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)，為自樣本空間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∀ω ∈ Ω，有一二維的點(x, y)與其對應，x, y ∈ R，则定義出一個二維的隨機向量。

例3.1 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。對一様本點ω = (i, j) ∈ Ω，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 \( \Omega \) 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 \( X = 7 \) 與 \( Y = 1 \) 之機率為何？即要求 \( P((X,Y) = (7,1)) \)，或是 \( P(X = 7 \text{且} Y = 1) \)。為了簡便，我們常將後者寫成 \( P(X = 7, Y = 1) \)，即以逗號“，”表“且”。

不難驗證 \( \Omega \) 中，使 \( X = 7 \text{且} Y = 1 \) 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為 \( 1/36 \)。故

\[ P(X = 7, Y = 1) = P(\{(3,4),(4,3)\}) = \frac{2}{36} = \frac{1}{18}. \]

另外，

\[ P(X = 7, Y \leq 4) \]
\[ = P(\{(4,3),(3,4),(5,2),(2,5)\}) = \frac{4}{36} = \frac{1}{9}. \]

表1.1 例1.1中 \( (X,Y) \) 之聯合p.d.f.及邊際p.d.f.

<table>
<thead>
<tr>
<th>( y ) ( x )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>( f_Y(y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>( \frac{1}{36} )</td>
<td>0</td>
<td>( \frac{1}{36} )</td>
<td>0</td>
<td>( \frac{1}{36} )</td>
<td>0</td>
<td>( \frac{1}{36} )</td>
<td>0</td>
<td>( \frac{1}{36} )</td>
<td>0</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{6} )</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{5}{18} )</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>( \frac{2}{9} )</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
<td>0</td>
<td>( \frac{1}{18} )</td>
</tr>
<tr>
<td>( f_X(x) )</td>
<td>( \frac{1}{36} )</td>
<td>( \frac{1}{18} )</td>
<td>( \frac{1}{12} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{1}{6} )</td>
<td>( \frac{5}{36} )</td>
<td>( \frac{1}{5} )</td>
<td>( \frac{1}{12} )</td>
<td>( \frac{1}{18} )</td>
<td>( \frac{1}{36} )</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
3.1 聯合及邊際分佈 163

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）。對離散型的隨機變量，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 \((X, Y)\) 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 \(\forall A \subset \mathbb{R}^2\)，

\[ P((x, y) \in A) = \sum_{(x, y) \in A} f(x, y). \]

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 \(\mathbb{R}^2\) 上一長方形，(1.1)式的右側，只需是一可數個 \(f(x, y)\) 的和。例如，取 \(A = \{(x, y) | x = 7, y \leq 4\}\)。這是 \(\mathbb{R}^2\) 中一準線，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。若求出 \((X, Y)\) 之聯合p.d.f.後，在求關於 \((X, Y)\) 之機率時，通常會對回到樣本空間上（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 \(C \subset \mathbb{R}^2\)，使得

\[ \sum_{(x, y) \in C} f(x, y) = 1. \]

反之，若 \(f(x, y)\) 為一由 \(\mathbb{R}^2\) 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，可以有不同的聯合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，或稱邊際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) \[ f_X(x) = \sum_{y \in R} f_{X,Y}(x,y), \]

(1.4) \[ f_Y(y) = \sum_{x \in R} f_{X,Y}(x,y). \]

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

\[ f_X(x) = P(X = x) \]
\[ = P(X = x, -\infty < Y < \infty) \]
\[ = P((X,Y) \in \{(x,y)|-\infty < y < \infty\}) \]
\[ = \sum_{y \in R} f_{X,Y}(x,y). \]

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

(1.6) \[ \sum_{y \in R} f_Y(y) = \sum_{y \in R} \sum_{x \in R} f_{X,Y}(x,y) = \sum_{(x,y) \in R^2} f(x,y) = 1. \]
因只是對所有可能的值相加，為了簡便，常以

\[
\sum_x f_X(x), \ \sum_y f_Y(y), \ \sum_{x,y} f(x,y)
\]

分別取代

\[
\sum_{x \in R} f_X(x), \ \sum_{y \in R} f_Y(y), \ \sum_{(x,y) \in \mathbb{R}^2} f(x,y).
\]

例1.2 利用定理1.1，可求出例1.1中X,Y之邊際p.d.f.: 

\[
f_X(2) = \sum_{j=0}^{5} f_{X,Y}(2,j) = f_{X,Y}(2,0) = \frac{1}{36},
\]

\[
f_X(3) = \sum_{j=0}^{5} f_{X,Y}(3,j) = f_{X,Y}(3,1) = \frac{1}{18},
\]

\[
\vdots
\]

\[
f_Y(0) = \sum_{i=2}^{12} f_{X,Y}(i,0) = \frac{1}{6},
\]

\[
\vdots
\]

我們將X,Y之邊際p.d.f.亦給在表1.1。可驗證

\[
\sum_x f_X(x) = \sum_{x=2}^{12} f_X(x) = 1,
\]

\[
\sum_y f_Y(y) = \sum_{y=0}^{5} f_Y(y) = 1.
\]

由(X,Y)之邊緣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。

再來我們考慮連続型的隨機向量。一由R^2映至R的非負函數f(x,y)，
若滿足對∀A ⊂ R^2，

\[
P((X, Y) \in A) = \int_A f(x,y)dxdy,
\]

(1.7)
便稱為連續型隨機變量$(X,Y)$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，任一隨變數的非負函數$f(x,y)$，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變量$(X,Y)$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$X \overset{d}{=} U, Y \overset{d}{=} V$。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.:

\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy \\
= \int_{0}^{1} \int_{0}^{1-y} 24xy \, dx \, dy \\
= \int_{0}^{1} 12x^2y \left[ 1-y \right]_0 \, dy \\
= \int_{0}^{1} 12y(1-y)^2 \, dy \\
= 12 \left( \frac{1}{2}y^2 - \frac{2}{3}y^3 + \frac{1}{4}y^4 \right)_0^1 \\
= 1. 
\]

其次\( X \)之邊際p.d.f.為

\[
f_X(x) = \int_{0}^{1-x} f(x, y) \, dy \\
= \int_{0}^{1-x} 24xy \, dy \\
= 12x(1-x)^2, 0 \leq x \leq 1. 
\]

故

\[
P(X \leq \frac{1}{2}) = \int_{0}^{\frac{1}{2}} 12x(1-x)^2 \, dx = \frac{11}{16}. 
\]

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：

\[
\int_{0}^{\frac{1}{2}} \int_{0}^{1-x} 24xy \, dy \, dx = \frac{11}{16}. 
\]
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 ≤ 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：

\[ f(x, y) = e^{-y}I_{0 < y > \infty}. \]

底下求事件X + Y ≥ 1之機率。由圖可看出X + Y ≤ 1的機率較好求。又因這是連續型的隨機變數，故X + Y ≤ 1與X + Y < 1之機率相同。

由上述說明得

\[ P(X + Y \geq 1) = 1 - P(X + Y < 1) \]

\[ = 1 - P(X + Y \leq 1) \]
\[
F(x, y) = P(X \leq x, Y \leq y).
\]

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

\[
F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v) \, dv \, du.
\]

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

\[
\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y).
\]

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

\[
E(g(X, Y)) = \sum_{x, y} g(x, y) f(x, y),
\]

\[
E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) \, dx \, dy.
\]
例1.5 承例1.3。我們有

\[ E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy \cdot f(x, y) dx dy \]
\[ = \int_{0}^{1} \int_{0}^{1-y} 24x^2y^2 dx dy \]
\[ = \int_{0}^{1} 8y^2(1 - y)^3 dy \]
\[ = \left( \frac{8}{3} y^3 - 6y^4 + \frac{24}{5} y^5 - \frac{4}{3} y^6 \right) \bigg|_0^1 \]
\[ = \frac{8}{3} - 6 + \frac{24}{5} - \frac{4}{3} \]
\[ = \frac{2}{15} \]

另外，

\[ E(X) = \int_{0}^{1} \int_{0}^{1-y} x \cdot f(x, y) dx dy = \int_{0}^{1} \int_{0}^{1-y} 24x^2y dx dy = \frac{2}{5} \]

\[ E(X) \]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

\[
1 = \int_{0}^{2} \int_{0}^{x} f(x, y) dy dx
= \int_{0}^{2} \int_{0}^{x} c dy dx
= \int_{0}^{2} cxdx = 2c.
\]

\[
y = x
\]

\[
O \quad 2
\]

\[
x = y
\]
得$c = \frac{1}{2}$。

(ii)  
\[
P(X > Y + 1) = \int_{1}^{2} \int_{0}^{x-1} \frac{1}{2} dy \, dx \\
= \frac{1}{2} \int_{1}^{2} (x-1) \, dx \\
= \frac{1}{2} \left[ \frac{1}{2}x^2 - x \right]_{1}^{2} = \frac{1}{4}.
\]

(iii) 設$0 < x < 2$ 之
\[
P(X > x) = \int_{x}^{2} \int_{0}^{u} \frac{1}{2} dy \, du \\
= \int_{x}^{2} \frac{1}{2} u \, du \\
= \frac{1}{4} \left[ u^2 \right]_{x}^{2} = 1 - \frac{1}{4}x^2.
\]

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$(X,Y)$為離散型的隨機變量，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
\[
P(Y = y|X = x) = \frac{P(Y = y, X = x)}{P(X = x)}.
\]

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 \(f_Y(x|y)\) 表在給定 \(X = x\) 之下，\(Y\) 的條件 \(p.d.f.\)。也常簡化地以 \(Y|X\) 之分佈，或 \(X|Y\) 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：

<table>
<thead>
<tr>
<th>(y) (x)</th>
<th>1</th>
<th>2</th>
<th>(f_Y(y))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(\frac{1}{8})</td>
<td>(\frac{1}{4})</td>
<td>(\frac{3}{8})</td>
</tr>
<tr>
<td>2</td>
<td>(\frac{2}{8})</td>
<td>(\frac{1}{4})</td>
<td>(\frac{5}{8})</td>
</tr>
<tr>
<td>(f_X(x))</td>
<td>(\frac{1}{16})</td>
<td>(\frac{5}{16})</td>
<td>1</td>
</tr>
</tbody>
</table>

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

\[
f_Y|X(1)|_1 = \frac{f(1,1)}{f_X(1)} = \frac{1/8}{11/16} = \frac{2}{11},
\]

\[
f_Y|X(2)|_1 = \frac{f(1,2)}{f_X(1)} = \frac{9/16}{11/16} = \frac{9}{11},
\]

\[
f_Y|X(1)|_2 = \frac{f(2,1)}{f_X(2)} = \frac{1/4}{5/16} = \frac{4}{5},
\]

\[
f_Y|X(2)|_2 = \frac{f(2,2)}{f_X(2)} = \frac{1/16}{5/16} = \frac{1}{5}.
\]
3.2 條件分佈及獨立性 173

至於 \( f_{X|Y}(x|y) \) 留給各位自行求出。可看出

\[
\sum_y f_{Y|X}(y|1) = f_{Y|X}(1|1) + f_{Y|X}(2|1) = \frac{2}{11} + \frac{9}{11} = 1,
\]

\[
\sum_y f_{Y|X}(y|2) = f_{Y|X}(1|2) + f_{Y|X}(2|2) = \frac{4}{5} + \frac{1}{5} = 1.
\]

但

\[
\sum_x f_{Y|X}(1|x) = f_{Y|X}(1|1) + f_{Y|X}(1|2) = \frac{2}{11} + \frac{4}{5} \neq 1,
\]

\[
\sum_x f_{Y|X}(2|x) = f_{Y|X}(2|1) + f_{Y|X}(2|2) = \frac{9}{11} + \frac{1}{5} \neq 1.
\]

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，

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

若 \((X, Y)\) 為連続型的隨機向量, 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 \((X, Y)\) 為連続型的隨機向量, 以 \( f(x, y) \) 爲聯合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 \( f(x | y) \) 皆為p.d.f.
例2.2 承例1.3，可求出

\[
f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{24xy}{12x(1-x)^2} = \frac{2y}{(1-x)^2}, 0 \leq y \leq 1-x, 0 < x < 1.
\]

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，\(y \neq 0,1\)。又可验证 \(f(y|x)\) 及 \(f(x|y)\) 都是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）

\[
P(Y \leq \frac{1}{2}|x) = \int_0^{\min(1/2,1-x)} f(y|x)dy
\]

\[
= \int_0^{\min(1/2,1-x)} \frac{2y}{(1-x)^2}dy
\]

\[
= \frac{(\min(1/2,1-x))^2}{(1-x)^2}
\]

\[
= \begin{cases} 
\frac{1}{4(1-x)^2}, & 0 < x \leq \frac{1}{2}, \\
\frac{1}{4(1-x)^2} = 1, & \frac{1}{2} < x < 1.
\end{cases}
\]

條件p.d.f.亦可用來求條件期望值。對一函数 \(g\)，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

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

\[(2.8)\quad E(E(Y|x)|x) = E(Y|x),\]
\[(2.9)\quad E(Y E(Y|x)|x) = E(Y|x) \cdot E(Y|x) = (E(Y|x))^2.\]

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** 拟例2.1。則

\[E(Y|X = 1) = \sum_y yf(y|1) = 1 \cdot \frac{2}{11} + 2 \cdot \frac{9}{11} = \frac{20}{11},\]
\[E(Y|X = 2) = \sum_y yf(y|2) = 1 \cdot \frac{4}{5} + 2 \cdot \frac{1}{5} = \frac{6}{5}.\]

**例2.4** 拟例2.2。

\[E(Y|X = x) = \int_0^{1-x} yf(y|x)dy = \int_0^{1-x} \frac{2y^2}{(1-x)^2}dy = \frac{2(1-x)^3}{3(1-x)^2} = \frac{2}{3}(1-x), 0 < x < 1.\]

又

\[E(Y^2|X = x) = \int_0^{1-x} y^2f(y|x)dy = \frac{1}{(1-x)^2} \cdot \frac{1}{2}(1-x)^4 = \frac{1}{2}(1-x)^2, 0 < x < 1.\]

因此
\[
E(2Y^2 + 3Y|X = x) \\
= 2E(Y^2|X = x) + 3E(Y|X = x) \\
= (1 - x)^2 + 2(1 - x) = (1 - x)(3 - x), 0 < x < 1.
\]

在给定 \(X = x\)之下，除了可求条件期望值 \(E(Y|X = x)\)，亦可求条件变

异数 \(\text{Var}(Y|X = x)\)，或以 \(\text{Var}(Y|x)\) 表之，其定义为

\[
(2.10) \quad \text{Var}(Y|x) = E((Y - E(Y|x))^2|x).
\]

以上式来定义条件变异数是有道理的。变异数是量测一变数（此处为 \(Y\)）与

其期望值（此处为 \(E(Y|x)\）之離差平方（此处为 \((Y - E(Y|x))^2\)）的期望值（此

处为给定 \(X = x\)之下的期望值），故得(2.10)式。由(2.10)式立即导致（证明

留在习题第3题）

\[
(2.11) \quad \text{Var}(Y|x) = E(Y^2|x) - (E(Y|x))^2.
\]

例2.5 承例1.3, 2.2, 及2.4。可得

\[
\text{Var}(Y|x) = E(Y^2|x) - (E(Y|x))^2 \\
= \frac{1}{2}(1 - x)^2 - \frac{4}{9}(1 - x)^2 \\
= \frac{1}{18}(1 - x)^2, 0 < x < 1.
\]

变异数与 \(x\) 有关。当 \(x \rightarrow 1\) 時，变异数 \(\rightarrow 0\)，這是合理的，由例1.3中的图形

可看出，当 \(x \rightarrow 1\) 時，\(y \rightarrow 0\)，变异数愈来愈小，变异数自然也愈来愈小。甚至

因 \(x\) 愈大時，\(y\)的活动範圍愈小 \((0 \leq y \leq 1 - x)\)，变异数也就愈小。事實上，我

们所求出的 \(\text{Var}(Y|x)\) 在 \(0 < x < 1\) 间，的確为 \(x\) 的消減函数。

例2.6 设 \(X, Y\) 之联合p.d.f.为

\[
f(x, y) = \frac{1}{64}(x + y), 0 \leq x, y \leq 4.
\]

试求(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) 依定义，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
\[
E(Y|x) = \int_0^4 y \cdot f(y|x)dy \\
= \int_0^4 y \cdot \frac{x + y}{4x + 8}dy \\
= \frac{1}{4x + 8} \int_0^4 (xy + y^2)dy \\
= \frac{8x + 64/3}{4x + 8} = \frac{6x + 16}{3x + 6}, 0 \leq x \leq 4.
\]

(iv) 先求出
\[
E(Y^2|x) = \int_0^4 y^2 \cdot f(y|x)dx \\
= \int_0^4 y^2 \cdot \frac{x + y}{4x + 8}dy \\
= \frac{1}{4x + 8} \int_0^4 (xy^2 + y^3)dy \\
= \frac{64x/3 + 64}{4x + 8} = \frac{16x + 48}{3x + 6}, 0 \leq x \leq 4.
\]

