 則獨立的 gamma 隨機變數與 beta 隨機變數相乘後, 仍有 gamma 分佈。這是一有趣的結果。另外, 二獨立的 beta 隨機變數相乘後, 亦可能仍有 beta 分佈, 我們留在習題中讓各位自行練習。

再給一個例子。

例 3.3 設 X 與 Y 為二獨立的隨機變數, 且皆有 $\mathcal{U}(1, \beta)$ 分佈, $\beta > 1$ 。令 $U = XY$ 。試求 U 之 p.d.f.。

解. 首先 (X, Y) 之聯合 p.d.f. 為

$$f(x, y) = \frac{1}{(\beta - 1)^2}, 1 < x, y < \beta.$$

令 $V = Y$ 。由 $u = xy, v = y$, 解出

$$x = h_1(u, v) = u/v, \quad y = h_2(u, v) = v.$$

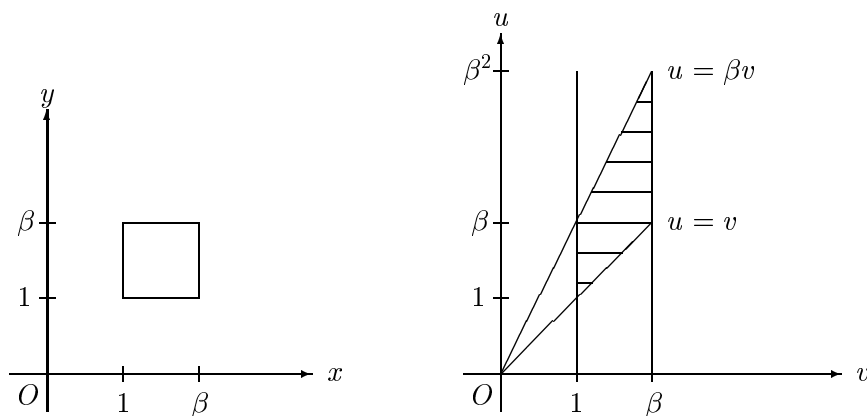
則

$$J = \begin{vmatrix} \frac{1}{v} & -\frac{u}{v^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{v}.$$

故得

$$\begin{aligned} (3.7) \quad f_{U,V}(u, v) &= |J| f_{X,Y}(h_1(u, v), h_2(u, v)) \\ &= \frac{1}{(\beta - 1)^2} \frac{1}{v}, 1 < \frac{u}{v} < \beta, 1 < v < \beta. \end{aligned}$$

我們分別給出 x, y 及 u, v 取值的範圍如下二圖。



由(3.7)式得

$$f_U(u) = \begin{cases} \frac{1}{(\beta-1)^2} \int_1^u \frac{1}{v} dv = \frac{\log u}{(\beta-1)^2}, & 1 < u < \beta, \\ \frac{1}{(\beta-1)^2} \int_{u/\beta}^{\beta} \frac{1}{v} dv = \frac{2\log \beta - \log u}{(\beta-1)^2}, & \beta \leq u < \beta^2. \end{cases}$$

可驗證 $f_U(u)$ 確為一p.d.f.(習題第13題)。

上例顯示, 在做變數代換時, 要小心新變數取值的範圍, 必要時畫一簡圖。底下再給一自表面上看不像獨立, 但卻是獨立的例子。

例3.4 設 X, Y 為獨立, 且 X 有 $\mathcal{N}(\mu, \sigma^2)$ 分佈, Y 有 $\mathcal{N}(\gamma, \sigma^2)$ 分佈。令 $U = X + Y, V = X - Y$ 。則 U 與 V 獨立, 且 U 有 $\mathcal{N}(\mu + \gamma, 2\sigma^2)$ 分佈, V 有 $\mathcal{N}(\mu - \gamma, 2\sigma^2)$ 分佈。

證明. 令 $u = x + y, v = x - y, u, v \in R$, 解出

$$x = h_1(u, v) = \frac{u + v}{2}, \quad y = h_2(u, v) = \frac{u - v}{2}.$$

則

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}.$$

故得

$$\begin{aligned}
f_{U,V}(u, v) &= |J|f_{X,Y}(h_1(u, v), h_2(u, v)) \\
&= \frac{1}{2} \frac{1}{2\pi\sigma^2} e^{-((u+v)/2-\mu)^2/(2\sigma^2) - ((u-v)/2-\gamma)^2/(2\sigma^2)} \\
&= \left(\frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} e^{-(u-\mu-\gamma)^2/(2\cdot 2\sigma^2)} \right) \cdot \left(\frac{1}{\sqrt{2\pi}\sqrt{2}\sigma} e^{-(v-\mu+\gamma)^2/(2\cdot 2\sigma^2)} \right),
\end{aligned}$$

中間的計算過程並不難，讓讀者自行補上。故得證 U 與 V 獨立，且分別有 $\mathcal{N}(\mu + \gamma, 2\sigma^2)$ 及 $\mathcal{N}(\mu - \gamma, 2\sigma^2)$ 分佈。

在上例中，若 $\mu = \gamma = 0$ ，則 U 與 V 不但獨立，且有相同的 $\mathcal{N}(0, 2\sigma^2)$ 分佈。底下的定理指出，對二獨立的隨機變數，其各自的函數亦獨立。

定理3.2 設 X, Y 為二獨立的隨機變數， $g(x), h(y)$ 分別為 x, y 之函數。則 $U = g(X)$ 與 $V = h(Y)$ 為二獨立的隨機變數。

證明. 對任意 $u, v \in R$ ，令

$$A_u = \{x | g(x) \leq u\}, \quad B_v = \{y | h(y) \leq v\}。$$

則

$$\begin{aligned}
F_{U,V}(u, v) &= P(U \leq u, V \leq v) \\
&= P(X \in A_u, Y \in B_v) \\
&= P(X \in A_u)P(Y \in B_v),
\end{aligned}$$

其中最後一等式成立是因 X 與 Y 獨立，故可利用定理2.2(i)。因 $P(X \in A_u)$ 及 $P(Y \in B_v)$ 分別為 u 及 v 之函數，故由系理2.1，即得 U 與 V 獨立。

例3.5 設 X, Y 為二獨立的隨機變數，分別有 $\Gamma(p, \lambda)$ 及 $\Gamma(q, \lambda)$ 分佈。令 $U = X + Y$ ， $V = X/(X + Y)$ 。由定理3.1，知 U 與 V 獨立。再由定理3.2知， U 與 $V^{-1} - 1 = Y/X$ 獨立。

有時由 (X, Y) 至 (U, V) 的變換並非 $1-1$ ，此時可仿單變數時的作法，可參考第一章(5.7)式。即設存在 $A = \{(x, y) | f(x, y) > 0\}$ 之一分割 $A_0, A_1, \dots, A_n$ ，其中 A_0 可能是空集合，滿足 $P((X, Y) \in A_0) = 0$ ，使得對

$\forall i = 1, \dots, k$, $U = g_1(X, Y)$, $V = g_2(X, Y)$ 爲由 A_i 映成至 B 之 $1-1$ 函數。則對 $\forall i = 1, \dots, k$, 由 B 至 A_i 之反函數存在, 以 $x = h_{1i}(u, v)$, $y = h_{2i}(u, v)$ 表第 i 個反函數。對第 i 個反函數, $\forall (u, v) \in B$, 有唯一的 $(x, y) \in A$, 使得

$$(u, v) = (g_1(x, y), g_2(x, y))。$$

令 J_i 表第 i 個反函數之雅可彬。假設 $J_1, J_2, \dots, J_k$ 在 B 中皆不爲 0。則 (U, V) 之聯合 p.d.f. 爲

$$(3.8) \quad f_{U,V}(u, v) = \sum_{i=1}^k |J_i| f_{X,Y}(h_{1i}(u, v), h_{2i}(u, v))。$$

例3.6 設 X 與 Y 爲二獨立的隨機變數, 皆有 $\mathcal{N}(0, 1)$ 分佈。令 $U = X/Y$, $V = |Y|$ 。當 $Y = 0$ 時, 可定義 $U = 1$ (或任何其他常數值)。由於對連續型的隨機變數, $P(Y = 0) = 0$, 所以倒不用擔心 $Y = 0$ 的情況。

因 (x, y) 與 $(-x, -y)$, 對應到相同的 (u, v) , 故由 (X, Y) 至 (U, V) 之變換並非 $1-1$ 。但若令

$$A_1 = \{(x, y) | x \in R, y > 0\}, \quad A_2 = \{(x, y) | x \in R, y < 0\},$$

$$A_0 = \{(x, y) | x \in R, y = 0\},$$

則 A_0, A_1, A_2 形成 $A = R^2$ 之一分割, 且 $P((X, Y) \in A_0) = P(Y = 0) = 0$ 。又 $B = \{(u, v) | u \in R, v > 0\}$ 。在 A_1 或 A_2 上, 由 (X, Y) 至 (U, V) 之變換皆爲 $1-1$ 。由 B 至 A_1 之反函數爲

$$x = h_{11}(u, v) = uv,$$

$$y = h_{21}(u, v) = v。$$

由 B 至 A_2 之反函數爲

$$x = h_{12}(u, v) = -uv,$$

$$y = h_{22}(u, v) = -v。$$

$J_1 = J_2 = v$ 。又

$$f(x, y) = \frac{1}{2\pi} e^{-x^2/2 - y^2/2}。$$

故

$$\begin{aligned} f_{U,V}(u, v) &= \frac{1}{2\pi} |v| e^{-u^2 v^2/2 - v^2/2} + \frac{1}{2\pi} |v| e^{-u^2 v^2/2 - v^2/2} \\ &= \frac{v}{\pi} e^{-(1+u^2)v^2/2}, u \in R, v > 0。 \end{aligned}$$

由此即得

$$\begin{aligned} f_U(u) &= \int_0^\infty \frac{v}{\pi} e^{-(1+u^2)v^2/2} dv \quad (\text{令 } v^2 = z) \\ &= \frac{1}{2\pi} \int_0^\infty e^{-(1+u^2)z/2} dz \\ &= \frac{1}{\pi(1+u^2)}, u \in R。 \end{aligned}$$

即 X/Y 有 $C(0, 1)$ 分佈。

在2.3節介紹柯西分佈時，我們曾提到此分佈有時在我們想不到的地方出現。此即爲一例。習題中尚有一些例子， X 與 Y 之變異數也不須爲0。

在上例中，若僅是要求 X/Y 之分佈，則可令 $U = X/Y$ ， $V = Y$ 。則此爲由 (X, Y) 至 (U, V) 之1-1變換。求出 (U, V) 之聯合p.d.f.

$$f_{U,V}(u, v) = \frac{v}{2\pi} e^{-(1+u^2)v^2/2}, u, v \in R,$$

之後，再求 U 之邊際p.d.f.，仍得 U 有 $C(0, 1)$ 分佈。

3.4 混合分佈

有時一隨機變數之參數，亦允許爲隨機變數。我們先看一個例子。

例4.1 假設高雄市每天發生之交通事故的次數 Y 有 $\mathcal{P}(\lambda)$ 分佈。每次交通事故有 p 的機率會有死亡。令 X 表每天有死亡之交通事件。則 $X|Y$ 有 $\mathcal{B}(Y, p)$ 分佈，而 Y 有 $\mathcal{P}(\lambda)$ 分佈。本模式即二項分佈試驗的次數亦爲一隨機變數 Y 。給定 $Y = y$ ，則 X 有 $\mathcal{B}(y, p)$ 分佈。至於 Y 則有 $\mathcal{P}(\lambda)$ 分佈。當一隨機

變數 X 之參數亦為隨機變數時，我們便說 X 有一**混合分佈**(mixture distribution)。

在本例中， X 之**非條件分佈**如何求？由假設，且利用第一章定理2.1，得

$$\begin{aligned}
 P(X = x) &= \sum_{y=0}^{\infty} P(X = x|Y = y)P(Y = y) \\
 &= \sum_{y=0}^{\infty} \binom{y}{x} p^x (1-p)^{y-x} \frac{e^{-\lambda} \lambda^y}{y!} \\
 &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{y=x}^{\infty} \frac{((1-p)\lambda)^{y-x}}{(y-x)!} \\
 &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \sum_{i=0}^{\infty} \frac{((1-p)\lambda)^i}{i!} \\
 &= \frac{(\lambda p)^x e^{-\lambda}}{x!} \cdot e^{\lambda(1-p)} = \frac{e^{-\lambda p} (\lambda p)^x}{x!}。
 \end{aligned}$$

因此 X 之非條件分佈為 $\mathcal{P}(\lambda p)$ 。

在上例中，每天死亡人數之期望值為 $E(X) = \lambda p$ 。又顯然 $E(X|Y) = Yp$, $E(Y) = \lambda$, 故得 $E(X) = E(E(X|Y)) = E(Yp) = pE(Y) = \lambda p$ 。這就是**總機率法則**(Law of total probability): 條件期望值之期望值，等於原期望值。我們列成如下的定理。

定理4.1 對任二隨機變數 X, Y ，只要下述各期望值皆為有限，便有

$$(4.1) \quad E(X) = E(E(X|Y))。$$

證明.我們只證連續型隨機變數的情況。設 (X, Y) 之聯合p.d.f.為 $f(x, y)$ 。由定義

$$\begin{aligned}
 E(X) &= \int \int x f(x, y) dx dy \\
 &= \int \left(\int x f(x|y) dx \right) f_Y(y) dy \\
 &= \int E(X|y) f_Y(y) dy \\
 &= E(E(X|Y)),
 \end{aligned}$$

其中第一個等號右側的積分是在 $A = \{(x, y) | f(x, y) > 0\}$, 其餘積分也分別在適當的範圍內。

由例4.1或定理4.1知, 所謂混合分佈, 乃是分層來看隨機變數。有時這樣子描述一機率模式, 會較容易理解。像例4.1, 對每天發生的交通事故, 給它一個分佈。每次事故又有一定的機率會有死亡。這樣子分兩層來想, 似乎容易多了。有時還可分好幾層, 見下例。

例4.2 承例4.1。假設不同的城市, 每天發生交通事故的次數雖皆可以波松分佈當模式, 但參數可能不同。即設 $Y|\Lambda$ 有 $\mathcal{P}(\Lambda)$ 分佈, 又設 Λ 有 $\mathcal{E}(\beta)$ 分佈。則

$$\begin{aligned} E(X) &= E(E(X|Y)) = E(pY) \\ &= pE(Y) = pE(E(Y|\Lambda)) \\ &= pE(\Lambda) = p/\beta。 \end{aligned}$$

此一三層的模式, 亦可表為兩層的模式。我們來求 Y 之非條件分佈如下。對 $\forall y = 0, 1, 2, \dots$,

$$\begin{aligned} P(Y = y) &= \int_0^\infty f_{Y,\Lambda}(y, \lambda) d\lambda \\ &= \int_0^\infty f_{Y|\Lambda}(y|\lambda) f_\Lambda(\lambda) d\lambda \\ &= \int_0^\infty \frac{e^{-\lambda} \lambda^y}{y!} \beta e^{-\beta\lambda} d\lambda \\ &= \frac{\beta}{y!} \int_0^\infty \lambda^y e^{-(1+\beta)\lambda} d\lambda \\ &= \frac{\beta}{y!} \frac{\Gamma(y+1)}{(1+\beta)^{y+1}} \\ &= \frac{\beta}{1+\beta} \left(\frac{1}{1+\beta} \right)^y。 \end{aligned}$$

可看出 Y 有自0開始之幾何分佈, 每次成功的機率為 $\beta/(1+\beta)$ 。知道 Y 的分佈, 就可以不用管 Λ 了(也可以說就不知道有 Λ 存在)。但以三層的方式來描述此機率模式, 顯然讓人較容易理解。

大家是否注意到, 在上例中, 我們引進了一與以前不太一樣的模式, 即二隨機變數一為連續一為離散。不過倒不至於有問題。例如, 在定義 $f_{Y,\Lambda}(y, \lambda)$ 時, 只要記住在求機率或期望值時, 若是離散型的隨機變數便用加的, 連續型的隨機變數則用積分。

底下再給一例。

例4.3 設 $X|P = p$ 有 $Ber(p)$ 分佈, 而 P 有 $Be(\alpha, \beta)$ 分佈。則仿例4.2, 對 $x = 0, 1$, 得

$$\begin{aligned} P(X = x) &= \int_0^1 p^x (1-p)^{1-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p^{\alpha+x-1} (1-p)^{\beta-x+1-1} dp \\ &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha+x)\Gamma(\beta-x+1)}{\Gamma(\alpha+\beta+1)} \\ &= \frac{\Gamma(\alpha+x)\Gamma(\beta-x+1)}{(\alpha+\beta)\Gamma(\alpha)\Gamma(\beta)} \\ &= \begin{cases} \frac{\alpha}{\alpha+\beta}, & x=1, \\ \frac{\beta}{\alpha+\beta}, & x=0. \end{cases} \end{aligned}$$

可看出 X 之非條件分佈為 $Ber(\alpha/(\alpha+\beta))$, 即仍有伯努力分佈, 只是參數改變了: 由 P 變成 P 的期望值 $\alpha/(\alpha+\beta)$ 。由此即得

$$E(X) = \frac{\alpha}{\alpha+\beta}。$$

但若只是要求 $E(X)$, 則可不用求出 X 之非條件分佈, 而利用總機率法則:

$$E(X) = E(E(X|P)) = E(P) = \frac{\alpha}{\alpha+\beta},$$

所得相同。

總機率法則以分兩層來求期望值。對於求變異數則有下述定理。

定理4.2 對任二隨機變數 X, Y , 只要下述各期望值及變異數皆存在, 便有

$$(4.2) \quad \text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(E(X|Y))。$$

證明.由定義

$$\begin{aligned}
 (4.3) \quad \text{Var}(X) &= E((X - E(X))^2) \\
 &= E((X - E(X|Y) + E(X|Y) - E(X))^2) \\
 &= E((X - E(X|Y))^2) + E((E(X|Y) - E(X))^2) \\
 &\quad + 2E((X - E(X|Y))(E(X|Y) - E(X)))。
 \end{aligned}$$

我們先看上式第三個等號右側最後一項，它必須是0，原因如下：

$$\begin{aligned}
 (4.4) \quad &E((X - E(X|Y))(E(X|Y) - E(X))) \\
 &= E(E((X - E(X|Y))(E(X|Y) - E(X))|Y))。
 \end{aligned}$$

但給定 Y 之下， $E(X|Y)$ 及 $E(X)$ 皆為常數。故

$$\begin{aligned}
 (4.5) \quad &E((X - E(X|Y))(E(X|Y) - E(X))|Y) \\
 &= (E(X|Y) - E(X))E(X - E(X|Y)|Y) \\
 &= (E(X|Y) - E(X))(E(X|Y) - E(X|Y)) \\
 &= 0。
 \end{aligned}$$

代入(4.4)式，即得

$$E((X - E(X|Y))(E(X|Y) - E(X))) = E(0) = 0。$$

回到(4.3)式，因

$$(4.6) \quad \text{Var}(X|Y) = E((X - E(X|Y))^2|Y),$$

上式也是 $\text{Var}(X|Y)$ 之定義，故

$$\begin{aligned}
 (4.7) \quad E((X - E(X|Y))^2) &= E(E((X - E(X|Y))^2|Y)) \\
 &= E(\text{Var}(X|Y))。
 \end{aligned}$$

又因 $E(E(X|Y)) = E(X)$ ，故

$$(4.8) \quad E((E(X|Y) - E(X))^2) = \text{Var}(E(X|Y))。$$

得證(4.2)式。

上述證明稍嫌冗長，底下的推導較簡潔些：

由(4.6)式，得

$$(4.9) \quad \text{Var}(X|Y) = E(X^2|Y) - (E(X|Y))^2。$$

因此

$$\begin{aligned} E(\text{Var}(X|Y)) &= E(E(X^2|Y) - (E(X|Y))^2) \\ &= E(E(X^2|Y)) - E((E(X|Y))^2) \\ &= E(X^2) - (E(X))^2 - (E((E(X|Y))^2) - (E(X))^2) \\ &= \text{Var}(X) - \text{Var}(E(X|Y))。 \end{aligned}$$

要注意的是，不要將總機率法則誤用在 $\text{Var}(X)$ 。即雖然 $E(X) = E(E(X|Y))$ ，但由(4.2)式，一般而言

$$\text{Var}(X) \neq E(\text{Var}(X|Y)),$$

除非 $\text{Var}(E(X|Y)) = 0$ 。

例4.4 承例4.3，欲求 X 的變異數。

首先，設未求出 X 之非條件分佈。由(4.2)式，

$$(4.10) \quad \text{Var}(X) = \text{Var}(E(X|P)) + E(\text{Var}(X|P))。$$

而因 $X|P$ 有 $\mathcal{Ber}(P)$ 分佈，故

$$E(X|P) = P, \quad \text{Var}(X|P) = P(1 - P)。$$

因此

$$\begin{aligned} \text{Var}(X) &= \text{Var}(P) + E(P(1 - P)) \\ &= E(P^2) - (E(P))^2 + E(P) - E(P^2) \end{aligned}$$

$$\begin{aligned}
&= E(P)(1 - E(P)) \\
&= \frac{\alpha}{\alpha + \beta} \left(1 - \frac{\alpha}{\alpha + \beta}\right) \\
&= \frac{\alpha\beta}{\alpha + \beta}。
\end{aligned}$$

另外，如果已求出 X 之非條件分佈為 $\text{Ber}(\alpha/(\alpha + \beta))$ ，則得到相同的 $\text{Var}(X)$ 。

3.5 共變異數及相關係數

兩個隨機變數可能獨立也可能不獨立。所謂獨立，是隨機式的獨立，表知其中之一的值，對另一變數的值毫無影響。如果二隨機變數不獨立，則二變數間便有關係。此關係可能強可能弱，本節便要介紹兩種度量兩隨機變數關係強弱的方式。

隨機變數關係的確是有強弱的。如果是獨立，則關係當然是最弱的。若樣本為水，令 X 表其體積， Y 表其重量。顯然 X 與 Y 之關係很強。如果取樣多次，將所得的 (X, Y) 數據畫在座標平面上，則所有的點很可能落在一直線上或直線的附近。這是因為水的重量與體積有線性關係。有些數據不在直線上，可能是因量測的誤差，或水質不純所致。其次，若以 X 表某人之身高， Y 表其體重， X 與 Y 顯然亦有關係，但可能不是那麼強。量測一些不同的人所得之 (X, Y) ，大約不會形成一直線，雖然我們仍會預期圖形是向上增長，即較大的 X 有較大的 Y 之傾向。共變異數(covariance，又稱協方差)及相關係數(correlation coefficient，或只稱correlation)，都可用來度量兩隨機變數關係(特別是線性關係)的強弱。

在本節中，令 $\mu_X = E(X)$ ， $\mu_Y = E(Y)$ ， $\sigma_X^2 = \text{Var}(X)$ ， $\sigma_Y^2 = \text{Var}(Y)$ ，如果存在的話。

定義5.1 二隨機變數 X, Y 之共變異數為

$$(5.1) \quad \text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y))。$$

設 $0 < \sigma_X^2, \sigma_Y^2 < \infty$ 。則 X, Y 之相關係數為

$$(5.2) \quad \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}。$$

$\text{Cov}(X, Y)$ 可以 σ_{XY} 表之, $\rho(X, Y)$ 有時以 ρ_{XY} 表之。

當 σ_X 或 $\sigma_Y = 0$, 此時 $\text{Cov}(X, Y) = 0$ (習題第1題), 至於 $\rho(X, Y)$ 則不定義。由定義5.1知, 只要我們提到相關係數, 皆隱含二隨機變數之期望值及變異數皆存在, 且變異數皆不為0。

