CHƯƠNG III: VÉC TƠ NGẪU NHIÊN

3.1 KHÁI NIỆM VÉC TƠ NGẪU NHIÊN

3.1.1 Khái niệm và phân loại véc tơ ngẫu nhiên

Trong chương trước ta xét các biến ngẫu nhiên mà giá trị chúng nhận được có thể biểu diễn bằng giá trị số, đó là các biến ngẫu nhiên một chiều. Tuy nhiên trong thực tế có thể gặp các đại lượng ngẫu nhiên mà giá trị nhận được là các bộ gồm hai, ba, … , n số. Những đại lượng này được gọi một cách tương ứng là biến ngẫu nhiên hai chiều, ba chiều, … , n chiều và được gọi chung là véc tơ ngẫu nhiên.

Định nghĩa 3.1: *Một véc tơ ngẫu nhiên n chiều là một bộ có thứ tự $\left(X_1, X_2, ..., X_n\right)$ với các thành phần $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên xác định trong cùng một phép thử.*

Ta ký hiệu véc tơ ngẫu nhiên hai chiều là (X, Y), trong đó X là biến ngẫu nhiên thành phần thứ nhất và Y là biến ngẫu nhiên thành phần thứ hai.

Ví dụ 3.1: Một nhà máy sản xuất một loại sản phẩm. Nếu kích thước của sản phẩm được đo bằng chiều dài X và chiều rộng Y thì ta có biến ngẫu nhiên hai chiều, còn nếu xét thêm cả chiều cao Z nữa thì ta có biến ngẫu nhiên ba chiều. Ngoài ra nếu quan tâm thêm trọng lượng của sản phẩm thì ta được biến ngẫu nhiên bốn chiều …

Véc tơ ngẫu nhiên n chiều $\left(X_1, X_2, ..., X_n\right)$ là liên tục hay rời rạc nếu tất cả các biến ngẫu nhiên thành phần $X_1, X_2, ..., X_n$ là liên tục hay rời rạc. Tồn tại véc tơ ngẫu nhiên có một thành phần rời rạc và một số thành phần liên tục, tuy nhiên giáo trình này không xét trường hợp đó.

3.1.2 Hàm phân bố xác suất

Định nghĩa 3.2: Hàm n biến $F_{X_1, ..., X_n}(x_1, x_2, ..., x_n)$ xác định bởi:

$$F_{X_1, ..., X_n}(x_1, x_2, ..., x_n) = P\left\{X_1 \le x_1, X_2 \le x_2, ..., X_n \le x_n\right\} \qquad \textbf{(3.1)}$$

được gọi là hàm phân bố xác suất của véc tơ ngẫu nhiên $X = \left(X_1, X_2, ..., X_n\right)$ và còn được gọi là hàm phân bố xác suất đồng thời của các biến ngẫu nhiên $X_1, X_2, ..., X_n$.

Hàm phân bố xác suất đồng thời có các tính chất:

1. $0 \le F_{X_1, ..., X_n}(x_1, ..., x_n) \le 1$. **(3.2)**

2. $\lim\limits_{x_k \to -\infty} F_{X_1, ..., X_n}(x_1, ..., x_n) = 0$, với k nào đó thuộc $\{1, ..., n\}$. **(3.3)**

3. $\lim\limits_{(x_1, ..., x_n) \to (\infty, ..., \infty)} F_{X_1, ..., X_n}(x_1, ..., x_n) = 1$. **(3.4)**

4. $F_{X_1,\dots,X_n}(x_1,\dots,x_n)$ không giảm theo từng biến. **(3.5)**

5. $\lim\limits_{x_1\to\infty} F_{X_1,\dots,X_n}(x_1,\dots,x_n) = F_{X_2,\dots,X_n}(x_2,\dots,x_n)$.

Như vậy nếu lấy giới hạn của hàm phân bố xác suất đồng thời của n biến ngẫu nhiên $X_1, X_2,\dots,X_n$ khi biến x_1 tiến đến vô cùng thì được hàm phân bố xác suất đồng thời của $n-1$ biến ngẫu nhiên $X_2,\dots,X_n$.

Tương tự nếu lấy giới hạn của hàm phân bố xác suất đồng thời của $X_1, X_2,\dots,X_n$ khi biến x_k tiến đến vô cùng, với k nào đó thuộc $\{1,2,\dots,n\}$, thì được hàm phân bố xác suất đồng thời của $n-1$ biến ngẫu nhiên $X_1,\dots,X_{k-1},X_{k+1},\dots,X_n$.

6. Đặc biệt nếu $F_{X,Y}(x,y)$ là hàm phân bố xác suất của véc tơ ngẫu nhiên hai chiều (X,Y) thì:

$$\lim_{y\to\infty} F_{X,Y}(x,y) = P\{X\le x\} = F_X(x);\ \lim_{x\to\infty} F_{X,Y}(x,y) = P\{Y\le y\} = F_Y(y) \qquad \textbf{(3.6)}$$

trong đó $F_X(x)$, $F_Y(y)$ là các hàm phân bố xác suất của biến ngẫu nhiên X, Y và được gọi là các hàm phân bố xác suất thành phần của véc tơ ngẫu nhiên (X,Y), hay còn gọi là hàm phân bố xác suất biên của hàm phân bố xác suất đồng thời $F_{X,Y}(x,y)$.

7. $P\{x_1 < X \le x_2; y_1 < Y \le y_2\} = F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1)$ **(3.7)**

Ví dụ 3.2: Hàm số $F(x,y) = \begin{cases} 1-e^{-(x+y)} & 0\le x<\infty, 0\le y<\infty \\ 0 & \text{nÕu ng-îc l¹i} \end{cases}$

Xét $x_2 = y_2 = 2$ và $x_1 = y_1 = 1$, ta có

$$F(2,2) - F(1,2) - F(2,1) + F(1,1) = (1-e^{-4}) - (1-e^{-3}) - (1-e^{-3}) + (1-e^{-2})$$

$$= -e^{-4} + 2e^{-3} - e^{-2} = -(e^{-2}-e^{-1})^2 < 0.$$

Vậy $F(x,y)$ không thể là hàm phân bố của véc tơ ngẫu nhiên vì không thỏa mãn (3.7).

Ví dụ 3.3: Cho hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y) xác định như sau

$$F_{X,Y}(x,y) = \begin{cases} (1-e^{-\alpha x})(1-e^{-\beta y}) & x\ge 0, y\ge 0; \alpha,\beta>0 \\ 0 & \text{nÕu ng-îc l¹i} \end{cases}$$

a) Tìm hai hàm phân bố xác suất biên $F_X(x)$, $F_Y(y)$.

b) Tính $P\{X\le 1, Y\le 1\}$, $P\{X\le 1\}$, $P\{Y>1\}$ và $P\{X>x, Y>y\}$.

Giải: a) $F_X(x) = F_{X,Y}(x,\infty) = \begin{cases} 1-e^{-\alpha x} & x\ge 0 \\ 0 & x<0 \end{cases}$; $F_Y(y) = F_{X,Y}(\infty,y) = \begin{cases} 1-e^{-\beta y} & y\ge 0 \\ 0 & y<0 \end{cases}$.

b) $P\{X \le 1, Y \le 1\} = F_{X,Y}(1,1) = (1-e^{-\alpha})(1-e^{-\beta})$;

$P\{X \le 1\} = F_X(1) = 1-e^{-\alpha}$; $P\{Y > 1\} = 1 - P\{Y \le 1\} = 1 - F_Y(1) = e^{-\beta}$;

Áp dụng luật De Morgan ta có

$$\overline{\{X > x\}\{Y > y\}} = \overline{\{X > x\}} \cup \overline{\{Y > y\}} = \{X \le x\} \cup \{Y \le y\}$$

Do đó

$$P\left(\overline{\{X > x\}\{Y > y\}}\right) = P\left(\{X \le x\} \cup \{Y \le y\}\right) = P\{X \le x\} + P\{Y \le y\} - P\{X \le x; Y \le y\}$$

$$= F_X(x) + F_Y(y) - F_{X,Y}(x,y)$$

$$= (1-e^{-\alpha x}) + (1-e^{-\beta y}) - (1-e^{-\alpha x})(1-e^{-\beta y}) = 1 - e^{-\alpha x}e^{-\beta y}.$$

Vậy $P\{X > x, Y > y\} = 1 - P\left(\overline{\{X > x\}\{Y > y\}}\right) = e^{-\alpha x}e^{-\beta y}$.