因此
\[
\text{Var}(Y|x) = E(Y^2|x) - (E(Y|x))^2 \\
= \frac{16x + 48}{3x + 6} - \left(\frac{6x + 16}{3x + 6}\right)^2 \\
= \frac{12x^2 + 48x + 32}{(3x + 6)^2}, 0 \leq x \leq 4.
\]

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
$$f(x,y) = \begin{cases} 
\frac{1}{16}, & 0 < x < 8, 0 < y < 4 - x/2, \\
0, & \text{其他。} 
\end{cases}$$

由此得

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

则

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

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

$$E(Y|x) = \frac{1}{2} \left( 4 - \frac{x}{2} \right) = 2 - \frac{x}{4}, 0 < x < 8,$$

$$\operatorname{Var}(Y|x) = \frac{1}{12} \left( 4 - \frac{x}{2} \right)^2 = \frac{(8-x)^2}{48}, 0 < x < 8.$$ 

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

**注2.1** 設 $A$ 爲平面上一區域，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 $\operatorname{Var}(Y|x)$。

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
\[ f(y|x) = \frac{x^3 e^{-x[(1+y)/2]}}{x^2 e^{-x/2}} = xe^{-xy}, y > 0, x > 0. \]

即 \( Y | X = x \) 有 \( \mathcal{E}(x) \) 分布。因此

\[ E(Y|x) = \frac{1}{x}, \ Var(Y|x) = \frac{1}{x^2}, x > 0. \]

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

例2.9 设有一函数

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

(i) 試證 \( f \) 為一二维之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 \( \Var(Y|x) \)。

解. (i) 對 \( 0 < x, y < 1, \)

\[ 4xy - 2x - 2y + 2 = 2xy + 2(1 - x)(1 - y) > 0, \]

即 \( f \) 為非負。又

\[ \begin{align*}
\int_0^1 \int_0^1 f(x, y) dy dx &= \int_0^1 \int_0^1 (4xy - 2x - 2y + 2) dy dx \\
&= 1 - 1 - 1 + 2 = 1,
\end{align*} \]

故 \( f \) 為一二维之p.d.f。{}

(ii) 首先

\[ f_X(x) = \int_0^1 (4xy - 2x - 2y + 2) dy \]

\[ = 2x - 2x - 1 + 2 = 1, 0 < x < 1. \]

因此

\[ f(y|x) = 4xy - 2x - 2y + 2, 0 < y < 1, 0 < x < 1. \]
利用上述条件p.d.f., 即得

\[ E(Y|x) = \int_{0}^{1} y(4xy - 2x - 2y + 2) dy = \frac{1}{3}(1 + x), 0 < x < 1, \]

\[ \text{Var}(Y|x) = \int_{0}^{1} y^2(4xy - 2x - 2y + 2) dy - \left( \frac{1}{3}(1 + x) \right)^2 \]

\[ = \frac{1}{6}(1 + 2x) - \frac{1}{9}(1 + x)^2 \]

\[ = \frac{1}{18}(1 + 2x - 2x^2), 0 < x < 1. \]

讀者可否看出X,Y之邊際分佈皆為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的分佈可能會隨著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B(x,p)分佈, 則可以說Y|X有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) \[ f(x,y) = f_X(x)f_Y(y), \forall x,y \in R, \]

便稱X與Y獨立。
3.2 條件分佈及獨立性 181

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

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

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

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

\[
f_Y(y) = \frac{3}{(1+y)^3}, y > 0.
\]

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. 以及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∀x, y ∈ R，

\[
    f(x, y) = g(x)h(y) \quad (2.16)
\]

之成立。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“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

\[
    ab = \int_{-\infty}^{\infty} g(x)dx \cdot \int_{-\infty}^{\infty} h(y)dy
\]

\[
    = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)dxdy
\]

\[
    = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy
\]

\[
    = 1,
\]

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

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

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

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。
3.2 條件分佈及獨立性 183

下述係理為定理2.1之一立即的推論。

**定理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

\[
F(x,y) = F_X(x)F_Y(y), x, y \in R,
\]

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

\[
F(x,y) = G(x)H(y).
\]

利用定理2.1，立即可判別例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

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

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“積疊”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 \(\{0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1\}\)，為平面上一三角形，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

\[
f(x,y) = 24xyI_{\{0 \leq x \leq 1, 0 \leq y \leq 1, x + y \leq 1\}}.
\]

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$\forall A,B \subseteq \mathbb{R}$，

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

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

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

証明. 我們仍只證連続型隨機變數的情況。先證(ii)。由於$X$與$Y$獨立，

$$f(x,y) = f_X(x)f_Y(y).$$

故

$$E(g(X)h(Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f(x,y)dxdy$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{\infty} h(y)f_Y(y)\int_{-\infty}^{\infty} g(x)f_X(x)dxdy$$

$$= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy$$

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

得証(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$\mathbb{R}^2$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

\[ P(X \in A, Y \in B) = P((X, Y) \in C) \]
\[ = E(g(X)h(Y)) \]
\[ = E(g(X))E(h(Y)) \]
\[ = P(X \in A)P(Y \in B). \]

証畢。

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

\[ P(X \geq 3, Y \geq 2) = P(X \geq 3)P(Y \geq 2) \]
\[ = \int_3^\infty 2e^{-2x} dx \cdot \int_2^3 \frac{1}{3} dy \]
\[ = \frac{1}{3}e^{-6}. \]

又

\[ E(XY^2) = E(X) \cdot E(Y^2) = \frac{1}{2} \cdot \int_0^3 \frac{1}{3} y^2 dy \]
\[ = \frac{1}{2} \cdot 3 = \frac{3}{2}. \]

例2.14 設\(U_1, U_2\)為二獨立的隨機變數, 且皆有\(U(0, 1)\)分佈。又設平面上有一隨機的正方形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

\[ E(U_1U_2) = E(U_1)E(U_2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}. \]

次求面積\(U_1U_2\)之中位數。對\(0 < y \leq 1\),

\[ P(U_1U_2 \geq y) = \int_y^1 \left( \int_{y/u_1}^1 du_2 \right) du_1 \]
\[ = \int_y^1 (1 - \frac{y}{u_1}) du_1 \]
第三章 多維隨機變數

\[
\begin{align*}
&= (u_1 - y \log u_1)^1_y \\
&= 1 - y(1 - \log y).
\end{align*}
\]

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

\[
P(U_1 U_2 \geq \frac{1}{4}) = 1 - \frac{1}{4}(1 + \log 4) = 0.403.
\]

故 \(U_1 U_2\) 中位數小於 \(1/4\)。即 \(U_1 U_2\) 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. 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

\[
f(x, y) = 1, 0 < x, y < 1.
\]

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

\[
P(X/Y < 0.5) = P(2X < Y) = \int_{0}^{0.5} \int_{2x}^{1} dydx = \frac{1}{4},
\]

又

\[
P(2n - 0.5 < X/Y < 2n + 0.5)
= P\left( \frac{2X}{4n+1} < Y < \frac{2X}{4n-1} \right)
= \int_{0}^{1} \left( \frac{2x}{4n-1} - \frac{2x}{4n+1} \right) dx
= \frac{1}{4n-1} - \frac{1}{4n+1},
\]

故

\[
P(\text{最接近 } X/Y \text{ 的整數恰為一偶數})
\]
\[
\begin{align*}
&= \frac{1}{4} + \sum_{n=1}^{\infty} \left( \frac{1}{4n-1} - \frac{1}{4n+1} \right) \\
&= \frac{1}{4} + \left( \frac{1}{3} - \frac{1}{5} + \frac{1}{7} - \frac{1}{9} + \cdots \right) \\
&= \frac{1}{4} + 1 - \arctan 1 \\
&= \frac{1}{4} + 1 - \frac{\pi}{4} = \frac{5 - \pi}{4} < \frac{1}{2}.
\end{align*}
\]

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。

記 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) \[\psi(s, t) = E(e^{isX+itY}) = E(e^{isX})E(e^{itY}), s, t \in R.\]

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多變數)及動差母函數之版本。只要這些轉換至少在一維的開區間存在(母函數及拉普拉斯轉換沒問題, 動差母函數 \(M(s, t) = E(e^{isX+itY})\) 有時只有在 \(s = t = 0\) 才存在)。

定理 2.3 有下列重要推論。
系理2.2 設 \( X, Y \) 為二獨立的隨機變數。則

\[
E(e^{it(X+Y)}) = E(e^{itX})E(e^{itY}), t \in R.
\]

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 \( P(\lambda) \) 及 \( P(\mu) \) 的分佈。則由系理2.2，\( X + Y \) 之拉普拉斯轉換為

\[
\phi(s, t) = E(e^{-s(X+Y)}) = E(e^{-sX})E(e^{-sY})
= e^{-\lambda[1-e^{-s}]}e^{-\mu[1-e^{-s}]}
= e^{-(\lambda+\mu)[1-e^{-s}]}, s \geq 0.
\]

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

利用適當的轉換，可求出一些有常見分佈(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 \( P(\lambda_i) \) 分佈，則 \( S \) 有 \( P(\lambda_1 + \lambda_2) \) 分佈；
(ii) 若 \( X_i \) 有 \( B(n_i, p) \) 分佈，則 \( S \) 有 \( B(n_1 + n_2, p) \) 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 \( \chi^2_{n_i} \) 分佈，則 \( S \) 有 \( \chi^2_{n_1+n_2} \) 分佈；
(v) 若 \( X_i \) 有 \( N(\mu_i, \sigma^2_i) \) 分佈，則 \( S \) 有 \( N(\mu_1 + \mu_2, \sigma^2_1 + \sigma^2_2) \) 分佈；
(vi) 若 \( X_i \) 有 0 開始之 \( Ge(p) \) 分佈，則 \( S \) 有 \( NB(2, p) \) 分佈；
(vii) 若 \( X_i \) 有 \( NB(r_i, p) \) 分佈，則 \( S \) 有 \( NB(r_1 + r_2, p) \) 分佈。

並非所有獨立的同一類分佈之隨機變數的和仍有一同一類分佈(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 \( U(0, 1) \) 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，原因是讓大家自行構想。利用

$$
\sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}.
$$

試求任選二正整數互質的概率$p$。

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{$i, 2i, 3i, \ldots$}，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

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

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$A$成立，但$\int_A dxdy = 0$，則$X$與$Y$仍為獨立的隨機變數。例如，若

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

令$A = \{(x,y)|x,y > 0, x = y\}$。由於

$$
\int \int_A dxdy = \int_{0}^{\infty} \int_{y}^{\infty} 1dxdy = \int_{0}^{\infty} 0dy = 0,
$$

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

圖2.1給出兩個$SN(\lambda)$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 $SN(2)$及$SN(5)$p.d.f.之圖形](image)

我們列出一些偏斜常態分佈簡單的性質如下。

(i) $SN(0) = N(0,1)$;
(ii) 若 $Z \sim SN(\lambda)$，則 $-Z \sim SN(-\lambda)$;
(iii) 若 $Z \sim SN(\lambda)$，則 $Z^2 \sim \chi^2$;
(iv) 若 \( Z \sim \mathcal{N}(\lambda) \)，則 \( Z \)之分佈函數為
\[
\Phi(z; \lambda) = P(Z \leq z) = 2 \int_{-\infty}^{\lambda} \phi(t) \phi(u) \, du \, dt;
\]
(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因難，因此留在題中，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
\[
\lim_{z \to \infty} \Phi(z; \lambda) = 2 \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda} \phi(t) \phi(u) \, du \, dt = 1.
\]
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 \( X \overset{d}{=} (-X) \),
\( Y \overset{d}{=} (-Y) \)，故
\[
-(X - \lambda Y) = (-X) - \lambda(-Y) \overset{d}{=} X - \lambda Y.
\]
因此
\[
P(X - \lambda Y < 0) = \frac{1}{2}.
\]
又
\[
\frac{1}{2} = P(X - \lambda Y < 0) = P(X < \lambda Y)
\]
\[
= \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda y} \phi(x) \phi(y) \, dx \, dy
\]
\[
= \int_{-\infty}^{\infty} \phi(y) \Phi(\lambda y) \, dy.
\]
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.。習題中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 \( \mathcal{N}(0, 1) \) 分布。令

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

試證對 \( \forall \lambda \in \mathbb{R} \)，\( Z \) 有 \( \mathcal{N}(\lambda) \) 分布。

証明。我們只證 \( \lambda > 0 \) 的情況，\( \lambda \leq 0 \) 的情況可類似地證出。底下計算 \( Z \) 之分

佈函數。

(2.26) \( F(z) = P(Z \leq z) \)

\[
P(Z \leq z, \lambda X > Y) + P(Z \leq z, \lambda X \leq Y) \\
P(X \leq z, X > Y/\lambda) + P(-X \leq z, \lambda X \leq Y) \\
P(Y/\lambda < X \leq z, Y \leq \lambda z) + P(-z \leq X \leq Y/\lambda, Y \geq -\lambda z) \\
= \int_{-\infty}^{\lambda z} \int_{y/\lambda}^{z} \phi(x)\phi(y)dx\,dy + \int_{-\lambda z}^{\infty} \int_{-z}^{y/\lambda} \phi(x)\phi(y)\,dx\,dy \\
= \int_{-\infty}^{\lambda z} [\Phi(z) - \Phi(y/\lambda)]\phi(y)\,dy + \int_{-\lambda z}^{\infty} (\Phi(y/\lambda) - \Phi(-z))\phi(y)\,dy \\
= \Phi(z)\Phi(\lambda z) - \int_{-\infty}^{\lambda z} \Phi(y/\lambda)\phi(y)\,dy \\
+ \int_{-\lambda z}^{\infty} \Phi(y/\lambda)\phi(y)\,dy - \Phi(-z)(1 - \Phi(-\lambda z)).
\]

此處用到

\[
P(X \leq z, X > Y/\lambda, Y > \lambda z) = 0, \\
P(-X \leq z, \lambda X < Y, Y < -\lambda z) = 0, \\
P(X \leq z, X > Y/\lambda, Y \leq \lambda z) = P(Y/\lambda < X \leq z, Y \leq \lambda z),
\]

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

得证Z有SN(λ)分布。

3.3 变数代换

在1.5节我们曾介绍单变数的变数代换。对随机向量(X, Y)，有时我们会对X, Y的二函数，如

(3.1) \[U = g_1(X, Y),\]

(3.2) \[V = g_2(X, Y),\]

有兴趣。如果令A = \{(x, y)|f(x, y) > 0\}，其中f(x, y)为X, Y之联合密度，

\[B = \{(u, v)|u = g_1(x, y), v = g_2(x, y), (x, y) \in A\} \]。

则(U, V)便是由A映至B之二变数代换。有时也可能只对一个函数W = g(X, Y)有兴趣。这两种情况都属于两个变数的变数代换。

例如，令X表某位学生第一次考试的成绩，Y表第二次考试的成绩。我们可能想对其进阶成绩U = Y - X，与总分V = X + Y有兴趣，也可能只想对其进阶的比率W = (Y - X)/X有兴趣。(U, V)或是单独的U, V，

W皆是将原来的随机向量(X, Y)变换(或说转换)至新的变数。

设由(X, Y)经由(3.1)及(3.2)式变换至(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)，则我们说由(X, Y)至(U, V)的变换，为1 - 1。此时有

(3.3) \[u = g_1(x, y), v = g_2(x, y) \iff x = h_1(u, v), y = h_2(u, v).\]

例如，设

\[U = X + Y, \quad V = X - Y,\]

则

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

故(X, Y)至(U, V)为1 - 1变换。另外，设X, Y为二正的随机变数，且今
\[ U = XY, \quad V = X/Y, \]

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

\[ f_{U,V}(u,v) = f(h_1(u,v), h_2(u,v)). \]

此因

\[ f_{U,V}(u,v) = P(g_1(X,Y) = u, g_2(X,Y) = v) \]

\[ = P(X = h_1(u,v), Y = h_2(u,v)) \]

\[ = f(h_1(u,v), h_2(u,v)). \]

做變數代換，務必要將新的支持(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 \((U, V)\) 仍為 \(1-1\) 聯 合 p.d.f.，因 此 \((U, V)\) 之 聯 合 p.d.f. 為

\[ f_{U,V}(u,v) = f(uv, u(1-v)) \]

\[ = \frac{n!}{(uv)!(u(1-v))!(n-u)!} p_1^{uv} p_2^{u(1-v)} (1-p_1-p_2)^{n-u}, \]

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$P(\lambda)$及$P(\mu)$分佈。則$(X,Y)$之
聯合p.d.f.為

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

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

$$f_{U,V}(u,v) = f(u-v,v)$$

$$= \frac{e^{-\lambda} \lambda^{u-v} e^{-\mu} \mu^v}{(u-v)! \frac{v!}{v}}, \ u, v = 0, 1, 2, \cdots, u = v, v + 1, v + 2, \cdots.$$