顧名思義, 共變異數量測兩個變數同時變化的情況。正如變異數乃量測一隨機變數變化的大小 ($\text{Cov}(X, X) = \text{Var}(X)$)。如果較大的 X 傾向於伴隨較大的 Y , 且較小的 X 傾向於伴隨較小的 Y , 則 $\text{Cov}(X, Y)$ 將是正的。因此若 $X > \mu_X$ 時, 較可能有 $Y > \mu_Y$, 則 $(X - \mu_X)(Y - \mu_Y)$ 較可能為正; 若 $X < \mu_X$ 時, 亦較可能有 $Y < \mu_Y$, 則 $(X - \mu_X)(Y - \mu_Y)$ 也將較可能為正。如此一來, $\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)) > 0$ 。但若較大的 X 傾向於伴隨較小的 Y , 且較小的 X 傾向於伴隨較大的 Y , 則 $(X - \mu_X)(Y - \mu_Y)$ 較可能為負, 如此一來, $\text{Cov}(X, Y) < 0$ 。故 $\text{Cov}(X, Y)$ 之正負, 反映 X, Y 增長方向之相同或相反。若以 X, Y 分別表父親的身高及兒子的身高, 我們會預期 $\text{Cov}(X, Y)$ 為正。若以 X, Y 分別表成人的體重及跳高的高度, 我們會預期 $\text{Cov}(X, Y)$ 為負。

較大且為正的共變異數, 顯示兩個變數, 大致以線性的方式同時增大或減小; 較大且為負的共變異數, 顯示兩個變數, 大致以線性的方式一增大時另一減小。共變異數很接近0時, 表示 X 與 Y 的線性關係很弱。特別地, 當 X 與 Y 獨立時, $\text{Cov}(X, Y) = 0$ (見底下系理5.1)。不過 $\text{Cov}(X, Y) = 0$ 時, X 與 Y 卻不一定獨立。另外, $\rho(X, Y)$ 與 $\text{Cov}(X, Y)$ 之符號相同, 且當 $\text{Var}(X)$ 及 $\text{Var}(Y)$ 皆存在時, $\rho(X, Y) = 0$, 若且唯若 $\text{Cov}(X, Y) = 0$ 。當 $\rho(X, Y) = 0$ 則稱 X 與 Y 無相關 (uncorrelated)。

做為量測二隨機變數 X 及 Y 之變化情況, 共變異數有一缺點, 就是其值與量測 X, Y 之尺度有關。例如, X 以公升為單位, Y 以公斤為單位, 跟 X 以公撮 (立方公分) 為單位, Y 以公克為單位, 則前者求出之共變數為後者之 10^{-6} 倍。相關係數便可消除這種困擾, 它永遠取值在區間 $[-1, 1]$ (稍後

會證明)。而相關係數等於1或-1時，表 X 與 Y 有完美的線性關係。我們先給一定理以簡化共變異數之計算。

定理5.1 對二隨機變數 X, Y ,

$$(5.3) \quad \text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y。$$

證明. 由下述推導立即得證。

$$\begin{aligned} \text{Cov}(X, Y) &= E((X - \mu_X)(Y - \mu_Y)) \\ &= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\ &= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y \\ &= E(XY) - \mu_X \mu_Y。 \end{aligned}$$

系理5.1 設 X 與 Y 為二獨立的隨機變數，則 $\text{Cov}(X, Y) = 0$ 且 $\rho(X, Y) = 0$ 。

證明. 由於 X 與 Y 獨立，故 $E(XY) = E(X)E(Y)$ 。因此

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= E(X)E(Y) - E(X)E(Y) = 0。 \end{aligned}$$

其次，

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{0}{\sigma_X \sigma_Y} = 0。$$

證畢。

底下的推論也是很顯然的。

系理5.2 對二隨機變數 X, Y ,

$$\rho(X, Y) = \text{Cov}\left(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}\right) = E\left(\frac{(X - \mu_X)(Y - \mu_Y)}{\sigma_X \sigma_Y}\right)。$$

我們知道, 對一隨機變數 X , 減去期望值, 再除以標準差, 所得之 $(X - \mu_X)/\sigma_X$, 就是將其標準化。由系理5.2知, 相關係數即二標準化後的隨機變數之共變異數。標準化後的隨機變數, 稱為無維數的(dimensionless)。

下述定理的證明不難, 我們留在習題中(第14題)。

定理5.2 對任二隨機變數 X, Y , 及常數 a, b, c, d ,

$$(5.4) \quad \text{Cov}(aX + b, cY + d) = ac\text{Cov}(X, Y),$$

$$(5.5) \quad \rho(aX + b, cY + d) = \frac{ac}{|ac|} \rho(X, Y), ac \neq 0。$$

(5.4)式印證我們前面所說的共變異數與所取的尺度(指 a, c 的效應)有關。不過與座標的平移(指 b, d 的效應)無關。(5.5)式指出, 只要尺度之改變, 並未使 X, Y 之值的方向變成相反(即 $ac > 0$), 則相關係數不變。

例5.1 假設 X 表某電腦廠商拜訪客戶之次數, Y 表客戶買的電腦數目, 且 (X, Y) 之聯合p.d.f. $f(x, y)$ 為

$x \setminus y$	0	1	2	$f_X(x)$
1	0.20	0.10	0.01	0.31
2	0.15	0.30	0.06	0.51
3	0.03	0.05	0.10	0.18
$f_Y(y)$	0.38	0.45	0.17	1

由上表得

$$\mu_X = 1 \cdot 0.31 + 2 \cdot 0.51 + 3 \cdot 0.18 = 1.87,$$

$$\mu_Y = 0 \cdot 0.38 + 1 \cdot 0.45 + 2 \cdot 0.17 = 0.79。$$

又

$$E(XY) = \sum_{x=1}^3 \sum_{y=0}^2 xyf(x, y) = 1.71。$$

故

$$\text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y = 1.71 - 1.87 \cdot 0.79 = 0.2327。$$

因共變異數為正，顯示廠商拜訪客戶愈頻繁，客戶買的電腦數有愈多的傾向。

例5.2 設 (X, Y) 之聯合p.d.f.為

$$f(x, y) = 1, 0 < x < 1, x < y < x + 1。$$

則 X, Y 之邊際p.d.f.分別為

$$f_X(x) = 1, 0 < x < 1,$$

$$f_Y(y) = \begin{cases} y & , 0 < y < 1, \\ 2 - y & , 1 \leq y < 2。 \end{cases}$$

則 $\mu_X = 1/2, \sigma_X^2 = 1/12, \mu_Y = 1, \sigma_Y^2 = 1/6$ 。又

$$\begin{aligned} E(XY) &= \int_0^1 \int_x^{x+1} xy dy dx \\ &= \int_0^1 \frac{1}{2} xy^2 \Big|_x^{x+1} dx \\ &= \int_0^1 (x^2 + \frac{1}{2}x) dx = \frac{7}{12}。 \end{aligned}$$

因此

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - \mu_X \mu_Y = \frac{7}{12} - \frac{1}{2} \cdot 1 = \frac{1}{12}, \\ \rho(X, Y) &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/12}{\sqrt{1/12} \sqrt{1/6}} = \frac{\sqrt{2}}{2}。 \end{aligned}$$

相關係數作為一個指標，主要是反映二隨機變數之分佈的線性關係之強度及符號。因此相關係數為0，僅表示二隨機變數之線性關係很低，而非表示二變數**機率上無關**(probabilistically unrelated, 即獨立)。見下例。

例5.3 設 (X, Y) 為離散型之隨機向量，以 $f(x, y)$ 為其聯合p.d.f.，且

$$\begin{aligned} f(-1, 1) &= f(1, 1) = \frac{1}{3}, \\ f(-2, 4) &= f(2, 4) = \frac{1}{6}. \end{aligned}$$

則

$$\begin{aligned} f_X(-1) &= f_X(1) = \frac{1}{3}, & f_X(-2) &= f_X(2) = \frac{1}{6}, \\ f_Y(1) &= \frac{2}{3}, & f_Y(4) &= \frac{1}{3}. \end{aligned}$$

故

$$\begin{aligned} \mu_X &= 0, & \sigma_X^2 &= 2, \\ \mu_Y &= 2, & \sigma_Y^2 &= 2, \\ E(XY) &= 0, \end{aligned}$$

因此

$$\text{Cov}(X, Y) = 0, \quad \rho(X, Y) = 0.$$

但因

$$f_X(-1)f_Y(1) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9} \neq \frac{1}{3} = f(-1, 1),$$

故 X 與 Y 並不獨立。

另一方面， X 與 Y 之間有一很強的非線性關係。因對 $\forall(x, y)$ ，只要 $P(X = x, Y = y) \neq 0$ ，則 $y = x^2$ ，即

$$P(Y = X^2) = 1.$$

亦即雖由 $\rho(X, Y) = 0$ ，得知 X 與 Y 缺乏機率上的線性關係。但並未排除 X 與 Y 間，有非線性的關係。換句話說，“無相關”並非“不相關”。

例5.4 設 (X, Y) 之聯合p.d.f. 為

$$f(x, y) = \frac{1}{2}, 0 \leq y \leq x \leq 2.$$

試求 $\text{Cov}(X, Y)$ 及 $\rho(X, Y)$ 。

解. 首先 X, Y 之邊際 p.d.f. 分別為

$$f_X(x) = \int_0^x \frac{1}{2} dy = \frac{x}{2}, 0 \leq x \leq 2,$$

$$f_Y(y) = \int_y^2 \frac{1}{2} dx = 1 - \frac{y}{2}, 0 \leq y \leq 2。$$

由此得

$$\mu_X = \int_0^2 x \cdot \frac{x}{2} dx = \frac{4}{3},$$

$$E(X^2) = \int_0^2 x^2 \cdot \frac{x}{2} dx = 2,$$

故

$$\sigma_X^2 = E(X^2) - (\mu_X)^2 = 2 - \frac{16}{9} = \frac{2}{9}。$$

同理可得

$$\mu_Y = \frac{2}{3}, \quad \sigma_Y^2 = \frac{2}{9}。$$

又

$$\begin{aligned} E(XY) &= \int_0^2 \int_0^x \frac{1}{2} xy dy dx \\ &= \int_0^2 \frac{x}{2} \left(\frac{1}{2} y^2 \right) \Big|_0^x dx \\ &= \int_0^2 \frac{x^3}{4} dx = 1。 \end{aligned}$$

故得

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - \mu_X \mu_Y = 1 - \frac{4}{3} \cdot \frac{2}{3} = \frac{1}{9}, \\ \rho(X, Y) &= \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/9}{\sqrt{2/9} \sqrt{2/9}} = \frac{1}{2}。 \end{aligned}$$

再給一例。

例5.5 設 X 與 Z 為二獨立的隨機變數，且 X 有 $\mathcal{U}(0, 1)$ 分佈， Z 有 $\mathcal{U}(0, 1/10)$ 分佈。令 $Y = X + Z$ ，並考慮隨機向量 (X, Y) 。利用變數代換，可得 X, Y 之聯合p.d.f.為

$$f(x, y) = 10, 0 < x < 1, x < y < x + \frac{1}{10},$$

如此便可求出 $\text{Cov}(X, Y)$ 及 $\rho(X, Y)$ 。不過因 Y 為 X 與 Z 之和，且 X 與 Z 獨立，我們可經由 X 與 Z ，如下地求出 $\text{Cov}(X, Y)$ 與 $\rho(X, Y)$ 。

$$E(X) = \frac{1}{2}, \text{Var}(X) = \frac{1}{12},$$

$$E(Y) = E(X + Z) = E(X) + E(Z) = \frac{1}{2} + \frac{1}{20} = \frac{11}{20},$$

$$\text{Var}(Y) = \text{Var}(X + Z) = \text{Var}(X) + \text{Var}(Z) = \frac{1}{12} + \frac{1}{1,200} = \frac{101}{1,200}。$$

$$\begin{aligned} \text{Cov}(X, Y) &= E(XY) - E(X)E(Y) \\ &= E(X(X + Z)) - E(X)(E(X) + E(Z)) \\ &= E(X^2) + E(XZ) - (E(X))^2 - E(X)E(Z) \\ &= E(X^2) - (E(X))^2 \\ &= \text{Var}(X) = \frac{1}{12}。 \end{aligned}$$

此處用到因 X 與 Z 獨立，所以 $E(XZ) = E(X)E(Z)$ 。由此又得

$$\rho(X, Y) = \frac{\frac{1}{12}}{\sqrt{\frac{1}{12}}\sqrt{\frac{101}{1,200}}} = \frac{10}{\sqrt{101}}。$$

與例5.2比較，在例5.5中， X 與 Y 的相關係數很接近1。我們將兩例中使 $f(x, y)$ 為正之 (x, y) 的圖形繪出，見圖5.1。

圖5.1.a及5.1.b，顯示 X 與 Y 大致有線性的關係，但在圖5.1.b中線性關係較強(區域較窄)。也可以另一方式來看此線性關係的強弱。在例5.5中，因 $Y = X + Z$ ，故給定 $X = x$ ， Y 有 $\mathcal{U}(x, x + 1/10)$ 分佈。而在例5.2中，也可驗證給定 $X = x$ ， Y 有 $\mathcal{U}(x, x + 1)$ 分佈。故知道 $X = x$ ，在例5.5中，較在例5.2中，提供較多關於 Y 之資訊(前者誤差不超過0.1，後者誤差不超過1)。因此在例5.5中， X 與 Y 的相關係數較大。

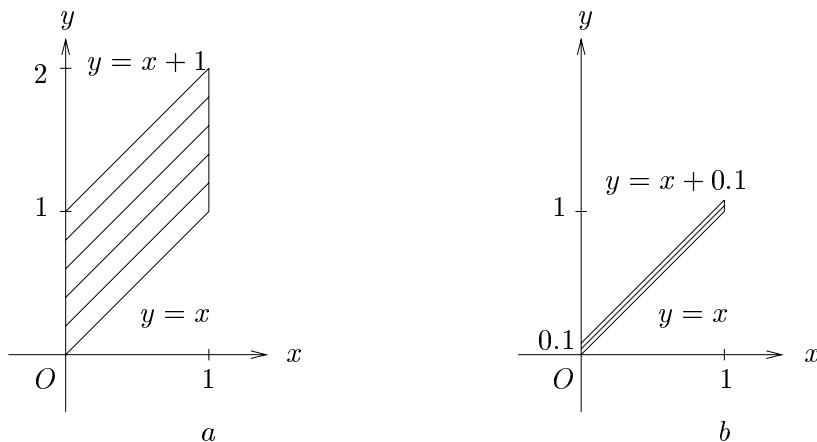


圖 5.1 圖 *a* 為例 5.2 中使 $f(x, y) > 0$ 之區域，圖 *b* 為例 5.5 中使 $f(x, y) > 0$ 之區域。

底下為一連續型隨機向量 (X, Y) ， X 與 Y 之間有強烈的關係，但因該關係並非線性，所以相關係數很小之例，可與例 5.5 比較。

例 5.6 設 X 與 Z 為二獨立的隨機變數，且 X 有 $\mathcal{U}(-1, 1)$ 分佈， Z 有 $\mathcal{U}(0, 1/10)$ 分佈。令 $Y = X^2 + Z$ ，並考慮隨機向量 (X, Y) 。可看出給定 $X = x$ ， Y 有 $\mathcal{U}(x^2, x^2 + 1/10)$ 分佈，因此 X 與 Y 有很強的關係，只是並非線性。又可求出 X, Y 之聯合 p.d.f. 為

$$f(x, y) = 5, \quad -1 < x < 1, x^2 < y < x^2 + \frac{1}{10}.$$

如此便可求出 $\text{Cov}(X, Y)$ 及 $\rho(X, Y)$ 。不過如同例 5.5，我們以下述方式求之。由於 X 有 $\mathcal{U}(-1, 1)$ 分佈，故

$$E(X) = E(X^3) = 0.$$

又因 X 與 Z 獨立，故

$$E(XZ) = E(X)E(Z) = 0.$$

因此

$$\text{Cov}(X, Y) = E(X(X^2 + Z)) - E(X)(E(X^2 + Z))$$

$$= E(X^3) + E(XZ) = 0,$$

且 $\rho(X, Y) = 0$ 。

再提醒一次，相關係數是量測線性關係的強弱，而非量測其他非線性關係之強弱。

例5.7 設 X 有 $\mathcal{U}(-1, 1)$ 分佈。令 $Y = X^2$, $Z = 2X + 3$, $W = -2X + 3$ 。則

$$E(X) = 0,$$

$$E(XY) = E(X^3) = 0,$$

故

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = 0,$$

且

$$\rho(X, Y) = 0。$$

即 X 與 Y 為無相關，雖二者看起來關係如此密切。

另外，

$$E(XZ) = E(X(2X + 3)) = 2E(X^2) + 3E(X) = \frac{2}{3},$$

$$E(X)E(Z) = 0,$$

$$\text{Var}(X) = E(X^2) = \frac{1}{3},$$

$$\text{Var}(Z) = 4\text{Var}(X) = \frac{4}{3},$$

故

$$\text{Cov}(X, Z) = E(XZ) - E(X)E(Z) = \frac{2}{3},$$

且

$$\rho(X, Z) = \frac{\text{Cov}(X, Z)}{\sigma_X \sigma_Z} = \frac{2/3}{\sqrt{1/3} \cdot \sqrt{4/3}} = 1。$$

同理可得

$$\rho(X, W) = -1。$$

X 與 Z 有線性關係，且增長方向相同，其相關係數為1。 X 與 W 亦有線性關係，只是增長方向相反，所以相關係數為-1。一般而言是對的，見底下系理5.4。

共變異數對了解隨機變數和之變異有很大幫助。底下定理為第一章定理6.3之推廣。一般在這類定理中， $\text{Var}(X)$ 及 $\text{Var}(Y)$ 皆須假設存在，但常略去陳述此條件。

定理5.3 設 X, Y 為二隨機變數。則對任二常數 a, b ,

$$(5.6) \quad \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)。$$

特別地，若 X 與 Y 獨立，則

$$(5.7) \quad \text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y)。$$

證明. 首先

$$E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y。$$

故

$$\begin{aligned} \text{Var}(aX + bY) &= E((aX + bY - (a\mu_X + b\mu_Y))^2) \\ &= E((a(X - \mu_X) + b(Y - \mu_Y))^2) \\ &= E(a^2(X - \mu_X)^2 + b^2(Y - \mu_Y)^2 + 2ab(X - \mu_X)(Y - \mu_Y)) \\ &= a^2 E((X - \mu_X)^2) + b^2 E((Y - \mu_Y)^2) + 2ab E((X - \mu_X)(Y - \mu_Y)) \\ &= a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y)。$$

其次，若 X 與 Y 獨立，則由系理5.1， $\text{Cov}(X, Y) = 0$ ，因此(5.7)式成立。證畢。

由定理5.3得

$$(5.8) \quad \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)。$$

因此若 X 與 Y 為**正相關**(positively correlated), 即 $\rho(X, Y) > 0$ (因此 $\text{Cov}(X, Y) > 0$), 則 $X + Y$ 之變異數, 比 X, Y 變異數之和還大。反之, 若 X 與 Y 為**負相關**(negatively correlated), 則 $X + Y$ 之變異數, 比 X, Y 變異數之和還小。當負相關時, X 與 Y 一個值較大, 另一值傾向於較小, 而其和就不至於那麼極端, 因此 $X + Y$ 之變異減小。同理可對 $\text{Var}(X - Y)$ 做討論。

底下我們證明對任二隨機變數 X, Y , $|\rho(X, Y)| \leq 1$, 且 $|\rho(X, Y)| = 1$, 若且唯若 $P(Y = aX + b) = 1$, 其中 a, b 為二常數, 且 $a \neq 0$ 。由於相關係數的值介於正負1之間, 且當其值等於+1或-1時, 二變數有一線性關係之機率為1, 這便說明了如前所述, 相關係數是用來量測二變數之線性相依情形。當 $|\rho(X, Y)| = 1$ 時, 我們稱 X 與 Y 為**完全相關**(completely correlated)。

我們先證明底下著名的**史瓦茲不等式**(Schwarz's inequality), 又稱**柯西-史瓦茲不等式**(Cauchy-Schwarz's inequality), 此不等式出現於數學中很多不同的地方。Schwarz(1843-1921)為近代德國著名的數學家。

定理5.4 設 X, Y 為二隨機變數, 且 $E(X^2), E(Y^2) < \infty$ 。則

$$(5.9) \quad (E(XY))^2 \leq E(X^2)E(Y^2)。$$

又(5.9)式中等式成立, 若且唯若 $P(X = 0) = 1$, 或存在一常數 a , 使得 $P(Y = aX) = 1$ 。

證明. 若 $P(X = 0) = 1$, 則 $P(XY = 0) = 1$, $E(XY) = 0$ 且 $E(X^2) = 0$, 此時(5.9)式左右皆為0, 故等式成立。另外, 若 $P(Y = aX) = 1$, 亦可證明此時(5.9)式之左右皆等於 $(aE(X^2))^2$, 故等式仍成立。底下設 $P(X = 0) < 1$, 因此 $E(X^2) > 0$ 。

現對 $\forall \lambda > 0$,

$$(5.10) \quad E((Y - \lambda X)^2) = \lambda^2 E(X^2) - 2\lambda E(XY) + E(Y^2)$$

為一恆非負之 λ 的二次式。又因首項 λ^2 之係數 $E(X^2) > 0$, 故此二次式在 $\lambda = \lambda_0 = E(XY)/E(X^2)$ 有極小值, 且

$$0 \leq E((Y - \lambda_0 X)^2) = E(Y^2) - (E(XY))^2/E(X^2),$$

故(5.9)式成立。若(5.9)式中等式成立，則 $E((Y - \lambda_0 X)^2) = 0$ ，此時 $P(Y = \lambda_0 X) = 1$ 。證畢。

將史瓦茲不等式應用至隨機變數 $X - \mu_X$ 及 $Y - \mu_Y$ ，得

$$(E((X - \mu_X)(Y - \mu_Y)))^2 \leq E((X - \mu_X)^2)E((Y - \mu_Y)^2),$$

上式即為

$$(5.11) \quad (\text{Cov}(X, Y))^2 \leq \text{Var}(X)\text{Var}(Y)。$$

因此下述結果便得證了。

系理5.3 對任二隨機變數 X, Y ，只要 $E(X^2), E(Y^2) < \infty$ ，便有

$$(5.12) \quad |\rho(X, Y)| \leq 1。$$

我們又有下述重要的推論。

系理5.4 設 X, Y 為二隨機變數，且 $E(X^2), E(Y^2) < \infty$ 。則 $|\rho(X, Y)| = 1$ ，若且唯若存在二常數 a, b ，且 $a \neq 0$ ，使得 $P(Y = aX + b) = 1$ 。

證明。 首先，由相關係數之定義知， $\text{Var}(X)$ 及 $\text{Var}(Y)$ 皆不能為0，因此 $P(X = c) = 1$ 不能成立，其中 c 為一常數。利用定理5.4於隨機變數 $X - \mu_X$ 及 $Y - \mu_Y$ ，即得 $|\rho(X, Y)| = 1$ ，若且唯若存在一常數 a ，使得 $P(Y - \mu_Y = a(X - \mu_X)) = 1$ 。又 $a \neq 0$ ，否則 $P(Y = \mu_Y) = 1$ ，因此 $\text{Var}(Y) = 0$ 。取 $b = \mu_Y - a\mu_X$ ，可看出此即我們要證的結果。

在系理5.4中，若 $a > 0$ ，則 $\rho(X, Y) = 1$ ；若 $a < 0$ ，則 $\rho(X, Y) = -1$ 。又 a, b 也可以 X, Y 之期望值及標準差表示，見下系理，證明則留在習題中。

系理5.5 設 X, Y 為二隨機變數，且 $E(X^2), E(Y^2) < \infty$ 。則當 $\rho(X, Y) = 1$ ，

$$(5.13) \quad P(Y = \mu_Y + (\sigma_Y/\sigma_X)(X - \mu_X)) = 1;$$

當 $\rho(X, Y) = -1$ ，

$$(5.14) \quad P(Y = \mu_Y - (\sigma_Y/\sigma_X)(X - \mu_X)) = 1。$$

當 $E(X^2) < \infty$, 我們注意到 $\rho(X, X) = 1$ 。自己與自己的相關係數為1, 自然是合理的。而這也與系理5.4及5.5之推論吻合。事實上, 對任二常數 a, b , 且 $a \neq 0$, $\rho(X, aX + b) = 1$ (此亦為系理5.4之推論)。