3.2 BẢNG PHÂN BỐ XÁC SUẤT CỦA VÉC TƠ NGẪU NHIÊN RỜI RẠC HAI CHIỀU

Tương tự trường hợp biến ngẫu nhiên rời rạc, quy luật phân bố xác suất của véc tơ ngẫu nhiên rời rạc 2 chiều có thể được xác định thông qua bảng phân bố xác suất đồng thời hoặc hàm khối lượng xác suất đồng thời.

3.2.1 Bảng phân bố xác suất đồng thời, hàm khối lượng xác suất đồng thời

Bảng phân bố xác suất đồng thời của biến ngẫu nhiên rời rạc hai chiều (X,Y) là bảng liệt kê tất cả các giá trị của X theo cột, Y theo hàng và các xác suất tương ứng có dạng sau.

$X \backslash Y$	y_1	y_2	…	y_j	…	y_m
x_1	$p_{XY}(x_1,y_1)$	$p_{XY}(x_1,y_2)$	…	$p_{XY}(x_1,y_j)$	…	$p_{XY}(x_1,y_m)$
x_2	$p_{XY}(x_2,y_1)$	$p_{XY}(x_2,y_2)$	…	$p_{XY}(x_2,y_j)$	…	$p_{XY}(x_2,y_m)$
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$	…	$\vdots$
x_i	$p_{XY}(x_i,y_1)$	$p_{XY}(x_i,y_2)$	…	$p_{XY}(x_i,y_j)$	…	$p_{XY}(x_i,y_m)$
$\vdots$	$\vdots$	$\vdots$	…	$\vdots$	…	$\vdots$
x_n	$p_{XY}(x_n,y_1)$	$p_{XY}(x_n,y_2)$	…	$p_{XY}(x_n,y_j)$	…	$p_{XY}(x_n,y_m)$

Trong đó x_i ($i=1,...,n$) là các giá trị có thể có của thành phần X; y_j ($j=1,...,m$) là các giá trị có thể có của thành phần Y. $R_{X,Y}=\left\{(x_i,y_j)\middle| i=1,...,n;\, j=1,...,m\right\}$ gọi là miền giá trị của véc tơ ngẫu nhiên (X,Y).

Hàm khối lượng xác suất đồng thời của biến ngẫu nhiên hai chiều (X,Y) ký hiệu và xác định bởi:

$$p_{X,Y}(x_i,y_j)=P\left\{X=x_i, Y=y_j\right\}, \tag{3.8}$$

thỏa mãn điều kiện

$$\begin{cases} p_{X,Y}(x_i,y_j)\ge 0,\ \forall i=1,...,n,\ j=1,...,m \\ \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} p_{X,Y}(x_i,y_j)=1 \end{cases} \tag{3.9}$$

Hàm phân bố xác suất đồng thời xác định như sau:

$$F_{X,Y}(x,y)=\sum_{x_i\le x}\sum_{y_j\le y} p_{X,Y}(x_i,y_j) \tag{3.10}$$

3.2.2 Bảng phân bố xác suất biên

Áp dụng công thức xác suất đầy đủ (1.19) cho hệ $\{X=x_1\}, \{X=x_2\}, ..., \{X=x_n\}$ (xem công thức 2.14) ta có:

$$p_Y(y_j)=P\left\{Y=y_j\right\}=\sum_{i=1}^{n}P\left\{X=x_i,Y=y_j\right\}=\sum_{i=1}^{n}p_{X,Y}(x_i,y_j)\ ;\quad j=1,...,m \tag{3.11}$$

$$p_X(x_i)=P\left\{X=x_i\right\}=\sum_{j=1}^{m}P\left\{X=x_i,Y=y_j\right\}=\sum_{j=1}^{m}p_{X,Y}(x_i,y_j)\ ;\quad i=1,...,n \tag{3.12}$$

Nhận xét 3.1: Từ công thức (3.11) và (3.12), nếu ta cộng các xác suất của bảng phân bố xác suất đồng thời của (X,Y) theo cột thì ta được các xác suất tương ứng với các giá trị của Y, nếu ta cộng các xác suất theo hàng ta được các xác suất tương ứng với giá trị của X. TỪ đó nhận được phân bố xác suất của biến ngẫu nhiên thành phần Y và biến ngẫu nhiên thành phần X.

X	x_1	x_2	...	x_i	...	x_n
P	$p_X(x_1)$	$p_X(x_2)$	...	$p_X(x_i)$	...	$p_X(x_n)$

Y	y_1	y_2	...	y_j	...	y_m
P	$p_Y(y_1)$	$p_Y(y_2)$	...	$p_Y(y_j)$	...	$p_Y(y_m)$

Ví dụ 3.4: Gieo 3 đồng tiền cân đối A, B, C. Gọi X là số mặt ngửa xuất hiện của 2 đồng tiền A, B và Y là số mặt ngửa xuất hiện của cả 3 đồng tiền A, B, C. Hãy lập bảng phân bố xác suất đồng thời của X,Y.

Giải: Chúng ta có bảng 8 kết quả đồng khả năng của việc gieo 3 đồng tiền cân đối và tính các giá trị của X,Y tương ứng, trong đó N là ký hiệu mặt ngửa xuất hiện còn S là mặt sấp.

A	B	C	X	Y
N	N	N	2	3
N	N	S	2	2
N	S	N	1	2
N	S	S	1	1
S	N	N	1	2
S	N	S	1	1
S	S	N	0	1
S	S	S	0	0

Sử dụng công thức tính xác suất cổ điển (1.1a) ta có:

$$P\{X=2,Y=3\}=\frac{1}{8}\ ;\ P\{X=2,Y=2\}=\frac{1}{8}\ ;\ P\{X=1,Y=2\}=\frac{2}{8}\ \ldots$$

Vậy bảng phân bố xác suất đồng thời của X và Y là

X \ Y	0	1	2	3
0	1/8	1/8	0	0
1	0	2/8	2/8	0
2	0	0	1/8	1/8

Phân bố xác suất của hai biến ngẫu nhiên thành phần:

Cộng các xác suất theo hàng ta được:

X	0	1	2
P	2/8	4/8	2/8

Cộng các xác suất theo cột ta được:

Y	0	1	2	3
P	1/8	3/8	3/8	1/8

Ví dụ 3.5: Có hai hộp, mỗi hộp đựng 6 bi.

Hộp I có 1 bi mang số 1, 2 bi mang số 2, 3 bi mang số 3.

Hộp II có 2 bi mang số 1, 3 bi mang số 2, 1 bi mang số 3.

Rút ngẫu nhiên từ mỗi hộp 1 bi. Gọi X, Y lần lượt là số ghi trên bi rút được từ hộp I và hộp II. Hãy lập bảng phân bố xác suất đồng thời của X, Y.

Giải: Mỗi hộp có 6 bi cho nên số các trường hợp có thể của phép thử là $6 \cdot 6 = 36$, trong đó có 2 trường hợp $(1,1)$, 3 trường hợp $(1,2)$, 4 trường hợp $(2,1)$, …

Vậy bảng phân bố xác suất đồng thời của X, Y như sau:

X \ Y	1	2	3
1	2/36	3/36	1/36
2	4/36	6/36	2/36
3	6/36	9/36	3/36

Ví dụ 3.6: (Phân bố đa thức, multinomial) Véc tơ ngẫu nhiên k chiều $X = (X_1, X_2, ..., X_k)$ được gọi là có phân bố đa thức với các tham số $(n; p_1, ..., p_k)$ ký hiệu $X \sim \text{MUT}(n; p_1, ..., p_k)$ nếu:

$$P\{X_1 = m_1, ..., X_k = m_k\} = \frac{n!}{m_1!m_2!...m_k!} p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k} \quad \textbf{(3.13)}$$

trong đó $0 \le m_i \le n,\ i = 1, ..., n;\ m_1 + \cdots + m_k = n;\ p_i > 0,\ i = 1, ..., n;\ p_1 + \cdots + p_k = 1$

Trường hợp $n = 2$ ta được phân bố nhị thức.

Xét n phép thử độc lập, thuần nhất và mỗi phép thử có k kết quả ngẫu nhiên A_1, ..., A_k; giả sử xác suất xuất hiện kết quả thứ A_i là p_i; $p_1 + \cdots + p_k = 1$.

Gọi X_i là số lần xuất hiện của biến cố A_i trong n phép thử thì $X = (X_1, X_2, ..., X_k)$ có phân bố đa thức $X \sim \text{MUT}(n; p_1, ..., p_k)$.