有了聯合p.d.f.，便可求出$U$的邊際p.d.f.:

$$f_U(u) = \sum_{v=0}^{u} \frac{e^{-\lambda} \lambda^{u-v} e^{-\mu} \mu^v}{(u-v)! \frac{v!}{v}}$$

$$= e^{-(\lambda+\mu)} \sum_{v=0}^{u} \frac{\lambda^{u-v} \mu^v}{(u-v)! \frac{v!}{v}}$$

$$= e^{-(\lambda+\mu)} \sum_{v=0}^{u} \frac{\lambda^{u-v} \mu^v}{u! \frac{v!}{v}}$$

$$= e^{-(\lambda+\mu)} \frac{(\lambda + \mu)^u}{u!}.$$

其中例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$P(\lambda + \mu)$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$B(m,p)$及$B(n,p)$分佈，導出$X + Y$有$B(m + n,p)$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 便提供一種典型的作法：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 \((X,Y)\) 為連続型的隨機變量，仍以 \(f(x,y)\) 為其聯合p.d.f.，則 \((U,V)\)之聯合p.d.f.，可如單變數的情形(見第一章(5.5)式)得到。即

\[
f_{U,V}(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} f(h_1(u,v), h_2(u,v)),
\]

其中

\[
J(u,v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u} \frac{\partial y}{\partial v} - \frac{\partial y}{\partial u} \frac{\partial x}{\partial v},
\]

而

\[
\frac{\partial x}{\partial u} = \frac{\partial h_1(u,v)}{\partial u}, \quad \frac{\partial x}{\partial v} = \frac{\partial h_1(u,v)}{\partial v},
\]

\[
\frac{\partial y}{\partial u} = \frac{\partial h_2(u,v)}{\partial u}, \quad \frac{\partial y}{\partial v} = \frac{\partial h_2(u,v)}{\partial v}.
\]

\(J\) 稱為變換 \( x = h_1(u,v), y = h_2(u,v) \) 之雅可比 (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

\[
F_{U,V}(u,v) = \Pr(U \leq u, V \leq v)
= \Pr(g_1(X,Y) \leq u, g_2(X,Y) \leq v)
= \int_A \int_A f(x,y)dxdy
= \int^v_u \int^u_{-\infty} |J(s,t)| f(h_1(s,t), h_2(s,t))dsdtdt,
\]

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。
3.3 變數對換 197

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 \(B\epsilon(p, q)\) 分佈，\(V\) 有 \(\Gamma(p + q, \lambda)\) 分佈。

證明. 首先 \(X, Y\) 之聯合 p.d.f. 為

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

由

\[
u = x/(x + y), \ v = x + y, \ 0 < u < 1, \ v > 0 .
\]

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

\[
f_{U, V}(u,v) = \left|J\right| f(h_1(u, v), h_2(u, v))
\]

\[
= v \frac{(uv)^{p-1}e^{-uv/\lambda}(v(1 - u))^{q-1}e^{-v(1-u)/\lambda}}{\Gamma(p)\Gamma(q)\lambda^{p+q}} ,
\]

\[0 < u < \infty, 0 < v < 1 .\]

經化簡後得

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

由定義2.3知，\(U\) 與 \(V\) 獨立，且分別有 \(B\epsilon(p, q)\) 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\( V = X + Y \) 則是 \( U \) 的分母。但對 \( X, Y \) 皆為 gamma 分布，且有相同的尺度參數，則 \( U \) 與 \( V \) 獨立。事實上此為 gamma 分布之一特性。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 分布，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。

**定理 3.1** 設 \( U \) 與 \( V \) 獨立，且 \( U \) 有 \( \mathcal{B}(p, q) \) 分布，\( V \) 有 \( \Gamma(p + q, \lambda) \) 分布。則 \( UV \) 有 \( \Gamma(p, \lambda) \) 分布。

定理 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 分布。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 分布，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 \( \mathcal{U}(1, \beta) \) 分布，\( \beta > 1 \)。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

\[
f_{U,V}(u, v) = |J|f_{X,Y}(h_1(u, v), h_2(u, v))
\]

\[
= \frac{1}{(\beta - 1)^2 v}, 1 < \frac{u}{v} < \beta, 1 < v < \beta.
\]
我們分別給出$x, y$及$u, v$取值的範圍如下二圖。

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

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。底下再給一自在表面上看不像獨立，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$u = x + y, v = x - y, u, v \in \mathbb{R}$，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
\[ f_{U,V}(u,v) = \left| \frac{1}{2\pi\sigma^2} \right| \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u+\mu)^2}{2\sigma^2}} \cdot \left( \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(v-\mu+\gamma)^2}{2\sigma^2}} \right) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(v-\mu-\gamma)^2}{2\sigma^2}} \right) \]

中间的计算过程并不难，让读者自行补上。故得证\(U\)与\(V\)独立，且分别有\(\mathcal{N}(\mu + \gamma, 2\sigma^2)\)及\(\mathcal{N}(\mu - \gamma, 2\sigma^2)\)分佈。

在上例中，若\(\mu = \gamma = 0\)，则\(U\)与\(V\)不但独立，且有相同的\(\mathcal{N}(0, 2\sigma^2)\)分佈。底下的定理指出，对二独立的随机变数，其各自的函数亦独立。

**定理 3.2** 设\(X, Y\)为二独立的随机变数，\(g(x), h(y)\)分别为\(x, y\)之函数。则\(U = g(X)\)与\(V = h(Y)\)为二独立的随机变数。

**证明** 对任意\(u, v \in \mathbb{R}\)令

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

则

\[ F_{U,V}(u,v) = P(U \leq u, V \leq v) = P(X \in A_u, Y \in B_v) = P(X \in A_u)P(Y \in B_v) \]

其中最后一等式成立是因\(X\)与\(Y\)独立，故可利用定理 2.2(i)。因\(P(X \in A_u)\)及\(P(Y \in B_v)\)分别为其函数，故由系理 2.1，即得\(U\)与\(V\)独立。

**例 3.5** 设\(X, Y\)为二独立的随机变数，分别有\(\Gamma(p, \lambda)\)及\(\Gamma(q, \lambda)\)分佈。令\(U = X + Y, V = X/(X + Y)\)。由定理 3.1，知\(U\)与\(V\)独立。再由定理 3.2知，\(U\)与\(V^{-1} - 1 = Y/X\)独立。

有时由\((X, Y)\)至\((U, V)\)的变换并非\(1-1\)，此时可仿单变数时作法，可参考第一章(5.7)式。即设存在\(A = \{(x, y) | f(x, y) > 0\}\)之一分割\(A_0, A_1, \cdots, A_n\)，其中\(A_0\)可能是空集合，满足\(P((X, Y) \in A_0) = 0\)，使得对
3.3 变数代换 201

∀i = 1, ⋯, k, U = g_1(X, Y), V = g_2(X, Y) 为由 A_i 映成至 B 之 1 − 1 函数。则对 ∀i = 1, ⋯, k, 由 B 至 A_i 之反函数存在，以 x = h_{1i}(u, v), y = h_{2i}(u, v) 表第 i 个反函数。对第 i 个反函数, ∀(u, v) ∈ B, 有唯一的(x, y) ∈ A, 使得

\[(u, v) = (g_1(x, y), g_2(x, y))\]。

令 J_i 表第 i 个反函数之雅可比。假设 J_1, J_2, ⋯, J_k 在 B 中皆不为 0。则(U, V)之联合 p.d.f. 为

\[f_{U, V}(u, v) = \sum_{i=1}^{k} \left| J_i \right| f_{X, Y}(h_{1i}(u, v), h_{2i}(u, v))\]。

例 3.6 設 X 与 Y 為二独立的随机变量，皆有 N(0, 1) 分布。令 U = X/Y, V = |Y|。当 Y = 0 时，可定义 U = 1(或任何其他常数值)。由于对连续型的随机变量，P(Y = 0) = 0, 所以亦不用担心 Y = 0 的情况。

因 (x, y) 与 (−x, −y)，对称到相同的(u, v)，故由 (X, Y) 至 (U, V) 之变换

并非 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\},

则 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) 之变换

皆为 1 − 1。由 B 至 A_1 之反函数

\[x = h_{11}(x, y) = uv,\]

\[y = h_{21}(x, y) = v.\]

由 B 至 A_2 之反函数

\[x = h_{12}(x, y) = -uv,\]

\[y = h_{22}(x, y) = -v.\]

J_1 = J_2 = v。又
\[ f(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}. \]

故

\[ f_{U,V}(u,v) = \frac{1}{2\pi} |v| e^{-\frac{u^2v^2}{2v^2} + \frac{1}{2\pi} |v| e^{-\frac{u^2v^2}{2v^2}}} \]
\[ = \frac{v}{\pi} e^{-\left(1+u^2\right)v^2/2}, u \in R, v > 0. \]

由此即得

\[ f_U(u) = \int_0^\infty \frac{v}{\pi} e^{-\left(1+u^2\right)v^2/2} dv \quad (v^2 = z) \]
\[ = \frac{1}{2\pi} \int_0^\infty e^{-\left(1+u^2\right)z/2} dz \]
\[ = \frac{1}{\pi(1+u^2)}, u \in R. \]

即 \( X/Y \) 有 \( \mathcal{C}(0, 1) \) 分布。

在2.3 裂位矩形分布时，我们曾提到此分布有时在我们想不到的地方出现。此即为一例。题目中尚有一些例子，\( X \) 与 \( Y \) 之变数也不须为 0。

在上例中，若仅是要求 \( X/Y \) 之分布，则可令 \( U = X/Y, V = Y \)。则此为由 \( (X, Y) \) 至 \( (U, V) \) 之 1 − 1 变换。求出 \( (U, V) \) 之联合 p.d.f.

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

之后，再求 \( U \) 之边缘 p.d.f. 仍得 \( U \) 有 \( \mathcal{C}(0, 1) \) 分布。

### 3.4 混合分布

有时一随机变量之参数，亦允许为随机变量。我们先看一个例子。

**例 4.1** 假设高雄市每天发生之交通事故的次数 \( Y \) 有 \( \mathcal{P}(\lambda) \) 分布。每次交通事故有 p 的机会会有死亡。令 \( X \) 表示每天有死亡之交通事件。则 \( X \mid Y \) 有 \( \mathcal{B}(Y, p) \) 分布，而 \( Y \) 有 \( \mathcal{P}(\lambda) \) 分布。本模式即二项分佈试验之次数亦为一随机变量 \( Y \)。给定 \( Y = y \)，则 \( X \) 有 \( \mathcal{B}(y, p) \) 分布。至 \( Y \) 则有 \( \mathcal{P}(\lambda) \) 分布。当一随机
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

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

因此 $X$ 之非條件分佈為 $P(\lambda p)$。

在上例中，每天死亡人數之期望值為 $E(X) = \lambda p$。又顯然 $E(X|Y) = Y p$, $E(Y) = \lambda$，故得 $E(X) = E(E(X | Y)) = E(Y p) = p E(Y) = \lambda p$。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**。我們只證明連統型隨機變數的情況。設 $(X, Y)$ 之聯合p.d.f. 為 $f(x, y)$。

由定義

$$E(X) = \int \int xf(x, y) dxdy$$
$$= \int (\int xf(x | y) dx) f_Y(y) dy$$
$$= \int E(X | y) f_Y(y) dy$$
$$= E(E(X | Y)).$$
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\(P(\Lambda)\)分佈，又設\(\Lambda\)有\(E(\beta)\)分佈。則

\[
E(X) = E(E(X|Y)) = E(pY) = pE(Y) = pE(E(Y|\Lambda)) = pE(\Lambda) = p/\beta.
\]

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

\[
P(Y = y) = \int_0^\infty f_{Y|\Lambda}(y, \lambda) d\lambda
= \int_0^\infty f_Y(y|\lambda) f_\Lambda(\lambda) d\lambda
= \int_0^\infty \frac{e^{-\lambda y}}{y!} \beta e^{-\beta \lambda} d\lambda
= \frac{\beta}{y!} \int_0^\infty \lambda^y e^{-(1+\beta)\lambda} d\lambda
= \frac{\beta}{y!} \frac{\Gamma(y + 1)}{(1 + \beta)^{y + 1}}
= \frac{\beta}{1 + \beta} \left(\frac{1}{1 + \beta}\right)^y.
\]

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\(\Lambda\)了(也可以說就不知道有\(\Lambda\)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$P(x,y)$時，只要記住在求機率或期望值時，若離散型的隨機變數使
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

$$P(X = x) = \int_0^1 p^x(1 - p)^{1-x} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1 - p)^{\beta-1} dp$$

$$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 p^{\alpha+x-1}(1 - p)^{\beta-x+1-1} dp$$

$$= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(\alpha + x)\Gamma(\beta - x + 1)}{\Gamma(\alpha + x + 1)}$$

$$= \frac{\Gamma(\alpha + x)\Gamma(\beta - x + 1)}{(\alpha + \beta)\Gamma(\alpha)\Gamma(\beta)}$$

$$= \left\{ \begin{array}{ll}
\frac{\alpha}{\alpha + \beta}, & x = 1, \\
\frac{\alpha}{\alpha + \beta}, & x = 0.
\end{array} \right.$$

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) = \text{E}(\text{Var}(X|Y)) + \text{Var}(E(X|Y)).
證明. 由定義

(4.3) \[ \text{Var}(X) = E((X - E(X))^2) \]

\[ = E((X - E(X|Y) + E(X|Y) - E(X))^2) \]

\[ = E((X - E(X|Y))^2) + E((E(X|Y) - E(X))^2) \]

\[ + 2E((X - E(X|Y))(E(X|Y) - E(X))) \]

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:

(4.4) \[ E((X - E(X|Y))(E(X|Y) - E(X))) \]

\[ = E(E((X - E(X|Y))(E(X|Y) - E(X)))|Y) \]

但給定Y之下, E(X|Y)及E(X)皆為常數。故

(4.5) \[ E((X - E(X|Y))(E(X|Y) - E(X)))|Y) \]

\[ = (E(X|Y) - E(X))E(X - E(X|Y)|Y) \]

\[ = (E(X|Y) - E(X))(E(X|Y) - E(X|Y)) \]

\[ = 0 \]

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) \[ \text{Var}(X|Y) = E((X - E(X|Y))^2|Y), \]

上式也是Var(X|Y)之定義, 故

(4.7) \[ E((X - E(X|Y))^2) = E(E((X - E(X|Y))^2 |Y)) \]

\[ = E(\text{Var}(X|Y)) \]

又因E(E(X|Y)) = E(X), 故

(4.8) \[ E((E(X|Y) - E(X))^2) = \text{Var}(E(X|Y)) \]
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

\[
E(\text{Var}(X|Y)) = E(E(X^2|Y) - (E(X|Y))^2)
= E(E(X^2|Y)) - E((E(X|Y))^2)
= E(X^2) - (E(X))^2 - (E((E(X|Y))^2) - (E(X))^2)
= \text{Var}(X) - \text{Var}(E(X|Y)).
\]

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** 指例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\(\text{Ber}(P)\)分佈，故

\[
E(X|P) = P, \quad \text{Var}(X|P) = P(1 - P).
\]

因此

\[
\text{Var}(X) = \text{Var}(P) + E(P(1 - P))
= E(P^2) - (E(P))^2 + E(P) - E(P^2)
\]
\[ 
E(P)(1 - E(P)) = \frac{\alpha}{\alpha + \beta}(1 - \frac{\alpha}{\alpha + \beta}) = \frac{\alpha\beta}{\alpha + \beta}.
\]

另外，如果已求出 \( X \) 之非條件分佈為 \( Ber(\alpha/(\alpha + \beta)) \)，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，則係數當然是最弱的。若樣本為水，令 \( X \) 表其體積，\( Y \) 表其重量。顯然\( X \) 與 \( Y \) 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，又稱變異方差）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) \[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}. \]

Cov(X, Y)可以σXY 表之, ρ(X, Y)有時以ρXY 表之。

當σX 或σY = 0, 此時Cov(X, Y) = 0(習題第1題), 至於ρ(X, Y)則不定義。由定義5.1知, 只要我們提到相関係數, 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(Cov(X, X) = 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Cov(X, Y)將是正的。因此若X > μX時, 應可能有Y > μY, 則(X - μX)(Y - μY)較可能為正; 若X < μX時, 亦較可能有Y < μY, 則(X - μX)(Y - μY)也將較可能為正。如此一來, Cov(X, Y) = E((X - μX)(Y - μY)) > 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 - μX)(Y - μY)較可能為負, 如此一來, Cov(X, Y) < 0。故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一減。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, Cov(X, Y) = 0(見底下系理5.1)。不過Cov(X, Y) = 0時, X與Y卻不一定獨立。另外, ρ(X, Y)與Cov(X, Y)之符號相同, 且當Var(X)及Var(Y)皆存在時, ρ(X, Y) = 0, 若且唯若Cov(X, Y) = 0。當ρ(X, Y) = 0則稱X與Y無相關(uncorrelated)。

作為量測二隨機變數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, 若X以公升為單位, Y以公斤為單位, 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
給定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。

\[
\text{Cov}(X, Y) = E((X - \mu_X)(Y - \mu_Y)) \\
= E(XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y) \\
= E(XY) - \mu_X E(Y) - \mu_Y E(X) + \mu_X \mu_Y \\
= E(XY) - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y \\
= E(XY) - \mu_X \mu_Y.
\]