二隨機變數即使相關係數很高, 只是其分佈上所顯示的現象, 並不表二隨機現象間一定有因果關係。例如, 令 X 表某地可樂之月銷售量, Y 表該地每月至醫院腸胃科就診的人數。則 X 與 Y 之相關係數可能很大。但這並不是說喝可樂後易拉肚子, 而很可能是夏天天氣炎熱, 可樂需求量因而增大, 而天氣炎熱時病蟲又易滋長, 因此腸胃有問題的人也變多。如此導致 X 與 Y 之相關係數很高。至於喝可樂與拉肚子間倒不見得有因果關係。

現以 ρ 表二隨機變數 X, Y 之相關係數, 則 ρ 介於 -1 與 1 之間。我們已指出, 由相關係數之定義, 正相關 ($\rho > 0$), 顯示較大的 X 傾向於對應較大的 Y , 且較小的 X 傾向於對應較小的 Y ; 負相關 ($\rho < 0$), 則顯示較大的 X 傾向於對應於較小的 Y , 且較小的 X 傾向於對應於較大的 Y ; ρ 接近 0 時, 顯示前述兩種傾向皆很弱。雖然由系理5.4知, $|\rho|$ 接近 1 時, X 與 Y 應有較強的線性關係。比較例5.2與例5.5, 似乎較強的線性關係, $|\rho|$ 值亦應較大。但這些並未能完全解釋本節一開始所說的, 相關係數是用來度量兩隨機變數線性關係之強度。

底下我們引用 Roussas(1997)pp.130-131 所提供的說明, 應有助於了解為何採用相關係數來量測二隨機變數的共線性。為了簡便, 以 μ_1 表 μ_X , μ_2 表 μ_Y , σ_1 表 σ_X , σ_2 表 σ_Y 。

先考慮 xy 平面上的直線

$$(5.15) \quad y = \mu_2 + (\sigma_2/\sigma_1)(x - \mu_1)。$$

令 D_1 表一隨機的點 (X, Y) 與上述直線之距離。在高中數學裡, 大家學過平面上一點 (x_0, y_0) 與直線 $ax + by + c = 0$ 之距離為 $|ax_0 + by_0 + c|/\sqrt{a^2 + b^2}$ 。因此(注意 $a = \sigma_2/\sigma_1$, $b = -1$, $c = \mu_2 - (\sigma_2/\sigma_1)\mu_1$)

$$(5.16) \quad D_1 = |(\sigma_2/\sigma_1)X - Y + (\mu_2 - (\sigma_2/\sigma_1)\mu_1)|/\sqrt{\sigma_2^2/\sigma_1^2 + 1},$$

且

$$(5.17) \quad D_1^2 = ((\sigma_2/\sigma_1)X - Y + (\mu_2 - (\sigma_2/\sigma_1)\mu_1))^2 / (\sigma_2^2/\sigma_1^2 + 1)。$$

由(5.17)式得

$$(5.18) \quad \begin{aligned} (\sigma_1^2 + \sigma_2^2)D_1^2 &= \sigma_2^2 X^2 + \sigma_1^2 Y^2 - 2\sigma_1\sigma_2 XY \\ &\quad + 2\sigma_2(\sigma_1\mu_2 - \sigma_2\mu_1)X - 2\sigma_1(\sigma_1\mu_2 - \sigma_2\mu_1)Y \\ &\quad + (\sigma_1\mu_2 - \sigma_2\mu_1)^2。 \end{aligned}$$

對上式兩側取期望值, 且利用

$$E(X^2) = \sigma_1^2 + \mu_1^2, \quad E(Y^2) = \sigma_2^2 + \mu_2^2,$$

$$E(XY) = \rho\sigma_1\sigma_2 + \mu_1\mu_2,$$

經化簡後, 得

$$(5.19) \quad E(D_1^2) = \frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}(1 - \rho)。$$

其次考慮 xy 平面上的直線

$$(5.20) \quad y = \mu_2 - (\sigma_2/\sigma_1)(x - \mu_1),$$

同理可得點 (X, Y) 與此直線距離 D_2 平方的期望值為

$$(5.21) \quad E(D_2^2) = \frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}(1 + \rho)。$$

我們為什麼會想到考慮這二條直線呢? 由系理5.5, $\rho = 1$ 時, (X, Y) 落在直線 $y = \mu_2 + (\sigma_2/\sigma_1)(x - \mu_1)$ 上的機率為1; $\rho = -1$ 時, (X, Y) 落在直線 $y = \mu_2 - (\sigma_2/\sigma_1)(x - \mu_1)$ 上的機率為1。所以此二直線與 (X, Y) 是有密切關係的。此二直線前者斜率為正, 後者為負, 二直線相交於點 (μ_1, μ_2) , 對稱於直線 $y = \mu_2$, 也對稱於直線 $x = \mu_1$ 。由於期望值有平均的意味, (5.19)及(5.21)式顯示:

若 $\rho > 0$ (正相關), 比較與二直線的距離平方的期望值 $E(D_1^2)$ 與 $E(D_2^2)$, 可看出點 (X, Y) 較有沿著直線 $y = \mu_2 + (\sigma_2/\sigma_1)(x - \mu_1)$ 的傾向。且當 ρ 愈來愈

愈接近1時, $E(D_1^2)$ 愈來愈接近0, 亦即 (X, Y) 愈來愈接近此直線。 $\rho = 1$ 時, 則 (X, Y) 恰巧落在此直線(機率為1)。

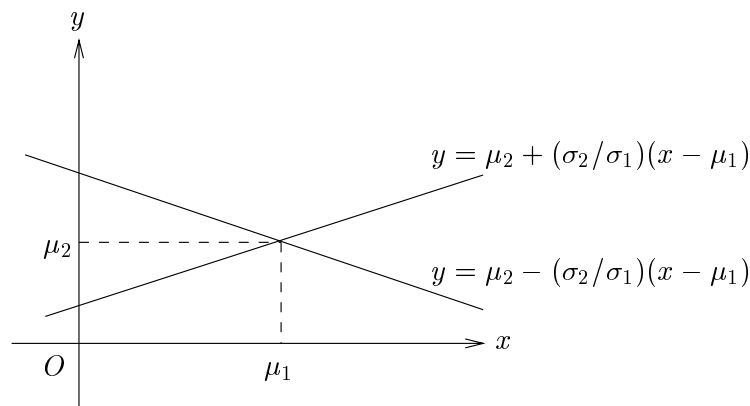


圖5.2 相關係數 $\rho = \pm 1$ 時, (X, Y) 所落在的直線

若 $\rho < 0$ (負相關), 同理點 (X, Y) 較有沿著直線 $y = \mu_2 - (\sigma_2/\sigma_1)(x - \mu_1)$ 的傾向。且當 ρ 愈來愈接近 -1 時, $E(D_2^2)$ 愈來愈接近0, 亦即 (X, Y) 愈來愈接近此直線。 $\rho = -1$ 時, 則 (X, Y) 恰巧落在此直線(機率為1)。

若 $\rho = 0$, 點 (X, Y) 與前述兩直線任一之距離平方的期望值皆為 $2\sigma_1^2 \sigma_2^2 / (\sigma_1^2 + \sigma_2^2)$ 。所以點 (X, Y) 並無較有沿著此二直線之任一的傾向。點 (X, Y) 就只是散佈在 xy 平面上。

相關係數 ρ 用來量測隨機變數 X 與 Y 之共線性, 就是基於以上的涵意。

下述定理顯示相關係數與條件期望值關係密切。此定理亦可視為系理5.5之推廣。

定理5.5 設隨機變數 X, Y 之相關係數 ρ 存在。若 $E(Y|X = x)$ 為 x 之線性函數, 則

$$(5.22) \quad E(Y|X = x) = \mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1),$$

$$(5.23) \quad E(\text{Var}(Y|X)) = \sigma_2^2(1 - \rho^2)。$$

證明. 由假設知

$$E(Y|X = x) = ax + b,$$

其中 a, b 為常數。又由總機率法則,

$$\begin{aligned}\mu_2 &= E(Y) = E(E(Y|X)) \\ &= E(aX + b) \\ &= a\mu_1 + b.\end{aligned}$$

故得 $b = \mu_2 - a\mu_1$, 且

$$(5.24) \quad E(Y|X = x) = ax + b = ax + \mu_2 - a\mu_1 = \mu_2 + a(x - \mu_1)。$$

若能證出 $a = \rho(\sigma_2/\sigma_1)$ 便得證了。

$$\begin{aligned}\rho\sigma_1\sigma_2 &= \text{Cov}(X, Y) \\ &= E((X - \mu_1)(Y - \mu_2)) \\ &= E(E((X - \mu_1)(Y - \mu_2)|X)) \\ &= E((X - \mu_1)(E(Y|X) - \mu_2)) \\ &= E((X - \mu_1) \cdot a(X - \mu_1)) \\ &= a\text{Var}(X) = a\sigma_1^2,\end{aligned}$$

其中倒數第二等式的成立用到(5.24)式。即得 $a = \rho(\sigma_2/\sigma_1)$ 。

其次由定理4.2及(5.22)式, 得

$$\begin{aligned}E(\text{Var}(Y|X)) &= \text{Var}(Y) - \text{Var}(E(Y|X)) \\ &= \sigma_2^2 - \text{Var}(\mu_2 + \rho(\sigma_2/\sigma_1)(X - \mu_1)) \\ &= \sigma_2^2 - \rho^2(\sigma_2^2/\sigma_1^2)\text{Var}(X) \\ &= \sigma_2^2 - \rho^2(\sigma_2^2/\sigma_1^2)\sigma_1^2 = \sigma_2^2(1 - \rho^2)。$$

證畢。

本節最後我們給一定理5.3之推廣, 其證明留在習題中(第18題)。

定理5.6 對二隨機變數 X, Y , 及常數 $a_1, b_1, c_1, a_2, b_2, c_2$,

$$(5.25) \quad \text{Cov}(a_1X + b_1Y + c_1, a_2X + b_2Y + c_2)$$

$$= a_1 a_2 \text{Var}(X) + b_1 b_2 \text{Var}(Y) + (a_1 b_2 + a_2 b_1) \text{Cov}(X, Y)。$$

只要知道 $\text{Var}(X)$, $\text{Var}(Y)$ 及 $\text{Cov}(X, Y)$, 則任二 X, Y 之線性函數的共變異數皆可求出。又當 $a_1 = a_2 = a$, $b_1 = b_2 = b$, $c_1 = c_2 = 0$, 則(5.25)式成爲(5.6)式。

例5.8 對二隨機變數 X, Y , 試決定 $X + Y$ 與 $X - Y$ 無相關之條件。

解.因

$$\begin{aligned} \text{Cov}(X + Y, X - Y) \\ &= \text{Var}(X) - \text{Var}(Y) + \text{Cov}(X, Y) - \text{Cov}(X, Y) \\ &= \text{Var}(X) - \text{Var}(Y), \end{aligned}$$

故 $X + Y$ 與 $X - Y$ 無相關, 若且唯若 $\text{Var}(X) = \text{Var}(Y)$ 。

3.6 二變數常態分佈

在機率及統計裡, 兩個變數以上之常態分佈很重要, 本節我們介紹二變數常態分佈。此分佈具有大部分多維常態分佈(multivariate normal distribution)的特性, 不但是一很重要的多維常態分佈, 且二變數之相關係數, 亦爲其聯合p.d.f之一參數。先給一定義。

定義6.1 設隨機向量 (X, Y) 爲有參數 $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$ 之二變數常態分佈, 以 $\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ (或 $BV\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$)表之, 其中 $\mu_1, \mu_2 \in R$, $\sigma_1, \sigma_2 > 0$, $-1 < \rho < 1$, 若其聯合p.d.f.爲

$$(6.1) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-q/2}, x, y \in R,$$

其中

$$(6.2) \quad q = \frac{1}{1-\rho^2} \left(\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho \frac{x-\mu_1}{\sigma_1} \frac{y-\mu_2}{\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right)。$$

雖然上述p.d.f.看起來是如此不可親, 不過此為一極常出現之二變數的分佈。此分佈並有不少好的性質, 我們列出一些如下。

(i) X, Y 之邊際分佈, 分別為 $\mathcal{N}(\mu_1, \sigma_1^2), \mathcal{N}(\mu_2, \sigma_2^2)$ 。

(ii) $\rho(X, Y) = \rho$ 。

(iii) 對任二常數 $a, b \in R$, 且 $a^2 + b^2 \neq 0$, $aX + bY$ 有 $\mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$ 分佈。

(iv) X 與 Y 獨立, 若且唯若 $\rho = 0$ 。

性質(iii)為二變數常態分佈之一特性。即只有二變數常態分佈才有性質(iii)。因此性質(iii)可刻劃(characterize)二變數常態分佈。

性質(iv)的證明可藉助定理2.1。由於 (X, Y) 之聯合p.d.f.如(6.1)式, 可看出要能寫成 $g(x)h(y)$ 的型式, 若且唯若 $\rho = 0$ 。性質(iv)是比較特別的, 因 ρ 為 X, Y 之相關係數, 而我們已數次強調, 二隨機變數之相關係數為0, 不表此二變數獨立。但對二變數之常態分佈, 其獨立與否, 就依 ρ 是否為0而定。可見 ρ 在此分佈中之重要角色。

底下我們來證明(6.1)式所定義之 $f(x, y)$ 確為一p.d.f.。首先 $f(x, y)$ 可改寫為

$$\begin{aligned} (6.3) \quad f(x, y) &= \frac{1}{\sigma_2 \sqrt{1 - \rho^2} \sqrt{2\pi}} e^{-\frac{1}{2\sigma_2^2(1-\rho^2)}(y - (\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1)))^2} \\ &\quad \cdot \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x - \mu_1)^2} \\ &= A(x, y)B(x), \end{aligned}$$

這只要經過一些代數運算即可, 其證明留在習題中(第2題)。可看出在給定 x 之下, $A(x, y)$ 為 $\mathcal{N}(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2))$ 分佈之p.d.f., 而 $B(x)$ 為 $\mathcal{N}(\mu_1, \sigma_1^2)$ 分佈之p.d.f.。即

$$\int_{-\infty}^{\infty} A(x, y)dy = \int_{-\infty}^{\infty} B(x)dx = 1。$$

故

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dydx$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} A(x, y) B(x) dy dx \\
&= \int_{-\infty}^{\infty} B(x) \left(\int_{-\infty}^{\infty} A(x, y) dy \right) dx \\
&= \int_{-\infty}^{\infty} B(x) dx = 1.
\end{aligned}$$

得證 $f(x, y)$ 爲一 p.d.f.。

在上述推導中, 因 X 之邊際 p.d.f. 爲

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy = B(x) \int_{-\infty}^{\infty} A(x, y) dy = B(x),$$

故 X 之邊際分佈爲 $\mathcal{N}(\mu_1, \sigma_1^2)$ 。同理可得 Y 之邊際分佈爲 $\mathcal{N}(\mu_2, \sigma_2^2)$ 。得證性質(i)。

由於 $Y|X = x$ 之 p.d.f. 爲

$$f(y|x) = \frac{f(x, y)}{f_X(x)} = \frac{A(x, y)B(x)}{B(x)} = A(x, y).$$

即得二變數常態分佈之另一性質:

(v) $Y|X = x$ 及 $X|Y = y$ 等二條件分佈, 皆仍爲常態分佈, 且 $Y|X = x \sim \mathcal{N}(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2))$, $X|Y = y \sim \mathcal{N}(\mu_1 + \rho(\sigma_1/\sigma_2)(y - \mu_2), \sigma_1^2(1 - \rho^2))$ 。

由性質(v)再度看出, 若 X 與 Y 獨立, 則 ρ 須爲 0。又

$$(6.4) \quad E(Y|X = x) = \mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1),$$

$$(6.5) \quad \text{Var}(Y|X = x) = \sigma_2^2(1 - \rho^2).$$

此二式可與定理 5.5 相對應。當然亦可寫出 $E(X|Y = y)$ 及 $\text{Var}(X|Y = y)$ 。注意此處條件變異數 $\text{Var}(Y|X = x)$ 與 x 無關。由 (6.5) 式即得

$$E(\text{Var}(Y|X)) = \text{Var}(Y|X = x) = \sigma_2^2(1 - \rho^2).$$

故 (6.5) 式較 (5.23) 式的結果還強。又在統計學裡, $E(Y|X = x)$ 有時稱爲**迴歸函數**(regression function)。而 $E(Y|X = x)$ 之圖形則稱爲給定 $X = x$ 下

之迴歸曲線(regression curve)。(6.4)式指出, 對二變數常態分佈, 有線性迴歸函數。

我們常以(6.4)式右側, 做為給定 $X = x$, Y 之預測值。當 $\sigma_1 = \sigma_2$, 且 $0 < \rho < 1$, 則 Y 之預測值 $y = \mu_2 + \rho(x - \mu_1)$ 滿足

$$0 < y - \mu_2 < x - \mu_1, \quad x > \mu_1,$$

$$0 < \mu_2 - y < \mu_1 - x, \quad x < \mu_1.$$

即 Y 之預測值 y 有向 Y 之期望值 μ_2 靠近之傾向。此現象稱為迴歸效應(regression effect)。第一次考試考較好的學生, 第二次考試成績較可能下降; 第一次考試考較差的學生, 第二次考試成績則較可能上升。不過原先成績較高的, 其預測值仍然較高。又只有當 $\rho = 1$ 時, $y - \mu_2 = x - \mu_1$, 即 x, y 與期望值之差距不會改變。

這並不是太奇怪的事。考試考得好, 除了實力外, 總是運氣也不錯, 下一次考試不見得又有這麼好的運氣, 因此成績下降之可能性較大。考試考差的情況則反過來。

上述討論是在正相關之下。至於若負相關($\rho < 0$), 則第一次考高分, 第二次考低分就反而是較可能發生了。又上述討論雖是在 $\sigma_1 = \sigma_2$ 之下, 當 $0 < \rho < 1$, 即使 $\sigma_1 \neq \sigma_2$, 仍有

$$0 < \frac{y - \mu_2}{\sigma_2} < \frac{x - \mu_1}{\sigma_1}, \quad x > \mu_1,$$

$$0 < \frac{\mu_2 - y}{\sigma_2} < \frac{\mu_1 - x}{\sigma_1}, \quad x < \mu_1.$$

到底 x 或 y , 何者較接近期望值, 考慮相對差距(即各除以其標準差)是較合理的。

由於 $|\rho| < 1$, 除非 $\rho = 0$ (此時 X 與 Y 獨立), 否則

$$(6.6) \quad \text{Var}(Y|X = x) \leq \text{Var}(Y),$$

顯示知道 X 之值後, 對 Y 提供了一些資訊。由於對 Y 了解較多, 因此使變異數 $\text{Var}(Y|X = x)$ 減小。又當 $|\rho|$ 接近1時, 條件變異數 $\sigma_2^2(1 - \rho^2)$ 接近0。此

時條件分佈 $Y|X = x$ 集中在直線 $y = \mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1)$ 附近。此印證之前我們所說的，當相關係數較接近1或-1時， X 與 Y 有較強的線性關係。

利用性質(v)及總機率法則，可證明前述性質(ii)：

$$\begin{aligned}
 E(XY) &= E(E(XY|X)) \\
 &= E(XE(Y|X)) \\
 &= E(X(\mu_2 + \rho(\sigma_2/\sigma_1)(X - \mu_1))) \\
 &= \mu_2 E(X) + \rho(\sigma_2/\sigma_1)E(X^2) - \rho\mu_1(\sigma_2/\sigma_1)E(X) \\
 &= \mu_1\mu_2 + \rho(\sigma_2/\sigma_1)(\sigma_1^2 + \mu_1^2) - \rho\mu_1^2(\sigma_2/\sigma_1) \\
 &= \mu_1\mu_2 + \rho\sigma_1\sigma_2,
 \end{aligned}$$

故

$$\begin{aligned}
 \text{Cov}(X, Y) &= E(XY) - \mu_1\mu_2 \\
 &= (\mu_1\mu_2 + \rho\sigma_1\sigma_2) - \mu_1\mu_2 \\
 &= \rho\sigma_1\sigma_2.
 \end{aligned}$$

由此即得 $\rho(X, Y) = \rho$ 。

(vi) (X, Y) 之聯合特徵函數為

$$\begin{aligned}
 (6.7) \quad \phi(s, t) &= E(e^{isX+itY}) \\
 &= \exp\{i\mu_1s + i\mu_2t - (\sigma_1^2s^2 + \sigma_2^2t^2 + 2\sigma_1\sigma_2\rho st)/2\}, s, t \in R,
 \end{aligned}$$

動差母函數則為

$$\begin{aligned}
 (6.8) \quad M(s, t) &= E(e^{sX+tY}) \\
 &= \exp\{\mu_1s + \mu_2t + (\sigma_1^2s^2 + \sigma_2^2t^2 + 2\sigma_1\sigma_2\rho st)/2\}, s, t \in R.
 \end{aligned}$$

(6.7)式與(6.8)式等價。底下我們利用性質(v)來證明(6.8)式：

首先由總機率法則，

$$\begin{aligned}
 M(s, t) &= E(e^{sX+tY}) \\
 &= E(E(e^{sX+tY}|X)) \\
 &= E(e^{sX}E(e^{tY}|X)).
 \end{aligned}$$

而因 $Y|X = x \sim \mathcal{N}(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2))$, 利用第二章(3.25)式, 得

$$E(e^{tY}|X = x) = \exp\{(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1))t + \sigma_2^2(1 - \rho^2)t^2/2\}。$$

故

$$\begin{aligned} M(s, t) &= E(\exp\{sX + (\mu_2 + \rho(\sigma_2/\sigma_1)(X - \mu_1))t + \sigma_2^2(1 - \rho^2)t^2/2\}) \\ &= \exp\{\mu_2 t - \rho(\sigma_2/\sigma_1)\mu_1 t + \sigma_2^2(1 - \rho^2)t^2/2\} E(\exp\{(s + \rho(\sigma_2/\sigma_1)t)X\}) \\ &= \exp\{\mu_2 t - \rho(\sigma_2/\sigma_1)\mu_1 t + \sigma_2^2(1 - \rho^2)t^2/2\} \\ &\quad \cdot \exp\{\mu_1(s + \rho(\sigma_2/\sigma_1)t) + \sigma_1^2(s + \rho(\sigma_2/\sigma_1)t)^2/2\} \\ &= \exp\{\mu_1 s + \mu_2 t + (\sigma_1^2 s^2 + \sigma_2^2 t^2 + 2\sigma_1\sigma_2\rho st)/2\}。 \end{aligned}$$

證畢。

由性質(vi)可得性質(ii): $aX + bY$ 之動差母函數為

$$\begin{aligned} &E(e^{t(aX+bY)}) \\ &= E(e^{atX+btY}) \\ &= \exp\{\mu_1 at + \mu_2 bt + (\sigma_1^2 a^2 t^2 + \sigma_2^2 b^2 t^2 + 2\sigma_1\sigma_2\rho abt^2)/2\} \\ &= \exp\{(a\mu_1 + b\mu_2)t + (a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)t^2/2\}。 \end{aligned}$$

與第二章(3.25)式比較, 即得 $aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2)$ 。

利用(6.8)式, 及二變數之聯合動差母函數可唯一決定其聯合分佈, 即得下述性質, 其證明留在習題中(第8題)。此亦為性質(iii)之推廣。

(vii) 令 $Z_1 = a_1X + b_1Y$, $Z_2 = a_2X + b_2Y$, 且 $a_1b_2 - a_2b_1 \neq 0$, 則 (Z_1, Z_2) 有 $\mathcal{N}(\eta_1, \eta_2, \tau_1^2, \tau_2^2, \gamma)$ 分佈, 其中

$$\begin{aligned} (6.9) \quad \eta_1 &= a_1\mu_1 + b_1\mu_2, \quad \eta_2 = a_2\mu_1 + b_2\mu_2, \\ \tau_1^2 &= a_1^2\sigma_1^2 + b_1^2\sigma_2^2 + 2a_1b_1\rho\sigma_1\sigma_2, \quad \tau_2^2 = a_2^2\sigma_1^2 + b_2^2\sigma_2^2 + 2a_2b_2\rho\sigma_1\sigma_2, \\ \gamma &= (a_1a_2\sigma_1^2 + b_1b_2\sigma_2^2 + (a_1b_2 + a_2b_1)\rho\sigma_1\sigma_2)/(\tau_1\tau_2)。 \end{aligned}$$