Chẳng hạn, gieo một con xúc xắc cân đối 10 lần. Tính các xác suất:

1) Có đúng 3 lần xuất hiện mặt 5 chấm (biến cố A).

2) Có 2 lần xuất hiện mặt 1 chấm, 4 lần mặt 3 chấm, 1 lần mặt 4 chấm và 3 lần mặt 6 chấm (biến cố B).

Giải: 1) Xét phép thử Bernoulli với thành công của mỗi lần thử là xuất hiện mặt có 5 chấm, vậy xác suất thành công mỗi lần thử là 1/6. Gọi X là số lần xuất hiện mặt 5 trong 10 lần thử thì X có phân bố nhị thức tham số $(10;1/6)$, do đó

$$P(A)=P\{X=3\}=C_{10}^{3}\left(\frac{1}{6}\right)^{3}\left(\frac{5}{6}\right)^{7}=0{,}155.$$

2) Gọi X_i là số lần xuất hiện mặt i chấm trong 10 phép thử thì $(X_1,X_2,X_3,X_4,X_5,X_6)$ có phân bố đa thức $\mathrm{MUT}(10;1/6,\,1/6,\,1/6,\,1/6,\,1/6,\,1/6)$.

$$P(B)=P\{X_1=2,X_2=0,X_3=4,X_4=1,X_5=0,X_6=3\}=\frac{10!}{2!0!4!1!0!3!}\left(\frac{1}{6}\right)^{10}\approx 0{,}0002.$$

3.3 HÀM MẬT ĐỘ XÁC SUẤT CỦA VÉC TƠ NGẪU NHIÊN LIÊN TỤC

3.3.1 Hàm mật độ xác suất đồng thời

Định nghĩa 3.3: *Hàm mật độ xác suất của véc tơ ngẫu nhiên liên tục* $(X_1,X_2,...,X_n)$ *là hàm* n *biến* $f_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)\ge 0$ *thoả mãn:*

$$F_{X_1,X_2,...,X_n}(x_1,x_2,...,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}f_{X_1,X_2,...,X_n}(t_1,t_2,...,t_n)\,dt_1dt_2\,...\,dt_n \quad \textbf{(3.14)}$$

$f_{X_1,...,X_n}(x_1,...,x_n)$ *còn được gọi là hàm mật độ xác suất đồng thời của* các biến ngẫu nhiên $X_1,X_2,...,X_n$.

Tính chất 3.1: Để đơn giản cho cách biểu diễn ta xét trường hợp véc tơ ngẫu nhiên hai chiều (X,Y) có hàm mật độ xác suất $f_{XY}(x,y)$.

1) $f_{X,Y}(x,y)\ge 0$ với mọi (x,y) và $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(x,y)dxdy=1$. **(3.15)**

2) $P\{(X,Y)\in A\}=\iint\limits_{(x,y)\in A\cap R_{X,Y}}f_{X,Y}(x,y)dxdy$ với mọi $A\subset \mathrm{R}^2$, **(3.16)**

$R_{X,Y}$ là miền giá trị của (X,Y).

3) $$f_{X,Y}(x,y)=\begin{cases}\dfrac{\partial^2}{\partial x\partial y}F_{X,Y}(x,y) & \text{nÕu tån t¹i đ¹o hµm t¹i } (x,y)\\ 0 & \text{nÕu ng-îc l¹i}\end{cases} \quad \textbf{(3.17)}$$

3.3.2 Hàm mật độ xác suất biên

Từ bảng phân bố xác suất đồng thời của véc tơ ngẫu nhiên rời rạc (X,Y), nếu ta cộng giá trị của hàm khối lượng xác suất theo cột tức là cộng tất cả các giá trị của hàm khối lượng xác suất theo mọi giá trị của X thì ta được giá trị của hàm khối lượng xác suất với giá trị tương ứng của Y, tương tự nếu ta cộng các giá trị của hàm khối lượng xác suất theo hàng ta được giá trị của hàm khối lượng xác suất với giá trị tương ứng của X.

Trường hợp véc tơ ngẫu nhiên (X,Y) liên tục thay vì cộng ta thực hiện lấy tích phân của hàm mật độ đồng thời $f_{X,Y}(x,y)$ theo biến y ta được hàm mật độ của biến ngẫu nhiên thành phần X và lấy tích phân của hàm mật độ đồng thời $f_{X,Y}(x,y)$ theo biến x ta được hàm mật độ của biến ngẫu nhiên thành phần Y, cụ thể

$$\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = f_X(x) \text{ hàm mật độ xác suất của biến ngẫu nhiên } X. \tag{3.18}$$

$$\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = f_Y(y) \text{ hàm mật độ xác suất của biến ngẫu nhiên } Y. \tag{3.19}$$

Ví dụ 3.7: Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{X,Y}(x,y) = \begin{cases} k & \text{nếu } 0 < y \le x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

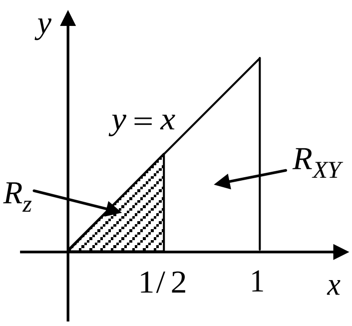
a) Tìm k.

b) Tìm các hàm mật độ xác suất của các biến ngẫu nhiên thành phần X, Y.

c) Tìm xác suất $P\{0 < X \le 1/2; 0 < Y \le 1/2\}$.

Giải: a) Miền giá trị của X, Y là tam giác R_{XY} (xem hình 3.1). Do đó

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = k\mathrm{dt}(R_{X,Y}) = k\frac{1}{2} \Rightarrow k = 2$$



Hình 3.1

b) $$f_X(x) = \begin{cases} \int_0^x 2dy = 2x & \text{nếu } 0 < x < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

$$f_Y(y) = \begin{cases} \int_y^1 2dx = 2(1-y) & \text{nếu } 0 < y < 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

c) $P\{0 < X \le 1/2; 0 < Y \le 1/2\} = P\{0 < X \le 1/2; 0 < Y < X\} = \iint_{R_z} f_{X,Y}(x,y)dxdy = 2\left(\frac{1}{8}\right) = \frac{1}{4}.$

Ví dụ 3.8: Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{X,Y}(x,y)=\begin{cases} k & \text{nÕu } |x|+|y|\le 1 \\ 0 & \text{nÕu ng-îc l¹i} \end{cases}.$$

a) Tìm k.

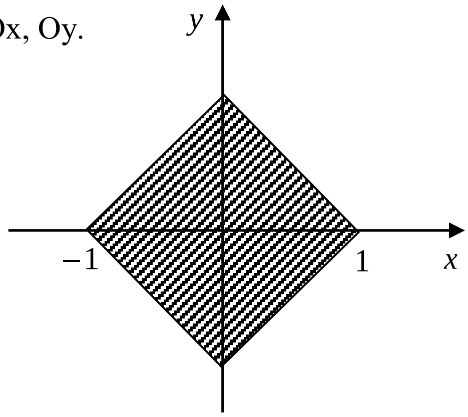
b) Tìm các hàm mật độ xác suất của các biến ngẫu nhiên X, Y.

Giải: a) Miền D: $|x|+|y|\le 1$ đối xứng qua hai trục toạ độ Ox, Oy.

Phần của D nằm trong góc phần tư thứ nhất là tam giác vuông cân $0\le x, 0\le y;\ x+y\le 1$.