系理5.1 設X與Y為二獨立的隨機變數, 則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

\[
\text{Cov}(X, Y) = E(XY) - E(X)E(Y) \\
= E(X)E(Y) - E(X)E(Y) = 0.
\]

其次,

\[
\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{0}{\sigma_X \sigma_Y} = 0.
\]

證明。

底下的推論也是很顯然的。

系理5.2 對二隨機變數X, Y,

\[
\rho(X, Y) = \text{Cov}(\frac{X - \mu_X}{\sigma_X}, \frac{Y - \mu_Y}{\sigma_Y}) = E\left(\frac{(X - \mu_X)(Y - \mu_Y)}{\sigma_X \sigma_Y}\right).
\]
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。由係數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，只要尺標之改變，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

<table>
<thead>
<tr>
<th>$x \setminus y$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>$f_X(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20 &amp; 0.10 &amp; 0.01 &amp; 0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.15 &amp; 0.30 &amp; 0.06 &amp; 0.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.03 &amp; 0.05 &amp; 0.10 &amp; 0.18</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$f_Y(y)$ | 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

\[ f_X(x) = \begin{cases} 
1, & 0 < x < 1, \\
0, & \text{其他} 
\end{cases} \]

\[ f_Y(y) = \begin{cases} 
1, & 0 < y < 1, \\
2 - y, & 1 \leq y < 2.
\end{cases} \]

則μX = 1/2, σX^2 = 1/12, μY = 1, σY^2 = 1/6。又

\[ E(XY) = \int_0^1 \int_x^{x+1} xydydx = \int_0^1 \frac{1}{2} x^2 \left[ x^{x+1} \right]_x^1 = \int_0^1 (x^2 + \frac{1}{2} x)dx = \frac{7}{12}. \]

因此

\[ \text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y = \frac{7}{12} - \frac{1}{2} \cdot 1 = \frac{1}{12}, \]

\[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/12}{\sqrt{1/12 \cdot 1/6}} = \frac{\sqrt{2}}{2}. \]

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 设$(X,Y)$为离散型之随机变量，以$f(x,y)$为其联合p.d.f.，且
\[
    f(-1,1) = f(1,1) = \frac{1}{3}, \\
    f(-2,4) = f(2,4) = \frac{1}{6}.
\]
则
\[
    f_X(-1) = f_X(1) = \frac{1}{3}, \quad f_X(-2) = f_X(2) = \frac{1}{6}, \\
    f_Y(1) = \frac{2}{3}, \quad f_Y(4) = \frac{1}{3}.
\]
故
\[
    \mu_X = 0, \quad \sigma_X^2 = 2, \\
    \mu_Y = 2, \quad \sigma_Y^2 = 2, \\
    E(XY) = 0,
\]
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$X$与$Y$不独立。
另一方面，$X$与$Y$之间有很强的非线性关系。因对$(x,y)$，只要$P(X = x, Y = y) \neq 0$，则$y = x^2$，即
\[
    P(Y = X^2) = 1.
\]
亦即虽有$\rho(X,Y) = 0$，得知$X$与$Y$缺乏线性上的线性关系。但并未排除$X$与$Y$间有非线性的关系。换句话说，“无相关”并非“不相关”。

例5.4 设$(X,Y)$之联合p.d.f.为
\[
    f(x,y) = \frac{1}{2}, \quad 0 \leq y \leq x \leq 2.
\]
試求 Cov(X, Y) 及 ρ(X, Y)。
解. 首先 X, Y 之連機 p.d.f. 分別為

\[ f_X(x) = \int_0^x \frac{1}{2} dy = \frac{x}{2}, \quad 0 \leq x \leq 2, \]
\[ f_Y(y) = \int_y^2 \frac{1}{2} dx = 1 - \frac{y}{2}, \quad 0 \leq y \leq 2. \]

由此得

\[ \mu_X = \int_0^2 x \cdot \frac{1}{2} dx = \frac{4}{3}, \]
\[ E(X^2) = \int_0^2 x^2 \cdot \frac{1}{2} dx = 2, \]

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

\[ E(XY) = \int_0^2 \int_0^x \frac{1}{2} y dx dy \]
\[ = \int_0^2 \frac{1}{2} x y^2 \bigg|_0^x dx \]
\[ = \int_0^2 \frac{x^3}{4} dx = 1. \]

故得

\[ \text{Cov}(X, Y) = E(XY) - \mu_X \mu_Y = 1 - \frac{4}{3} \cdot \frac{2}{3} = \frac{1}{9}. \]
\[ \rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{1/9}{\sqrt{2/9}} = \frac{1}{2}. \]

再給一例。
例5.5  设$X$与$Z$为二独立的随机变数，且$X$有$\mathcal{U}(0, 1)$分佈，$Z$有$\mathcal{U}(0, 1/10)$分佈。令$Y = X + Z$，並考慮随机向量$(X, Y)$。利用变数代换，可得$X, Y$之联合p.d.f.为

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

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$Z$独立，我們可經由$X$與$Z$，如下地求出$\text{Cov}(X, Y)$與$\rho(X, Y)$。

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

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

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

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

$$= E(X(X + Z)) - E(X)(E(X) + E(Z))$$

$$= E(X^2) + E(XZ) - (E(X))^2 - E(X)E(Z)$$

$$= E(X^2) - (E(X))^2$$

$$= \text{Var}(X) = \frac{1}{12}.$$

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

$$\rho(X, Y) = \frac{\frac{11}{20}}{\sqrt{\frac{1}{12} \cdot \frac{101}{1,200}}} = \frac{10}{\sqrt{101}}.$$

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5.1.0及5.1.1，顯示$X$與$Y$大致有線性的關係，但在圖5.1.1中線性關係較強（區域較窄）。也可以另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$\mathcal{U}(x, x + 1/10)$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$cov(X, Y)$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

$$cov(X, Y) = E(X(X^2 + Z)) - E(X)(E(X^2 + Z)).$$
\[ E(X^3) + E(X Z) = 0, \]

且 \( \rho(X, Y) = 0 \)。

再提醒一次，相關係數是度量線性關係的強弱，而非度量其他非線性關係之強弱。

例5.7 设 \( X \sim \mathcal{U}(-1, 1) \) 分布。令 \( Y = X^2, Z = 2X + 3, W = -2X + 3 \)。则

\[
E(X) = 0, \\
E(X^2) = 2E(X) + 3E(X) = \frac{2}{3}, \\
E(X)E(Z) = 0, \\
Var(X) = E(X^2) = \frac{1}{3}, \\
Var(Z) = 4Var(X) = \frac{4}{3}.
\]

故

\[
Cov(X, Z) = E(X Z) - E(X)E(Z) = \frac{2}{3},
\]

且

\[
\rho(X, Z) = \frac{Cov(X, Z)}{\sigma_X \sigma_Z} = \frac{2/3}{\sqrt{1/3} \cdot \sqrt{4/3}} = 1.
\]

同理可得
\[ \rho(X, W) = -1. \]

\(X\)与\(Z\)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，見底下定理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** 设\(X, Y\)为二随机变数。则对任二常数\(a, b\),

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

特别地，若\(X\)与\(Y\)独立，则

\[ (5.7) \quad \text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y). \]

**证明.** 首先

\[ E(aX + bY) = aE(X) + bE(Y) = a\mu_X + b\mu_Y. \]

故

\[
\begin{align*}
\text{Var}(aX + bY) &= E((aX + bY - (a\mu_X + b\mu_Y))^2) \\
&= E((a(X - \mu_X) + b(Y - \mu_Y))^2) \\
&= a^2E((X - \mu_X)^2) + b^2E((Y - \mu_Y)^2) + 2abE((X - \mu_X)(Y - \mu_Y)) \\
&= a^2\text{Var}(X) + b^2\text{Var}(Y) + 2ab\text{Cov}(X, Y). 
\end{align*}
\]

其次，若\(X\)与\(Y\)独立，则由定理5.1，\(\text{Cov}(X, Y) = 0\)，因此(5.7)式成立。证毕。

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 **正相関**（positively correlated），即 \( \rho(X, Y) > 0 \)（因此 \( \text{Cov}(X, Y) > 0 \)），則 \( X + Y \) 之變異數，比 \( X \) 之變異數之和還大。反之，若 \( X \) 與 \( Y \) 為 **負相関**（negatively correlated），則 \( X + Y \) 之變異數，比 \( X \) 之變異數之和還小。當負相関時，\( X \) 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，\( a \neq 0 \)。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，二變數有一線性相關係之機率為 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 **完全相関**（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
\[(E(XY))^2 \leq E(X^2)E(Y^2) \cdot \tag{5.9}\]
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 \( \forall \lambda > 0 \),
\[E((Y - \lambda X)^2) = \lambda^2 E(X^2) - 2 \lambda E(XY) + E(Y^2) \tag{5.10}\]
為一恆非負之 \( \lambda \) 的二次式。又因首項 \( \lambda^2 \) 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，则 \(E((Y - \lambda_0 X)^2) = 0\)，此时 \(P(Y = \lambda_0 X) = 1\)。证毕。

将史瓦茲不等式应用至随机变量 \(X - \mu_X\) 与 \(Y - \mu_Y\)，得

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

上式即为

\[(5.11)\quad (\text{Cov}(X, Y))^2 \leq \text{Var}(X)\text{Var}(Y).\]

因此上述结果便得到证了。

**系理5.3** 任随机变量 \(X, Y\)，只要 \(E(X^2), E(Y^2) < \infty\)，便有

\[(5.12)\quad |\rho(X, Y)| \leq 1.\]

我们又有下述重要的推论。

**系理5.4** 设 \(X, Y\) 为随机变量，且 \(E(X^2), E(Y^2) < \infty\)。则 \(|\rho(X, Y)| = 1\)，若且唯若存在二常数 \(a, b\)，且 \(a \neq 0\)，使得 \(P(Y = aX + b) = 1\)。

**证明** 首先，由相关系数之定义知，\(\text{Var}(X)\) 及 \(\text{Var}(Y)\) 皆不能为 0。因此 \(P(X = c) = 1\) 不能成立，其中 \(c\) 为一常数，利用系理5.4 关于随机变量 \(X - \mu_X\) 及 \(Y - \mu_Y\)，即 \(|\rho(X, Y)| = 1\)，若且唯若存在一常数 \(a\)，使得 \(P(Y - \mu_Y = a(X - \mu_X)) = 1\)，又 \(a \neq 0\)，否则 \(P(Y = \mu_Y) = 1\)，因此 \(\text{Var}(Y) = 0\)。取 \(b = \mu_Y - a\mu_X\)，可看出此即我们要证的结果。

在系理5.4 中，若 \(a > 0\)，则 \(\rho(X, Y) = 1\)；若 \(a < 0\)，则 \(\rho(X, Y) = -1\)。又 \(a, b\) 也可 \(X, Y\) 之期望值及标准差表示，见下系理，证明则留在习题中。

**系理5.5** 设 \(X, Y\) 为随机变量，且 \(E(X^2), E(Y^2) < \infty\)。则当 \(\rho(X, Y) = 1\)，

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

当 \(\rho(X, Y) = -1\)，

\[(5.14)\quad P(Y = \mu_Y - (\sigma_Y / \sigma_X)(X - \mu_X)) = 1.\]
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現出二隨機變數間有因果關係。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，因此腸胃病有問題的人也變多。如此導致$X$與$Y$之相關係數很大。至於喝可樂與拉肚子間倒不見得有因果關係。

現以$\rho$表示隨機變數$X, Y$之相關係數，則$\rho$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$Y$；$\rho$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，$|\rho|$值愈大。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Rooussas(1977)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之相關係數。為免簡便，已$\mu_1\mu_X, \mu_2 \mu_Y, \sigma_1 \sigma_X, \sigma_2 \sigma_Y$。

先考慮$xy$平面上的直線

\begin{equation}
(5.15)
y = \mu_2 + (\sigma_2/\sigma_1)(x - \mu_1)。
\end{equation}

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$a = \sigma_2/\sigma_1, b = -1, c = \mu_2 - (\sigma_2/\sigma_1)\mu_1$)

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

且
(5.17) \[ D_1^2 = ((\sigma_2/\sigma_1)X - Y + (\mu_2 - (\sigma_2/\sigma_1)\mu_1))²/(\sigma_2^2/\sigma_1^2 + 1) . \]

由(5.17)式得

(5.18) \( (\sigma_1^2 + \sigma_2^2)D_1^2 = \sigma_2^2X^2 + \sigma_1^2Y^2 - 2\sigma_1\sigma_2XY \]
\[ + 2\sigma_2(\sigma_1\mu_2 - \sigma_2\mu_1)X - 2\sigma_1(\sigma_1\mu_2 - \sigma_2\mu_1)Y \]
\[ + (\sigma_1\mu_2 - \sigma_2\mu_1)^2 . \]

對上式兩側取期望值，且利用

\[ E(X^2) = \sigma_1^2, \quad E(Y^2) = \sigma_2^2, \]
\[ E(XY) = \rho \sigma_1 \sigma_2 + \mu_1 \mu_2 , \]

化簡後，得

(5.19) \[ E(D_1^2) = \frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}(1 - \rho) . \]

其次考慮xy平面上的直線

(5.20) \[ y = \mu_2 - (\sigma_2/\sigma_1)(x - \mu_1) , \]

同理可得點(X, Y)與此直線距離D_2平方的期望值為

(5.21) \[ E(D_2^2) = \frac{2\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}(1 + \rho) . \]

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(\( \mu_1, \mu_2 \))，在直線\( y = \mu_2 \)，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。當\( \rho \)愈來
3.5 共變異數及相關係數 223

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）。

\[
y = \mu_2 + \frac{\sigma_2}{\sigma_1}(x - \mu_1)
\]

\[
y = \mu_2 - \frac{\sigma_2}{\sigma_1}(x - \mu_1)
\]

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\(y = \mu_2 - \frac{\sigma_2}{\sigma_1}(x - \mu_1)\)的傾向。且當\(\rho\)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\(2\sigma_1^2\)
\(\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\)。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\(\rho\)用來量測隨機變數\(X\)與\(Y\)之共線性，就是基於以上的涵義。

下述定理顯示相關係數與條件期望值關係密切。此定理亦可視為係
理5.5之推廣。

定理5.5 設隨機變數\(X,Y\)之相關係數\(\rho\)存在。若\(E(Y|X = x)\)為一\(x\)之線性

t 函數，則

\[
E(Y|X = x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1),
\]

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

證明，由假設知

\[
E(Y|X = x) = ax + b,
\]

\(\rho \neq -1\)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）。
其中\( a, b \)为常数。又由總機率法則，

\[
\mu_2 = E(Y) = E(E(Y|X)) = E(aX + b) = a\mu_1 + b.
\]

故得 \( b = \mu_2 - a\mu_1 \)，且

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

若能揲出 \( a = \rho(\sigma_2/\sigma_1) \) 使得證了。

\[
\rho\sigma_1\sigma_2 = \text{Cov}(X,Y)
\]

\[
= E((X - \mu_1)(Y - \mu_2))
\]

\[
= E(E((X - \mu_1)(Y - \mu_2)|X))
\]

\[
= E((X - \mu_1)(E(Y|X) - \mu_2))
\]

\[
= E((X - \mu_1) \cdot a(X - \mu_1))
\]

\[
= a\text{Var}(X) = a\sigma_1^2,
\]

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

\[
E(\text{Var}(Y|X)) = \text{Var}(Y) - \text{Var}(E(Y|X))
\]

\[
= \sigma_2^2 - \text{Var}(\mu_2 + \rho(\sigma_2/\sigma_1)(X - \mu_1))
\]

\[
= \sigma_2^2 - \rho^2(\sigma_2^2/\sigma_1^2)\text{Var}(X)
\]

\[
= \sigma_2^2 - \rho^2(\sigma_2^2/\sigma_1^2)\sigma_1^2 = \sigma_2^2(1 - \rho^2).
\]

證畢。

本節最後我們給定理5.3之推廣，其證明留在習題中(第18題)。

**定理5.6** 二随机变数 \( X, Y \)，及常数 \( a_1, b_1, c_1, a_2, b_2, c_2 \)，

(5.25) \[ \text{Cov}(a_1X + b_1Y + c_1, a_2X + b_2Y + c_2) \]
\[= a_1a_2 \text{Var}(X) + b_1b_2 \text{Var}(Y) + (a_1b_2 + a_2b_1) \text{Cov}(X, Y)\].