由此也可看出 Z_1 與 Z_2 獨立, 若且唯若

$$(6.10) \quad a_1 a_2 \sigma_1^2 + b_1 b_2 \sigma_2^2 + (a_1 b_2 + a_2 b_1) \rho \sigma_1 \sigma_2 = 0。$$

又若 X 與 Y 獨立, 即 $\rho = 0$, 則上述 τ_1, τ_2, γ 可簡化不少。

利用定理5.5, 此處檢驗 γ 確如(6.9)式中所給。首先

$$\begin{aligned} \text{Cov}(Z_1, Z_2) &= \text{Cov}(a_1 X + b_1 Y, a_2 X + b_2 Y) \\ &= a_1 a_2 \text{Var}(X) + b_1 b_2 \text{Var}(Y) + (a_1 b_2 + a_2 b_1) \text{Cov}(X, Y) \\ &= a_1 a_2 \sigma_1^2 + b_1 b_2 \sigma_2^2 + (a_1 b_2 + a_2 b_1) \rho \sigma_1 \sigma_2。 \end{aligned}$$

而

$$\rho(Z_1, Z_2) = \frac{\text{Cov}(Z_1, Z_2)}{\sqrt{\text{Var}(Z_1)} \sqrt{\text{Var}(Z_2)}} = \frac{\text{Cov}(Z_1, Z_2)}{\tau_1 \tau_2}。$$

故 $\gamma = \rho(Z_1, Z_2)$ 確如(6.9)式中所給。

雖然二變數常態分佈, 其邊際分佈必皆亦為常態分佈, 但其逆不真。即若 X, Y 之邊際分佈皆有常態分佈, 並不一定導致 (X, Y) 有二變數常態分佈。當然此時這兩個變數必不獨立。原因是兩獨立的常態分佈, 其聯合 p.d.f. 為邊際 p.d.f. 之乘積, 故有(6.1)式的型式(且 $\rho = 0$), 見習題第6題。

例如, 設隨機變數 X, Y 之聯合 p.d.f. 為

$$(6.11) \quad f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} \left(1 - \frac{xy}{(1+x^2)(1+y^2)}\right), x, y \in R。$$

則 (X, Y) 並無二變數常態分佈, 但可證明 X, Y 之邊際分佈皆為常態(習題第14題)。由(6.11)式, 讀者不難寫出許多邊際分佈為常態, 但並非二變數常態分佈之聯合 p.d.f.。

Anderson(1984)曾以幾何的方法給下述著名的例子。設 (X, Y) 有二變數常態分佈, 且期望值皆為0。在圖6.1中, 有四個正方形, A與D, C與B分別對稱於 y 軸; A與C, D與B分別對稱於 x 軸。將A中之機率移至C中, B中之機率移至D中, 則得到的兩個變數 U, V , 其邊際分佈仍分別與 X, Y 相同, 也就是皆有常態分佈。但 (U, V) 之聯合分佈不為二變數常態分佈。

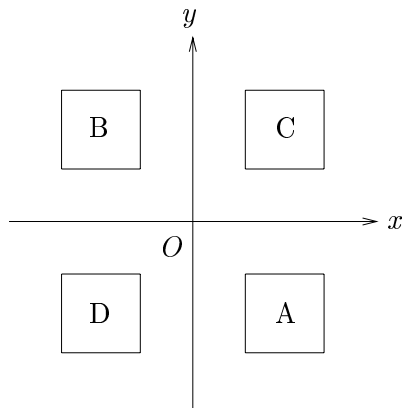


圖6.1 非二變數之常態分佈但邊際分佈為常態

另外，一般而言，若 X, Y 為二無相關之隨機變數，且皆有常態分佈，則 X 與 Y 並不一定獨立。由性質(iv)知，此時 (X, Y) 必非二變數之常態分佈。下例為Broffitt(1986)所給。

例6.1 設隨機變數 X 有 $\mathcal{N}(0, 1)$ 分佈， Z 與 X 獨立，且 $P(Z = 1) = P(Z = -1) = 1/2$ ，又令 $Y = ZX$ 。由

$$\begin{aligned} P(Y \leq y | Z = 1) &= P(ZX \leq y | Z = 1) = P(X \leq y | Z = 1) \\ &= P(X \leq y), \\ P(Y \leq y | Z = -1) &= P(ZX \leq y | Z = -1) = P(-X \leq y | Z = -1) \\ &= P(-X \leq y) = P(X \leq y), \end{aligned}$$

注意 $X \stackrel{d}{=} (-X)$ ，得

$$P(Y \leq y | Z = 1) = P(Y \leq y | Z = -1), y \in R。$$

故 Y 與 Z 獨立(此亦為一兩個變數隨機的獨立，但不是函數的獨立(functionally independent)之例)。因此

$$P(Y \leq y) = P(Y \leq y | Z = 1) = P(X \leq y),$$

即 Y 亦有 $\mathcal{N}(0, 1)$ 分佈。現因 $E(X) = E(Y) = E(Z) = 0$ ，且

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(ZX^2) = E(Z)E(X^2) = 0,$$

故 X 與 Y 無相關, 但明顯地 X 與 Y 並不獨立。因此 (X, Y) 無二變數常態分佈, 雖然其邊際分佈皆為常態。

在一維的情況, 若 $X \sim \mathcal{N}(\mu, \sigma^2)$, 令 $Z = (X - \mu)/\sigma$, 則 $Z \sim \mathcal{N}(0, 1)$ 。因此只需一個標準常態分佈之機率值表, 便能求出任一常態分佈之機率值。在二維的情況, 設 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。令 $Z_1 = (X - \mu_1)/\sigma_1$, $Z_2 = (Y - \mu_2)/\sigma_2$, 則 $(Z_1, Z_2) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ (習題第12題)。 Z_1, Z_2 皆有標準常態分佈, 而 Z_1, Z_2 之相關係數不變仍為 ρ 。因此

$$\begin{aligned} F_{X,Y}(a, b) &= P(X \leq a, Y \leq b) \\ &= P\left(\frac{X - \mu_1}{\sigma_1} \leq \frac{a - \mu_1}{\sigma_1}, \frac{Y - \mu_2}{\sigma_2} \leq \frac{b - \mu_2}{\sigma_2}\right) \\ &= F_{Z_1, Z_2}\left(\frac{a - \mu_1}{\sigma_1}, \frac{b - \mu_2}{\sigma_2}\right). \end{aligned}$$

任一二變數常態之隨機變數 X, Y , 經由

$$Z_1 = \frac{X - \mu_1}{\sigma_1}, \quad Z_2 = \frac{Y - \mu_2}{\sigma_2}$$

之變數代換, 可轉為標準二變數常態隨機變數 Z_1, Z_2 。雖然只有一個一維的標準常態分佈, 卻有無限多個標準二變數常態分佈, 每一相關係數 ρ 便有一個。所以二變數標準常態分佈之機率值表是較麻煩的, 因 $F_{Z_1, Z_2}(h, k)$ 有三個變數: h, k, ρ 。本書並不給出這種表。但對有些特殊情況, 可以求出有關二變數之機率。見下二例。

例6.2 某地區男性的身高可以 $\mathcal{N}(170, 7.2^2)$ 當做其分佈, 女性身高可以 $\mathcal{N}(165, 5.4^2)$ 當做其分佈。男性與女性的身高假設為相互獨立。任選一男性及一女性。試求女性身高高於男性之機率。

解. 設 $X \sim \mathcal{N}(170, 7.2^2)$, $Y \sim \mathcal{N}(165, 5.4^2)$, X 與 Y 獨立。本例即要求 $P(Y > X)$ 。由定理2.4-(v), 得 $Y - X \sim \mathcal{N}(-5, 9^2)$ 。故

$$P(Y > X) = P(Y - X > 0)$$

$$\begin{aligned}
&= P\left(\frac{Y - X - (-5)}{9} > \frac{0 - (-5)}{9}\right) \\
&= P\left(Z > \frac{5}{9}\right) \doteq 1 - P(Z \leq 0.556) \\
&\doteq 1 - 0.7109 = 0.2891,
\end{aligned}$$

其中 $Z \sim \mathcal{N}(0, 1)$ 。

例6.3 承上例。設 $\rho(X, Y) = 0.6$ 。例如 X 可表一家庭中男孩之身高, Y 表女孩之身高, 通常 X 與 Y 不獨立。試求 $P(Y > X)$ 。

解. 由性質 (iii), $Y - X \sim \mathcal{N}(-5, 7.2^2 + 5.4^2 - 2 \cdot 0.6 \cdot 7.2 \cdot 5.4) = \mathcal{N}(-5, 34.344)$ 。因此

$$\begin{aligned}
P(Y > X) &= P(Y - X > 0) \\
&= P\left(\frac{Y - X - (-5)}{\sqrt{34.344}} > \frac{0 - (-5)}{\sqrt{34.344}}\right) \\
&\doteq P(Z > 0.853) \doteq 1 - P(Z \leq 0.853) \\
&\doteq 1 - 0.8031 = 0.1969。
\end{aligned}$$

一般對二變數常態分佈, 若涉及的本質上只是一維的常態分佈, 如 $aX + bY$, 便可求出其機率值。例6.3中所求即為關於一維變數 $-X + Y$ 之機率。

3.7 多維分佈

前幾節我們主要討論兩個變數之多維分佈。一般當然有 n 維的隨機向量 $(X_1, \dots, X_n)$, $n \geq 2$ 。我們以粗體的 $\mathbf{X}$ 表隨機向量 $(X_1, \dots, X_n)$, 而以粗體的 $\mathbf{x}$ 表觀測值 $(x_1, \dots, x_n)$ 。

隨機向量 $\mathbf{X} = (X_1, \dots, X_n)$ 以 R^n 之一子集為其樣本空間。若 $\mathbf{X}$ 為一離散型的隨機向量(樣本空間假設為可數的), 則其聯合 p.d.f. 為

$$\begin{aligned}
f(\mathbf{x}) &= f(x_1, \dots, x_n) \\
&= P(X_1 = x_1, \dots, X_n = x_n), (x_1, \dots, x_n) \in R^n。
\end{aligned}$$

因此對任意 $A \subset R^n$,

$$(7.1) \quad P(\mathbf{X} \in A) = \sum_{\mathbf{x} \in A} f(\mathbf{x}).$$

若 $\mathbf{X}$ 爲一連續型的隨機向量, 則其聯合 p.d.f. 爲一滿足下述條件的函數 $f(x_1, \dots, x_n)$:

$$(7.2) \quad P(\mathbf{X} \in A) = \int \cdots \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

令 $g(\mathbf{x}) = g(x_1, \dots, x_n)$ 爲一實值函數。則 $g(\mathbf{X})$ 爲一隨機變數, 且對連續型及離散型, 其期望值分別爲

$$(7.3) \quad E(g(\mathbf{X})) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\mathbf{x}) f(\mathbf{x}) d\mathbf{x},$$

$$(7.4) \quad E(g(\mathbf{X})) = \sum_{\mathbf{x} \in R^n} g(\mathbf{x}) f(\mathbf{x}).$$

除了積分改爲在 R^n 上, 上述這些定義與二變數時的情況類似。另外, 有時爲了簡便, 僅寫成“ $X_1, \dots, X_n$ 之聯合 p.d.f.”, 省去 $\mathbf{X}$, 對 $X_1, \dots, X_n$ 也可不用加括號。

$(X_1, \dots, X_n)$ 之任一子集的邊際 p.d.f., 可將聯合 p.d.f. 對其餘變數之所有可能的值積分或相加而得。例如, 首 k 個變數 $(X_1, \dots, X_k)$, $1 \leq k < n$, 之邊際 p.d.f. 爲

$$(7.5) \quad f(x_1, \dots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_{k+1} \cdots dx_n,$$

或

$$(7.6) \quad f(x_1, \dots, x_k) = \sum_{(x_{k+1}, \dots, x_n) \in R^{n-k}} f(x_1, \dots, x_n).$$

條件分佈也可類似地定義。如

$$(7.7) \quad f(x_{k+1}, \dots, x_n | x_1, \dots, x_k) = \frac{f(x_1, \dots, x_n)}{f(x_1, \dots, x_k)}.$$

如果爲了更清楚, 可將上式左側寫成

$$f_{X_{k+1}, \dots, X_n | X_1, \dots, X_k}(x_{k+1}, \dots, x_n | x_1, \dots, x_k),$$

特別是當 x_i 's 為給定數值的時候。

例7.1 設取整數值隨機變數 X, Y, Z 之聯合 p.d.f. 為

$$f(x, y, z) = \frac{1}{216}, 1 \leq x \leq 6, x+1 \leq y \leq x+6, y+1 \leq z \leq y+6。$$

試驗證 $f(x, y, z)$ 為一 p.d.f., 並求 $f_{Z|X,Y}(z|x, y)$ 及 $f_{Z|Y}(z|y)$ 。

解. 因

$$\sum_{x=1}^6 \sum_{y=x+1}^{x+6} \sum_{z=y+1}^{y+6} f(x, y, z) = \sum_{x=1}^6 \sum_{y=x+1}^{x+6} \frac{6}{216} = \sum_{x=1}^6 \frac{6}{36} = 1,$$

故 $f(x, y, z)$ 確為一 p.d.f.。

其次

$$f_{X,Y}(x, y) = \sum_{z=y+1}^{y+6} f(x, y, z) = \frac{6}{216} = \frac{1}{36}, 1 \leq x \leq 6, x+1 \leq y \leq x+6。$$

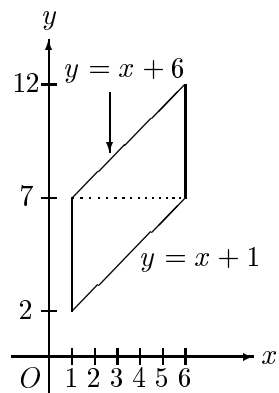
因此

$$f_{Z|X,Y}(z|x, y) = \frac{f(x, y, z)}{f_{X,Y}(x, y)} = \frac{1/216}{1/36} = \frac{1}{6},$$

$$y+1 \leq z \leq y+6, 1 \leq x \leq 6, x+1 \leq y \leq x+6。$$

又

$$f_{Y,Z}(y, z) = \begin{cases} \sum_{x=1}^{y-1} f(x, y, z) = \frac{y-1}{216}, & 2 \leq y \leq 7, y+1 \leq z \leq y+6, \\ \sum_{x=y-6}^6 f(x, y, z) = \frac{13-y}{216}, & 8 \leq y \leq 12, y+1 \leq z \leq y+6。 \end{cases}$$



因此

$$f_Y(y) = \begin{cases} \sum_{z=y+1}^{y+6} f_{Y,Z}(y, z) = \frac{6(y-1)}{216} = \frac{y-1}{36}, & 2 \leq y \leq 7, \\ \sum_{z=y+1}^{y+6} f_{Y,Z}(y, z) = \frac{6(13-y)}{216} = \frac{13-y}{36}, & 8 \leq y \leq 12. \end{cases}$$

故得

$$f_{Z|Y}(z|y) = \frac{f_{Y,Z}(y, z)}{f_Y(y)} = \frac{1}{6}, y+1 \leq z \leq y+6, 2 \leq y \leq 12.$$

比較 $f_{Z|X,Y}$ 與 $f_{Z|Y}$, 可看出在給定 $Y = y$ 之下, 多知道 X 之值, 並不影響 Z 之p.d.f.。我們稱此為給定 $Y = y$ 之下, Z 與 X 為**條件獨立**(conditionally independent)。事實上可證明

$$(7.8) \quad f_{Z|X,Y}(z|x, y) = f_{Z|Y}(z|y)$$

與

$$(7.9) \quad f_{X,Z|Y}(x, z|y) = f_{X|Y}(x|y)f_{Z|Y}(z|y)$$

等價。即可以(7.8)或(7.9)式當做給定 Y 之下, Z 與 X 為條件獨立之定義。

另外,

$$f_X(x) = \sum_{y=x+1}^{x+6} f_{X,Y}(x, y) = \frac{6}{36} = \frac{1}{6}.$$

因此 $X \sim \mathcal{D}\mathcal{U}(1, 6)$ 。讀者可以驗證 $Y \stackrel{d}{=} X_1 + X_2$, 其中 X_1, X_2 獨立且皆有 $\mathcal{D}\mathcal{U}(1, 6)$ 分佈。即 X 相當於投擲一公正的骰子所得點數, Y 相當於投擲二公正骰子之點數和。 Z 的邊際分佈亦可求出, 不過稍繁瑣些, 在此略去。

例7.2 設 X, Y, Z 之聯合p.d.f.為

$$f(x, y, z) = z^2 e^{-z(1+x+y)}, x, y, z \geq 0.$$

我們先證明 f 的確為一p.d.f.:

$$\begin{aligned}
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z) dx dy dz \\
&= \int_0^{\infty} \int_0^{\infty} \int_0^{\infty} z^2 e^{-z(1+x+y)} dx dy dz \\
&= \int_0^{\infty} \int_0^{\infty} -ze^{-z(1+x+y)} \Big|_0^{\infty} dy dz \\
&= \int_0^{\infty} \int_0^{\infty} ze^{-z(1+y)} dy dz \\
&= \int_0^{\infty} -e^{-z(1+y)} \Big|_0^{\infty} dz \\
&= \int_0^{\infty} e^{-z} dz = 1。
\end{aligned}$$

在上述過程中，也附帶求出

$$f_Z(z) = e^{-z}, z \geq 0,$$

即 Z 有 $\mathcal{E}(1)$ 分佈。因此在給定 $Z = z$ 之下， X, Y 之條件 p.d.f. 為

$$\begin{aligned}
f_{X,Y|Z}(x, y|z) &= \frac{f(x, y, z)}{f_Z(z)} = \frac{z^2 e^{-z(1+x+y)}}{e^{-z}} \\
&= z^2 e^{-z(x+y)}, x, y, z \geq 0。
\end{aligned}$$

由上式得

$$\begin{aligned}
f_{X|Z}(x|z) &= \int_0^{\infty} f_{X,Y|Z}(x, y|z) dy \\
&= \int_0^{\infty} z^2 e^{-z(x+y)} dy = ze^{-zx}, x \geq 0, z \geq 0。
\end{aligned}$$

同理有 $f_{Y|Z}(y|z) = ze^{-zy}, y \geq 0, z \geq 0$ 。因此

$$f_{X,Y|Z}(x, y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z)。$$

即在給定 $Z = z$ 之下， X 與 Y 為條件獨立，且皆有 $\mathcal{E}(z)$ 分佈。但若不給定 $Z = z$ ，則 X 與 Y 並不獨立： X, Y 之聯合 p.d.f. 為

$$\begin{aligned}
f_{X,Y}(x, y) &= \int_0^{\infty} f(x, y, z) dz = \int_0^{\infty} z^2 e^{-z(1+x+y)} dz, \\
&= \frac{2}{(1+x+y)^3}, x, y \geq 0。
\end{aligned}$$

因 $f_{X,Y}(x,y)$ 無法寫成 $g(x)h(y)$ 的型式, 故 X 與 Y 不獨立。

附帶一提, 由 $f_{X,Y}(x,y)$ 即得 X 之邊際 p.d.f. 為

$$f_X(x) = \int_0^\infty \frac{2}{(1+x+y)^3} dy = \frac{1}{(1+x)^2}, x \geq 0。$$

同理可得 Y 之邊際 p.d.f. 為

$$f_Y(y) = \frac{1}{(1+y)^2}, y \geq 0。$$

由此再度看出 X 與 Y 不獨立, 而且 $E(X), E(Y)$ 皆不存在:

$$\begin{aligned} E(X) &= \int_0^\infty x f_X(x) dx = \int_0^\infty \frac{x}{(1+x)^2} dx = \infty, \\ E(Y) &= \int_0^\infty y f_Y(y) dy = \int_0^\infty \frac{y}{(1+y)^2} dy = \infty。 \end{aligned}$$

不過 $E(Z) = 1$ 倒是存在的。

對於在給定 Z 之下, X 與 Y 為條件獨立, 底下以另一方式來檢驗。首先 X, Z 之聯合 p.d.f. 為

$$\begin{aligned} f_{X,Z}(x,z) &= \int_0^\infty f(x,y,z) dy \\ &= \int_0^\infty z^2 e^{-z(1+x+y)} dy \\ &= z e^{-z(1+x)}, x, z \geq 0。 \end{aligned}$$

因此

$$f_{X|Z}(x|z) = \frac{f_{X,Z}(x,z)}{f_Z(z)} = \frac{z e^{-z(1+x)}}{e^{-z}} = z e^{-zx}, x, z \geq 0,$$

與之前所指出的在給定 $Z = z$ 之下, X 有 $\mathcal{E}(z)$ 分佈吻合。同理可求出

$$f_{Y|Z}(y|z) = z e^{-zy}, y, z \geq 0。$$

故

$$f_{X,Y|Z}(x,y|z) = f_{X|Z}(x|z) f_{Y|Z}(y|z), x, y, z \geq 0。$$

讀者也可自行驗證 $f_{X|Y,Z}(x|y,z) = f_{X|Z}(x|z)$ (習題第1題)。

多項分佈(multinomial distribution)為二項分佈之一推廣,此為一重要的多維分佈。在此設 n, k 為正整數,且常數 $p_1, \dots, p_k$ 滿足 $0 \leq p_i \leq 1, i = 1, \dots, k, \sum_{i=1}^k p_i = 1$ 。則隨機向量 $(X_1, \dots, X_k)$ 稱為有參數 $n, p_1, \dots, p_k$ 之**多項分佈**,以 $\mathcal{M}(n, p_1, \dots, p_k)$ 表之,若 $(X_1, \dots, X_k)$ 之聯合p.d.f.為

$$(7.10) \quad f(x_1, \dots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},$$

其中 $x_1, \dots, x_k$ 為非負整數,且 $\sum_{i=1}^k x_i = n$ 。通常取 $0 < p_i < 1, i = 1, \dots, k$ 。

二項分佈可描述一有兩個不同的結果之試驗。有些試驗有超過兩個的結果,這時多項分佈便可派上用場。可看出二項分佈為多項分佈 $k = 2$ 之一特例:

$$f(x_1, x_2) = \frac{n!}{x_1! x_2!} p_1^{x_1} p_2^{x_2} = \binom{n}{x_1} p_1^{x_1} (1 - p_1)^{n - x_1}。$$

只是因 $k = 2$ 時,由於 $X_2 = n - X_1$,知道 X_1 後, X_2 便決定了,所以對 $k = 2$ 的情況,便僅簡單地以

$$f(x) = \binom{n}{x} p^x (1 - p)^{n - x}, x = 0, 1, \dots, n,$$

表 X_1 之p.d.f.。在第二章例2.11,將 n 個球隨機地丟進 n 個箱子中,若將箱子編號1至 n ,且令 $X_1, \dots, X_n$ 表各箱中之球數,則 $(X_1, \dots, X_n)$ 便有 $\mathcal{M}(n, 1/n, \dots, 1/n)$ 分佈。底下再給一例。

例7.3 投擲一四面體10次,設四面為 A, B, C, D ,其出現的機率分別為0.1, 0.2, 0.3及0.4。試求 A, B, C, D 面分別出現1次,2次,3次及4次之機率。
解. 令 X_i 表第 i 面出現的次數, $i = 1, 2, 3, 4$ 。則 (X_1, X_2, X_3, X_4) 有 $\mathcal{M}(10, 0.1, 0.2, 0.3, 0.4)$ 分佈。因此

$$\begin{aligned} & P(X_1 = 1, X_2 = 2, X_3 = 3, X_4 = 4) \\ &= \frac{10!}{1!2!3!4!} 0.1^1 0.2^2 0.3^3 0.4^4 \\ &= \frac{3,628,800}{288} \cdot 2.7648 \cdot 10^{-6} = 0.03483648。 \end{aligned}$$