Vậy D là hình vuông có độ dài cạnh bằng $\sqrt{2}$, do đó:

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = k\,\mathrm{dt}\,\mathrm{D}=2k \ \Rightarrow\ k=\frac{1}{2}.$$



Hình 3.2

b) $|x|+|y|\le 1 \Leftrightarrow |x|-1\le y\le |x|+1$

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy=\begin{cases} \dfrac{1}{2}\displaystyle\int_{|x|-1}^{1-|x|} dy & \text{nÕu } -1\le x\le 1 \\ 0 & \text{nÕu ng-îc l¹i} \end{cases} = \begin{cases} 1-|x| & \text{nÕu } |x|\le 1 \\ 0 & \text{nÕu } |x|>1 \end{cases}$$

Do tính chất đối xứng của X và Y nên ta cũng có:

$$f_Y(y)=\begin{cases} 1-|y| & \text{nÕu } |y|\le 1 \\ 0 & \text{nÕu } |y|>1 \end{cases}.$$

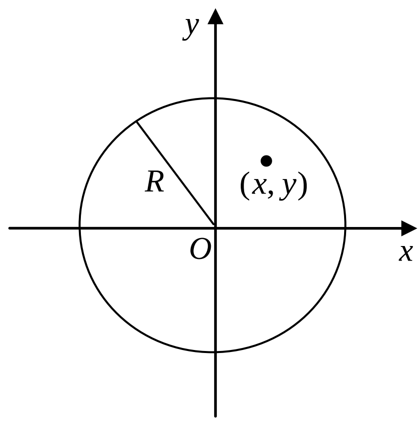
Ví dụ 3.9: Chọn ngẫu nhiên một điểm từ hình tròn tâm O bán kính R. Đặt X, Y là hoành độ và tung độ của điểm được chọn thì (X,Y) là véc tơ ngẫu nhiên liên tục có hàm mật độ

$$f_{X,Y}(x,y)=\begin{cases} k & x^2+y^2\le R^2 \\ 0 & x^2+y^2>R^2 \end{cases}.$$

a) Xác định k.

b) Tìm các hàm mật độ thành phần $f_X(x)$, $f_Y(y)$.

c) Tìm xác suất để khoảng cách từ gốc O đến điểm được chọn không lớn hơn a, với $0\le a\le R$.



Hình 3.3

Giải: a) $1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = k\iint_{x^2+y^2\le R^2} dxdy = k\pi R^2$

$$\Rightarrow\ k=\frac{1}{\pi R^2}$$

b) Khi $x^2 \le R^2$: $f_X(x) = \dfrac{1}{\pi R^2} \displaystyle\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}} dy = \dfrac{2}{\pi R^2}\sqrt{R^2-x^2}$.

Vậy
$$f_X(x) = \begin{cases} \dfrac{2}{\pi R^2}\sqrt{R^2-x^2} & |x| \le R \\ 0 & |x| > R \end{cases}$$

Tương tự
$$f_Y(y) = \begin{cases} \dfrac{2}{\pi R^2}\sqrt{R^2-y^2} & |y| \le R \\ 0 & |y| > R \end{cases}.$$

c) Với $0 \le a \le R$:

$$P\left\{X^2+Y^2 \le a^2\right\} = \iint_{x^2+y^2\le a^2} f_{X,Y}(x,y)dxdy = \frac{1}{\pi R^2}\iint_{x^2+y^2\le a^2} dxdy = \frac{\pi a^2}{\pi R^2} = \frac{a^2}{R^2}.$$

3.4 TÍNH ĐỘC LẬP CỦA CÁC BIẾN NGẪU NHIÊN

Từ định nghĩa 2.2 về tính độc lập của hai biến ngẫu nhiên và áp dụng các tính chất của hàm phân bố xác suất đồng thời, hàm khối lượng xác suất đồng thời, hàm mật độ xác suất đồng thời ta có các dấu hiệu sau đây để nhận biết tính độc lập của hai biến ngẫu nhiên.

Định lý 3.1: Giả sử $F_{X,Y}(x,y)$ là hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y). Khi đó X,Y độc lập khi và chỉ khi

$$F_{X,Y}(x,y) = F_X(x)F_Y(y) \qquad \textbf{(3.20)}$$

trong đó $F_X(x)$, $F_Y(y)$ lần lượt là hàm phân bố xác suất của X và Y.

Ví dụ 3.10: Hàm phân bố xác suất đồng thời của (X,Y) trong ví dụ 3.3 thỏa mãn (3.20) do đó độc lập.

Trường hợp véc tơ ngẫu nhiên (X,Y) rời rạc thì điều kiện (3.20) của hàm phân bố được thay bởi điều kiện của hàm khối lượng xác suất như sau.

Định lý 3.2: Véc tơ ngẫu nhiên hai chiều rời rạc (X,Y), X nhận các giá trị $x_1,...,x_n$, Y nhận các giá trị $y_1,...,y_m$ với hàm khối lượng xác suất đồng thời (3.8) là độc lập khi và chỉ khi

$$p_{X,Y}(x_i,y_j) = p_X(x_i)p_Y(y_j);\ \forall i=1,...,n,\ j=1,...,m. \qquad \textbf{(3.21)}$$

Tuy nhiên kiểm tra trực tiếp điều kiện (3.21) khá phức tạp và người ta thường sử dụng dấu hiệu sau để nhận biết hai biến ngẫu nhiên rời rạc độc lập: đó là bảng phân bố xác suất đồng thời có tính chất:

- Hai hàng bất kỳ tỉ lệ với nhau.
- Hai cột bất kỳ tỉ lệ với nhau. **(3.22)**

Ví dụ 3.11: Hai biến ngẫu nhiên X và Y của ví dụ 3.4 không độc lập vì $P\{X=2\}=\dfrac{2}{8}$, $P\{Y=1\}=\dfrac{3}{8}$ nhưng $P\{X=2, Y=1\}=0$. Hai hàng của bảng phân bố xác suất không tỉ lệ.

Hai biến ngẫu nhiên X và Y của ví dụ 3.5 độc lập vì thỏa mãn điều kiện (3.21), (3.22), các hàng của bảng phân bố xác suất tỉ lệ nhau theo tỉ lệ 1:2:3.

Trường hợp véc tơ ngẫu nhiên (X,Y) liên tục, bằng cách lấy đạo hàm hai vế của (3.20) (xem công thức 3.17) ta được:

Định lý 3.3: Giả sử véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất $f_{X,Y}(x,y)$. Khi đó X,Y độc lập khi và chỉ khi

$$f_{X,Y}(x,y)=f_X(x)f_Y(y) \qquad \textbf{(3.23)}$$

trong đó $f_X(x)$, $f_Y(y)$ lần lượt là hàm mật độ xác suất của X và Y.

Ví dụ 3.12: Véc tơ ngẫu nhiên (X,Y) của ví dụ 3.7, ví dụ 3.8, ví dụ 3.9 không độc lập vì hàm mật độ xác suất $f_{X,Y}(x,y)\neq f_X(x)f_Y(y)$.

Ví dụ 3.13: Véc tơ ngẫu nhiên (X,Y) có hàm mật độ:

$$f_{X,Y}(x,y)=\begin{cases}4xy & \text{nếu } 0<x<1,\ 0<y<1\\ 0 & \text{nếu ngược lại}\end{cases}$$

Có hai hàm mật độ thành phần

$$f_X(x)=\begin{cases}2x & \text{nếu } 0<x<1\\ 0 & \text{nếu ngược lại}\end{cases};\quad f_Y(y)=\begin{cases}2y & \text{nếu } 0<y<1\\ 0 & \text{nếu ngược lại}\end{cases}$$

Ta có $f_{X,Y}(x,y)=f_X(x)f_Y(y)$. Vậy X và Y độc lập.

3.5 CÁC THAM SỐ ĐẶC TRƯNG CỦA VÉC TƠ NGẪU NHIÊN

3.5.1 Kỳ vọng và phương sai của các biến ngẫu nhiên thành phần

Từ bảng phân bố xác suất thành phần (3.11), (3.12) và hàm mật độ xác suất thành phần (3.18), (3.19) ta có công thức tính kỳ vọng và phương sai của biến ngẫu nhiên thành phần X, Y của véc tơ ngẫu nhiên (X,Y):

a. Trường hợp X, Y rời rạc

$$\mathrm{E}X=\sum_{i=1}^{n}x_i p_X(x_i)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{X,Y}(x_i,y_j) \qquad \textbf{(3.24)}$$

$$\mathrm{E}Y=\sum_{j=1}^{m}y_j p_Y(y_j)=\sum_{j=1}^{m}\sum_{i=1}^{n}y_j p_{X,Y}(x_i,y_j) \qquad \textbf{(3.25)}$$

$$\mathrm{E}\,X^2=\sum_{i=1}^{n}x_i^2p_X(x_i)\,;\ \mathrm{D}\,X=\mathrm{E}\,X^2-\left(\mathrm{E}\,X\right)^2 \tag{3.26}$$

$$\mathrm{E}\,Y^2=\sum_{j=1}^{m}y_j^2p_Y(y_j)\,;\ \mathrm{D}\,Y=\mathrm{E}\,Y^2-\left(\mathrm{E}\,Y\right)^2 \tag{3.27}$$

b. Trường hợp X , Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$

$$\mathrm{E}\,X=\int_{-\infty}^{\infty}xf_X(x)dx=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{X,Y}(x,y)dxdy \tag{3.28}$$

$$\mathrm{E}\,Y=\int_{-\infty}^{\infty}yf_Y(y)dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf_{X,Y}(x,y)dxdy \tag{3.29}$$

$$\mathrm{E}\,X^2=\int_{-\infty}^{\infty}x^2f_X(x)dx=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^2f_{X,Y}(x,y)dxdy\,;\ \mathrm{D}\,X=\mathrm{E}\,X^2-\left(\mathrm{E}\,X\right)^2 \tag{3.30}$$

$$\mathrm{E}\,Y^2=\int_{-\infty}^{\infty}y^2f_Y(y)dy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}y^2f_{X,Y}(x,y)dydx\,;\ \mathrm{D}\,Y=\mathrm{E}\,Y^2-\left(\mathrm{E}\,Y\right)^2 \tag{3.31}$$