只要知道Var(X), Var(Y)及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)式成為(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
\[
\text{Cov}(X + Y, X - Y) \\
= \text{Var}(X) - \text{Var}(Y) + \text{Cov}(X, Y) - \text{Cov}(X, X) \\
= \text{Var}(X) - \text{Var}(Y),
\]
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變數常態分佈(multivariate normal distribution)的特性, 不但是一很重要的多變數常態分佈, 且二變數之相關係數, 亦為其聯合p.d.f.之一參數。先給一定義。

定義6.1 設隨機向量\((X, Y)\)為有參數\(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho\)之二變數常態分佈, 以\(\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)\)或\(B\mathcal{V}\mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)\)表之, 其中\(\mu_1, \mu_2 \in \mathbb{R}, \sigma_1 > 0, 0 < \rho < 1,\) 若其聯合p.d.f.為
\[
(6.1) \quad f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - \rho^2}} e^{-q/2}, x, y \in \mathbb{R},
\]
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pdf看起來是如此不可視，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) $X, Y$ 之連階分佈，分別為 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$。

(ii) $\rho(X, Y) = \rho$。

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

(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pdf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 $\rho$ 為 $X, Y$ 之相關係數，而我們已無次強調，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 $\rho$ 是否為 0 而定。可見 $\rho$ 在此分佈中之重要角色。

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

$$f(x, y) = \frac{1}{\sigma_2 \sqrt{1 - \rho^2} \sqrt{2\pi}} e^{-\frac{1}{2(1-\rho^2)}[y - (\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1))]^2} \cdot \frac{1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{1}{2\sigma_1^2}(x - \mu_1)^2} \cdot \frac{1}{\sigma_2 \sqrt{1 - \rho^2} \sqrt{2\pi}} e^{-\frac{1}{2(1-\rho^2)}[(y - (\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1))]^2} = A(x, y)B(x),$$

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

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

故

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

得证$f(x, y)$为一p.d.f.。

在上述推导中，因$X$之边际p.d.f.为

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

故$X$之边际分佈为$N(\mu_1, \sigma_1^2)$。同理可得$Y$之边际分佈为$N(\mu_2, \sigma_2^2)$。得证

性质(i)。

由于$Y|X = x$之p.d.f.为

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

即得

二变数常態分佈之另一性质:

(v) $Y|X = x$及$X|Y = y$等二条件分佈, 皆仍为常态分佈, $\forall Y|X = x$

$\sim N(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2))$, $X|Y = y \sim N(\mu_1 + \rho(\sigma_1/\sigma_2)(y - \mu_2), \sigma_1^2(1 - \rho^2))$。

由性质(v)再度看出, 若$X$与$Y$独立, 则$\rho$须为0。又

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

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

此二式可与定理5.5相對應。當然亦可寫出$E(X|Y = y)$及$\text{Var}(X|Y = y)$。注意此處条件變異數$\text{Var}(Y|X = x)$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)式的结果更强。又在统计学中,$E(Y|X = x)$有时称为迴歸函数(regression function)。而$E(Y|X = x)$之图形则称为给定$X = x$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$\mu_2$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

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

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

倒底$x$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，當關係係數較接近1或−1時，$X$與$Y$有較強的線性關係。

利用性質(iii)及總機率法則，可證明前述性質(ii):

$$
E(XY) = E(E(XY|X))
= E(XE(Y|X))
= E(X(\mu_2 + \rho(\sigma_2/\sigma_1)(X - \mu_1)))
= \mu_2 E(X) + \rho(\sigma_2/\sigma_1)E(X^2) - \rho \mu_1(\sigma_2/\sigma_1)E(X)
= \mu_1 \mu_2 + \rho(\sigma_2/\sigma_1)(\sigma_1^2 + \mu_1^2) - \rho \mu_1^2(\sigma_2/\sigma_1)
= \mu_1 \mu_2 + \rho \sigma_1 \sigma_2,
$$

故

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

由此即得$\rho(X, Y) = \rho$。

(vi) $(X, Y)$之聯合特徵函數為

$$
\phi(s, t) = E(e^{isX + itY})
= \exp\{i\mu_1 s + i\mu_2 t - (\sigma_1^2 s^2 + \sigma_2^2 t^2 + 2\sigma_1 \sigma_2 \rho st)/2\}, s, t \in R,
$$

動差母函數則為

$$
M(s, t) = E(e^{sX + tY})
= \exp\{\mu_1 s + \mu_2 t + (\sigma_1^2 s^2 + \sigma_2^2 t^2 + 2\sigma_1 \sigma_2 \rho st)/2\}, s, t \in R.
$$

(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，

$$
M(s, t) = E(e^{sX + tY})
= E(E(e^{sX + tY}|X))
= E(e^{sX} E(e^{tY}|X)).
$$
而因 \( Y \mid X = x \sim \mathcal{N}(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1), \sigma_2^2(1 - \rho^2)) \)，利用第二章(3.25)式，得

\[
E(e^{tY} \mid X = x) = \exp\{(\mu_2 + \rho(\sigma_2/\sigma_1)(x - \mu_1))t + \sigma_2^2(1 - \rho^2)t^2/2\}.
\]

故

\[
M(s, t) = E(\exp\{sX + (\mu_2 + \rho(\sigma_2/\sigma_1)(X - \mu_1))t + \sigma_2^2(1 - \rho^2)t^2/2\})
= \exp\{\mu_2 t - \rho(\sigma_2/\sigma_1)\mu_1 t + \frac{\sigma_2^2}{2} (1 - \rho^2)t^2/2\}E(\exp\{(s + \rho(\sigma_2/\sigma_1)t)X\})
= \exp\{\mu_2 t - \rho(\sigma_2/\sigma_1)\mu_1 t + \frac{\sigma_2^2}{2} (1 - \rho^2)t^2/2\}
\cdot \exp\{\mu_1 (s + \rho(\sigma_2/\sigma_1)t) + \frac{\sigma_2^2}{2} (s + \rho(\sigma_2/\sigma_1)t)^2/2\}
= \exp\{\mu_1 s + \mu_2 t + (\sigma_1^2 s^2 + \sigma_2^2 t^2 + 2\rho_1\rho_2 s t)/2\}.
\]

證畢。

由性質(vi)可得性質(ii): 令 \( aX + bY \) 之動差母函數為

\[
E(e^{t(aX + bY)}) = E(e^{aX + bY})
= \exp\{\mu_1 at + \mu_2 bt + (\sigma_1^2 a^2 t^2 + \sigma_2^2 b^2 t^2 + 2\rho_1\rho_2 ab t^2)/2\}
= \exp\{(a\mu_1 + b\mu_2)t + (a^2\sigma_1^2 + b^2\sigma_2^2 + 2\rho_1\rho_2\sigma_1\sigma_2)t^2/2\}.
\]

與第二章(3.25)式比較，即得 \( aX + bY \sim \mathcal{N}(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho_1\rho_2\sigma_1\sigma_2) \)。

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 \( a_1 b_2 - a_2 b_1 \neq 0, \) 則 \( Z_1, Z_2 \) 有 \( \mathcal{N}(\eta_1, \eta_2, \tau_1^2, \tau_2^2, \gamma) \) 分佈，其中

(6.9) 令 \( \eta_1 = a_1\mu_1 + b_1\mu_2, \quad \eta_2 = a_2\mu_1 + b_2\mu_2, \)

\[
\tau_1^2 = a_1^2\sigma_1^2 + b_1^2\sigma_2^2 + 2a_1 b_1 \rho_1 \sigma_1 \sigma_2, \quad \tau_2^2 = a_2^2\sigma_1^2 + b_2^2\sigma_2^2 + 2a_2 b_2 \rho_2 \sigma_1 \sigma_2, \\
\gamma = (a_1 a_2\sigma_1^2 + b_1 b_2\sigma_2^2 + (a_1 b_2 + a_2 b_1) \rho_1 \sigma_1 \sigma_2)/(\tau_1 \tau_2).
\]
由此也可看出$Z_1$與$Z_2$獨立，若且唯若

$$a_1a_2\sigma_1^2 + b_1b_2\sigma_2^2 + (a_1b_2 + a_2b_1)\rho\sigma_1\sigma_2 = 0.$$  

又若$X$與$Y$獨立，即$\rho = 0$，則上述$\tau_1, \tau_2, \gamma$可簡化不少。

利用定理5.5，此處檢驗$\gamma$確如(6.9)式中所給。首先

$$\Cov(Z_1, Z_2) = \Cov(a_1X + b_1Y, a_2X + b_2Y)$$
$$= a_1a_2\Var(X) + b_1b_2\Var(Y) + (a_1b_2 + a_2b_1)\Cov(X, Y)$$
$$= a_1a_2\sigma_1^2 + b_1b_2\sigma_2^2 + (a_1b_2 + a_2b_1)\rho\sigma_1\sigma_2.$$  

而

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

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。原因是由獨立的常態分佈，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

$$f(x, y) = \frac{1}{2\pi} e^{-(x^2+y^2)/2}(1 - \frac{xy}{(1+x^2)(1+y^2)}), x, y \in R.$$  

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.。

Ardelson(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隨際分佈為常態

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\(N(0, 1)\)分佈，\(Z\)與\(X\)獨立，且\(P(Z = 1) = P(Z = -1) = 1/2\)，又令\(Y = ZX\)。由

\[
P(Y \leq y|Z = 1) = P(ZX \leq y|Z = 1) = P(X \leq y|Z = 1)
\]

\[
= P(X \leq y),
\]

\[
P(Y \leq y|Z = -1) = P(ZX \leq y|Z = -1) = P(-X \leq y|Z = -1)
\]

\[
= P(-X \leq y) = P(X \leq y),
\]

注意\(X \overset{d}{=} (-X)\)，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\(N(0, 1)\)分佈。現因\(E(X) = E(Y) = E(Z) = 0\)，且
3.6 二變數常態分佈 233

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

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 \( \rho \) 。因此

\[
F_{X,Y}(a, b) = P(X \leq a, Y \leq b) = P\left( \frac{X - \mu_1}{\sigma_1} \leq \frac{a - \mu_1}{\sigma_1}, \frac{Y - \mu_2}{\sigma_2} \leq \frac{b - \mu_2}{\sigma_2} \right)
= F_{Z_1, Z_2}\left( \frac{a - \mu_1}{\sigma_1}, \frac{b - \mu_2}{\sigma_2} \right).
\]

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 \( \rho \) 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： \( h, k, \rho \) 。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 某地區男性身高可 \( \mathcal{N}(170, 7.2^2) \) 隨機分佈，女性身高可 \( \mathcal{N}(165, 5.4^2) \) 隨機分
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

\[
P(Y > X) = P(Y - X > 0)
\]
\[ P\left( \frac{Y - X - (-5)}{9} > \frac{0 - (-5)}{9} \right) = P(Z > \frac{5}{9}) = 1 - P(Z \leq 0.556) \]
\[ = 1 - 0.7109 = 0.2891, \]

其中 \( Z \sim N(0,1) \)。

**例6.3** 令上例。设 \( \rho(X,Y) = 0.6 \)。例如 \( X \) 代表一家庭中男孩之身高，\( Y \) 表女孩之身高。通常 \( X \) 与 \( Y \) 不独立。试求 \( P(Y > X) \)。

解。由性质(iii)，\( Y-X \sim N(-5,7.2^2+5.4^2-2\cdot0.6\cdot7.2\cdot5.4) = N(-5,34.344) \)。因此

\[
P(Y > X) = P(Y - X > 0)
= P\left( \frac{Y - X - (-5)}{\sqrt{34.344}} > \frac{0 - (-5)}{\sqrt{34.344}} \right)
= P(Z > 0.853) = 1 - P(Z \leq 0.853)
= 1 - 0.8031 = 0.1969.
\]

一般对二变数常态分布，若涉及的本质上只是一维的常态分布，如 \( aX + bY \)，便可求其概率值。例6.3中所求即为关于一维变数 \(-X+Y\) 之概率。

### 3.7 多维分佈

前几节我们主要讨论两个变数之多维分佈。一般当然有 \( n \) 维的随机向量 \( (X_1, \ldots, X_n) \), \( n \geq 2 \)。我们以祖体的 \( X \) 表随机向量 \( (X_1, \ldots, X_n) \)，而以祖体的 \( x \) 表观测值 \( (x_1, \ldots, x_n) \)。

随机向量 \( X = (X_1, \ldots, X_n) \) 以 \( \mathbb{R}^n \) 之一子集作为其样本空间。若 \( X \) 为一独立同型的随机向量 (样本空间假设为可数的)，则其联合pdf为

\[
f(x) = f(x_1, \ldots, x_n)
= P(X_1 = x_1, \ldots, X_n = x_n), (x_1, \ldots, x_n) \in \mathbb{R}^n.
\]

因此对任意\( A \subset \mathbb{R}^n \),

\[
P(X \in A) = \sum_{x \in A} f(x).
\]

若\( X \)为一连续型的随机向量，则其联合p.d.f.为一满足下述条件的函数\( f(x_1, \cdots, x_n) \):

\[
P(X \in A) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_1 \cdots dx_n.
\]

令\( g(x) = g(x_1, \cdots, x_n) \)为一实值函数，则\( g(X) \)为一随机变量，且对连续型及离散型，其期望值分别为

\[
E(g(X)) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x) f(x) dx,
\]

\[
E(g(X)) = \sum_{x \in \mathbb{R}^n} g(x) f(x).
\]

除了积分改在\( \mathbb{R}^n \)上，上述这些定义与二变数时的情况类似。另外，有时为了简便，仅写成“\( X_1, \cdots, X_n \)之联合p.d.f.”，省去\( X \)，对\( X_1, \cdots, X_n \)也可不用括号。

\((X_1, \cdots, X_n)\)之任一子集的边际p.d.f.，可将联合p.d.f.对其余变数之所有可能的值积分或相加而得。例如，首\( k \)个变数\((X_1, \cdots, X_k), 1 \leq k < n\)，之边际p.d.f.为

\[
f(x_1, \cdots, x_k) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_n) dx_{k+1} \cdots dx_n,
\]

或

\[
f(x_1, \cdots, x_k) = \sum_{(x_{k+1}, \cdots, x_n) \in \mathbb{R}^{n-k}} f(x_{k+1}, \cdots, x_n).
\]

条件分布也可类似地定义。如

\[
f(x_{k+1}, \cdots, x_n | x_1, \cdots, x_k) = \frac{f(x_1, \cdots, x_n)}{f(x_1, \cdots, x_k)}.
\]

如果为了更清楚，可将上式左侧写成

\[
f_{X_{k+1}, \cdots, X_n | X_1, \cdots, X_k}(x_{k+1}, \cdots, x_n | x_1, \cdots, x_k),
\]
特別是當 \( x_i \)'s \( i \) 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

\[
f(x, y, z) = \frac{1}{216}, \quad 1 \leq x \leq 6, x + 1 \leq y \leq x + 6, y + 1 \leq z \leq y + 6.
\]

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

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

故 \( f(x, y, z) \) 確為一 p.d.f.。

其次

\[
f_{X,Y}(x, y) = \sum_{z=y+1}^{y+6} f(x, y, z) = \frac{6}{216} = \frac{1}{36}, \quad 1 \leq x \leq 6, x + 1 \leq y \leq x + 6.
\]

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

\[
    f_Y(y) = \begin{cases} 
        \frac{y+6}{216} & 2 \leq y \leq 7, \\
        \frac{6(13-y)}{216} & 8 \leq y \leq 12.
    \end{cases}
\]

故得

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

比較\(f_{Z|X,Y}\)與\(f_{Z|Y}\)，可看出在給定\(Y = y\)之下，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) \hspace{1cm} f_{Z|X,Y}(z|x, y) = f_{Z|Y}(z|y)

與

(7.9) \hspace{1cm} 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\(X \sim \text{D-}\text{U}(1, 6)\)。讀者可以驗證\(Y \overset{d}{=} X_1 + X_2\)，其中\(X_1, X_2\)獨立且皆有\(\text{D-}\text{U}(1, 6)\)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，不過稍複雜些，在此略去。

例7.2 設\(X, Y, Z\)之聯合p.d.f.為

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

我們先證明\(f\)的確為一p.d.f.:


\[
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z) \, dx \, dy \, dz \\
= \int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} z^2 e^{-z(1+x+y)} \, dx \, dy \, dz \\
= \int_{0}^{\infty} \int_{0}^{\infty} -ze^{-z(1+x+y)} \bigg|_{0}^{\infty} \, dy \, dz \\
= \int_{0}^{\infty} \int_{0}^{\infty} ze^{-z(1+y)} \, dy \, dz \\
= \int_{0}^{\infty} -e^{-z(1+y)} \bigg|_{0}^{\infty} \, dz \\
= \int_{0}^{\infty} e^{-z} \, dz = 1.
\]

在上述過程中，也附帶求出

\[
f(z) = e^{-z}, z \geq 0,
\]

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

\[
f_{X,Y|Z}(x, y|z) = \frac{f(x, y, z)}{f_z(z)} = \frac{z^2 e^{-z(1+x+y)}}{e^{-z}} = z^2 e^{-z(x+y)}, x, y, z \geq 0.
\]

由上式得

\[
f_{X|Z}(x|z) = \int_{0}^{\infty} f_{X,Y|Z}(x, y|z) \, dy \\
= \int_{0}^{\infty} z^2 e^{-z(x+y)} \, dy = ze^{-zx}, x \geq 0, z \geq 0.
\]

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，\(X\) 与 \(Y\) 之条件独立，且皆有 \(\mathcal{E}(z)\) 分布。但若不給定 \(Z = z\)，則 \(X\) 与 \(Y\) 幾乎/>
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

\[
f_X(x) = \int_0^{\infty} \frac{2}{(1 + x + y)^3} dy = \frac{1}{(1 + x)^2}, \quad x \geq 0.
\]