在(7.10)式中,

$$(7.11) \quad \frac{n!}{x_1! \cdots x_k!} = \binom{n}{x_1, \dots, x_k}, \sum_{i=1}^k x_i = n, x_1, \dots, x_k \geq 0,$$

此即1.1節所定義的多項係數。多項係數就是 n 個事物分成 k 類, 第1類有 x_1 個, 第2類有 x_2 個, $\dots$, 第 k 類有 x_k 個, 總共之分法。而2.2.2節之二項式定理, 也可推廣至**多項式定理**(Multinomial theorem):

設 n, k 為二正整數, 則對任意實數 $a_1, \dots, a_k$,

$$(7.12) \quad (a_1 + \cdots + a_k)^n = \sum \frac{n!}{x_1! \cdots x_k!} a_1^{x_1} \cdots a_k^{x_k},$$

其中和是對所有 x_i 為非負整數, 且 $\sum_{i=1}^k x_i = n$ 而相加。有了多項式定理, 才保證(7.10)式所定義出的 f 確為 k 維的p.d.f.。

對多項分佈, 其邊際分佈, 及條件分佈, 直觀上便可以得到。說明如下:

考慮 X_k , 若將第 k 種結果視為成功, 其餘結果皆視為失敗, 則 X_k 為 n 個伯努力試驗成功的次數, 每次成功的機率皆為 p_k 。因此, X_k 應有 $\mathcal{B}(n, p_k)$ 分佈。這點利用(7.6)式便可證明。對 $x_k \in \{0, 1, \dots, n\}$, 我們必須將聯合p.d.f.對所有可能的 $(x_1, \dots, x_{k-1})$ 相加。即對所有非負整數的 $x_1, \dots, x_{k-1}$, 且 $\sum_{i=1}^{k-1} x_i = n - x_k$ 相加。以 B 表此集合。則

$$\begin{aligned} f(x_k) &= \sum_{(x_1, \dots, x_{k-1}) \in B} \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \\ &= \frac{n!}{x_k! (n - x_k)!} p_k^{x_k} \cdot \sum_{(x_1, \dots, x_{k-1}) \in B} \frac{(n - x_k)!}{x_1! \cdots x_{k-1}!} p_1^{x_1} \cdots p_{k-1}^{x_{k-1}}. \end{aligned}$$

因 $x_1 + \cdots + x_{k-1} = n - x_k$, 且 $p_1 + \cdots + p_{k-1} = 1 - p_k$, 利用多項式定理, 上述和 $= (p_1 + \cdots + p_{k-1})^{n-x_k} = (1 - p_k)^{n-x_k}$ 。故得

$$f(x_k) = \frac{n!}{x_k! (n - x_k)!} p_k^{x_k} (1 - p_k)^{n-x_k}.$$

因此 X_k 有 $\mathcal{B}(n, p_k)$ 分佈。

同理可得證每一 X_i 皆有 $\mathcal{B}(n, p_i)$ 分佈, $1 \leq i \leq k$ 。

給定 $X_k = x_k$, 則相當於有 $n - x_k$ 次試驗, 而可能有 $k - 1$ 個可能的結果, $(X_1, \dots, X_{k-1})$ 表 $n - x_k$ 次試驗後, 各結果出現的次數。因此直觀上, 給定 $X_k = x_k$ 之下, $(X_1, \dots, X_{k-1})$ 應仍有多項分佈。這是對的, 利用 (7.7) 式得

$$\begin{aligned} f(x_1, \dots, x_{k-1} | x_k) &= \frac{f(x_1, \dots, x_k)}{f(x_k)} \\ &= \frac{\frac{n!}{x_1! \dots x_k!} p_1^{x_1} \dots p_k^{x_k}}{\frac{n!}{x_k!(n-x_k)!} p_k^{x_k} (1-p_k)^{n-x_k}} \\ &= \frac{(n-x_k)!}{x_1! \dots x_{k-1}!} \left(\frac{p_1}{1-p_k}\right)^{x_1} \dots \left(\frac{p_{k-1}}{1-p_k}\right)^{x_{k-1}}。 \end{aligned}$$

故 $X_1, \dots, X_{k-1} | X_k = x_k$ 有 $\mathcal{M}(n - x_k, p_1/(1-p_k), \dots, p_{k-1}/(1-p_k))$ 。

事實上, 對任一 $(X_1, \dots, X_k)$ 之一子集合, 在給定其餘 X_i 之值下, 皆有多項分佈。例如, $X_1, \dots, X_i | X_{i+1} = x_{i+1}, \dots, X_k = x_k \sim \mathcal{M}(n - \sum_{j=i+1}^k x_j, p_1/(1 - \sum_{j=i+1}^k p_j), \dots, p_i/(1 - \sum_{j=i+1}^k p_j))$ 。

其次我們介紹隨機向量獨立的概念。

定義 7.1 設 $\mathbf{X}_1, \dots, \mathbf{X}_n$ 為隨機向量, 以 $f(\mathbf{x}_1, \dots, \mathbf{x}_n)$ 為其聯合 p.d.f., 且 $\mathbf{X}_i$ 之邊際 p.d.f. 為 $f_{\mathbf{X}_i}(\mathbf{x}_i)$, $i = 1, \dots, n$ 。則若對 $\forall(\mathbf{x}_1, \dots, \mathbf{x}_n)$, 皆有

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = \prod_{i=1}^n f_{\mathbf{X}_i}(\mathbf{x}_i),$$

便稱 $\mathbf{X}_1, \dots, \mathbf{X}_n$ 為相互獨立的隨機向量。若所有 $\mathbf{X}_i$ 皆為一維, 則稱 $X_1, \dots, X_n$ 為相互獨立的隨機變數。

如前, 相互獨立如果不會混淆的話, 有時只稱獨立。若 $X_1, \dots, X_n$ 為相互獨立, 則知道其中幾個變數之值, 對其餘變數, 不能提供任何資訊。即利用定義 7.1, 可證明給定 $(X_{i_1}, \dots, X_{i_k})$, 其中 $\{X_{i_1}, \dots, X_{i_k}\}$ 為 $\{X_1, \dots, X_n\}$ 之一子集, 其餘隨機變數之條件分佈, 與邊際分佈相同, 未做任何改變。

相互獨立導致每對獨立。即若 $X_1, \dots, X_n$ 相互獨立, 則任二 $X_i, X_j, i \neq j$, 亦獨立。但相互獨立比每對獨立還強, 即有可能每對獨立, 但不相互獨

立(見習題第15題)。亦可參考第一章例2.4, 事件的每對獨立, 卻不相互獨立的例子。

相互獨立的隨機變數, 有許多好的性質。底下幾個定理, 其證明均類似在3.2節及3.3節中, 兩個變數情況中對應的定理, 因此我們均略去證明。

定理7.1 (定理2.2(ii)之推廣). 設 $X_1, \dots, X_n$ 為相互獨立的隨機變數, 且函數 $g_j(X_j), j = 1, \dots, n$, 滿足 $E(g_j(X_j))$ 存在。則

$$(7.13) \quad E(g_1(X_1) \cdots g_n(X_n)) = \prod_{j=1}^n E(g_j(X_j))。$$

定理7.2 (定理2.3之推廣). $X_1, \dots, X_n$ 為相互獨立的隨機變數, 若且唯若

$$(7.14) \quad E(e^{it_1 X_1 + \cdots + it_n X_n}) = \prod_{j=1}^n E(e^{it_j X_j}), t_j \in R, j = 1, \dots, n。$$

系理7.1 設 $X_1, \dots, X_n$ 為相互獨立的隨機變數。令 $Z = X_1 + \cdots + X_n$, 則

$$(7.15) \quad E(e^{itZ}) = \prod_{j=1}^n E(e^{itX_j}), t \in R。$$

特別地, 若 $X_1, \dots, X_n$ 有相同的分佈, 則

$$E(e^{itZ}) = (E(e^{itX_1}))^n, t \in R。$$

定理7.2及系理7.1, 亦有母函數, 拉普拉斯轉換及動差母函數之版本。

定理2.4可以推廣至 n 個獨立隨機變數的情況。我們只給gamma分佈的結果。

例7.4 設 $X_1, \dots, X_n$ 為相互獨立的隨機變數, 且 X_i 有 $\Gamma(\alpha_i, \beta)$ 分佈, $i = 1, \dots, n$ 。因

$$E(e^{-sX_i}) = \frac{1}{(1 + \beta s)^{\alpha_i}}, s \geq 0,$$

故若令 $Z = X_1 + \cdots + X_n$, 則

$$\begin{aligned}
E(e^{-sZ}) &= \prod_{i=1}^n E(e^{-sX_i}) = \prod_{i=1}^n \frac{1}{(1+\beta s)^{\alpha_i}} \\
&= \frac{1}{(1+\beta s)^{\sum_{i=1}^n \alpha_i}}, s \geq 0,
\end{aligned}$$

此為 $\Gamma(\sum_{i=1}^n \alpha_i, \beta)$ 分佈之拉普拉斯轉換。故 $\sum_{i=1}^n X_i$ 有 $\Gamma(\sum_{i=1}^n \alpha_i, \beta)$ 分佈。即獨立且有 gamma 分佈之隨機變數，只要尺度參數皆相同，其和便仍有 gamma 分佈。

例7.5 設 $X_1, \dots, X_n$ 為相互獨立的隨機變數，分別有動差母函數 $M_{X_i}(t)$, $t \in R, i = 1, \dots, n$ 。又設 $a_1, \dots, a_n, b_1, \dots, b_n$ 為常數。令 $Z = \sum_{i=1}^n (a_i X_i + b_i)$ 。則 Z 之動差母函數為

$$\begin{aligned}
M_Z(t) &= E(e^{tZ}) \\
&= E(e^{t \sum_{i=1}^n (a_i X_i + b_i)}) \\
&= e^{t \sum_{i=1}^n b_i} \prod_{i=1}^n E(e^{a_i t X_i}) \\
&= e^{t \sum_{i=1}^n b_i} \prod_{i=1}^n M_{X_i}(a_i t), t \in R.
\end{aligned}$$

利用上例，可得獨立的常態分佈，其線性組合之和仍為常態分佈。

例7.6 設 $X_1, \dots, X_n$ 為相互獨立的隨機變數，分別有 $\mathcal{N}(\mu_i, \sigma_i^2)$ 分佈， $i = 1, \dots, n$ 。令 $a_1, \dots, a_n, b_1, \dots, b_n$ 為常數。則 $Z = \sum_{i=1}^n (a_i X_i + b_i)$ 有 $\mathcal{N}(\sum_{i=1}^n (a_i \mu_i + b_i), \sum_{i=1}^n a_i^2 \sigma_i^2)$ 分佈。

證明. 首先 X_i 之動差母函數為

$$M_{X_i}(t) = E(e^{tX_i}) = e^{\mu_i t + \sigma_i^2 t^2 / 2}, t \in R.$$

因此 Z 之動差母函數為

$$\begin{aligned}
M_Z(t) &= E(e^{tZ}) \\
&= e^{t \sum_{i=1}^n b_i} \prod_{i=1}^n M_{X_i}(a_i t)
\end{aligned}$$

$$\begin{aligned}
&= e^{t \sum_{i=1}^n b_i} \prod_{i=1}^n e^{\mu_i a_i t + \sigma_i^2 a_i^2 t^2 / 2} \\
&= e^{(\sum_{i=1}^n (a_i \mu_i + b_i))t + (\sum_{i=1}^n a_i^2 \sigma_i^2) t^2 / 2}, t \in R.
\end{aligned}$$

而此為 $\mathcal{N}(\sum_{i=1}^n (a_i \mu_i + b_i), \sum_{i=1}^n a_i^2 \sigma_i^2)$ 之動差母函數。證畢。

底下為判定 n 個隨機向量相互獨立的一個定理，為定理 2.1 之推廣。

定理 7.3 設 $\mathbf{X}_1, \dots, \mathbf{X}_n$ 為隨機向量。則 $\mathbf{X}_1, \dots, \mathbf{X}_n$ 為相互獨立，若且唯若存在函數 $g_i(\mathbf{x}_i), i = 1, \dots, n$ ，使得 $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ 之聯合 p.d.f. 可寫成

$$f(\mathbf{x}_1, \dots, \mathbf{x}_n) = g_1(\mathbf{x}_1) \cdots g_n(\mathbf{x}_n)。$$

底下為定理 3.2 之推廣。

定理 7.4 設 $\mathbf{X}_1, \dots, \mathbf{X}_n$ 為獨立的隨機向量。令 $g_i(\mathbf{x}_i)$ 為 $\mathbf{x}_i$ 的函數， $i = 1, \dots, n$ 。則 $U_i = g_i(\mathbf{X}_i), i = 1, \dots, n$ ，為相互獨立。

n 維隨機向量的變數代換，只是二維隨機向量變數代換之一立即的推廣。例如，在做變換時，雅可彬為一 $n \times n$ 矩陣之行列式的值。

底下我們給一定理 3.1 之推廣。

例 7.7 設 X, Y, Z 為相互獨立，分別有 $\Gamma(p, \lambda), \Gamma(q, \lambda)$ 及 $\Gamma(r, \lambda)$ 分佈。令

$$U = \frac{X}{X+Y}, \quad V = \frac{X+Y}{X+Y+Z}, \quad W = X+Y+Z。$$

則 U, V, W 相互獨立，且分別有 $Be(p, q), Be(p+q, r)$ 及 $\Gamma(p+q+r, \lambda)$ 分佈。

證明。 首先 (X, Y, Z) 之聯合 p.d.f. 為

$$f(x, y, z) = \frac{x^{p-1} e^{-x/\lambda}}{\Gamma(p) \lambda^p} \frac{y^{q-1} e^{-y/\lambda}}{\Gamma(q) \lambda^q} \frac{z^{r-1} e^{-z/\lambda}}{\Gamma(r) \lambda^r}, x, y, z > 0。$$

令

$$u = \frac{x}{x+y}, \quad v = \frac{x+y}{x+y+z}, \quad w = x+y+z, 0 < u, v < 1, w > 0,$$

解出

$$x = uvw,$$

$$y = -uvw + vw = vw(1 - u),$$

$$z = -vw + w = w(1 - v)。$$

則

$$J = \begin{vmatrix} vw & uw & uv \\ -vw & -uw + w & -uv + v \\ 0 & -w & -v + 1 \end{vmatrix} = vw^2。$$

故得

$$\begin{aligned} & f_{U,V,W}(u, v, w) \\ &= \frac{(uvw)^{p-1} e^{-uvw/p} (vw(1-u))^{q-1} e^{(uvw-vw)/\lambda} (w(1-v))^{r-1} e^{(vw-w)/\lambda}}{\Gamma(p)\Gamma(q)\Gamma(r)\lambda^{p+q+r}} \\ &= \frac{vw^2}{\Gamma(p)\Gamma(q)\Gamma(r)\lambda^{p+q+r}} \\ &= \frac{u^{p-1}(1-u)^{q-1}v^{p+q-1}(1-v)^{r-1}w^{p+q+r-1}e^{-w/\lambda}}{\Gamma(p)\Gamma(q)\Gamma(r)\lambda^{p+q+r}} \\ &= \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} u^{p-1}(1-u)^{q-1} \frac{\Gamma(p+q+r)}{\Gamma(p+q)\Gamma(r)} v^{p+q-1}(1-v)^{r-1} \\ &\quad \cdot \frac{w^{p+q+r-1}e^{-w/\lambda}}{\Gamma(p+q+r)\lambda^{p+q+r}}。 \end{aligned}$$

由此即知 U, V, W 獨立，且分別有 $\mathcal{B}e(p, q), \mathcal{B}e(p+q, r)$ 及 $\Gamma(p+q+r, \lambda)$ 分佈。

事實上此結果可推廣到 n 維的情況，見第五章例3.11。

底下為一有關隨機變數之和的期望值及變異數的定理。證明並不困難，因此留在習題中(第11題)。

定理7.5 設有隨機變數 $X_1, \dots, X_n$ 。

(i) 若 $E(X_i)$ 皆存在， $i = 1, \dots, n$ ，則對任意常數 $a_1, \dots, a_n$ ，

$$(7.16) \quad E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i)。$$

(ii) 若 $\text{Var}(X_i)$ 皆存在, $i = 1, \dots, n$, 則對任意常數 $a_1, \dots, a_n$,

$$(7.17) \quad \text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j);$$

若 $X_1, \dots, X_n$ 為每對獨立, 則

$$(7.18) \quad \text{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i)。$$

(iii) 若 $\text{Var}(X_i)$ 及 $\text{Var}(Y_j)$ 皆存在, $i = 1, \dots, n, j = 1, \dots, m$, 則對任意常數 $a_1, \dots, a_n, b_1, \dots, b_m$,

$$(7.19) \quad \text{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{Cov}(X_i, Y_j)。$$

有些本來期望值或變異數不太容易計算的隨機變數, 若能表示成一些隨機變數的和, 則可利用定理7.5求之。我們給一些例子如下。

例7.8 在2.2.2節我們曾指出, 若 $Y_1, \dots, Y_n$ 為有獨立的 $\text{Ber}(p)$ 分佈之隨機變數, 則 $X = Y_1 + \dots + Y_n$ 有 $\mathcal{B}(n, p)$ 分佈。由於

$$E(Y_i) = p, \quad \text{Var}(Y_i) = p(1-p), \quad i = 1, \dots, n,$$

利用(7.16)及(7.18)式, 可得

$$\begin{aligned} E(X) &= \sum_{i=1}^n E(Y_i) = np, \\ \text{Var}(X) &= \sum_{i=1}^n \text{Var}(Y_i) = np(1-p)。 \end{aligned}$$

與2.2.2節所得相同, 但此處計算顯然容易多了。

例7.9 利用機器將 n 個人的薪資單放進 n 封貼有名字的信封, 由於作業錯誤, 變成隨機地將薪資單放進信封, 每個信封恰放進一份薪資單。令 N 表放對的信封數, 求 $E(N)$ 及 $\text{Var}(N)$ 。

解. 首先 N 之p.d.f.並不易給出。如果 k 較接近 n , $P(N = k)$ 尚還好寫出(見習題第19題), 但當 k 離 n 較遠, $P(N = k)$ 就不是很好表示出來。幸好我們只是要求 N 的期望值及變異數, 倒不用知道 N 的p.d.f.。現先將 N 表為 n 個指示函數之和:

$$N = I_{A_1} + \cdots + I_{A_n}。$$

$I_{A_i} = 1$ 或 0 , 就視事件 A_i 是否發生, 其中 A_i 表第 i 個信封放的是正確的薪資單之事件。注意 $A_1, \cdots, A_n$ 既不互斥亦不獨立。但 $P(A_i) = 1/n$, 因此 $E(I_{A_i}) = P(A_i) = 1/n, \forall i = 1, \cdots, n$ 。由(7.16)式得

$$E(N) = \sum_{i=1}^n E(I_{A_i}) = n \cdot \frac{1}{n} = 1。$$

其次

$$E(N^2) = \sum_{i=1}^n E(I_{A_i}^2) + 2 \sum_{i < j} E(I_{A_i} I_{A_j})。$$

又 $E(I_{A_i}^2) = E(I_{A_i}) = 1/n$, 且對 $i \neq j$,

$$E(I_{A_i} I_{A_j}) = P(A_i \cap A_j) = \frac{1}{n(n-1)}。$$

故

$$E(N^2) = n \cdot \frac{1}{n} + 2 \binom{n}{2} \frac{1}{n(n-1)} = 2。$$

因此

$$\text{Var}(N) = E(N^2) - (E(N))^2 = 2 - 1 = 1。$$

故在“亂放”的情況下, 平均會放對一封, 標準差亦不大, 也是1。

讀者也可試以(7.17)式求 $\text{Var}(N)$ (習題第20題)。

例7.10 將 n 個球隨機地放進 n 個箱子中。則空箱數 N 之期望值為何?

解. 首先 N 可表示為 n 個指示函數之和:

$$N = I_1 + \cdots + I_n,$$

其中 $I_i = 1$ 或 0 , 就視第 i 箱是否為空箱。則由(7.16)式, 得

$$E(N) = \sum_{i=1}^n E(I_i) = n \cdot P(\text{第 } i \text{ 箱為空箱})。$$

第 i 箱為空箱表 n 個球全要放進其餘的 $n-1$ 箱中，故

$$P(\text{第 } i \text{ 箱為空箱}) = (1 - \frac{1}{n})^n。$$

因此

$$E(N) = n \cdot (1 - \frac{1}{n})^n。$$

例如，若 $n = 10$ ，則

$$E(N) = 10 \cdot (\frac{9}{10})^{10} \doteq 3.486(\text{箱})。$$

這就是第二章例2.11中，10個球隨機地放進10箱子中，空箱數之期望值。但此處的推導顯然容易多了。

空箱數佔全部箱數之比例 N/n 的期望值為

$$\frac{E(N)}{n} = (1 - \frac{1}{n})^n。$$

此即每一箱子會是空箱的機率。附帶一提，在第二章例2.11之後，我們提過每一箱的球數有 $\mathcal{B}(n, 1/n)$ 分佈。 $(1 - 1/n)^n$ 為此分佈取值0之機率。當 $n \rightarrow \infty$ 時，此機率趨近至 $e^{-1} \doteq 0.367879$ 。此值剛好是 $\mathcal{P}(1)$ 分佈取值0之機率。此點由波松近似亦可看出。

最後利用(7.17)式，可求出

$$\begin{aligned} (7.20) \quad \text{Var}(N) &= n \cdot ((1 - \frac{1}{n})^n - (1 - \frac{1}{n})^{2n}) \\ &\quad + n(n-1)((1 - \frac{2}{n})^n - (1 - \frac{1}{n})^{2n}), \end{aligned}$$

這部分留給讀者自行完成(習題第21題)。

例7.11 某食品公司為了促銷其產品，在每盒餅乾中均隨機地放一張風景卡片，且設總共有 N 張不同的卡片。

我們先求若每次只買一盒, 欲收集齊全 N 張卡片所需買一盒數之期望值。可看出當已有 n 張不同卡片時, $0 \leq n < N$, 欲得一張尚未有的卡片, 所需買一盒數 X_n 有自1開始的幾何分佈, 且參數為 $(N-n)/N$, 因此期望值為 $N/(N-n)$ 。故收集全部 N 張卡片所需買一盒數 $X_0 + X_1 + \cdots + X_{N-1}$ 之期望值為

$$E(X_0 + X_1 + \cdots + X_{N-1}) = \sum_{n=0}^{N-1} \frac{N}{N-n} = N(1 + \frac{1}{2} + \cdots + \frac{1}{N})。$$

由微積分裡的結果知, 當 $N \rightarrow \infty$ 時,

$$(7.21) \quad 1 + \frac{1}{2} + \cdots + \frac{1}{N} - \log N \rightarrow \gamma,$$

其中 γ 為**歐拉常數**(Euler's constant, Euler(1707-1783)為有史以來最多產的數學家, 其生平介紹見黃文璋(1999)第十一章), 其值約為0.5772。因此 N 很大時, 所需買一盒數之期望值可以 $N(\log N + \gamma)$ 來估計。不過要注意 $N(1 + \frac{1}{2} + \cdots + \frac{1}{N})$ 與 $N(\log N + \gamma)$ 之差可能會很大。