3.5.2 Hiệp phương sai

Định nghĩa 3.4: *Hiệp phương sai (hay còn gọi là Covariance) của hai biến ngẫu nhiên* X,Y *, ký hiệu* $\mathrm{cov}(X,Y)$*, là kỳ vọng toán của tích các sai lệch của hai biến ngẫu nhiên đó với kỳ vọng toán của chúng*:

$$\mathrm{cov}(X,Y)=\mathrm{E}\left[\left(X-\mathrm{E}\,X\right)\left(Y-\mathrm{E}\,Y\right)\right] \tag{3.32a}$$

Khai triển vế phải và áp dụng tính chất của kỳ vọng ta được

$$\mathrm{cov}(X,Y)=\mathrm{E}(XY)-(\mathrm{E}\,X)(\mathrm{E}\,Y) \tag{3.32b}$$

Nếu X,Y rời rạc có hàm khối lượng xác suất đồng thời theo công thức (3.8) thì

$$\mathrm{E}\,(XY)=\sum_{j=1}^{m}\sum_{i=1}^{n}x_iy_jp_{X,Y}(x_i,y_j) \tag{3.33}$$

Nếu X,Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$ thì

$$\mathrm{E}\,(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_{X,Y}(x,y)dxdy \tag{3.34}$$

Tính chất 3.2: Từ định nghĩa ta có thể chứng minh được các tính chất sau

1) $\mathrm{cov}(X,Y)=\mathrm{cov}(Y,X)$. **(3.35)**

2) $\mathrm{cov}(X,X)=\mathrm{D}\,X$. **(3.36)**

3) $\operatorname{cov}(aX+c, bY+d) = ab\operatorname{cov}(Y, X)$ với mọi hằng số a, b, c, d. **(3.37)**

4) Nếu X, Y độc lập thì $\operatorname{cov}(X, Y) = 0$ nhưng ngược lại chưa chắc đúng.

3.5.3 Ma trận hiệp phương sai

Định nghĩa 3.5: Cho véc tơ ngẫu nhiên $X = (X_1, X_2, ..., X_n)$.

$$\text{Ma trận } M = \left[C_{ij}\right]_{n\times n} \text{ với } C_{ij} = \operatorname{cov}(X_i, X_j) \tag{3.38}$$

được gọi là ma trận hiệp phương sai (ma trận covariance) của véc tơ ngẫu nhiên X.

Tính chất 3.3:

1) Ma trận hiệp phương sai là ma trận đối xứng.

2) Với mọi $t_1, t_2, ..., t_n \in \mathrm{R}^n$ luôn có $\sum\limits_j \sum\limits_i C_{ij} t_i t_j \ge 0$.

3) Các định thức con chính của M không âm.

Nói cách khác ma trận hiệp phương sai M là ma trận của dạng toàn phương không âm.

3.5.4 Hệ số tương quan

Định nghĩa 3.6: *Hệ số tương quan của hai biến ngẫu nhiên* X, Y *với phương sai khác 0 được ký hiệu và định nghĩa bởi công thức:*

$$\rho_{X,Y} = \frac{\operatorname{cov}(X, Y)}{\sqrt{\mathrm{D}X}\sqrt{\mathrm{D}Y}} \tag{3.39}$$

Tính chất 3.4:

1) $-1 \le \rho_{X,Y} \le 1$ với mọi X, Y. **(3.40)**

2) Nếu X, Y độc lập thì $\rho_{X,Y} = 0$, nhưng điều ngược lại chưa chắc đúng.

3) Từ công thức (3.37) ta suy ra: với mọi hằng số a, b, c, d

$$\rho_{aX+c, bY+d} = \begin{cases} \rho_{X,Y} & \text{nÕu } ab > 0 \\ -\rho_{X,Y} & \text{nÕu } ab < 0 \end{cases} \tag{3.41a}$$

4) $Y = aX + b$, $a \ne 0$ khi và chỉ khi

$$\rho_{X,Y} = \begin{cases} 1 & \text{nÕu } a > 0 \\ -1 & \text{nÕu } a < 0 \end{cases} \tag{3.41b}$$

Nhận xét 3.2: Ý nghĩa của hệ số tương quan

Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa X và Y. Khi $|\rho_{X,Y}|$ càng gần 1 thì tính chất tương quan tuyến tính càng chặt, khi $|\rho_{X,Y}|$ càng gần 0 thì sự phụ thuộc tuyến tính càng ít, càng lỏng lẻo.

Khi $\rho_{X,Y} = 0$ ta nói X và Y không tương quan.

Như vậy hai biến ngẫu nhiên độc lập thì không tương quan, nhưng ngược lại chưa chắc đúng (xem ví dụ 3.14).

Ví dụ 3.14: Xét véc tơ ngẫu nhiên (X,Y) có bảng phân bố xác suất đồng thời

$X \backslash Y$	-1	0	1
-1	4/15	1/15	4/15
0	1/15	2/15	1/15
1	0	2/15	0

Có bảng phân bố xác suất biên

X	-1	0	1
P	9/15	4/15	2/15

Y	-1	0	1
P	1/3	1/3	1/3

$$\mathrm{E}\,X = -1.\frac{9}{15} + 0.\frac{4}{15} + 1.\frac{2}{15} = -\frac{7}{15};$$

$$\mathrm{E}\,X = (-1)^2.\frac{9}{15} + 0^2.\frac{4}{15} + 1^2.\frac{2}{15} = \frac{11}{15} \Rightarrow \mathrm{D}\,X = \mathrm{E}\,X^2 - (\mathrm{E}\,X)^2 = \frac{116}{225}$$

$$\mathrm{E}Y = (-1).\frac{1}{3} + 0.\frac{1}{3} + 1.\frac{1}{3} = 0;\ \mathrm{E}Y^2 = \frac{2}{3} \Rightarrow \mathrm{D}Y = \frac{2}{3}.$$

$$\mathrm{E}(XY) = (-1).(-1).\frac{4}{15} + (-1).0.\frac{1}{15} + (-1).1.\frac{4}{15}$$

$$+0.(-1).\frac{1}{15} + 0.0.\frac{2}{15} + 0.1.\frac{1}{15} + 1.(-1).0 + 1.0.\frac{2}{15} + 1.1.0 = 0$$

Hiệp phương sai $\quad \mathrm{cov}(X,Y) = \mathrm{E}(XY) - (\mathrm{E}\,X)(\mathrm{E}\,Y) = 0 - \frac{-7}{15}.0 = 0.$

Hệ số tương quan $\quad \rho_{X,Y} = \dfrac{\mathrm{cov}(X,Y)}{\sqrt{\mathrm{D}\,X}\sqrt{\mathrm{D}Y}} = \dfrac{0}{\sqrt{(116/225)(2/3)}} = 0.$

Ma trận hiệp phương sai $\quad M = \begin{bmatrix} 116/225 & 0 \\ 0 & 2/3 \end{bmatrix}.$

Vì các hàng của bảng phân bố xác suất đồng thời không tỉ lệ nên hai biến ngẫu nhiên X,Y không độc lập, mặc dù hiệp phương sai $\operatorname{cov}(X,Y)=0$.

Ví dụ 3.15: Xét kênh viễn thông nhị phân được cho trong ví dụ 1.31 chương 1. Ký hiệu X là đầu vào và Y đầu ra của kênh. (X,Y) là véc tơ ngẫu nhiên thỏa mãn

$$P\{X=0\}=0,5\,;\; P\{Y=1|X=0\}=0,1 \text{ và } P\{Y=0|X=1\}=0,2\,.$$

a) Tìm bảng phân bố xác suất đồng thời của (X,Y).

b) Tìm bảng phân bố xác suất thành phần X và Y. X và Y có độc lập không?

c) Tính $\mathrm{E}X$, $\mathrm{E}Y$; $\mathrm{D}X$, $\mathrm{D}Y$

d) Tìm hiệp phương sai, hệ số tương quan và ma trận hiệp phương sai.