同理可得 \( Y \) 之邊際 \( p.d.f. \) 為

\[
f_Y(y) = \frac{1}{(1 + y)^2}, \quad y \geq 0.
\]

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

\[
E(X) = \int_0^{\infty} x f_X(x) dx = \int_0^{\infty} \frac{x}{(1 + x)^2} dx = \infty,
\]

\[
E(Y) = \int_0^{\infty} y f_Y(y) dy = \int_0^{\infty} \frac{y}{(1 + y)^2} dy = \infty.
\]

不過 \( E(Z) = 1 \) 倒是存在的。

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

\[
f_{X,Z}(x,z) = \int_0^{\infty} f(x,y,z) dy = \int_0^{\infty} z^2 e^{-z(1+x+y)} dy = ze^{z(1+x)}, \quad x, z \geq 0.
\]

因此

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

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

\[
f_{Y|Z}(y|z) = ze^{-zy}, \quad y, z \geq 0.
\]

\[
f_{X,Y|Z}(x,y|z) = f_{X|Z}(x|z)f_{Y|Z}(y|z), \quad x, y, z \geq 0.
\]

讀者也可自行驗証 \( f_{X,Y|Z}(x,y,z) = f_{X|Z}(x|z)(\text{習題第1題}) \)。
多項分佈 (multinomial distribution) 為二項分佈之一推廣，此為重要之多維分佈。在此設 \( n, k \) 為正整數，且常數 \( p_1, \ldots, p_k \) 質足 \( 0 \leq p_i \leq 1, i = 1, \ldots, k, \sum_{i=1}^{k} p_i = 1 \)。則隨機向量 \( (X_1, \ldots, X_k) \) 記為有參數 \( n, p_1, \ldots, p_k \) 之多項分佈，以 \( \mathcal{M}(n, p_1, \ldots, p_k) \) 表之，若 \( (X_1, \ldots, X_k) \) 之聯合 p.d.f. 為

\[
 f(x_1, \ldots, x_k) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k},
\]

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

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:

\[
 f(x_1, x_2) = \binom{n}{x_1, x_2} p_1^{x_1} p_2^{x_2} = \binom{n}{x_1} p_1^{x_1} (1 - p_1)^{n-x_1}.
\]

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

\[
 f(x) = \binom{n}{x} p^x (1 - p)^{n-x}, x = 0, 1, \ldots, n,
\]

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 \( X_1, \ldots, X_n \) 表各箱中之球數，則 \( (X_1, \ldots, X_n) \) 表有 \( \mathcal{M}(n, 1/n, \ldots, 1/n) \) 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

\[
 P(X_1 = 1, X_2 = 2, X_3 = 3, X_4 = 4)
 = \frac{10!}{1!2!3!4!} 0.1^1 0.2^2 0.3^3 0.4^4
 = \frac{3628800}{288} \cdot 2.7648 \cdot 10^{-6}
 = 0.03483648.
\]
在(7.10)式中，

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

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，\ldots，第k類有x_k個，總共有n個。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, \ldots, 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確為一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B(n, p_k) 分佈。這點利用(7.6)式便可證明。對x_k \in \{0, 1, \ldots, n\}，我們必須將聯合p.d.f對所有可能的(x_1, \ldots, x_{k-1})相加。即對所有非負整數的x_1, \ldots, x_{k-1}, 且\sum_{i=1}^{k-1} x_i = n - x_k相加。以B表此集合。則

\[
f(x_k) = \sum_{(x_1, \ldots, x_{k-1}) \in B} \frac{n!}{x_1! \cdots x_{k-1}! x_k!} p_1^{x_1} \cdots p_k^{x_k},
\]

因\sum_{i=1}^{k-1} x_i = n - x_k，且p_1 + \cdots + p_{k-1} = 1 - p_k，利用多項式定理，上述和是(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B(n, p_k)分佈。

同理可得每一X_i皆有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。

\((X_1, \cdots, X_k)\) 表 \( n - x_k \) 次試驗後，各結果出現的次數。因此直觀上，給定 \( X_k = x_k \) 下，\((X_1, \cdots, X_k)\) 應仍有多項分佈。這是對的，利用(7.7)式得

\[
f(x_1, \cdots, x_{k-1}|x_k) = \frac{f(x_1, \cdots, x_k)}{f(x_k)} = \frac{n!}{x_1! \cdots x_k! p_1^{x_1} \cdots p_k^{x_k}} \frac{1}{n! (n - x_k) p_k^{x_k} (1 - p_k)^{n-x_k}} \]

\[
= \frac{(n - x_k)!}{x_1! \cdots x_{k-1}!} \left( \frac{p_1}{1 - p_k} \right)^{x_1} \cdots \left( \frac{p_k}{1 - p_k} \right)^{x_k}.
\]

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

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

其次我們介紹隨機向量獨立的概念。

**定義7.1** 設 \( X_1, \cdots, X_n \) 爲隨機向量，以 \( f(x_1, \cdots, x_n) \) 爲其聯合p.d.f.，且 \( X_i \) 之

邊際p.d.f. 爲 \( f_{X_i}(x_i), i = 1, \cdots, n \)。則若對 \( \forall (x_1, \cdots, x_n) \)，皆有

\[
f(x_1, \cdots, x_n) = \prod_{i=1}^n f_{X_i}(x_i),
\]

便稱 \( X_1, \cdots, X_n \) 爲相應獨立的隨機向量。若所有 \( X_i \) 皆為一維，則稱 \( X_1, \cdots, X_n \) 爲相應獨立的隨機變數。

如前，相應獨立如果不會混淆的話，有時只稱相應。若 \( X_1, \cdots, X_n \) 爲相應相應，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 \((X_1, \cdots, X_k)\)，其中 \( \{X_1, \cdots, X_k\} \) 爲 \( \{X_1, \cdots, X_n\} \) 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。

相應相應導致每對相應，若 \( X_1, \cdots, X_n \) 相應相應，則任 \( X_i, X_j, i \neq j \)，亦相應。但相應相應比每對相應邊強，即有可能每對相應，但不相應


立（见习题第15题）。亦可参考第一章例2.4，事件的每对独立，却互不相互独立的例子。

相互独立的随机变量，有许多好的性质。底下三个定理，其证明均类似在3.2节及3.3节中，两个变量情况下对应的定理，因此我们略去证明。

**定理7.1**（定理2.2(ii)之推广）。设 \( X_1, \ldots, X_n \) 为相互独立的随机变量，且函数 \( g_j(X_j), j = 1, \ldots, n \)，满足 \( E(g_j(X_j)) \) 存在。则

\[
E(g_1(X_1) \cdots g_n(X_n)) = \prod_{j=1}^{n} E(g_j(X_j)).
\]

**定理7.2**（定理2.3之推广）。\( X_1, \ldots, X_n \) 为相互独立的随机变量，若且唯若

\[
E(e^{it_1X_1 + \cdots + it_nX_n}) = \prod_{j=1}^{n} E(e^{it_jX_j}), t_j \in R, j = 1, \ldots, n.
\]

**系理7.1**。设 \( X_1, \ldots, X_n \) 为相互独立的随机变量。令 \( Z = X_1 + \cdots + X_n \)，则

\[
E(e^{itZ}) = \prod_{j=1}^{n} E(e^{itX_j}), t \in R.
\]

特别地，若 \( X_1, \ldots, X_n \) 有相同的分布，则

\[
E(e^{itZ}) = (E(e^{itX_1}))^n, t \in R.
\]

定理7.2及系理7.1，亦有母函数、拉普拉斯变换及动差母函数之版本。

定理2.4可以推广到 \( n \) 个独立随机变量的情况。我们只给 gamma 分布的结果。

**例7.4**。设 \( X_1, \ldots, X_n \) 为相互独立的随机变量，且 \( X_i \) 有 \( \Gamma(\alpha_i, \beta) \) 分布，\( i = 1, \ldots, n \)。因

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

故若令 \( Z = X_1 + \cdots + X_n \)，则
\[ E(e^{-sZ}) = \prod_{i=1}^{n} E(e^{-sX_i}) = \prod_{i=1}^{n} \frac{1}{(1 + \beta s)^{\alpha_i}} \]

\[ = \frac{1}{(1 + \beta s)^{\sum_{i=1}^{n} \alpha_i}}, \quad s \geq 0, \]

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，只要尺度係數皆相同，其和便仍為 gamma 分佈。

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

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 \(X_1, \ldots, X_n\) 為相互獨立的隨機變數，分別有 \(\mathcal{N}(\mu_i, \sigma_i^2)\) 分佈， \(i = 1, \ldots, n\)。又設 \(a_1, \ldots, a_n, b_1, \ldots, b_n\) 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

首先 \(X_i\) 之動差母函數為
\[ M_{X_i}(t) = E(e^{tX_i}) = e^{\mu_i t + \sigma_i^2 t^2/2}, \quad t \in R. \]

因此 \(Z\) 之動差母函數為
\[ M_Z(t) = E(e^{tZ}) = e^{t \sum_{i=1}^{n} b_i} \prod_{i=1}^{n} M_{X_i}(a_i t) \]
$$
= e^{t \sum_{i=1}^{n} b_i} \prod_{i=1}^{n} e^{\mu_i a_i t + \sigma_i^2 t^2 / 2} \\
= e^{\left( \sum_{i=1}^{n} (a_i \mu_i + b_i) \right) t + \left( \sum_{i=1}^{n} \frac{a_i^2 \sigma_i^2}{2} \right) t^2}, t \in \mathbb{R}.
$$

而此為 $\mathcal{N} \left( \sum_{i=1}^{n} (a_i \mu_i + b_i), \sum_{i=1}^{n} a_i^2 \sigma_i^2 \right)$ 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 $X_1, \cdots, X_n$ 爲隨機向量。則 $X_1, \cdots, X_n$ 爲相互獨立，若且唯若存在函數 $g_i(x_i), i = 1, \cdots, n$，使得 $(X_1, \cdots, X_n)$ 之聯合p.d.f.可寫成

$$
f(x_1, \cdots, x_n) = g_1(x_1) \cdots g_n(x_n).
$$

底下為定理3.2之推廣。

定理7.4 設 $X_1, \cdots, X_n$ 爲獨立的隨機向量。令 $g_i(x_i)$ 為 $X_i$ 的函數，$i = 1, \cdots, n$。則 $U_i = g_i(X_i), i = 1, \cdots, n$，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的立即的推廣。例如，在做變換時，雅可比為 $n \times n$ 矩陣之行列式的值。

底下我們給定定理3.1之推廣。

例7.7 設 $X, Y, Z$ 爲相互獨立，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 $\mathcal{B}(p, q), \mathcal{B}(p+q, r)$ 及 $\Gamma(p+q+r, \lambda)$ 分佈。

證明。首先 $(X, Y, Z)$ 之聯合p.d.f.為

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

令

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

分別有 $\mathcal{B}(p, q), \mathcal{B}(p+q, r)$ 及 $\Gamma(p+q+r, \lambda)$ 分佈。
第三章 多维随机变量

解出

\[ x = uvw, \]
\[ y = -uvw + vw = vw(1-u), \]
\[ z = -uvw + w = u(1-v). \]

则

\[
J = \begin{vmatrix}
    vw & uv & uv \\
    -uvw & -uvw + w & -uv + v \\
    0 & -w & -v + 1
\end{vmatrix} = uv^2.
\]

故得

\[
f_{U,V,W}(u,v,w) = (uvw)^{p-1}e^{-uvw/p} \gamma \left( \frac{uv(1-u)}{\lambda}, r-1, \frac{uv-w}{\lambda}, \lambda \right)
\]
\[
= \frac{\lambda^{q+r}e^{-w/\lambda}}{\Gamma(p)\Gamma(q)\Gamma(r)}
\]
\[
= \frac{\lambda^{q+r}e^{-w/\lambda}}{\Gamma(p)\Gamma(q)\Gamma(r)}
\]
\[
\Gamma(p+q+1) \Gamma(p+q+r) \Gamma(p+r)
\]
\[
\Gamma(p+q+1) \Gamma(p+q+r) \Gamma(p+r)
\]

由此即知 \(U, V, W\) 独立，且分别有 \(\text{Be}(p,q), \text{Be}(p+q,r)\) 及 \(\Gamma(p+q+r, \lambda)\) 分布。

事实上此结果可推广到 \(n\) 维的情况，见第五章例 3.11。

底下为进一步有关随机变量之和的期望值及变异数的定理。证明并不困难，因此留在习题中（第 11 题）。

**定理 7.5** 设有随机变量 \(X_1, \ldots, X_n\)。

(i) 若 \(E(X_i)\) 存在，\(i = 1, \ldots, n\)，则对任意常数 \(a_1, \ldots, a_n\)，

\[
E(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i E(X_i).
\]

(7.16)
(ii) 若 Var\((X_i)\) 皆存在, \(i = 1, \cdots, n\), 則對任意常數\(a_1, \cdots, a_n\),

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

若 \(X_1, \cdots, X_n\) 為每對獨立, 則

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

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

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

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, 則可用利用定理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 \(Y_1, \cdots, Y_n\) 為有獨立的\(\text{Ber}(p)\) 分佈之隨機變數, 則 \(X = Y_1 + \cdots + Y_n\) 有 \(\text{B}(n, p)\) 分佈。由於

\[
E(Y_i) = p, \quad \text{Var}(Y_i) = p(1-p), \quad i = 1, \cdots, n,
\]

利用(7.16)及(7.18)式, 可得

\[
E(X) = \sum_{i=1}^{n} E(Y_i) = np,
\]

\[
\text{Var}(X) = \sum_{i=1}^{n} \text{Var}(Y_i) = np(1-p) .
\]

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$k$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$A_1, \ldots, A_n$既不互斥亦不獨立。但$P(A_i) = 1/n$, 因此$E(I_{A_i}) = P(A_i) = 1/n, \forall i = 1, \ldots, n$. 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$i$個箱是否为空箱。則由(7.16)式，得
\[ E(N) = \sum_{i=1}^{n} E(I_i) = n \cdot P(\text{第i箱为空箱}) \]

第i箱为空箱表n个球全要放進其餘的n − 1箱中，故
\[ P(\text{第i箱为空箱}) = (1 - \frac{1}{n})^n \]

因此
\[ E(N) = n \cdot (1 - \frac{1}{n})^n \]

例如，若n = 10，则
\[ E(N) = 10 \cdot (\frac{9}{10})^{10} \approx 3.486(箱) \]

這就是第二章例2.11中，10个球随机地放进10箱子中，空箱數之期望值。但此处的推導显然容易多了。

空箱数佔全部箱数之比例N/n的期望值为
\[ \frac{E(N)}{n} = (1 - \frac{1}{n})^n \]

此即每一箱子會是空箱的機率。附帯一提，在第二章例2.11之後，我們提過每一箱的球数有\(\text{B}(n, 1/n)\)分佈 \((1 - 1/n)^n\)为此分布取值0之機率。當n → ∞時，此機率趨近至\(e^{-1} = 0.367879\)。此值刚好是\(\text{P}(1)\)分佈取值0之機率。此點由泊松近似亦可看出。

最後利用(7.17)式，可求出
\[
(7.20) \quad \text{Var}(N) = n \cdot ((1 - \frac{1}{n})^n - (1 - \frac{1}{n})^{2n}) \\
+ n(n - 1)((1 - \frac{2}{n})^n - (1 - \frac{1}{n})^{2n})
\]

這部分留給讀者自行完成(習題第21题)。

例7.11 某食品公司为了促销其产品，在每盒餅乾中均隨機地放一张风景卡片，且設總共有N张不同的卡片。
我們先求若每次只買一盒，欲收集齊全N張卡片所需買的盒數之期望值。可看出當已有n張不同卡片時，0 ≤ n < N，欲得一張尚未有的卡片，所需買的盒數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張卡片所需買的盒數X_0 + X_1 + ⋯ + X_{N−1}之期望值為

\[ E(X_0 + X_1 + ⋯ + X_{N−1}) = \sum_{n=0}^{N−1} \frac{N}{N−n} = N(1 + \frac{1}{2} + ⋯ + \frac{1}{N}) \]

由微積分計算結果知，當N → ∞時，

(7.21) \[ 1 + \frac{1}{2} + ⋯ + \frac{1}{N} = \log N \longrightarrow \gamma, \]

其中γ為歐拉常數(Euler's constant, Euler(1707-1783)之有史以來最多產的數學家，其生平介紹見黃文華(1999)第十一章)，其值約為0.5772。因此N很大時，所需買的盒數之期望值可以N(log N + γ)來估計。不過要注意N(1 + 1/2 + ⋯ + 1/N)與N(log N + γ)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, ⋯, Y_N，其中Y_i = 1，若第i張卡片在這T盒中，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 + Nγ。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

\[ y = (1 - \frac{1}{N})^N \log N. \]