其次, 若一次買很多盒, 譬如說 T 盒, 則其中會包含多少張不同的卡片呢? 以 Z 表此值, 求其期望值 $E(Z)$ 。為了方便, 引進隨機變數 $Y_1, \cdots, Y_N$, 其中 $Y_i = 1$, 若第 i 張卡片在這 T 盒裡, 否則 $Y_i = 0$ 。則

$$P(Y_i = 0) = (1 - \frac{1}{N})^T = 1 - P(Y_i = 1)。$$

因此

$$E(Y_i) = 1 - (1 - \frac{1}{N})^T。$$

故

$$E(Z) = E(\sum_{i=1}^N Y_i) = N(1 - (1 - \frac{1}{N})^T)。$$

明顯地上述期望值小於 N 。而之前我們已推導出, 若一次只買一盒, 則當 N 很大時, 欲收集全部 N 張卡片, 所需盒數之期望值約為 $N \log N + N\gamma$ 。由此結果, 引導我們考慮 $T = N \log N$ 。當然此 T 值並非整數, 不過我們只是想了解這時 $N(1 - (1 - 1/N)^T)$ 的大小。

令

$$y = \left(1 - \frac{1}{N}\right)^{N \log N}。$$

則

$$\begin{aligned} \log y &= N \log N \cdot \log\left(1 - \frac{1}{N}\right) \\ &= N \log N \cdot \left(-\frac{1}{N} - \frac{1}{2N^2} - \frac{1}{3N^3} - \cdots\right) \\ &= -\log N \cdot \left(1 + \frac{1}{2N} + \frac{1}{3N^2} + \cdots\right) \\ &< -\log N。 \end{aligned}$$

故得 $y < 1/N$ 。因此若一次買 $[N \log N] + 1$ 盒, 其中 $[x]$ 表小於或等於 x 之最大整數, 則所獲得不同卡片張數之期望值

$$\begin{aligned} &N\left(1 - \left(1 - \frac{1}{N}\right)^{[N \log N] + 1}\right) \\ &> N\left(1 - \left(1 - \frac{1}{N}\right)^{N \log N}\right) > N\left(1 - \frac{1}{N}\right) = N - 1。 \end{aligned}$$

由於最多只能得到 N 張不同的卡片, 故得此時期望值已算是很接近 N 了。

在 2.2.5 節我們曾介紹超幾何分佈, 但並未說明期望值及變異數是如何推導出來的。直接計算可能較繁瑣, 利用多變數倒是可求出來。見下例。

例 7.12 設 X 有 $\mathcal{H}(N, D, n)$ 分佈, 即 X 之 p.d.f. 為

$$(7.22) \quad f(x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}}, \max\{0, n - N + D\} \leq x \leq \min\{n, D\}。$$

底下求 X 之期望值及變異數。

仍以 2.2.5 節取球為例。令 $Y_i = 1$ 或 0 , 依取出的第 i 個球是否為白球而定, $i = 1, \dots, n$ (這些球的編號並不重要, 因每次取出後不放回, 所以可以一口氣取出 n 個, 再任意編號)。則每一 Y_i 皆有 $\text{Ber}(D/N)$ 分佈, 且 $X = Y_1 + \cdots + Y_n$ 。因此 $E(Y_i) = D/N$, 且 $\text{Var}(Y_i) = D(N-D)/N^2$, $i = 1, \dots, n$ 。再利用 (7.16) 式, 便可求出 X 之期望值:

$$(7.23) \quad E(X) = \sum_{i=1}^n E(Y_i) = nE(Y_1) = nD/N.$$

$Y_1, \dots, Y_n$ 並不獨立, 我們利用(7.17)式來求 $\text{Var}(X)$ 。首先對 $i \neq j$,

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= E(Y_i Y_j) - E(Y_i)E(Y_j) \\ &= P(Y_i = 1, Y_j = 1) - (D/N)^2 \\ &= \frac{\binom{D}{2}}{\binom{N}{2}} - \left(\frac{D}{N}\right)^2 \\ &= -\frac{D(N-D)}{N^2(N-1)}, \end{aligned}$$

其中用到 $Y_i = 1$ 且 $Y_j = 1$, 表第 i 及第 j 個球皆為白球, 此二球為 N 個球中的任二球, 共有 $\binom{N}{2}$ 個可能, 若要皆為白球, 表此二球皆要屬於 D 個白球中, 共有 $\binom{D}{2}$ 個可能, 因此 $Y_i = 1$ 且 $Y_j = 1$ 之機率為 $\binom{D}{2}/\binom{N}{2}$ (讀者可否想通為什麼共變異數為負)。由是得

$$\begin{aligned} (7.24) \quad \text{Var}(X) &= n\text{Var}(Y_1) + 2\binom{n}{2}\text{Cov}(Y_1, Y_2) \\ &= \frac{nD(N-1)}{N^2} - \frac{n(n-1)D(N-D)}{N^2(N-1)} \\ &= \frac{nD(N-D)(N-n)}{N^2(N-1)}. \end{aligned}$$

其次考慮每次取出的球要放回, 且以 Z 表 n 次後共得之白球數。則 Z 有 $\mathcal{B}(n, D/N)$ 分佈, $E(Z) = nD/N$ 與 $E(X)$ 相等, 但此時 $\text{Var}(Z) = nD(N-D)/N^2$ 。故 $\text{Var}(X)/\text{Var}(Z) = (N-n)/(N-1)$ 。若樣本數 n 與母體數 N 之比值很小, 則二變數之比便很接近 1。

若令 $p = D/N$, 則 X 有 $\mathcal{H}(N, Np, n)$ 分佈, Z 有 $\mathcal{B}(n, p)$ 分佈。由上推導知, 當 n, p 固定且令 $N \rightarrow \infty$, 則 $\text{Var}(X) \rightarrow np(1-p) = \text{Var}(Z)$ 。這是可以預期的。因當母體數 N 很大, $Y_1, \dots, Y_n$ 便近似獨立, 因此 X 可以 $\mathcal{B}(n, p)$ 為其近似分佈。在做各種民意調查時, 選民總數 N 往往很大, 且樣本數 n 與 N 相比很小, 而想估計選民中對某候選人之支持率 p 。此時常近似地以較簡單的二項分佈 $\mathcal{B}(n, p)$, 當做樣本中對該候選人之

支持人數 X 的分佈, 而 X 原本是有 $\mathcal{H}(N, Np, n)$ 分佈的。事實上可以證明 $N \rightarrow \infty$ 時(且 n, p 固定)

$$P(X = k) \rightarrow P(Z = k)。$$

當然若 n 與 N 相比並非很小, 就不適合以 $\mathcal{B}(n, p)$ 分佈當做 $\mathcal{H}(N, Np, n)$ 分佈之近似。

對隨機向量 $\mathbf{X} = (X_1, \dots, X_n)$, 令 $\boldsymbol{\mu}_{\mathbf{X}}$ 及 $\boldsymbol{\Sigma}_{\mathbf{X}}$ 分別表其期望向量(mean vector)及共變異數矩陣(covariance matrix), 也可僅以 $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ 表之, 其中

$$\boldsymbol{\mu} = (\mu_1, \dots, \mu_n),$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix},$$

而

$$\mu_i = E(X_i), i = 1, \dots, n,$$

$$\sigma_i^2 = \text{Var}(X_i), i = 1, \dots, n,$$

$$\sigma_{ij} = \sigma_{ji} = \text{Cov}(X_i, X_j), i, j = 1, \dots, n。$$

令

$$\rho_{ij} = \rho(X_i, X_j), i, j = 1, \dots, n。$$

則

$$\sigma_{ij} = \text{Cov}(X_i, X_j) = \rho_{ij} \sigma_i \sigma_j, i, j = 1, \dots, n。$$

故 $\boldsymbol{\Sigma}$ 亦可寫成

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 & \cdots & \rho_{1n} \sigma_1 \sigma_n \\ \rho_{12} \sigma_1 \sigma_2 & \sigma_2^2 & \cdots & \rho_{2n} \sigma_2 \sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1n} \sigma_1 \sigma_n & \rho_{2n} \sigma_2 \sigma_n & \cdots & \sigma_n^2 \end{pmatrix}。$$

對一列向量 $\mathbf{a} = (a_1, \dots, a_n)$,

$$\mathbf{a}' = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix},$$

$\mathbf{a}'$ 即為 $\mathbf{a}$ 之轉置 (transpose), 為一行向量。則

$$\begin{aligned} & (\mathbf{X} - \boldsymbol{\mu})'(\mathbf{X} - \boldsymbol{\mu}) \\ &= \begin{pmatrix} X_1 - \mu_1 \\ X_2 - \mu_2 \\ \vdots \\ X_n - \mu_n \end{pmatrix} (X_1 - \mu_1, X_2 - \mu_2, \dots, X_n - \mu_n) \\ &= \begin{pmatrix} (X_1 - \mu_1)^2 & \cdot & \cdots & (X_1 - \mu_1)(X_n - \mu_n) \\ (X_2 - \mu_2)(X_1 - \mu_1) & (X_2 - \mu_2)^2 & \cdots & (X_2 - \mu_2)(X_n - \mu_n) \\ \vdots & \vdots & \ddots & \vdots \\ (X_n - \mu_n)(X_1 - \mu_1) & \cdot & \cdots & (X_n - \mu_n)^2 \end{pmatrix}. \end{aligned}$$

上述 $\boldsymbol{\mu}$ 及 $\boldsymbol{\Sigma}$ 便可分別寫成

$$\boldsymbol{\mu} = E(\mathbf{X}),$$

$$\boldsymbol{\Sigma} = E((\mathbf{X} - \boldsymbol{\mu})'(\mathbf{X} - \boldsymbol{\mu})).$$

例7.13 設 $(X_1, X_2) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。則

$$\begin{aligned} \boldsymbol{\mu} &= (\mu_1, \mu_2), \\ \boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \end{aligned}$$

例7.14 設 (X_1, X_2, X_3) 之共變異數矩陣為

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & -0.5 & 0.2 \\ -0.5 & 6 & -0.4 \\ 0.2 & -0.4 & 1 \end{pmatrix}.$$

試求 $\text{Cov}(2X_1 + X_3, X_1 - X_2 - X_3)$ 。

解. 由定理 7.5,

$$\begin{aligned}
 & \text{Cov}(2X_1 + X_3, X_1 - X_2 - X_3) \\
 &= 2\text{Cov}(X_1, X_1) - 2\text{Cov}(X_1, X_2) - 2\text{Cov}(X_1, X_3) + \text{Cov}(X_1, X_3) \\
 &\quad - \text{Cov}(X_2, X_3) - \text{Cov}(X_3, X_3) \\
 &= 2 \cdot 2 - 2 \cdot (-0.5) - 0.2 - (-0.4) - 1 \\
 &= 4 + 1 - 0.2 + 0.4 - 1 = 4.2。
 \end{aligned}$$

3.8 不等式

在數學裡, 不等式的角色很重要。機率及統計裡, 也有一些不等式常出現。本節我們介紹一些。

引理 8.1 設 $u, v, p, q > 0$, 且

$$(8.1) \quad \frac{1}{p} + \frac{1}{q} = 1。$$

則

$$(8.2) \quad \frac{u^p}{p} + \frac{v^q}{q} \geq uv,$$

且等號只有在 $u = v^{1/(p-1)}$ 才成立。

證明. 首先由 (8.1) 式及 $p, q > 0$, 立即得 $p > 1, q > 1$ 。現固定 v , 並考慮函數

$$g(u) = \frac{u^p}{p} + \frac{v^q}{q} - uv。$$

欲求 $g(u)$ 之絕對極小值。令 $g'(u) = 0$, 得

$$u^{p-1} - v = 0 \Rightarrow u = v^{1/(p-1)}。$$

又 $g''(u) = (p-1)u^{p-2} > 0, \forall u > 0$, 故 $g(u)$ 在 $u = v^{1/(p-1)}$ 有唯一的相對極小值。且極小值為 (由 (8.1) 式得 $p/(p-1) = q$)

$$g(v^{1/(p-1)}) = \frac{1}{p}v^{p/(p-1)} + \frac{1}{q}v^q - v^{1/(p-1)}v$$

$$= \left(\frac{1}{p} + \frac{1}{q} - 1\right)v^q = 0。$$

由微積分中的結果知, 一函數的絕對極小值, 只能發生在該函數導數等於0的點、不可微的點, 或定義域的端點。而 $g(u)$ 定義在 $(0, \infty)$, 到處可微, 只有一導數為0的點, 且 $\lim_{u \rightarrow 0} g(u) = v^q/q > 0$, $\lim_{u \rightarrow \infty} g(u) = \infty$ 。故唯一的絕對極小值, 發生在 $u = v^{1/(p-1)}$, 且極小值為0。故 $g(u) \geq 0, \forall u \in R$, 且 $g(u) > 0, \forall u \neq v^{1/(p-1)}$ 。得證(8.2)式, 且等號只有在 $u = v^{1/(p-1)}$ 才成立。

利用上述引理即得下述不等式:

赫德爾不等式(Hölder's inequality). 設 X, Y 為二隨機變數, p, q 滿足(8.1)式, 且設 $E(|X|^p)$ 及 $E(|X|^q)$ 皆存在。則

$$(8.3) \quad E(|XY|) \leq (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q}。$$

證明. 令

$$u = \frac{|X|}{(E(|X|^p))^{1/p}}, \quad v = \frac{|Y|}{(E(|Y|^q))^{1/q}},$$

則由引理8.1, 得

$$\frac{1}{p} \frac{|X|^p}{E(|X|^p)} + \frac{1}{q} \frac{|Y|^q}{E(|Y|^q)} \geq \frac{|X||Y|}{(E|X|^p)^{1/p} (E|Y|^q)^{1/q}}。$$

將上式兩側取期望值, 則左側成為1, 再將右側的分母移至左側即得(8.3)式。證畢。

Hölder(1859-1937)為德國數學家。赫德爾不等式之最著名的特例, 很可能是 $p = q = 2$ 的情況。此即在定理5.4中我們證過的**史瓦茲不等式**。

史瓦茲不等式. 對任二隨機變數 X, Y , 只要 $E(X^2), E(Y^2) < \infty$, 則

$$(8.4) \quad E(|XY|) \leq (E(X^2))^{1/2} (E(Y^2))^{1/2}。$$

赫德爾不等式尚有一些特例也常出現。例如，在(8.3)式中令 $Y \equiv 1$ ，得

$$(8.5) \quad E(|X|) \leq (E(|X|^p))^{1/p}, 1 < p < \infty。$$

在此對一隨機變數 Y ，及常數 c ， $Y \equiv c$ ，讀做“ Y 恆等於 c ”，其意義為 $P(Y = c) = 1$ ，即 Y 為一退化隨機變數，取值在 c 。

對 $1 < r < p$ ，若以 $|X|^r$ 取代(8.5)式中之 $|X|$ ，得

$$E(|X|^r) \leq (E(|X|^{pr}))^{1/p}。$$

次令 $s = rp$ (注意因 $p > 1$ ，故 $s > r$)，再將兩側開 r 次方，得

$$(8.6) \quad E(|X|^r)^{1/r} \leq (E(|X|^s))^{1/s}, 1 < r < s < \infty,$$

此即**黎波諾夫不等式**(Liapounov's inequality, Liapounov(1857-1918)，為俄國機率學家)。由黎波諾夫不等式知，若一隨機變數之高次動差存在，則較低次的動差亦存在。例如，若知 $E(|X|^4) = E(X^4)$ 存在，則知 $E(|X|)$ ， $E(|X|^2)$ 及 $E(|X|^3)$ 也皆存在，因此 $E(X)$ ， $E(X^2)$ 及 $E(X^3)$ 也皆存在。

底下的不等式精神上與赫德爾不等式類似，也可藉助赫德爾不等式得到。

明考夫斯基不等式(Minkowski's inequality)。設 X, Y 為二隨機變數，且 $1 \leq p < \infty$ 。則

$$(8.7) \quad (E(|X + Y|^p))^{1/p} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p}。$$

證明。首先利用**三角形不等式**(Triangle inequality)，

$$|X + Y| \leq |X| + |Y|,$$

可得

$$\begin{aligned} E(|X + Y|^p) &= E(|X + Y||X + Y|^{p-1}) \\ &\leq E(|X||X + Y|^{p-1}) + E(|Y||X + Y|^{p-1})。 \end{aligned}$$

再利用赫德爾不等式於上述右側之每一期望值，得

$$E(|X+Y|^p) \leq E(|X|^p)^{1/p} (E(|X+Y|^{(p-1)q}))^{1/q} \\ + (E(|Y|^p))^{1/p} (E(|X+Y|^{(p-1)q}))^{1/q},$$

其中 q 滿足 $1/p + 1/q = 1$ 。上式兩側各除以 $(E(|X+Y|^{(p-1)q}))^{1/q}$ ，且因 $(p-1)q = p$, $1 - 1/q = 1/p$ ，即得(8.7)式。證畢。

Minkowski(1864-1909)亦為著名的俄國數學家。上述這些不等式，也有對應之數值的版本，即將期望值改為和。例如，對任意實數 $a_i, b_i, i = 1, \dots, n$,

$$(8.8) \quad \sum_{i=1}^n |a_i b_i| \leq \left(\sum_{i=1}^n |a_i|^p \right)^{1/p} \left(\sum_{i=1}^n |b_i|^q \right)^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1, p, q > 0,$$

便對應赫德爾不等式。這只要在赫德爾不等式中，定義隨機變數 X, Y ，分別以 $1/n$ 的機率取值在 $(a_1, b_1), \dots, (a_n, b_n)$ ，便導致(8.8)式(習題第8題)。

明考夫斯基之數值的版本則為：對任意實數 $a_i, b_i, i = 1, \dots, n$ ，及 $p \geq 1$,

$$(8.9) \quad \left(\sum_{i=1}^n |a_i + b_i|^p \right)^{1/p} \leq \left(\sum_{i=1}^n |a_i|^p \right)^{1/p} + \left(\sum_{i=1}^n |b_i|^p \right)^{1/p}.$$

(8.8)式一重要的特例是 $p = q = 2$ 的情況：

$$(8.10) \quad \left(\sum_{i=1}^n a_i b_i \right)^2 \leq \left(\sum_{i=1}^n |a_i b_i| \right)^2 \leq \left(\sum_{i=1}^n a_i^2 \right) \left(\sum_{i=1}^n b_i^2 \right).$$

這是在數學中常見的關於數值之柯西-史瓦茲不等式。(8.10)式中若取 $b_i = 1, i = 1, \dots, n$ ，則得

$$(8.11) \quad \frac{1}{n} \left(\sum_{i=1}^n a_i \right)^2 \leq \frac{1}{n} \left(\sum_{i=1}^n |a_i| \right)^2 \leq \sum_{i=1}^n a_i^2.$$

有些不等式主要是利用實值函數的性質，而不是仰賴機率的性質。其中一個很有用的是間森不等式(Jensen's inequality)，適用於凸函數(convex function)。

定義8.1 函數 g 若滿足對 $\forall x, y \in R$ ，及 $0 < \lambda < 1$,

$$(8.12) \quad g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y),$$

便稱為**凸函數**。而若(8.12)式中之不等式反過來, g 便稱為**凹函數**(concave function)。可看出 g 為凹函數若且唯若 $-g$ 為凸函數。

(8.12)式成立, 表 $g(x)$ 之圖形由 u 至 v , 在 $(u, g(u)), (v, g(v))$ 連線的下方。至於凹函數則是函數圖形在前述連線段上方。例如, $g(x) = x^2$ 為凸函數, $g(x) = \log x$ 為凹函數。見圖8.1。

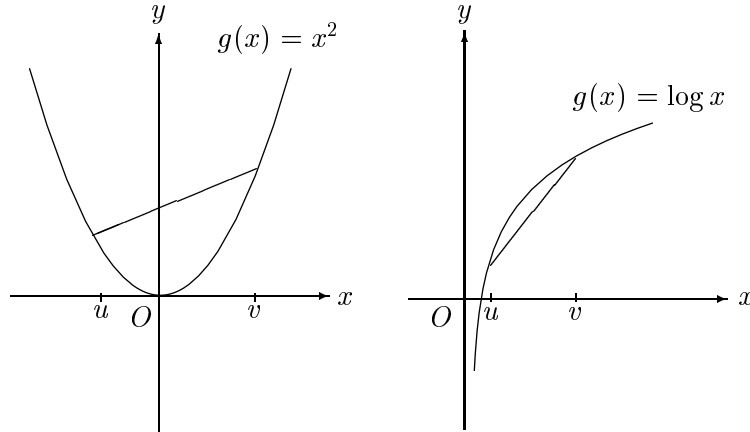


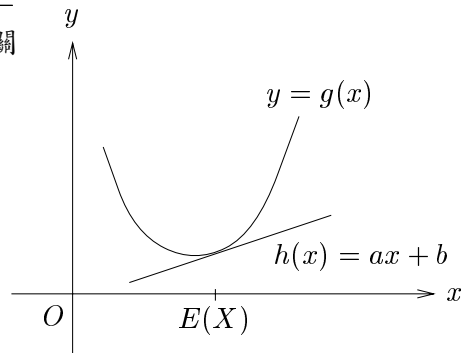
圖8.1 凸函數及凹函數之圖示

間森不等式. 設 $g(x)$ 為一凸函數, 且隨機變數 X 之 $E(X)$ 及 $E(g(X))$ 皆存在。則

$$(8.13) \quad E(g(X)) \geq g(E(X)).$$

證明. 考慮 $y = g(x)$ 之圖形, 令 $h(x) = ax + b$ 表圖形在 $(E(X), g(E(X)))$ 之切線。由於 $g(x)$ 為一凸函數, $g(x) \geq ax + b$ 。又因期望值會保持大小關係, 故

$$\begin{aligned} E(g(X)) &\geq E(ax + b) \\ &= aE(X) + b \\ &= h(E(X)) \\ &= g(E(X)), \end{aligned}$$



其中最後一等式成立乃是因直線 $y = h(x)$ 切圖形於 $(E(X), g(E(X)))$, 故 h 與 g 在 $E(X)$ 的值相等。

底下為間森不等式之一立即的應用(取 $g(x) = x^2$):

$$E(X^2) \geq (E(X))^2,$$

我們在1.6節便提過了。另外, 對 $x > 0$, 利用 $g(x) = 1/x$ 為凸函數, 則得

$$(8.14) \quad E(1/X) \geq 1/E(X)。$$

由微積分裡的知識得, 對一個二次可微的函數 g , 若 $g''(x) \geq 0, \forall x \in R$, 則 $g(x)$ 為凸函數; 而若 $g''(x) < 0, \forall x \in R$, $g(x)$ 便為凹函數。又對於凹函數 g , 因 $-g$ 為凸函數, 故

$$(8.15) \quad E(g(X)) \leq g(E(X))。$$

例8.1 設 $a_1, \dots, a_n$ 皆大於0, 令

$$\begin{aligned} M_A &= \frac{1}{n}(a_1 + \dots + a_n) = \frac{1}{n} \sum_{i=1}^n a_i, \\ M_G &= (a_1 \cdots a_n)^{1/n} = \left(\prod_{i=1}^n a_i \right)^{1/n}, \\ M_H &= \frac{1}{\frac{1}{n}(\frac{1}{a_1} + \dots + \frac{1}{a_n})} = \left(\frac{1}{n} \sum_{i=1}^n a_i^{-1} \right)^{-1}, \end{aligned}$$

分別表其**算術平均**(arithmetic mean)、**幾何平均**(geometric mean), 及**調和平均**(harmonic mean)。此三不等式之關係為

$$(8.16) \quad M_A \geq M_G \geq M_H。$$

底下我們看如何以間森不等式證明上述不等式。

設隨機變數 X 滿足 $P(X = a_i) = 1/n, i = 1, \dots, n$ 。因 $y = \log x$, $x > 0$, 為一凹函數, 間森不等式導致

$$E(\log X) \leq \log(E(X))。$$

故

$$\begin{aligned}\log M_G &= \frac{1}{n} \sum_{i=1}^n \log a_i = E(\log X) \leq \log(E(X)) \\ &= \log\left(\frac{1}{n} \sum_{i=1}^n a_i\right) = \log M_A.\end{aligned}$$

因此 $M_A \geq M_G$ 。其次因 $y = 1/x, x > 0$, 為凸函數, 故

$$\frac{1}{M_H} = \frac{1}{n} \sum_{i=1}^n \frac{1}{a_i} = E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)} = \frac{1}{\frac{1}{n} \sum_{i=1}^n a_i} = \frac{1}{M_A}。$$