Giải: a)

$X \backslash Y$	0	1	Σ
0	0,45	0,05	0,5
1	0,10	0,40	0,5
Σ	0,55	0,45	1

b) Bảng phân bố xác suất thành phần X và Y

X	0	1
P	0,5	0,5

Y	0	1
P	0,55	0,45

Hai hàng của bảng phân bố xác suất đồng thời của (X,Y) không tỉ lệ, do đó X, Y không độc lập.

Ta cũng thấy rằng $P\{X=0;Y=0\}=0,45\neq P\{X=0\}P\{Y=0\}=0,55\times 0,5=0,275$.

c) Áp dụng các công thức (3.24)-(3.27) ta có

$$\mathrm{E}X=0,5,\; \mathrm{E}X^2=0,5 \;\Rightarrow \mathrm{D}X=0,25$$

$$\mathrm{E}Y=0,45,\; \mathrm{E}Y^2=0,45 \;\Rightarrow \mathrm{D}Y=0,2475\,.$$

d) Áp dụng công thức (3.32), (3.33), (3.38), (3.39) ta được

Hiệp phương sai $\quad \mathrm{E}(XY)=0,4 \Rightarrow \operatorname{cov}(X,Y)=0,4-0,5\times 0,45=0,175$.

Hệ số tương quan $\quad \rho_{X,Y}=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{\mathrm{D}X}\sqrt{\mathrm{D}Y}}=\dfrac{0,175}{\sqrt{0,25\times 0,2475}}=0,704$.

Ma trận hiệp phương sai $M=\begin{bmatrix}0,25 & 0,175\\ 0,175 & 0,2475\end{bmatrix}$.

Ta thấy giá trị $\rho_{X,Y}=0,704$ khá xa 1, do đó Y không phụ thuộc tuyến tính đối với X.

3.6 HÀM CỦA CÁC BIẾN NGẪU NHIÊN

3.6.1 Hàm của một biến ngẫu nhiên

Giả sử X là một biến ngẫu nhiên và $g(x)$ là hàm liên tục một biến số. Khi đó

$$Y=g(X) \tag{3.42}$$

cũng là một biến ngẫu nhiên và gọi là hàm của biến ngẫu nhiên X.

Điều kiện liên tục của hàm $g(x)$ có thể thay bởi điều kiện đo được. Tuy nhiên vấn đề này vượt khỏi nội dung của giáo trình.

A. Trường hợp rời rạc:

Nếu X là biến ngẫu nhiên rời rạc với miền giá trị R_X và hàm khối lượng xác suất $p_X(x_k)$ thì Y có miền giá trị và hàm khối lượng xác suất xác định như sau:

$$R_Y=\left\{g(x)\,\middle|\,x\in R_X\right\};\ p_Y(y)=\sum_{x_k;\,x_k\in g^{-1}(y)} p_X(x_k) \tag{3.43}$$

B. Trường hợp liên tục:

Giả sử X liên tục có hàm phân bố xác suất $F_X(x)$ và hàm mật độ xác suất $f_X(x)$, bằng phương pháp giải tích ta có thể tìm được hàm phân bố xác suất và hàm mật độ xác suất của biến ngẫu nhiên $Y=g(X)$.

- ❖ Trường hợp $g(x)$ là một song ánh, khi đó $g(x)$ hoặc đơn điệu tăng hoặc đơn điệu giảm:

- ♦ $g(x)$ đơn điệu tăng

$$F_Y(y)=P\{g(X)\le y\}=P\left\{X\le g^{-1}(y)\right\}=F_X\left(g^{-1}(y)\right)\ \Rightarrow\ F_X(x)=F_Y\left(g(x)\right).$$

Lấy đạo hàm theo x ta được

$$f_X(x)=f_Y\left(g(x)\right)g'(x)\ \text{hay}\ f_Y(y)=\frac{f_X\left(g^{-1}(y)\right)}{g'\left(g^{-1}(y)\right)}=\frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|}=f_X(x)\left|\frac{dx}{dy}\right|. \tag{3.44a}$$

- ♦ $g(x)$ đơn điệu giảm

$$F_Y(y)=P\{g(X)\le y\}=P\left\{X>g^{-1}(y)\right\}=1-F_X\left(g^{-1}(y)\right)\ \Rightarrow\ F_X(x)=1-F_Y\left(g(x)\right)$$

$$f_X(x) = -f_Y\left(g(x)\right)g'(x) \text{ hay } f_Y(y) = -\frac{f_X\left(g^{-1}(y)\right)}{g'\left(g^{-1}(y)\right)} = \frac{f_X\left(g^{-1}(y)\right)}{\left|g'\left(g^{-1}(y)\right)\right|} = f_X(x)\left|\frac{dx}{dy}\right|. \quad \textbf{(3.44b)}$$

❖ Trường hợp $g(x)$ là một hàm liên tục bất kỳ (không đơn điệu) ta chia miền xác định của $g(x)$ thành các miền mà $g(x)$ đơn điệu và áp dụng công thức (3.43) hoặc (3.44a), (3.44b) trong các miền này. Ngoài ra cũng có thể tính trực tiếp xác suất $P\{Y = g(X) \le y\}$, từ đó suy ra hàm phân bố xác suất và hàm mật độ xác suất của Y.

Ví dụ 3.16: Giả sử $g_1(x) = x^3$, $g_2(x) = x^2$, $g_3(x) = e^{-x}$. X là một biến ngẫu nhiên liên tục có hàm phân bố xác suất $F_X(x)$ và hàm mật độ xác suất $f_X(x)$.

a) Tìm hàm mật độ xác suất của $Y = g_1(X)$.

b) Tìm hàm mật độ xác suất của $Z = g_2(X)$.

c) Tìm hàm mật độ xác suất của $T = g_3(X)$.

d) Tìm hàm mật độ xác suất $f_Y(y)$, hàm phân bố xác suất $F_Y(y)$ và xác suất biến cố $P\{2{,}5 < Y < 5\}$ nếu hàm mật độ xác suất $f_X(x)$ của X xác định bởi:

$$f_X(x) = \begin{cases} \dfrac{2}{x^2} & \text{nÕu } 1 \le x \le 2 \\ 0 & \text{nÕu ng-îc l¹i} \end{cases}$$

Giải: a) $y = g_1(x) = x^3$ đơn điệu tăng. $x = \sqrt[3]{y}$, $\dfrac{dx}{dy} = \dfrac{1}{3\sqrt[3]{y^2}}$. Áp dụng công thức (3.43) ta có:

$$f_Y(y) = f_X(x)\frac{dx}{dy} = \frac{1}{3}\frac{f_X(\sqrt[3]{y})}{\sqrt[3]{y^2}}.$$

b) $g_2(x) = x^2$ không đơn điệu. Ta tính trực tiếp $P\{Z \le z\}$ như sau:

- Với $z < 0$ thì $P\{Z \le z\} = 0$ vì $Z = X^2$ không nhận giá trị âm.
- Với $z \ge 0$ ta có:

$$F_Z(z) = P\{Z \le z\} = P\left\{X^2 \le z\right\} = P\left\{-\sqrt{z} \le X \le \sqrt{z}\right\} = F_X(\sqrt{z}) - F_X(-\sqrt{z}).$$

Vậy hàm mật độ xác suất của Z.

$$f_Z(z) = \begin{cases} \dfrac{1}{2\sqrt{z}}\left(f_X(\sqrt{z}) + f_X(-\sqrt{z})\right) & \text{nÕu } z > 0 \\ 0 & \text{nÕu } z \le 0 \end{cases}$$

c) $g_3(x) = e^{-x}$ đơn điệu giảm. $t = e^{-x} \Leftrightarrow x = -\ln t$; $\dfrac{dx}{dt} = -\dfrac{1}{t}$.

Áp dụng công thức (3.44) ta có: $f_T(t) = f_X(x)\left|\dfrac{dx}{dt}\right| = f_X(-\ln t)\dfrac{1}{t}$.