则

\[
\log y = N \log N \cdot \log(1 - \frac{1}{N}) \\
= N \log N \cdot (-\frac{1}{N} - \frac{1}{2N^2} - \frac{1}{3N^3} - \cdots) \\
= -\log N \cdot (1 + \frac{1}{2N} + \frac{1}{3N^2} + \cdots) \\
< -\log N.
\]

故得 \( y < 1/N \)。因此若一次买 \([N \log N] + 1\) 盒，其中 \([x]\) 表示小于或等于 \(x\) 的最大整数，则所获得不同卡片张数之期望值

\[
N(1 - (1 - \frac{1}{N})^{[N \log N]+1}) \\
> N(1 - (1 - \frac{1}{N})^{N \log N}) > N(1 - \frac{1}{N}) = N - 1.
\]

由于最多只能得到 \(N\) 张不同的卡片，故得此时期望值已算是很接近 \(N\) 了。

在2.2.5节我们曾介绍超几何分布，但并未说明期望值及变异数是如何推导出来的。直接计算可能较繁琐，利用多变数倒是可求出来。见下例。

例7.12 设有 \(H(N, D, n)\) 分布，即 \(X\) 之 pdf 为

\[
(7.22) \quad f(x) = \binom{D}{x} \binom{N-D}{n-x}, \max\{0, n - N + D\} \leq x \leq \min\{n, D\}.
\]

底下求 \(X\) 之期期值及变异数。

仍以2.2.5节取球为例。令 \(Y_i = 1\) 或 0，依取出的第 \(i\) 个球是否为白球而定，\(i = 1, \cdots, n\)（这些球的编号并不重要，因每次取出后不放回，所以可以一口气取出 \(n\) 个，再任意编号）。则每一 \(Y_i\) 皆有 \(Ber(D/N)\) 分布，\(\forall X = Y_1 + \cdots + Y_n\)。因此 \(E(Y_i) = D/N, \forall \text{Var}(Y_i) = D(N-D)/N^2, i = 1, \cdots, n\)。再利用(7.16) 式，便可求出 \(X\) 之期望值：
\[(7.23) \quad E(X) = \sum_{i=1}^{n} E(Y_i) = nE(Y_1) = nD/N.\]

\[Y_1, \ldots, Y_n\]並不獨立，我們利用(7.17)式來求\(\text{Var}(X)\)。首先對\(i \neq j\)，

\[
\text{Cov}(Y_i, Y_j) = E(Y_i Y_j) - E(Y_i)E(Y_j) = P(Y_i = 1, Y_j = 1) - (D/N)^2
\]

\[
= \frac{D}{N} - \frac{D}{N}^2
\]

\[
= -\frac{D(N-D)}{N^2(N-1)}.
\]

其中用到\(Y_i = 1 \text{且} Y_j = 1\)，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\(Y_i = 1 \text{且} Y_j = 1\)之機率為\(\frac{\binom{D}{2}}{\binom{N}{2}}\)（讀者可否想通為什麼共變異數為負）。由是得

\[(7.24) \quad \text{Var}(X) = n\text{Var}(Y_1) + 2\binom{n}{2}\text{Cov}(Y_1, Y_2)
\]

\[
= \frac{nD(N-1)}{N^2} - \frac{n(n-1)D(N-D)}{N^2(N-1)}
\]

\[
= \frac{nD(N-D)(N-n)}{N^2(N-1)}.
\]

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\(B(n, D/N)\)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\(H(N, Np, n)\)分佈，\(Z\)有\(B(n, p)\)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，\(Y_1, \ldots, Y_n\)便近似獨立，因此\(X\)可以\(B(n, p)\)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\(B(n, p)\)，當做樣本中對該候選人之
支持人數$X$的分佈，而$X$原本是有多$(N, np, n)$分佈的。事實上可以證明$N \to \infty $時（且$n, p$固定）

$$P(X = k) \to P(Z = k).$$

當然若$n$與$N$相比並非很小，就不適合以$B(n, p)$分佈當做$H(N, np, n)$分佈之近似。

對隨機向量$\mathbf{X} = (X_1, \cdots , X_n)$，令$\mu_\mathbf{X}$及$\Sigma_\mathbf{X}$分別表其期望向量(mean vector)及共變異數矩陣(covariance matrix)，也可僅以$\mu, \Sigma$表之，其中

$$\mu = (\mu_1, \cdots , \mu_n),$$

$$\Sigma = \begin{pmatrix}
\sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\
\sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2
\end{pmatrix},$$

而

$$\mu_i = E(X_i), i = 1, \cdots , n,$$

$$\sigma_i^2 = Var(X_i), i = 1, \cdots , n,$$

$$\sigma_{ij} = \sigma_{ji} = Cov(X_i, X_j), i, j = 1, \cdots , n.$$ 令

$$\rho_{ij} = \rho(X_i, X_j), i, j = 1, \cdots , n.$$ 則

$$\sigma_{ij} = Cov(X_i, X_j) = \rho_{ij} \sigma_i \sigma_j, i, j = 1, \cdots , n.$$  故$\Sigma$亦可寫成

$$\Sigma = \begin{pmatrix}
\sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 & \cdots & \rho_{1n} \sigma_1 \sigma_n \\
\rho_{12} \sigma_1 \sigma_2 & \sigma_2^2 & \cdots & \rho_{2n} \sigma_2 \sigma_n \\
\vdots & \vdots & \ddots & \vdots \\
\rho_{1n} \sigma_1 \sigma_n & \rho_{2n} \sigma_2 \sigma_n & \cdots & \sigma_n^2
\end{pmatrix}.$$
對一列向量 \( \mathbf{a} = (a_1, \ldots, a_n) \),

\[
\mathbf{a}' = \begin{pmatrix}
  a_1 \\
  \vdots \\
  a_n
\end{pmatrix},
\]

\( \mathbf{a}' \) 即為 \( \mathbf{a} \) 之轉置 (transpose)，為一行向量。則

\[
(X - \mu)'(X - \mu)
= \begin{pmatrix}
  X_1 - \mu_1 \\
  X_2 - \mu_2 \\
  \vdots \\
  X_n - \mu_n
\end{pmatrix}
\begin{pmatrix}
  X_1 - \mu_1, X_2 - \mu_2, \ldots, X_n - \mu_n
\end{pmatrix}
= \begin{pmatrix}
  (X_1 - \mu_1)^2 & \cdots & (X_1 - \mu_1)(X_n - \mu_n) \\
  \vdots & \ddots & \vdots \\
  (X_n - \mu_n)(X_1 - \mu_1) & \cdots & (X_n - \mu_n)^2
\end{pmatrix}.
\]

上述 \( \mu \) 及 \( \Sigma \) 便可分別寫成

\[
\mu = E(X),
\Sigma = E((X - \mu)'(X - \mu)).
\]

例7.13 設 \((X_1, X_2) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) \)。則

\[
\mu = (\mu_1, \mu_2),
\Sigma = \begin{pmatrix}
  \sigma_1^2 & \rho \sigma_1 \sigma_2 \\
  \rho \sigma_1 \sigma_2 & \sigma_2^2
\end{pmatrix}.
\]

例7.14 設 \((X_1, X_2, X_3) \) 之共變異數矩陣為

\[
\Sigma = \begin{pmatrix}
  2 & -0.5 & 0.2 \\
  -0.5 & 6 & -0.4 \\
  0.2 & -0.4 & 1
\end{pmatrix}.
\]
試求 $\text{Cov}(2X_1 + X_3, X_1 - X_2 - X_3)$。

解. 由定理7.5，
\[
\text{Cov}(2X_1 + X_3, X_1 - X_2 - X_3) \\
= 2 \text{Cov}(X_1, X_1) - 2 \text{Cov}(X_1, X_2) - 2 \text{Cov}(X_1, X_3) + \text{Cov}(X_1, X_3) \\
- \text{Cov}(X_2, X_3) - \text{Cov}(X_3, X_3) \\
= 2 \cdot 2 - 2 \cdot (-0.5) - 0.2 - (-0.4) - 1 \\
= 4 + 1 - 0.2 + 0.4 - 1 = 4.2.
\]

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 $p, q > 0$，立得 $p > 1, q > 1$。現將定$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 $p/(p-1) = q$)
\[
g(v^{1/(p-1)}) = \frac{1}{p}v^{p/(p-1)} + \frac{1}{q}v^{q/(p-1)} - v^{1/(p-1)}v.
\]
\[
\left( \frac{1}{p} + \frac{1}{q} - 1 \right) v^q = 0.
\]

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\(\lim_{u \to 0} g(u) = v^q / q > 0\)，\(\lim_{u \to \infty} g(u) = \infty\)。故唯一的絕對極小值，發生在\(u = v^{1/(p-1)}\)，且極小值為0。故\(g(u) \geq 0, \forall u \in R\)，\(g(u) > 0, \forall u \neq v^{1/(p-1)}\)。得証(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

\[
E(|XY|) \leq (E(|X|^p))^{1/p}(E(|Y|^q))^{1/q}.
\]

**証明**. 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。証畢。

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

\[
E(|XY|) \leq (E(X^2))^{1/2}(E(Y^2))^{1/2}.
\]
3.8 不等式 257

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

$$E(|X|) \leq (E(|X|^p))^{1/p}, 1 < p < \infty.$$  

在此對一隨機變數$Y$及常數$c, Y \equiv c$, 請做“$Y$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

$$E(|X|^r) \leq (E(|X|^p))^{1/p}.$$  

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), 爲俄國機率學家)。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$E(|X|), E(|X|^2)及E(|X|^3)$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

$$E(|X + Y|^p)^{1/p} \leq (E(|X|^p))^{1/p} + (E(|Y|^p))^{1/p}.$$  

証明. 首先利用三角形不等式(Triangle inequality),

$$|X + Y| \leq |X| + |Y|,$$

可得

$$E(|X + Y|^p) = E(|X + Y||X + Y|^{p-1})$$

$$\leq E(|X||X + Y|^{p-1}) + E(|Y||X + Y|^{p-1}).$$

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 \( a_i, b_i, i = 1, \ldots, n, \)

\[(8.8) \quad \sum_{i=1}^{n} |a_i b_i| \leq (\sum_{i=1}^{n} |a_i|^p)^{1/p} (\sum_{i=1}^{n} |b_i|^q)^{1/q}, \quad \frac{1}{p} + \frac{1}{q} = 1, p, q > 0, \]

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 \( a_1, b_1, \ldots, (a_n, b_n) \)，便導致 (8.8) 式(習題第 8 题)

明考夫斯基之數值的版本則為：對任意實數 \( a_i, b_i, i = 1, \ldots, n, \) 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 \( b_i = 1, i = 1, \ldots, n \)，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瞬森不等式 (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為凸函數若且唯若−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) = x²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](image)

凹然不等式。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) \[ E(g(X)) \geq g(E(X)) \]  

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) ≥ ax + b。又因期望值會保持大小關係，故

\[
E(g(X)) \geq E(ax + b) \\
= aE(X) + b \\
= h(E(X)) \\
= g(E(X)),
\]
其中最後一等式成立乃是由函數 \( y = h(x) \) 切圖形於 \((E(X), g(E(X)))\)，故 \( h \) 與 \( g \) 在 \( E(X) \) 的值相等。

底下為柯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 \( a_1, \cdots, a_n \) 皆大於 0，令

\[
M_A = \frac{1}{n}(a_1 + \cdots + a_n) = \frac{1}{n}\sum_{i=1}^{n} a_i,
\]

\[
M_G = (a_1 \cdots a_n)^{1/n} = \left(\prod_{i=1}^{n} a_i\right)^{1/n},
\]

\[
M_H = \frac{1}{\frac{1}{n/a_1} + \cdots + \frac{1}{a_n}} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} a_i^{-1}}^{-1},
\]

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

\[
(8.16) \quad M_A \geq M_G \geq M_H.
\]

底下我們看如何以柯森不等式證明上述不等式。

設隨機變數 \( X \) 滿足 \( P(X = a_i) = 1/n, i = 1, \cdots, n \)。因 \( y = \log x \)，\( x > 0 \) 為一凸函數，柯森不等式導致

\[
E(\log X) \leq \log(E(X)).
\]

故
\[ \log M_G = \frac{1}{n} \sum_{i=1}^{n} \log a_i = E(\log X) \leq \log(E(X)) = \log\left(\frac{1}{n} \sum_{i=1}^{n} a_i\right) = \log M_A. \]

因此 \( M_A \geq M_G \) 。其次因 \( y = 1/x, x > 0 \) 为凸函数，故

\[ \frac{1}{M_H} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i} = E\left(\frac{1}{X}\right) \geq \frac{1}{E(X)} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i} = \frac{1}{M_A}. \]

因此 \( M_A \geq M_H \) 。

甚至 \( M_G \geq M_H \) ，可利用上述推导而得，

\[ \frac{1}{M_H} = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{a_i} \geq \frac{1}{(\prod_{i=1}^{n} a_i)^{1/n}} = \frac{1}{(\prod_{i=1}^{n} a_i)^{1/n}} = \frac{1}{M_G}. \]

其中不等式，我们是利用 \( M_A \geq M_G \) 于正数 \( \frac{1}{a_1}, \ldots, \frac{1}{a_n} \) 。

**柴比雪夫不等式** (Chebyshev's inequality) 可以说是概率里非常重要且有用的不等式。此不等式的重要，是因其适用性广泛，而非其精确性。Chebyshev(1821-1894，有时写成Chebychev, Shebychev, Tschebyshev等)，为十九世纪俄国极重要的一位概率学家。

**柴比雪夫不等式**。设 \( X \) 为一随机变量，\( g(x) \) 为一非负函数，且满足 \( E(g(X)) \) 存在。则对任意 \( c > 0 \)，

\[ P(g(X) \geq c) \leq \frac{E(g(X))}{c}. \]

**证明**。我们只证连续型的情况。设 \( X \) 为一连续型的随机变量，以 \( f(x) \) 为其 p.d.f.，则

\[ E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx \geq \int_{\{x \mid g(x) \geq c\}} g(x) f(x) dx \geq c \int_{\{x \mid g(x) \geq c\}} f(x) dx = cP(g(X) \geq c). \]
262 第三章 多維隨機變數

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) \[P(|X - \mu| \geq c) \leq \frac{E|x - \mu|^r}{c^r}, r > 0.
\]

由(8.18)式，又得（取\(c = k\sigma\)，\(r = 2\)，其中\(\sigma^2 = \text{Var}(X)\))

(8.19) \[P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}, k > 0.
\]

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) \[P(|X - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}, \varepsilon > 0.
\]

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\(P(1)\)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

\[P(|X - 1| \geq 2) = P(X \geq 3) = 1 - \sum_{i=0}^{2} P(X = i)
\]
\[= 1 - \sum_{i=0}^{2} \frac{e^{-1}}{i!} = 1 - e^{-1} \left(\frac{1}{1} + \frac{1}{1} + \frac{1}{2}\right)
\]
\[= 1 - 2.5e^{-1} \approx 0.080.
\]
的確是小於$1/4 = 0.25$，但顯然誤差不小。換句話說，雖然$(X - \mu) / \varepsilon$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$X$偏離期望值$k$個標準差及一固定距離$\mu$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

\begin{align*}
(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.
\end{align*}

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。

$$
P(|X - \mu| < \sigma) = P(|Z| < 1) \approx 0.682, \quad 1 - \frac{1}{1^2} = 0,$$
$$
P(|X - \mu| < 2\sigma) = P(|Z| < 2) \approx 0.954, \quad 1 - \frac{1}{2^2} = 0.75,$$
$$
P(|X - \mu| < 3\sigma) = P(|Z| < 3) \approx 0.997, \quad 1 - \frac{1}{3^2} \approx 0.889.
$$

習題

第3.1節

1. 試證下述二函數皆為聯合p.d.f。
(i) \( f(x, y) = \frac{x+y}{3}, \quad x = 1, 2, 3, y = 1, 2, 3, 4; \)

(ii) \( f(x, y) = e^{-x(y+1)^2}, \quad x, y > 0. \)

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 \( U \overset{d}{=} X, V \overset{d}{=} Y \)，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

1. 試求出例2.1中的各 \( f_{X|Y}(x|y) \) 之值。

2. 設 \( f(x,y) \) 如例1.4中所給。試求 \( f_X(x), f(y|x), E(Y|x) \) 及 \( \text{Var}(Y|x) \)。

3. 試證(2.11)式。

4. 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

   \[ f(x, y) = \begin{cases} 4e^{-2x-2y} - 2e^{-2x-y} - 2e^{-x-2y} & , 0 < x, y < \infty, \\ +2e^{-x-y} & , \text{其他} \end{cases} \]

   (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. 試求例2.12中，\(X, Y\) 之邊際分佈。

11. 設 \(X, Y\) 之聯合p.d.f.為 \(f(x, y) = 2e^{-x-y}, 0 \leq x < \infty\)。
   (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) 試驗 \(X\) 與 \(Y\) 是否獨立。
12. 设 \( X, Y \) 之联合p.d.f.为 \( f(x, y) = ye^{-xy}, x, y > 0 \)。
   (i) 求 \( X \) 與 \( Y \) 是否独立；
   (ii) 求 \( E(Y | x) \) 及 \( \text{Var}(Y | x) \)。