因此 $M_A \geq M_H$ 。

至於 $M_G \geq M_H$, 可利用上述推導而得:

$$\frac{1}{M_H} = \frac{1}{n} \sum_{i=1}^n \frac{1}{a_i} \geq \left(\prod_{i=1}^n \frac{1}{a_i}\right)^{1/n} = \frac{1}{\left(\prod_{i=1}^n a_i\right)^{1/n}} = \frac{1}{M_G},$$

其中不等式, 我們是利用 $M_A \geq M_G$ 於正數 $\frac{1}{a_1}, \dots, \frac{1}{a_n}$ 。

柴比雪夫不等式(Chebyshev's inequality)可以說是機率裡一極重要且有用的不等式。此不等式的重要, 是因其適用性廣泛, 而非其精確性。Chebyshev(1821-1894, 有時寫成Chebychev, Shebychev, Tschebischew等), 為十九世紀俄國極重要的一位機率學家。

柴比雪夫不等式. 設 X 為一隨機變數, $g(x)$ 為一非負函數, 且滿足 $E(g(X))$ 存在。則對任意 $c > 0$,

$$(8.17) \quad P(g(X) \geq c) \leq \frac{E(g(X))}{c}。$$

證明. 我們只證連續型的情況。設 X 為一連續型的隨機變數, 以 $f(x)$ 為其 p.d.f., 則

$$\begin{aligned}E(g(X)) &= \int_{-\infty}^{\infty} g(x)f(x)dx \\ &\geq \int_{\{x|g(x) \geq c\}} g(x)f(x)dx \\ &\geq c \int_{\{x|g(x) \geq c\}} f(x)dx \\ &= cP(g(X) \geq c)。$$

得證(8.17)式。

在柴比雪夫不等式中, 取 $g(x) = |x - \mu|^r$, 且以 c^r 取代 c , 其中 $\mu = E(X)$, $r > 0$, 則只要 $E(|X - \mu|^r)$ 存在, 便有

$$P(|X - \mu|^r \geq c^r) \leq \frac{E(|X - \mu|^r)}{c^r},$$

但 $P(|X - \mu|^r \geq c^r) = P(|X - \mu| \geq c)$ 。故得

$$(8.18) \quad P(|X - \mu| \geq c) \leq \frac{E|x - \mu|^r}{c^r}, r > 0。$$

由(8.18)式, 又得(取 $c = k\sigma$, $r = 2$, 其中 $\sigma^2 = \text{Var}(X)$)

$$(8.19) \quad P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, k > 0。$$

這是柴比雪夫不等式一極常出現的型式。可驗證(8.19)式與下式等價

$$(8.20) \quad P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \varepsilon > 0。$$

當然(8.19)及(8.20)式成立的先決條件是 X 的變異數要存在。(8.17)式是最一般的柴比雪夫不等式, 不過(8.18)、(8.19)及(8.20)三式也常被稱做柴比雪夫不等式。

例8.2 設 X 有 $\mathcal{P}(1)$ 分佈, 則 $E(X) = \text{Var}(X) = 1$ 。柴比雪夫不等式給出(利用(8.20)式)

$$P(|X - 1| \geq 2) \leq \frac{1}{4}。$$

另外, 經由實際計算, 可得

$$\begin{aligned} P(|X - 1| \geq 2) &= P(X \geq 3) = 1 - \sum_{i=0}^2 P(X = i) \\ &= 1 - \sum_{i=0}^2 \frac{e^{-1}}{i!} = 1 - e^{-1} \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{2} \right) \\ &= 1 - 2.5e^{-1} \doteq 0.080, \end{aligned}$$

的確是小於 $1/4 = 0.25$, 但顯然誤差不小。換句話說, 雖然 σ^2/ε^2 為 $P(|X - \mu| \geq \varepsilon)$ 之一上界, 卻不一定很精確。但這還不是最差的情況。事實上(8.19)及(8.20)式之左側皆為機率值, 不能超過1, 但右側卻有可能大於1。此時不等式便未提供任何資訊。不過如前所述, 柴比雪夫不等式並非因精確性而重要, 而是因其普遍性。對於(8.19)及(8.20)二式, 對於只要是變異數存在的隨機變數便皆成立。對一隨機變數 X , (8.19)及(8.20)式, 分別提供 X 會偏離期望值 k 個標準差及一固定距離 μ 之機率的一個上界。在1.6節我們說過, 期望值像是隨機變數分佈的一個核心, 隨機變數的可能值則散佈在期望值的左右。對不同的隨機變數, 其可能值有時偏離期望值較遠, 有時較近, 但都會滿足(8.19)及(8.20)式。

(8.19)與(8.20)式分別等價於

$$(8.21) \quad P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}, k > 0,$$

$$(8.22) \quad P(|X - \mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}, \varepsilon > 0。$$

我們再以常態分佈來檢驗柴比雪夫不等式之精確性。

例8.3 設 $X \sim \mathcal{N}(\mu, \sigma^2)$, $Z \sim \mathcal{N}(0, 1)$ 。我們比較實際的機率及(8.21)式所給之機率下界。

$$P(|X - \mu| < \sigma) = P(|Z| < 1) \doteq 0.682, \quad 1 - \frac{1}{1^2} = 0,$$

$$P(|X - \mu| < 2\sigma) = P(|Z| < 2) \doteq 0.954, \quad 1 - \frac{1}{2^2} = 0.75,$$

$$P(|X - \mu| < 3\sigma) = P(|Z| < 3) \doteq 0.997, \quad 1 - \frac{1}{3^2} \doteq 0.889。$$

習題

第3.1節

1. 試證下述二函數皆為聯合p.d.f.。

- (i) $f(x, y) = \frac{x+y}{54}, x = 1, 2, 3, y = 1, 2, 3, 4;$
- (ii) $f(x, y) = e^{-x(y+1)^2}, x, y > 0$ 。
2. 設 (X, Y) 之聯合p.d.f.為 $f(x, y) = 6xy^2, 0 < x, y < 1$ 。
- (i) 試驗證 f 為一p.d.f.;
- (ii) 試求 $P(X + Y \geq 1)$;
- (iii) 試求 $P(1/2 < X < 3/4)$ 。
3. 在例1.4中, 試求 $P(X > Y^2)$ 及 $P(Y > X^2)$ 。
4. 設 $f(x, y) = 1/4, (x, y) \in (-1, 1) \times (-1, 1)$ 。試分別求 $X^2 + Y^2 < 1$,
 $X + Y < 1, 2X - Y > 0$, 及 $|X + Y| < 2$ 之機率。
5. 設 (X, Y) 之聯合p.d.f.為

$$f(x, y) = c(x + 2y), 0 < x < 2, 0 < y < 1。$$

- (i) 試決定常數 c 之值;
- (ii) 試求 X 之邊際p.d.f.;
- (iii) 試求 $Z = 9/(X + 1)^2$ 之p.d.f.;
- (iv) 試求 (X, Y) 之聯合分佈函數。
6. 設 (X, Y) 之聯合p.d.f.為

$$f(x, y) = c, \quad x^2 + y^2 \leq r^2,$$

其中 $r > 0$ 為一常數。

- (i) 試決定常數 c 之值;
- (ii) 試求 $P(X^2 + Y^2 \leq r^2/2)$;
- (iii) 試求 $P(Y > \sqrt{3}X > 0)$ 。
7. 設 (X, Y) 之聯合p.d.f.為 $f(x, y) = 1/\pi, x^2 + y^2 \leq 1$ 。試分別求 X 及
 Y 之邊際p.d.f.。

8. 設 (X, Y) 之聯合 p.d.f. 為 $f(x, y) = x + y, 0 \leq x, y \leq 1$ 。試求 $P(X > \sqrt{Y})$ 。
9. 設 (X, Y) 之聯合 p.d.f. 為 $f(x, y) = 2x, 0 \leq x, y \leq 1$ 。試求 $P(X^2 < Y < X)$ 。
10. 設 (X, Y) 之聯合 p.d.f. 為 $f(x, y) = \lambda^2 e^{-\lambda(x+y)}, x, y \geq 0$ 。試求 $P(X \geq 2Y)$ 。
11. (i) 設 (X, Y) 之聯合 p.d.f. 為

$$f(x, y) = 360xy^2(1 - x - y), x, y \geq 0, x + y \leq 1。$$

試求 $E(X/Y)$ 及 X, Y 之邊際 p.d.f., 並指出 X, Y 各為那一常見分佈。

(ii) 設 (U, V) 之聯合 p.d.f. 為

$$g(u, v) = 1,800u(1 - u)^4v^2(1 - v)^3, 0 \leq u, v \leq 1。$$

試證 $U \stackrel{d}{=} X, V \stackrel{d}{=} Y$, 其中 X, Y 之分佈如(i)中所給。

12. 設非負函數 $g(x)$ 滿足 $\int_0^\infty g(x)dx = 1$ 。試證

$$f(x, y) = \frac{2g(\sqrt{x^2 + y^2})}{\pi\sqrt{x^2 + y^2}}, x, y > 0,$$

為一 p.d.f.。

第3.2節

- 試求出例2.1中的各 $f_{X|Y}(x|y)$ 之值。
- 設 $f(x, y)$ 如例1.4中所給。試求 $f_X(x), f(y|x), E(Y|x)$ 及 $\text{Var}(Y|x)$ 。
- 試證(2.11)式。
- 設 (X, Y) 之聯合 p.d.f. 為

$$f(x, y) = \frac{2 + x + y}{8}, -1 < x, y < 1。$$

試分別求出 X, Y 之邊際 p.d.f., 並因此驗證 X 與 Y 不獨立。

5. 設 (X, Y) 之聯合p.d.f.為 $f(x, y) = x + y, 0 \leq x, y \leq 1$ 。試求 $f_X(x), f(y|x), E(Y|x)$ 及 $\text{Var}(Y|x)$ 。
6. 設 (X, Y) 之聯合p.d.f.為 $360xy^2(1 - x - y), 0 \leq x, y, x + y \leq 1$ 。試求 $f(y|x), E(Y|x)$ 及 $\text{Var}(Y|x)$ 。
7. 設 $f(x, y) = 1/2, 0 \leq x, y, x + y \leq 1$ 。
- (i) 試求 $f(y|x)$;
- (ii) 試證 $\text{Var}(Y|x) = \frac{1}{12}(1 - x)^2, 0 \leq x < 1$, 且 $\text{Var}(Y) = 1/18$, 並驗證 x 夠接近0時, $\text{Var}(Y) < \text{Var}(Y|x)$ 。
8. 設 (X, Y) 之聯合p.d.f.為

$$f(x, y) = \frac{\tau^\beta}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} y^{\beta-1} e^{-(\tau+x)y}, x, y > 0,$$

其中 $\alpha, \beta, \tau > 0$ 為常數。試求 $f_X(x), f(y|x), f_Y(y)$ 及 $f(x|y)$ 。

9. 設函數

$$f(x, y) = \begin{cases} 4e^{-2x-2y} - 2e^{-2x-y} - 2e^{-x-2y} \\ \quad + 2e^{-x-y} & , 0 < x, y < \infty, \\ 0 & , \text{其他}。 \end{cases}$$

- (i) 試證 f 為一二變數之p.d.f.;
- (ii) 假設 (X, Y) 以 f 為聯合p.d.f., 試求 $f(y|x), E(Y|x)$ 及 $\text{Var}(Y|x)$ 。
10. 試決定例2.12中, X, Y 之邊際分佈。
11. 設 X, Y 之聯合p.d.f.為 $f(x, y) = 2e^{-x-y}, 0 \leq x < y < \infty$ 。
- (i) 試求 $P(Y > 2X)$;
- (ii) 試證 X 之邊際分佈為 $\mathcal{E}(1/2)$;
- (iii) 試求 $f_{Y|X}(y|x)$ 及 $E(Y|X = x)$;
- (iv) 試檢驗 X 與 Y 是否獨立。

12. 設 X, Y 之聯合 p.d.f. 為 $f(x, y) = ye^{-xy-y}, x, y > 0$ 。
 - (i) 試檢驗 X 與 Y 是否獨立;
 - (ii) 試求 $E(Y|x)$ 及 $\text{Var}(Y|x)$ 。
13. 設 X 有 $\mathcal{E}(1)$ 分佈。令 $Y = [X + 1]$, 其中 $[\cdot]$ 表最大整數函數。
 - (i) 試求 Y 之分佈;
 - (ii) 試求給定 $Y \geq 5, X - 4$ 之分佈。
14. 試證定理 2.4(ii)-(vii)。
15. 設 X, Y 為二獨立的隨機變數, 且皆有 $\mathcal{N}(0, 1)$ 分佈。試求 $P(X^2 + Y^2 < 1)$ 。
16. 設 X, Y 為二獨立的隨機變數, 且分別有 $\mathcal{P}(\lambda)$ 及 $\mathcal{P}(\mu)$ 分佈。試求給定 $X + Y = n$ 之下, $n \geq 1$, X 之條件 p.d.f.。並指出此為那一常見的分佈, 參數為何。
17. 設 X, Y 為二獨立的隨機變數, 且分別有 $\mathcal{B}(m, p)$ 及 $\mathcal{B}(n, p)$ 分佈。試求給定 $X + Y = k$ 之下, $k \geq 1$, X 的條件 p.d.f.。給出 k 的範圍, 並指出此為那一常見的分佈, 參數為何?
18. 設 X_1, X_2 為二獨立的隨機變數, 且皆有 $\mathcal{U}(0, 1)$ 分佈。試證 $X_1 + X_2$ 並無均勻分佈。
19. 設 X_1, X_2 為二獨立的隨機變數, 且分別有 $\mathcal{E}(\lambda_1)$ 及 $\mathcal{E}(\lambda_2)$ 分佈, $\lambda_1 \neq \lambda_2$ 。試求 $X_1 + X_2$ 之分佈, 並問此是否為 gamma 分佈。
20. 設 (X, Y) 之聯合 p.d.f. 為 $f(x, y) = 2e^{-x-y}, 0 < x < y < \infty$ 。試求 (X, Y) 之聯合動差母函數。(解. $M(s, t) = 2((1-t)(2-s-t))^{-1}, t < 1, s+t < 2$)
21. 設 (X, Y) 之聯合 p.d.f. 為 $f(x, y) = e^{-y}, 0 < x < y < \infty$ 。試求 (X, Y) 之聯合動差母函數。(解. $M(s, t) = ((t-1)^{-1} - (s+t-1)^{-1})/s, t < 1, s+t < 1$)

22. 設 X 與 Y 獨立, 且 $X \sim \chi_m^2$, $Y \sim \chi_n^2$ 。試問當 $m > n$ 時, $X - Y$ 是否有卡方分佈, 若是則給出其自由度。(解. 否)
23. 設 X 與 Y 獨立, 令 $S = X + Y$ 。又知 $X \sim \chi_m^2$, $S \sim \chi_{m+n}^2$ 。試問 Y 是否有卡方分佈, 若是則給出其自由度。(解. 是, $Y \sim \chi_n^2$)
24. 試證(2.25)式所定義的函數為一p.d.f.。
25. 試證偏斜常態分佈的性質(i)-(vi)。
26. 試證由(2.26)式可得(2.27)式。
27. 設 f 為一連續型的p.d.f., 且為偶函數, G 為一連續型的分佈函數, 以 G' 為p.d.f., G' 亦為偶函數。試證對 $\forall \lambda \in R$, $2G(\lambda y)f(y)$, $-\infty < y < \infty$, 為一p.d.f.。

第3.3節

1. 試以變數代換的方法, 證明定理2.4之(ii)-(vii)。
2. 設隨機向量 (X, Y) 之聯合p.d.f.為 $f(x, y)$ 。令 $U = aX + bY$, $V = cX + dY$, 其中 $ad - bc \neq 0$ 。試證 (U, V) 之聯合p.d.f.為

$$f_{U,V}(u, v) = \frac{1}{|ad - bc|} f\left(\frac{du - bv}{ad - bc}, \frac{av - cu}{ad - bc}\right)。$$

3. 設 X 與 Y 獨立, 且皆有 $\mathcal{N}(\mu, 1)$ 分佈。試求 $(X - Y)^2/2$ 之p.d.f.。
4. 設 X 與 Y 獨立, 且分別有 $\text{Be}(\alpha + \beta, \gamma)$ 分佈及 $\text{Be}(\alpha, \beta)$ 分佈。試證 XY 有 $\text{Be}(\alpha, \beta + \gamma)$ 分佈。
5. 設 $R > 0$, 且 R^2 有 χ_2^2 分佈。又設 θ 與 R 獨立, 且有 $\mathcal{U}(0, 2\pi)$ 分佈。令 $X = R \cos \theta$, $Y = R \sin \theta$ 。試求 (X, Y) 之聯合分佈。
6. 設 X 與 Y 獨立, 且皆有 $\mathcal{N}(0, \sigma^2)$ 分佈。令 $U = \sin^{-1}(X/\sqrt{X^2 + Y^2})$ 。試證

(i) $X^2 + Y^2$ 與 $X/\sqrt{X^2 + Y^2}$ 獨立;

(ii) $X^2 + Y^2$ 與 X/Y 獨立;

- (iii) U 有 $\mathcal{U}(-\pi/2, \pi/2)$ 分佈;
- (iv) X/Y 有柯西分佈;
- (v) $X/|Y|$ 有柯西分佈;
- (vi) $X/(X+Y)$ 有柯西分佈。

7. 設 X 與 Y 獨立, 且分別有 $\mathcal{E}(\lambda)$ 及 $\mathcal{E}(\mu)$ 分佈。令

$$Z = \min\{X, Y\}, \quad W = \begin{cases} 1, & Z = X, \\ 0, & Z = Y. \end{cases}$$

- (i) 試求 Z, W 之聯合 p.d.f. (注意 Z 為連續型, W 為離散型);
 - (ii) 試證 $P(Z \leq z | W = i) = P(Z \leq z), i = 0, 1$, 即 Z 與 W 獨立。
8. 設 X 與 Y 獨立, 且分別有 $\Gamma(p, \lambda)$ 及 $\Gamma(q, \lambda)$ 分佈。試求 Y/X 之分佈。
9. 設 (X, Y) 為連續型的隨機向量, 其聯合 p.d.f. 為

$$f(x, y) = \frac{\Gamma(a+b+c)}{\Gamma(a)\Gamma(b)\Gamma(c)} x^{a-1} y^{b-1} (1-x-y)^{c-1},$$

$x, y \geq 0, x+y \leq 1$, 其中 a, b, c 為大於 0 之常數。 (X, Y) 稱為有二變數之黎瑞希里分佈 (bivariate Dirichlet distribution, Dirichlet (1805-1859), 德國數學家)。令 $U = X + Y, V = X/Y, W = (b/a)V$ 。

- (i) 試求 U, V 之聯合 p.d.f.;
 - (ii) 試求 U, W 之聯合 p.d.f., 並檢驗 U 與 W 是否獨立。
10. 設 X 與 Y 獨立, 且皆有 $\mathcal{C}(\theta, 1)$ 分佈。試證若 $a, b > 0$ 且 $a+b=1$, 則 $W = aX + bY$ 亦有 $\mathcal{C}(\theta, 1)$ 分佈。
11. 設 X 與 Y 獨立, 且皆有 $\mathcal{U}(\alpha, \beta)$ 分佈, $\alpha < \beta$ 。令 $Z = X + Y$ 。
- (i) 試求 Z 之分佈函數, 並得 Z 之 p.d.f.;
 - (ii) 試利用變數代換, 再求邊際分佈的方式, 而得 Z 之 p.d.f.。
12. 設 X 與 Y 獨立, 且皆有 $\mathcal{U}(\alpha, \alpha+1)$ 分佈, 其中 $\alpha \in R$ 為一常數。令 $U = X + Y, V = X - Y$ 。

(i) 試求 U 與 V 之聯合p.d.f.;

(ii) 試檢驗 U 與 V 是否獨立。

13. 試驗證例3.3中的 $f_U(u)$ 確為一p.d.f.。

14. 設 X, Y 之聯合p.d.f.為

$$f(x, y) = \frac{1}{\pi}, x^2 + y^2 \leq 1。$$

令 $Z = X^2 + Y^2$ 。試求 Z 之p.d.f.。

15. 設 X 與 Y 獨立, 且皆有 $\mathcal{N}(0, 1)$ 分佈。試求 $P(X - Y < 0, X + Y > 0)$ 。

16. 設 X 與 Y 獨立, 且皆有p.d.f. $f(x) = 1/x^2, x > 1$ 。令 $U = X/Y$, 試求 U 之p.d.f.。

17. 設 X 與 Y 獨立, 且皆有 $\mathcal{U}[0, 1]$ 分佈。令 $U = \max\{X, Y\}$, $V = \min\{X, Y\}$, $R = U - V$ 。

(i) 試證

$$F_{U,V}(u, v) = \begin{cases} u^2 & , 0 \leq u \leq \min\{v, 1\}, \\ 2vu - v^2 & , 0 \leq v \leq u \leq 1, \\ 2v - v^2 & , 0 \leq v \leq 1 < u < \infty, \\ 1 & , 1 \leq v \leq u < \infty, \\ 0 & , \text{其他。} \end{cases}$$

(ii) 試求 V, R 之聯合p.d.f., 並證明 (V, R) 有二變數之黎瑞希里分佈。

(iii) 試求 V, R 之邊際p.d.f., 又問 V 與 R 是否獨立?