Ta cũng có thể tìm $f_T(t)$ trực tiếp bằng cách tính $P\{T < t\}$

Với $t > 0$ ta có:

$$F_T(t) = P\{T \le t\} = P\left\{e^{-X} \le t\right\} = P\left\{e^{-X} < t\right\} = P\{X > -\ln t\} = 1 - F_X(-\ln t)$$

Khi $t \le 0$ biến cố $\{T \le t\}$ là biến cố không thể vì $T = e^{-X}$ chỉ nhận giá trị dương, do đó $P\{T \le t\} = 0$.

Vậy hàm mật độ xác suất của T xác định như sau:

$$f_T(t) = \begin{cases} \dfrac{f_X(-\ln t)}{t} & \text{nếu } t > 0 \\ 0 & \text{nếu } t \le 0 \end{cases}$$

d) $$f_Y(y) = \frac{1}{3}\frac{f_X(\sqrt[3]{y})}{\sqrt[3]{y^2}} = \begin{cases} \dfrac{2}{3\sqrt[3]{y^4}} & \text{nếu } 1 \le y \le 8 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

$$F_Y(y) = \begin{cases} 0 & \text{nếu } y \le 1 \\ 2\left(1 - \dfrac{1}{\sqrt[3]{y}}\right) & \text{nếu } 1 < y \le 8 \\ 1 & \text{nếu } y > 8 \end{cases}.$$

$$P\{2,5 < Y < 5\} = P\left\{2,5 < X^3 < 5\right\} = P\left\{\sqrt[3]{2,5} < X < \sqrt[3]{5}\right\} = \int_{\sqrt[3]{2,5}}^{\sqrt[3]{5}} \frac{2}{x^2}dx = -\frac{2}{x}\Bigg|_{\sqrt[3]{2,5}}^{\sqrt[3]{5}} = 2\left(\frac{1}{\sqrt[3]{2,5}} - \frac{1}{\sqrt[3]{5}}\right)$$

hoặc $P\{2,5 < Y < 5\} = F_Y(5) - F_Y(2,5) = 2\left(1 - \dfrac{1}{\sqrt[3]{5}}\right) - 2\left(1 - \dfrac{1}{\sqrt[3]{2,5}}\right) = 2\left(\dfrac{1}{\sqrt[3]{2,5}} - \dfrac{1}{\sqrt[3]{5}}\right)$.

Ví dụ 3.17: Cho X là biến ngẫu nhiên với hàm phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$. Tìm hàm mật độ xác suất của $Y = aX + b$; a, b là hai hằng số thực, $a \ne 0$.

Tìm a, b sao cho $Y = aX + b$ có phân bố chuẩn tắc (xem công thức (2.33) chương 2).

Giải: Xét hàm số $y = g(x) = ax + b$, áp dụng công thức (3.44a), (3.44b) ta được:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right) = \frac{1}{\sqrt{2\pi}|a|\sigma}\exp\left\{-\frac{1}{2\sigma^2}\left[\left(\frac{y-b}{a}\right) - \mu\right]^2\right\}$$

$$= \frac{1}{\sqrt{2\pi}|a|\sigma}\exp\left\{-\frac{1}{2a^2\sigma^2}\left[y - (a\mu + b)\right]^2\right\}.$$

Vậy Y có phân bố chuẩn $\mathbf{N}(a\mu+b;a^2\sigma^2)$.

Chọn $a=1/\sigma$, $b=-\mu/\sigma$ thì Y có phân bố chuẩn tắc. Như vậy ta có kết quả sau

$$\text{Nếu } X \sim \mathbf{N}(\mu;\sigma^2) \text{ thì } Y=\frac{X-\mu}{\sigma} \sim \mathbf{N}(0;1).$$

3.6.2 Hàm của hai biến ngẫu nhiên

Giả sử X, Y là hai biến ngẫu nhiên và $g(x,y)$ là một hàm hai biến liên tục cho trước, khi đó $Z=g(X,Y)$ là một biến ngẫu nhiên được gọi là hàm của hai biến ngẫu nhiên X và Y.

Ta có thể xác định hàm khối lượng xác suất hoặc hàm mật độ xác suất của Z qua hàm khối lượng xác suất đồng thời hoặc hàm mật độ xác suất đồng thời của X, Y.

a) *Trường hợp rời rạc*:

Nếu X, Y có tập các giá trị $\{x_1,x_2,...,x_m\}$ và $\{y_1,y_2,...,y_n\}$ thì $Z=g(X,Y)$ có tập các giá trị là $\left\{g(x_i,y_j)\,\middle|\, i=1,...m;\ j=1,...,n\right\}$.

Định lý 3.4: Xét $Z=g(X,Y)$, khi đó:

$$P\{Z=z\}=\sum_{\{(i,j)|g(x_i,y_j)=z\}} P\{X=x_i;Y=y_j\}. \tag{3.45}$$

Ví dụ 3.18: Xét $Z=X+Y$, với véc tơ ngẫu nhiên (X,Y) cho trong hai trường hợp sau:

a) ví dụ 3.4.

X	0	0	0	0	1	1	1	1	2	2	2	2
Y	0	1	2	3	0	1	2	3	0	1	2	3
$X+Y$	0	1	2	3	1	2	3	4	2	3	4	5
P	1/8	1/8	0	0	0	2/8	2/8	0	0	0	1/8	1/8

Vậy bảng phân bố xác suất của $Z=X+Y$ xác định như sau

$Z=X+Y$	0	1	2	3	4	5
P	1/8	1/8	2/8	2/8	1/8	1/8

b) ví dụ 3.5.

X	1	1	1	2	2	2	3	3	3
Y	1	2	3	1	2	3	1	2	3
$X+Y$	2	3	4	3	4	5	4	5	6

P	2/36	3/36	1/36	4/36	6/36	2/36	6/36	9/36	3/36

Vậy bảng phân bố xác suất của $Z = X + Y$ xác định như sau

$Z = X + Y$	2	3	4	5	6
P	2/36	7/36	13/36	11/36	3/36

b) *Trường hợp liên tục*:

Giả sử véc tơ ngẫu nhiên (X,Y) liên tục có hàm mật độ xác suất đồng thời $f_{XY}(x,y)$. $Z = g(X,Y)$ sẽ là biến ngẫu nhiên liên tục.

$$\text{Với mỗi } z \in \mathrm{R}\text{, ký hiệu } D_z = \left\{(x,y) \in R_{X.Y} \middle| g(x,y) \le z\right\}, \qquad \textbf{(3.46)}$$

trong đó R_{XY} là tập giá trị của (X,Y).

Định lý 3.5: *Hàm phân bố xác suất của biến ngẫu nhiên $Z = g(X,Y)$ được xác định như sau*

$$F_Z(z) = \iint_{D_z} f_{X.Y}(x,y)\,dxdy \qquad \textbf{(3.47)}$$

3.6.3 Hàm phân bố của tổng hai biến ngẫu nhiên

Định lý 3.6: Giả sử $Z = X + Y$ là tổng của hai biến ngẫu nhiên X, Y.

a) Nếu X, Y rời rạc thì hàm khối lượng xác suất của Z xác định như sau:

$$P\{Z = z\} = \sum_{x \in R_X} P\{X = x; Y = z - x\} = \sum_{y \in R_Y} P\{Y = y; X = z - y\}. \qquad \textbf{(3.48)}$$

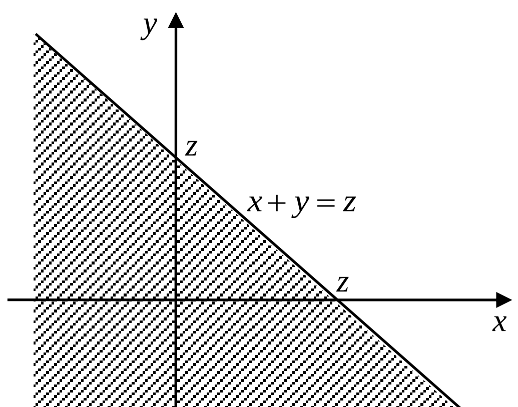
b) Nếu X, Y liên tục có hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$ thì hàm mật độ xác suất của Z:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)\,dx = \int_{-\infty}^{\infty} f_{X,Y}(z-y, y)\,dy. \qquad \textbf{(3.49)}$$

Chứng minh: a) Áp dụng công thức (3.45) ta có

$$P\{Z = z\} = \sum_{\{(i,j) | x_i + y_j = z\}} P\left\{X = x_i; Y = y_j\right\} = \sum_{x_i} P\left\{X = x_i; Y = z - x_i\right\} = \sum_{y_j} P\left\{X = z - y_j; Y = y_j\right\}$$

b) Miền D_z trong công thức (3.46) là tập các điểm $(x,y) \in R_{XY}$ nằm dưới đường thẳng $x + y = z$.