13. 设 \( X \) 有 \( \mathcal{E}(1) \) 分佈，令 \( Y = [X + 1] \)，其中 \([\cdot]\) 表最大整数函数。
   (i) 求 \( Y \) 之分佈；
   (ii) 求给定 \( Y \geq 5, X - 4 \) 之分佈。

14. 試證定理2.4(ii)-(vii)。

15. 设 \( X, Y \) 爲二独立的随机变数，且皆有 \( \mathcal{N}(0, 1) \) 分佈。求 \( P(X^2 + Y^2 < 1) \)。

16. 设 \( X, Y \) 爲二独立的随机变数，且分别有 \( \mathcal{P}(\lambda) \) 及 \( \mathcal{P}(\mu) \) 分佈。求给定 \( X + Y = n \) 之下， \( n \geq 1 \)，\( X \) 之条件p.d.f.。並指出此為那一常見的分佈，參數为何。

17. 设 \( X, Y \) 为二独立的随机变数，且分别有 \( \mathcal{B}(m, p) \) 及 \( \mathcal{B}(n, p) \) 分佈。求给定 \( X + Y = k \) 之下， \( k \geq 1 \)，\( X \) 的条件p.d.f.。给定k的範圍，並指出此為那一常見的分佈，參數为何？

18. 设 \( X_1, X_2 \) 爲二独立的随机变数，且皆有 \( \mathcal{U}(0, 1) \) 分佈。試證 \( X_1 + X_2 \) 並無均匀分佈。

19. 设 \( X_1, X_2 \) 爲二独立的随机变数，且分别有 \( \mathcal{E}(\lambda_1) \) 及 \( \mathcal{E}(\lambda_2) \) 分佈，\( \lambda_1 \neq \lambda_2 \)。試求 \( X_1 + X_2 \) 之分佈，並問此是否为gamma分佈。

20. 设 \( (X, Y) \) 之联合p.d.f.为 \( f(x, y) = 2e^{-x-y}, 0 < x < y < \infty \)。求 \( (X, Y) \) 之联合密度函数。(解。\( M(s, t) = 2((1 - t)(2 - s - t))^{-1}, t < 1, s + t < 2) \)

21. 设 \( (X, Y) \) 之联合p.d.f.为 \( f(x, y) = e^{-y}, 0 < x < y < \infty \)。求 \( (X, Y) \) 之联合密度函数。(解。\( M(s, t) = ((t - 1)^{-1} - (s + t - 1)^{-1})/s, t < 1, s + t < 1) \)
22. 设$X$与$Y$独立，且$X \sim \chi^2_m$, $Y \sim \chi^2_n$。试验当$m > n$时，$X - Y$是否有卡方分布，若不是则给出其自由度。(解：否)

23. 设$X$与$Y$独立，令$S = X + Y$。又知$X \sim \chi^2_m$, $S \sim \chi^2_{m+n}$。试验$Y$是否有卡方分布，若不是则给出其自由度。(解：是，$Y \sim \chi^2_n$)

24. 試證(2.25)式所定義的函數為$-p.d.f.$。

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. 设$f$为一连续型的$p.d.f.$，且为偶函数，$G$为一连续型的分佈函数，

以$G$为$p.d.f.$，$G$亦为偶函数。試證對$\forall \lambda \in R$, $2G(\lambda y) f(y), -\infty < y < \infty$, 為$-p.d.f.$。

第3.3節

1. 試以變數代換的方法，證明定理2.1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$\mathcal{B}(\alpha, \beta)$分佈及$\mathcal{B}(\alpha, \beta)$分佈。試證$XY$有$\mathcal{B}(\alpha, \beta + \gamma)$分佈。

5. 設$R > 0$，且$R^2$有$\chi^2$分佈。又設$\theta$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 \( Z \) 為連続型，\( W \) 為離散型)；
(ii) 試證 \( P(Z \leq z \mid W = i) = P(Z \leq z), i = 0, 1, \) 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 \( (X, Y) \) 為連続型之隨機向量，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.，並驗 \( U \) 與 \( W \) 是否獨立。

10. 設 \( X \) 與 \( Y \) 獨立，且皆有 \( \mathcal{U}(\theta, 1) \) 分佈。試證若 \( a, b \geq 0 \) 且 \( a + b = 1 \)，則 \( W = aX + bY \) 亦有 \( \mathcal{U}(\theta, 1) \) 分佈。

11. 設 \( X \) 與 \( Y \) 獨立，且皆有 \( \mathcal{U}(\alpha, \beta) \) 分佈，\( \alpha \leq \beta \)。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) 試驗$U$與$V$是否獨立。

13. 試驗例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$N(0,1)$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$U[0,1]$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$(V,R)$有二變數之裴瑞希里分佈。

(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$E(1)$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$U(0,1)$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$\Lambda$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$B(N, p)$分佈，$N|\Lambda$有$P(\Lambda)$分佈，且$\Lambda$有$\Gamma(\alpha, \beta)$分佈，試求$Y$之非條件分佈。

3. 設$Y|P$有$B(n, P)$分佈，$P$有$Be(\alpha, \beta)$分佈。
   (i) 試證$Y$之非條件分佈為
   
   $$P(Y = y) = \binom{n}{y} \frac{\Gamma(\alpha + \beta) \Gamma(y + \alpha) \Gamma(n - y + \beta)}{\Gamma(\alpha + \beta + n)}$$
   
   (ii) 試求$E(Y)$及$Var(Y)$。

4. 設$Var(X) = E(Var(X|Y))$。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$P(Y)$分佈，且$Y$亦有$P(\lambda)$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$N(Y, \sigma^2)$分佈，且$Y$有$N(\mu, \tau^2)$分佈，其中$\sigma^2, \mu, \tau^2$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$Cov(X, Y) = 0$；
   (ii) 設$\sigma_X = \sigma_Y$，試證$Cov(X + Y, X - Y) = 0$。

2. 在例5.1中，試求$\rho(X, Y)$，$Cov(2X + 3, -3Y + 4)$及$\rho(2X + 3, -3Y + 4)$。

3. 設隨機變數$X, Y$滿足$Var(X) = 1, Var(Y) = 2$，且$\rho(X, Y) = 1/2$。試求$Var(X - 2Y)$。
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ρ(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ρ(X, Y)。

6. 設有一公正的骰子，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ρ(X, Y)。

7. 設X之p.d.f.為一偶函數，且E(X^1) < ∞。試證X與X^2無相關。

8. 設X, Y之聯合p.d.f.為

\[ f(x, y) = p^2(1 - p)^y, x, y = 0, 1, 2, \ldots, \text{且} x \leq y, \]

其中0 < p < 1。

(i) 試證X, Y之邊際分佈，分別為自0開始之NB(1, p)及NB(2, p)分佈；
(ii) 試求Cov(X, Y)及ρ(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所示。試求ρ(X, Y)。

10. 設X, Y之聯合p.d.f.為f(x, y) = c, x^2 + y^2 ≤ 1。

(i) 試求c之值；
(ii) 試證μ_X = μ_Y = ρ(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 ≤ x, y ≤ 1, x + y ≤ 1。試求Cov(X, Y)，並對Cov(X, Y)之符號給一解釋。

12. 設X與Y獨立，且令E(X) = μ_X, E(Y) = μ_Y, Var(X) = σ_X^2, Var(Y) = σ_Y^2。試以μ_X, μ_Y, σ_X, 及σ_Y表示ρ(XY, Y)。

13. 設X, Y, Z為兩兩無相關之隨機變數，期待值皆為μ，變異數皆為σ^2。試以μ及σ^2表示Cov(X + Y, Y + Z)及Cov(X + Y, X - Y)。
14. 試證定理5.2。

15. 試證系理5.5。

16. 設\(X, Y\)為隨機變數，且\(E(X) = E(Y) = 0, \text{Var}(X) = \text{Var}(Y) = 1\)。試證

(i) \(|E(XY)| \leq 1; \\
(ii) E(XY) = 1, \text{若且唯若} P(Y = X) = 1; \\
(iii) E(XY) = -1, \text{若且唯若} P(Y = -X) = 1。

17. 設\(X, Y\)之聯合p.d.f.為\(f(x, y) = 24xy, x, y \geq 0, x + y \leq 1\)。

(i) 試求\(\text{Cov}(X, Y)\)及\(\rho(X, Y)\); \\
(ii) 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\(\varepsilon\)獨立。試求\(\text{Cov}(X, Y)\)及\(\text{Cov}(Y, \varepsilon)\)。(解. \(a\sigma_X^2, \sigma^2)\)

第3.6節

1. 設\((X, Y) \sim \mathcal{N}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)\)。試證\(X | Y = y \sim \mathcal{N}(\mu_1 + \rho(\sigma_1/\sigma_2)(y - \mu_2), \sigma_1^2(1 - \rho^2))\)。

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 \(\chi^2_2\)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 \( \chi^2_2 \) 分佈。

12. (i) 設 \( (Z_1, Z_2) \sim \mathcal{N}(0, 0, 1, \rho) \)。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 \( (Z_1, Z_2) \sim \mathcal{N}(0, 0, 1, \rho) \)。

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 变 数 之 p.d.f.；

(ii) \( X, Y \) 皆以 \( \mathcal{N}(0, 1) \) 爲 邊 際 分 布。

15. 設 \( (X, Y) \) 之聯合 p.d.f. 為

\[
f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}(1 + xy e^{-\frac{x^2+y^2-2}{2}})}, x, y \in R.
\]

試證

(i) \( f(x, y) \) 確為 一 二 变 数 之 p.d.f.；

(ii) \( X, Y \) 皆以 \( \mathcal{N}(0, 1) \) 爲 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 \(|\rho_1| < 1, |\rho_2| < 1, \text{且} \rho_1 \neq \rho_2 \) 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 \(N(0, 1)\) 為邊際分佈。

17. 設 \(X\) 與 \(Y\) 獨立，且皆有 \(N(0, 1)\) 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 \(Y\) 之符號恒相同)

18. 設 \(X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)\)，且 \(X\) 與 \(Y\) 獨立。試決 aX + bY 與 cX - dY 獨立之充要條件，其中 \(a, b, c, d\) 為常數。(解. \(aco_1^2 = bdo_2^2\))

19. 設 \((X, Y) \sim N(0, 0, 1, 1, \rho)\)。試證 \(X + Y \overset{d}{=} X - Y \sim N(0, 2(1 - \rho)), \) 且 \(X + Y\) 與 \(X - Y\) 獨立。

20. 設 \((X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)\)。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 \((X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)\)。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 \( X, Y, Z \) 之联合 p.d.f. 爲 \( f(x, y, z) = 6z^2e^{-x^2(y+3)}, x, y, z \geq 0 \)。試求下述各 p.d.f.:
(1) \( f_X(x), f_Y(y), f_Z(z), f_{X,Y}(x,y), f_{X,Z}(x,z), f_{Y,Z}(y,z) \)。
(2) \( P(Y < x) \)。試求各 p.d.f.。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 \( X, Y \) 之联合 p.d.f. 爲 \( f(x, y, z) = c, 0 < x < y < z < 1 \)。試求(i) \( c \) 之值, (ii) \( f_{X|Y,Z}(x|y,z) \), (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) \( 1/(1-y) \), \( 0 < x < y < z < 1 \), (ii) \( 1/(z-x) \), \( 0 < x < y < z < 1 \), (iv) \( 1/y \), \( 0 < x < y < z < 1 \), (v) \( 2/z^2 \), \( 0 < x < y < z < 1 \)

4. 設 \( (X_1, X_2, X_3, X_4) \) 之联合 p.d.f. 爲

\[
f(x_1, x_2, x_3, x_4) = \frac{3}{4}(x_1^2 + x_2^2 + x_3^2 + x_4^2), 0 < x_i < 1, i = 1, 2, 3, 4.
\]

(i) 試證 \( f \) 為一 p.d.f.；
(ii) 試求 \( P(X_1 < 1/2, X_2 < 3/4, X_4 > 1/2) \);
(iii) 試求 \( E(X_1X_2) \);
(iv) 試求 \( f(x_3, x_4|x_1, x_2) \);
(v) 試求 \( P(X_3 > 3/4, X_1 < 1/2|X_1 = 1/3, X_2 = 2/3) \)。

5. 設 \( X, Y, Z \) 之联合 p.d.f. 爲 \( f(x, y, z) = (x+2y+z)/2, x, y, z \in (0, 1) \)。試求

(i) \( f_{X|Y,Z}(x|y,z) \);
(ii) \( f_{X,Y|Z}(x,y|z) \);
(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 \( N(0, 1) \) 分佈。試求
(i) $P(X/Y \leq t)$ 及 $P(XY/Y \leq t)$；
(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 $N(0,1)$ 分布。若獨立地射兩次，試求兩次位置距離之分佈。

8. 設 $A, B, C$ 為相無互立的隨機變數，且皆有 $U(0,1)$ 分布。試求方程式 $Ax^2 + Bx + C = 0$ 有實根之機率。(提示：若 $X$ 有 $U(0,1)$ 分布，則 $-\log X$ 有指數分佈，而二獨立的 gamma 隨機變數之和仍為 gamma 分布)

9. 設 $X_1, \cdots, X_n$ 為相無互立的隨機變數，且皆有 $U(0,1)$ 分布。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) $\text{Cov}(X,Y) = 0 \Rightarrow E(\text{Cov}(X,Y|Z)) = E(\text{Cov}(X,Y|Z))$；
   (iv) $\text{Cov}(Z, E(Y|Z)) = \text{Cov}(Z, Y)$。

13. 設 $(X_1, X_2, X_3, X_4, X_5)$ 有 $M(n, p_1, p_2, p_3, p_4, p_5)$ 分布。試證
   (i) $X_3, X_4, X_5|X_1 = x_1, X_2 = x_2$ 有 $M(n - x_1 - x_2, p_3 + p_4 + p_5, p_4/(p_3 + p_4 + p_5), p_4/(p_3 + p_4 + p_5))$ 分布；
   (ii) $X_3, X_4, X_5|X_1 + X_2 = x$ 有 $M(n - x, p_3/(p_3 + p_4 + p_5), p_4/(p_3 + p_4 + p_5))$ 分布。

14. 設 $X_1, X_2, X_3$ 相互獨立且分別有 $P(\lambda_1)$, $P(\lambda_2)$, $P(\lambda_3)$ 分布。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 $M(n, \lambda_1/\lambda, \lambda_2/\lambda, \lambda_3/\lambda)$ 分布。
15. 設\(X, Y, Z\)之聯合p.d.f.為

\[
f(x, y, z) = 0.25, (x, y, z) \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 1)\}.
\]

(i) 試求\(f_{X,Y}, f_{X,Z}, f_{Y,Z}\) 並證明\(X\)與\(Y\), \(X\)與\(Z\), \(Y\)與\(Z\)皆互不獨立;

(ii) 試證\(X, Y, Z\)不相獨立;

(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.為

\[
f(x_1, x_2, x_3, x_4) = 24e^{-x_1-x_2-x_3-x_4}, 0 < x_1 < x_2 < x_3 < x_4\]。

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\(X_1, \cdots, X_{n+1}\)為互不相聯立的隨機變數, 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\(T = (X_1/U_{n+1}, \cdots, X_n/U_{n+1})\)之聯合p.d.f.為

\[
f_T(t_1, \cdots, t_n) = n!, \forall t_i > 0, \text{且} \sum_{i=1}^n t_i < 1;
\]

(提示: 先求\((X_1/U_{n+1}, \cdots, X_n/U_{n+1}, U_{n+1})\)之聯合p.d.f.)

(ii) 試證\(U = (U_1/U_{n+1}, \cdots, U_n/U_{n+1})\)之聯合p.d.f.為

\[
f_u(u_1, \cdots, u_n) = n!, 0 < u_1 < u_2 < \cdots < u_n < 1.
\]

18. 設\(X = (X_1, \cdots, X_k)\)有\(M(n, p_1, \cdots, p_k)\)分佈。試證\(E(X_i) = np_i, \text{Var}(X_i) = np_i(1-p_i), i = 1, \cdots, k, \text{且Cov}(X_i, X_j) = -np_ip_j, i, j = 1, \cdots, k, i \neq j\)。

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\( X_i = 0, i = 1, \ldots, r \)。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\( \text{Var}(X_0 + X_1 + \cdots + X_{N-1}) \)。(解: \( \sum_{n=0}^{N-1} n n/(N-n)^2 \))

24. 在例7.11中，試求\( \text{Var}(Z) \)。

第3.8節

1. 設\( X \)為\( \mathcal{E}(2) \)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上界。

2. 設\( X \)為\( \mathcal{N}(0,1) \)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\( X \)為\( \mathcal{E}(2) \)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\( P((X, Y) = (a_i, b_i)) = 1/n, i = 1, \ldots, n \)。試分別對\( X, Y \)之邊際p.d.f., 並證明此時赫德爾不等式導致(8.8)式。

9. 試用(8.7)式導出(8.9)式。
参考文献

1. 黃文璋(1999). 數學欣賞。華泰文化事業股份有限公司, 台北。