18. 設 X 與 Y 獨立, 且皆有 $\mathcal{E}(1)$ 分佈。令 $X_1 = 1 - e^{-X}$, $Y_1 = 1 - e^{-Y}$, $U = \max\{X, Y\}$, $V = \min\{X, Y\}$ 。

(i) 試證 X_1 與 Y_1 獨立, 且皆有 $\mathcal{U}(0, 1)$ 分佈;

(ii) 利用 $\max\{X_1, Y_1\} = 1 - e^{-U}$, $\min\{X_1, Y_1\} = 1 - e^{-V}$, 試求 U, V 之聯合p.d.f., 及 U, V 之邊際p.d.f.。

第3.4節

1. 設 $Y|\Lambda$ 有 $P(\Lambda)$ 分佈, 而 Λ 有 $\Gamma(\alpha, \beta)$ 分佈。試證 Y 有負二項分佈, 並決定其參數。
2. 設 $Y|N$ 有 $\mathcal{B}(N, p)$ 分佈, $N|\Lambda$ 有 $P(\Lambda)$ 分佈, 且 Λ 有 $\Gamma(\alpha, \beta)$ 分佈, 試求 Y 之非條件分佈。
3. 設 $Y|P$ 有 $\mathcal{B}(n, P)$ 分佈, P 有 $\mathcal{B}e(\alpha, \beta)$ 分佈。

(i) 試證 Y 之非條件分佈為

$$P(Y = y) = \binom{n}{y} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y + \alpha)\Gamma(n - y + \beta)}{\Gamma(\alpha + \beta + n)}, y = 0, 1, \dots, n;$$

(ii) 試求 $E(Y)$ 及 $\text{Var}(Y)$ 。

4. 設 $\text{Var}(X) = E(\text{Var}(X|Y))$ 。試證存在常數 c , 使得 $P(E(X|Y) = c) = 1$ 。
5. 試證或否證若 $E(Y|X) = c$, 其中 c 若為常數, 導致 X 與 Y 獨立。
6. 設 $X|Y$ 有 $\mathcal{P}(Y)$ 分佈, 且 Y 亦有 $\mathcal{P}(\lambda)$ 分佈, $\lambda > 0$ 為一常數。試以此組 (X, Y) 驗證(4.2)式成立。
7. 設 $X|Y$ 有 $\mathcal{N}(Y, \sigma^2)$ 分佈, 且 Y 有 $\mathcal{N}(\mu, \tau^2)$ 分佈, 其中 σ^2, μ, τ^2 為常數。試以此組 (X, Y) 驗證(4.2)式成立。

第3.5節

1. (i) 設 $\sigma_X = 0$, 試證 $\text{Cov}(X, Y) = 0$;
(ii) 設 $\sigma_X = \sigma_Y$, 試證 $\text{Cov}(X + Y, X - Y) = 0$ 。
2. 在例5.1中, 試求 $\rho(X, Y)$, $\text{Cov}(2X + 3, -3Y + 4)$ 及 $\rho(2X + 3, -3Y + 4)$ 。
3. 設隨機變數 X, Y 滿足 $\text{Var}(X) = 1$, $\text{Var}(Y) = 2$, 且 $\rho(X, Y) = 1/2$ 。試求 $\text{Var}(X - 2Y)$ 。

4. 設一盒中有3紅球及2黑球, 自此盒中依序隨機地取兩個球(每次取出後不放回), 令 U, V 分別表所得之紅球數及黑球數。試求 $\rho(U, V)$ 。
5. 設一盒中分別為編號1, 2, 3之三個球, 依序隨機地取二球(每次取出後不放回)。令 X, Y 分別表兩次所得之號碼。試求 $\rho(X, Y)$ 。
6. 設有一公正的銅板, 一面寫數字1, 一面寫2。投擲兩次, 所得數字之和以 X 表之, 極大值以 Y 表之。試求 $\rho(X, Y)$ 。
7. 設 X 之p.d.f.為一偶函數, 且 $E(X^4) < \infty$ 。試證 X 與 X^2 無相關。
8. 設 X, Y 之聯合p.d.f.為

$$f(x, y) = p^2(1-p)^y, x, y = 0, 1, 2, \dots, \text{且 } x \leq y,$$

其中 $0 < p < 1$ 。

- (i) 試證 X, Y 之邊際分佈, 分別為自0開始之 $\mathcal{NB}(1, p)$ 及 $\mathcal{NB}(2, p)$ 分佈;
 - (ii) 試求 $\text{Cov}(X, Y)$ 及 $\rho(X, Y)$, 並問 X 與 Y 是否獨立?
9. 設 (X, Y) 之聯合p.d.f.如例2.8所給。試求 $\rho(X, Y)$ 。
 10. 設 X, Y 之聯合p.d.f.為 $f(x, y) = c, x^2 + y^2 \leq 1$ 。
 - (i) 試求 c 之值;
 - (ii) 試證 $\mu_X = \mu_Y = \rho(X, Y) = 0$;
 - (iii) 試證 X 與 Y 不獨立。
 11. 設 X, Y 之聯合p.d.f.為 $f(x, y) = 360xy^2(1-x-y), 0 \leq x, y \leq 1, x+y \leq 1$ 。試求 $\text{Cov}(X, Y)$, 並對 $\text{Cov}(X, Y)$ 之符號給一解釋。
 12. 設 X 與 Y 獨立, 且令 $E(X) = \mu_X, E(Y) = \mu_Y, \text{Var}(X) = \sigma_X^2, \text{Var}(Y) = \sigma_Y^2$ 。試以 μ_X, μ_Y, σ_X , 及 σ_Y 表示 $\rho(XY, Y)$ 。
 13. 設 X, Y, Z 為兩兩無相關之隨機變數, 期望值皆為 μ , 變異數皆為 σ^2 。試以 μ 及 σ^2 表示 $\text{Cov}(X+Y, Y+Z)$ 及 $\text{Cov}(X+Y, X-Y)$ 。

14. 試證定理5.2。
15. 試證系理5.5。
16. 設 X, Y 為二隨機變數, 且 $E(X) = E(Y) = 0, \text{Var}(X) = \text{Var}(Y) = 1$ 。試證
- $|E(XY)| \leq 1$;
 - $E(XY) = 1$, 若且唯若 $P(Y = X) = 1$;
 - $E(XY) = -1$, 若且唯若 $P(Y = -X) = 1$ 。
17. 設 X, Y 之聯合p.d.f.為 $f(x, y) = 24xy, x, y \geq 0, x + y \leq 1$ 。
- 試求 $\text{Cov}(X, Y)$ 及 $\rho(X, Y)$;
 - 試證 $E(Y|X = x)$ 為 $-x$ 之線性函數, 並驗證此時(5.22)式成立。
- (解. (i) $\text{Cov}(X, Y) = -2/75, \rho(X, Y) = -2/3$, (ii) $E(Y|X = x) = (2/3)(1 - x)$)
18. 試證(5.25)式。
19. 設 X 與 Y 獨立, 且 $\text{Var}(X) = 2, \text{Var}(Y) = 16$ 。試求 $\text{Var}(5X - Y)$ 及 $\text{Cov}(3X + Y, 2X + 3Y)$ 。
20. 設 X, Y 之聯合p.d.f.為 $f(x, y) = 4(x - xy), 0 < x, y < 1$ 。試求 $E(X), E(Y), \text{Var}(X), \text{Var}(Y), E(X - Y), E(X^2Y), \text{Cov}(X, Y), \rho(X, Y), \text{Var}(X - Y), \text{Cov}(X + Y, X + 2Y), E(Y|X = x)$ 。
21. 設 $Y = aX + b + \varepsilon$, 其中 a, b 為常數, $X \sim \mathcal{N}(\mu_X, \sigma_X^2), \varepsilon \sim \mathcal{N}(0, \sigma^2)$, 且設 X 與 ε 獨立。試求 $\text{Cov}(X, Y)$ 及 $\text{Cov}(Y, \varepsilon)$ 。(解. $a\sigma_X^2, \sigma^2$)

第3.6節

- 設 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。試證 $X|Y = y \sim \mathcal{N}(\mu_1 + \rho(\sigma_1/\sigma_2) \cdot (y - \mu_2), \sigma_1^2(1 - \rho^2))$ 。
- 試證(6.3)式成立。

3. 設 $(X, Y) \sim \mathcal{N}(1, 3, 1, 4, 0.1)$ 。試分別決定 $2X$, $3Y$, 及 $2X + 3Y$ 之分佈。
4. 設 $(X, Y) \sim \mathcal{N}(3, 2, 4, 1, -0.6)$ 。試求 $P(X \leq 4 | Y = 3)$ 及 $P(|Y - 1| \geq 1.5 | X = 2)$ 。
5. 設 $(X, Y) \sim \mathcal{N}(1, 2, 4, 9, 0.5)$ 。
 - (i) 試給出其聯合動差母函數;
 - (ii) 試決定 $(X + Y, 2X - Y)$ 之分佈。
6. 設 X 與 Y 獨立, 且 $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ 。試證 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, 0)$ 。
7. 設 Z_1 與 Z_2 獨立, 且有共同分佈。令 $W_1 = (Z_1 + Z_2)/\sqrt{2}$, $W_2 = (Z_1 - Z_2)/\sqrt{2}$ 。試對下述二情況, 求 (W_1, W_2) 之聯合分佈。並問 W_1 與 W_2 是否獨立。
 - (i) $Z_1 \sim \mathcal{N}(0, 1)$;
 - (ii) $Z_2 \sim \mathcal{N}(\mu, \sigma^2)$ 。
8. 試證二變數常態分佈之性質(vii)。
9. 設 X, Y 之聯合 p.d.f. 為

$$f(x, y) = \frac{1}{2\pi(1 - \rho^2)^{1/2}} \exp\left\{-\frac{1}{2(1 - \rho^2)}(x^2 - 2\rho xy + y^2)\right\}。$$

試證 $\rho(X, Y) = \rho$, 且 $\rho(X^2, Y^2) = \rho^2$ 。

10. 設 $(X, Y) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ 。試求 (X^2, Y^2) 之聯合動差母函數。
11. (i) 設 $(X, Y) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ 。試求

$$V = \frac{X^2 + Y^2 - 2\rho XY}{1 - \rho^2}$$

之動差母函數, 並證明 V 有 χ_2^2 分佈。

- (ii) 試證(i)中之 V 可表為

$$V = \frac{(X - \rho Y)^2}{1 - \rho^2} + Y^2,$$

且 $Z_1 = (X - \rho Y)/\sqrt{1 - \rho^2}$ 與 $Z_2 = Y$ 獨立, 並皆有 $\mathcal{N}(0, 1)$ 分佈。又利用此結果證明 V 有 χ_2^2 分佈。

12. (i) 設 $(Z_1, Z_2) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ 。令 $X = \mu_1 + \sigma_1 Z_1$, $Y = \mu_2 + \sigma_2 Z_2$, $\sigma_1, \sigma_2 \neq 0$ 。試證 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。
- (ii) 設 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。令 $Z_1 = (X - \mu_1)/\sigma_1$, $Z_2 = (Y - \mu_2)/\sigma_2$ 。試證 $(Z_1, Z_2) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ 。
13. 設 Z_1 與 Z_2 獨立, 且皆有 $\mathcal{N}(0, 1)$ 分佈。令

$$X = \sigma_1 \sqrt{1 - \rho^2} Z_1 + \rho \sigma_1 Z_2 + \mu_1,$$

$$Y = \sigma_2 Z_2 + \mu_2,$$

其中 $\mu_1, \mu_2 \in R$, $\sigma_1, \sigma_2 > 0$, $|\rho| < 1$ 。試證 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。

14. 設 (X, Y) 之聯合 p.d.f. 如 (6.11) 式所給。試證

- (i) $f(x, y)$ 確為一二變數之 p.d.f.;
- (ii) X, Y 皆以 $\mathcal{N}(0, 1)$ 為邊際分佈。

15. 設 (X, Y) 之聯合 p.d.f. 為

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2} (1 + xy e^{-(x^2+y^2-2)/2}), x, y \in R。$$

試證

- (i) $f(x, y)$ 確為一二變數之 p.d.f.;
- (ii) X, Y 皆以 $\mathcal{N}(0, 1)$ 為邊際 p.d.f.。

16. 設 X, Y 之聯合 p.d.f. 為

$$f(x, y) = \frac{1}{2} g_1(x, y) + \frac{1}{2} g_2(x, y),$$

其中

$$g_1(x, y) = \frac{1}{2\pi\sqrt{1-\rho_1^2}} \exp\left\{-\frac{1}{2} \frac{x^2 + y^2 - 2\rho_1 xy}{1 - \rho_1^2}\right\}, x, y \in R,$$

$$g_2(x, y) = \frac{1}{2\pi\sqrt{1-\rho_2^2}} \exp\left\{-\frac{1}{2} \frac{x^2 + y^2 - 2\rho_2 xy}{1 - \rho_2^2}\right\}, x, y \in R。$$

其中 $|\rho_1| < 1$, $|\rho_2| < 1$, 且 $\rho_1 \neq \rho_2$ 。顯然 $f(x, y)$ 無法表示成如 (6.1) 式之型式。但試證

- (i) $f(x, y)$ 確為一二變數之 p.d.f.;
- (ii) X, Y 皆以 $\mathcal{N}(0, 1)$ 為邊際分佈。

17. 設 X 與 Y 獨立, 且皆有 $\mathcal{N}(0, 1)$ 分佈。令

$$Z = \begin{cases} X & , \text{若 } XY > 0, \\ 0 & , \text{若 } XY = 0, \\ -X & , \text{若 } XY < 0. \end{cases}$$

- (i) 試證 Z 有常態分佈;
- (ii) 試證 (Z, Y) 並非二變數之常態分佈。(提示: 證明 Z 與 Y 之符號恆相同)

18. 設 $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$, 且 X 與 Y 獨立。試決定 $aX + bY$ 與 $cX - dY$ 獨立之充要條件, 其中 a, b, c, d 為常數。(解. $ac\sigma_1^2 = bd\sigma_2^2$)

19. 設 $(X, Y) \sim \mathcal{N}(0, 0, 1, 1, \rho)$ 。試證 $X + Y \stackrel{d}{=} X - Y \sim \mathcal{N}(0, 2(1 - \rho))$, 且 $X + Y$ 與 $X - Y$ 獨立。

20. 設 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ 。令 $Z_1 = (X - \mu_1)/\sigma_1$, $Z_2 = (Y - \mu_2)/\sigma_2$ 。 $U_1 = Z_1 + Z_2$, $U_2 = Z_1 - Z_2$ 。

- (i) 試決定 Z_1, Z_2 之聯合分佈及邊際分佈;
- (ii) 試決定 U_1, U_2 之聯合分佈及邊際分佈, 並指出 U_1 與 U_2 是否獨立。

21. 設 $(X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma^2, \sigma^2, \rho)$ 。令 $V_1 = X + Y$, $V_2 = X - Y$ 。試決定 V_1, V_2 之聯合分佈, 並驗證 V_1 與 V_2 獨立。

第3.7節

1. 在例7.2中, 試證 $f_{X|Y,Z}(x|y, z) = f_{X|Z}(x|z)$ 。

2. 設 X, Y, Z 之聯合 p.d.f. 為 $f(x, y, z) = 6z^2 e^{-z(x+2y+3)}$, $x, y, z \geq 0$ 。試求下述各 p.d.f.: $f_X, f_Y, f_Z, f_{X,Y}, f_{X,Z}, f_{Y,Z}, f_{X,Y|Z}, f_{X|Z}, f_{Y|Z}$ 。並驗證 $f_{X,Y|Z}(x, y|z)$ 與 $f_{X|Z}(x|z)f_{Y|Z}(y|z)$, $x, y, z \geq 0$, 是否相等。

3. 設 X, Y 之聯合 p.d.f. 為 $f(x, y, z) = c$, $0 < x < y < z < 1$ 。試求 (i) c 之值, (ii) $f_{Z|X,Y}(z|x, y)$, (iii) $f_{Y|X,Z}(y|x, z)$, (iv) $f_{X|Y,Z}(x|y, z)$, (v) $f_{X,Y|Z}(x, y|z)$ 。 (解. (i) 6, (ii) $1/(1-y)$, $0 < x < y < z < 1$, (iii) $1/(z-x)$, $0 < x < y < z < 1$, (iv) $1/y$, $0 < x < y < z < 1$, (v) $2/z^2$, $0 < x < y < z < 1$)

4. 設 (X_1, X_2, X_3, X_4) 之聯合 p.d.f. 為

$$f(x_1, x_2, x_3, x_4) = \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2), 0 < x_i < 1, i = 1, 2, 3, 4。$$

(i) 試證 f 確為一 p.d.f.;

(ii) 試求 $P(X_1 < 1/2, X_2 < 3/4, X_4 > 1/2)$;

(iii) 試求 $E(X_1 X_2)$;

(iv) 試求 $f(x_3, x_4|x_1, x_2)$;

(v) 試求 $P(X_3 > 3/4, X_4 < 1/2|X_1 = 1/3, X_2 = 2/3)$ 。

5. 設 X, Y, Z 之聯合 p.d.f. 為 $f(x, y, z) = (x+2y+z)/2$, $x, y, z \in (0, 1)$ 。試求

(i) $f_{X|Y,Z}(x|1/2, 1/2)$;

(ii) $f_{X,Y|Z}(x, y|1/2)$;

(iii) $P(X < 1/4|Y = 1/2, Z = 1/2)$;

(iv) $E(X^2|Z = 1/2)$ 。

(解. (i) $x/2 + 3/4$, $x \in (0, 1)$, (ii) $x/2 + y + 1/4$, $x, y \in (0, 1)$, (iii) $13/64$, (iv) $3/8$)

6. 設 X, Y, Z 為相互獨立的隨機變數, 且皆有 $\mathcal{N}(0, 1)$ 分佈。試求

- (i) $P(X/Y \leq t)$ 及 $P(XY \leq t)$;
- (ii) $P(XY/Z \leq t)$ 。
7. 假設射飛標所中的點之座標為 (X, Y) , 設 X 與 Y 獨立, 且皆有 $\mathcal{N}(0, 1)$ 分佈。若獨立地射兩次, 試求兩次位置距離之分佈。
8. 設 A, B, C 為相互獨立的隨機變數, 且皆有 $\mathcal{U}(0, 1)$ 分佈。試求方程式 $Ax^2 + Bx + C = 0$ 有實根之機率。(提示: 若 X 有 $\mathcal{U}(0, 1)$ 分佈, 則 $-\log X$ 有指數分佈, 而二獨立的 gamma 隨機變數之和仍有 gamma 分佈)
9. 設 $X_1, \dots, X_n$ 為相互獨立的隨機變數, 且皆有 $\mathcal{U}(0, 1)$ 分佈。令 $Z = \prod_{i=1}^n X_i$ 。試求 Z 之 p.d.f.。
10. 試將定理 2.4 (除了 (iii) 之外), 推廣到 n 個隨機變數的情況。
11. 試證定理 7.5。
12. 對 X, Y, Z 三隨機變數, 假設底下提到的各量(期望值、變異數及共變異數等)皆存在。試證
- (i) X 與 $Y - E(Y|X)$ 為無相關;
- (ii) $\text{Var}(Y - E(Y|X)) = E(\text{Var}(Y|X))$;
- (iii) $\text{Cov}(X, Y) = 0 \Rightarrow E(\text{Cov}(X, Y|Z)) = -\text{Cov}(E(X|Z), E(Y|Z))$;
- (iv) $\text{Cov}(Z, E(Y|Z)) = \text{Cov}(Z, Y)$ 。
13. 設 $(X_1, X_2, X_3, X_4, X_5)$ 有 $\mathcal{M}(n, p_1, p_2, p_3, p_4, p_5)$ 分佈。試證
- (i) $X_3, X_4, X_5 | X_1 = x_1, X_2 = x_2$ 有 $\mathcal{M}(n - x_1 - x_2, p_3/(p_3 + p_4 + p_5), p_4/(p_3 + p_4 + p_5), p_5/(p_3 + p_4 + p_5))$ 分佈;
- (ii) $X_3, X_4, X_5 | X_1 + X_2 = x$ 有 $\mathcal{M}(n - x, p_3/(p_3 + p_4 + p_5), p_4/(p_3 + p_4 + p_5), p_5/(p_3 + p_4 + p_5))$ 分佈。
14. 設 X_1, X_2, X_3 相互獨立且分別有 $\mathcal{P}(\lambda_1), \mathcal{P}(\lambda_2), \mathcal{P}(\lambda_3)$ 分佈。令 $Y = X_1 + X_2 + X_3, \lambda = \lambda_1 + \lambda_2 + \lambda_3$ 。對任一正整數 n , 試證 $X_1, X_2, X_3 | Y = n$ 有 $\mathcal{M}(n, \lambda_1/\lambda, \lambda_2/\lambda, \lambda_3/\lambda)$ 分佈。

15. 設 X, Y, Z 之聯合 p.d.f. 爲

$$f(x, y, z) = 0.25, (x, y, z) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}。$$

(i) 試求 $f_{X,Y}, f_{X,Z}, f_{Y,Z}$, 並證明 X 與 Y, X 與 Z, Y 與 Z 皆每對獨立;

(ii) 試證 X, Y, Z 不相互獨立;

(iii) 試求 $f_{X|Y}(x|1)$, 及 $f_{X,Y|Z}(x, y|1)$ 。

16. 設 (X_1, X_2, X_3, X_4) 之聯合 p.d.f. 爲

$$f(x_1, x_2, x_3, x_4) = 24e^{-x_1-x_2-x_3-x_4}, 0 < x_1 < x_2 < x_3 < x_4。$$

令 $U_1 = X_1, U_2 = X_2 - X_1, U_3 = X_3 - X_2, U_4 = X_4 - X_3$ 。試求 (U_1, U_2, U_3, U_4) 之聯合 p.d.f.。

17. 設 $X_1, \dots, X_{n+1}$ 爲相互獨立的隨機變數, 且皆有 $\mathcal{E}(1)$ 分佈。令 $U_m = \sum_{i=1}^m X_i, 1 \leq m \leq n+1$ 。

(i) 試證 $\mathbf{T} = (X_1/U_{n+1}, \dots, X_n/U_{n+1})$ 之聯合 p.d.f. 爲

$$f_{\mathbf{T}}(t_1, \dots, t_n) = n!, \forall t_i > 0, \text{ 且 } \sum_{i=1}^n t_i < 1;$$

(提示: 先求 $(X_1/U_{n+1}, \dots, X_n/U_{n+1}, U_{n+1})$ 之聯合 p.d.f.)

(ii) 試證 $\mathbf{U} = (U_1/U_{n+1}, \dots, U_n/U_{n+1})$ 之聯合 p.d.f. 爲

$$f_{\mathbf{U}}(u_1, \dots, u_n) = n!, 0 < u_1 < u_2 < \dots < u_n < 1。$$

18. 設 $\mathbf{X} = (X_1, \dots, X_k)$ 有 $\mathcal{M}(n, p_1, \dots, p_k)$ 分佈。試證 $E(X_i) = np_i$, $\text{Var}(X_i) = np_i(1 - p_i), i = 1, \dots, k$, 且 $\text{Cov}(X_i, X_j) = -np_i p_j, i, j = 1, \dots, k, i \neq j$ 。

19. 在例 7.9 中, 試求 $P(N = k), k = n, n-1, n-2, n-3, n-4$ 。

20. 在例 7.9 中, 試以 (7.17) 式求 $\text{Var}(N)$ 。

21. 試證 (7.20) 式。

22. 將 n 個球隨機地放進 r 個箱子中, $r \geq 2$ 。令 $X_i = 1$ 表第 i 個箱子是空的, 否則令 $X_i = 0, i = 1, \dots, r$ 。又令 U_r 表總共之空箱子數。試求
- (i) $E(X_i)$, (ii) $\text{Var}(X_i)$, (iii) $E(X_i X_j)$, (iv) $\text{Cov}(X_i, X_j)$, (v) $\rho(X_i, X_j)$, (vi) $E(U_r)$, (vii) $\text{Var}(U_r)$, 其中 $i \neq j$ 。
23. 在例7.11中, 試求 $\text{Var}(X_0 + X_1 + \dots + X_{N-1})$ 。(解. $\sum_{n=0}^{N-1} Nn/(N-n)^2$)
24. 在例7.11中, 試求 $\text{Var}(Z)$ 。

第3.8節

1. 設 X 有 $\mathcal{E}(2)$ 分佈。試求 $P(|X| \geq 1)$ 之值, 並以柴比雪夫不等式給出此機率之一上界。
2. 設 X 有 $\mathcal{N}(0, 1)$ 分佈。試求出 $P(|X| \geq 2)$ 之值, 並以柴比雪夫不等式給出此機率之一上界。
3. 試分別對 X 有(i) $\mathcal{U}(0, 1)$ 分佈, (ii) $\mathcal{E}(\lambda)$ 分佈, 計算 $P(|X - \mu| \geq k\sigma)$ 並與柴比雪夫不等式所給之上界比較。
4. 某診所每天會來掛號之病人數有 $\mathcal{P}(50)$ 分佈。試利用柴比雪夫不等式, 求某日來掛號之病人數做介於40與60間之機率的一下界。
5. 設 X 有 $\mathcal{E}(2)$ 分佈。試對 $r = 2, s = 3$, 驗證(8.6)式成立。
6. 設 X 為一取正值之隨機變數。若 $E(\log X) = 0$, 試判定 $E(X)$ 與1之大小。
7. 試由定理5.4, 得到數值之柯西-史瓦茲不等式。
8. 設 $P((X, Y) = (a_i, b_i)) = 1/n, i = 1, \dots, n$ 。試分別求出 X, Y 之邊緣p.d.f., 並證明此時赫德爾不等式導致(8.8)式。
9. 設由(8.7)式導出(8.9)式。

參考文獻

1. 黃文章(1999). 數學欣賞。華泰文化事業股份有限公司, 台北。
2. Azzalini, A.(1985). A class of distributions which includes the normal ones. *Scandinavia Journal of Statistics*, 12, 171-178.
3. Anderson, T.W. (1984). *An Introduction to Multivariate Statistical Analysis*, 2nd ed. John Wiley & Sons, New York.
4. Broffitt, J.D. (1986). Zero correlation, independence, and normality. *The American Statistician*, 40, 276-277.
5. Lukacs, E. (1955). A characterization of the gamma distribution. *Annals of Mathematical Statistics*, 26, 319-324.
6. Roussas, G.G. (1997). *A Course in Mathematical Statistics*, 2nd ed. Academic Press, New York.