Do đó $$F_Z(z) = P\{X + Y \le z\} = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{z-x} f_{X,Y}(x,y)dy \right] dx$$

Vậy $$f_Z(z) = \frac{d}{dz} F_Z(z) = \int_{-\infty}^{\infty} \left[\frac{d}{dz} \int_{-\infty}^{z-x} f_{X,Y}(x,y)dy \right] dx = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx$$

Tương tự ta cũng có $f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}(z-y, y)dy$.

Hệ quả 3.1: Giả sử $Z = X + Y$ là tổng của hai biến ngẫu nhiên độc lập X, Y.

(i) Nếu X, Y rời rạc thì hàm khối lượng xác suất của Z có dạng

$$P\{Z = z\} = \sum_{x \in R_X} p_X(x) p_Y(z-x) = \sum_{y \in R_Y} p_Y(y) p_X(z-y). \tag{3.50}$$

Tổng $\sum_{x \in R_X}$ lấy theo x chạy trong miền giá trị của X và đồng thời $z-x$ thuộc miền giá trị của Y. Tương tự tổng $\sum_{y \in R_Y}$ lấy theo y chạy trong miền giá trị của Y và đồng thời $z-y$ thuộc miền giá trị của X.

(ii) Nếu X, Y liên tục có hàm mật độ xác suất đồng thời $f_X(x) f_Y(y)$ thì hàm mật độ xác suất của Z có dạng

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)dx = f_X(z) * f_Y(z) \tag{3.51a}$$

Như vậy biến ngẫu nhiên là tổng hai biến ngẫu nhiên độc lập liên tục X, Y có hàm mật độ xác suất là tích chập của hai hàm mật độ $f_X(z) * f_Y(z)$.

Vì tích chập có tính giao hoán nên ta cũng nhận được công thức

$$f_Z(z) = \int_{-\infty}^{\infty} f_Y(y) f_X(z-y)dy = f_Y(z) * f_X(z) \tag{3.51b}$$

Ví dụ 3.19: Giả sử X, Y là hai biến ngẫu nhiên độc lập có phân bố Poisson với các tham số λ_1, λ_2. Xét biến ngẫu nhiên $Z = X + Y$. Áp dụng công thức (3.50) ta có:

$$P\{Z=n\}=\sum_{k=0}^{n}P\{X=k\}P\{Y=n-k\}=\sum_{k=0}^{n}e^{-\lambda_1}\frac{\lambda_1^k}{k!}e^{-\lambda_2}\frac{\lambda_2^{n-k}}{(n-k)!}$$

$$=e^{-(\lambda_1+\lambda_2)}\sum_{k=0}^{n}\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}\sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\lambda_1^k\lambda_2^{n-k}=\frac{e^{-(\lambda_1+\lambda_2)}}{n!}(\lambda_1+\lambda_2)^n.$$

Vậy $Z=X+Y$ cũng có phân bố Poisson với tham số $\lambda_1+\lambda_2$.

Ví dụ 3.20: Giả sử X, Y là hai biến ngẫu nhiên độc lập có phân bố đều trong khoảng $(0;1)$. Xét biến ngẫu nhiên $Z=X+Y$.

Hàm mật độ xác suất của X và Y lần lượt là:

$$f_X(x)=\begin{cases}1 & 0<x<1\\ 0 & \text{nÕu ng-îc l¹i}\end{cases}\quad;\quad f_Y(y)=\begin{cases}1 & 0<y<1\\ 0 & \text{nÕu ng-îc l¹i}\end{cases}$$

Miền giá trị của X và Y là khoảng $(0;1)$, do đó miền giá trị của $Z=X+Y$ là khoảng $(0;2)$ và $R_{X,Y}=(0;1)\times(0;1)$.

Áp dụng công thức (4.51a) ta có

$$f_Z(z)=\int_{-\infty}^{\infty}f_X(x)f_Y(z-x)dx=\int_0^z f_X(x)f_Y(z-x)dx=\begin{cases}\int_0^z dx & 0<z<1\\ \int_{z-1}^{1}dx & 1<z<2\\ 0 & \text{nÕu ng-îc l¹i}\end{cases}$$

Vậy

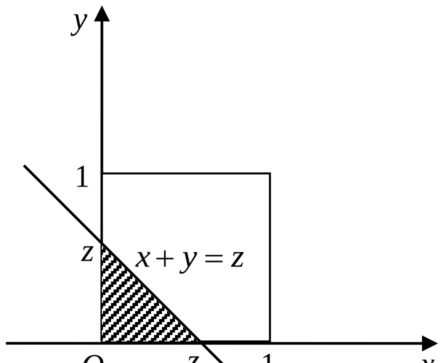
$$f_Z(z)=\begin{cases}z & 0<z<1\\ 2-z & 1<z<2\\ 0 & \text{nÕu ng-îc l¹i}\end{cases}$$

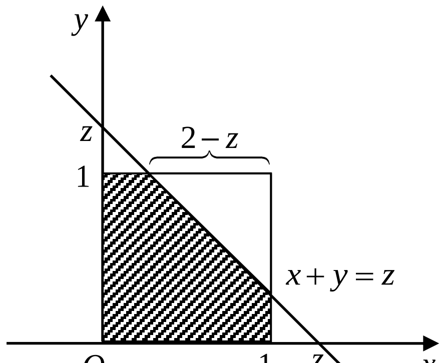
Ta cũng có thể tính trực tiếp bằng cách xét $F_Z(z)=P\{X+Y\le z\}$

Hàm mật độ xác suất đồng thời của X và Y:

$$f_{X,Y}(x,y)=\begin{cases}1 & 0<x<1, 0<y<1\\ 0 & \text{nÕu ng-îc l¹i}\end{cases}$$

$F_Z(z)=P\{X+Y\le z\}=\iint\limits_{x+y\le z}f_{X,Y}(x,y)dxdy=\iint\limits_{x+y\le z;(x,y)\in R_{X,Y}}dxdy=$ diện tích miền gạch chéo.





- Trường hợp $0 < z < 1$:

$$F_Z(z) = \iint\limits_{x+y\le z;(x,y)\in R_{X,Y}} dxdy = \text{diện tích miền gạch chéo} = \frac{z^2}{2},$$

$$f_Z(z) = \frac{d}{dz}F_Z(z) = z.$$

- Trường hợp $1 < z < 2$:

$$F_Z(z) = \iint\limits_{x+y\le z;(x,y)\in R_{X,Y}} dxdy = \text{diện tích miền gạch chéo} = 1 - \frac{(2-z)^2}{2},$$

$$f_Z(z) = \frac{d}{dz}F_Z(z) = 2 - z.$$

3.6.4 Hai hàm của hai biến ngẫu nhiên liên tục

Giả sử X, Y là hai biến ngẫu nhiên và $z = g(x,y),\ w = h(x,y)$ là hai hàm hai biến liên tục cho trước. Khi đó ta có thể thiết lập một cặp biến ngẫu nhiên mới Z, W bằng công thức:

$$\begin{cases} Z = g(X,Y) \\ W = h(X,Y) \end{cases} \quad \textbf{(3.52)}$$

Với mỗi $(z,w)\in \mathrm{R}^2$, ký hiệu $D_{zw} = \left\{(x,y)\in R_{XY} \,\middle|\, g(x,y)\le z; h(x,y)\le w\right\}$, trong đó R_{XY} là miền giá trị của (X,Y). Ta có

$$\left\{Z\le z; W\le w\right\} = \left\{g(X,Y)\le z; h(X,Y)\le w\right\} = \left\{(X,Y)\in D_{zw}\right\} \quad \textbf{(3.53)}$$

Áp dụng công thức (3.16) đối với véc tơ ngẫu nhiên liên tục (X,Y) với hàm mật độ xác suất $f_{X,Y}(x,y)$ ta được

$$F_{Z,W}(z,w) = \iint\limits_{D_{zw}} f_{X,Y}(x,y)dxdy \quad \textbf{(3.54)}$$

Công thức xác định hàm mật độ $f_{Z,W}(z,w)$ từ hàm mật độ $f_{X,Y}(x,y)$:

Giả sử T là ánh xạ từ R^2 vào R^2 xác định bởi:

$$T(x,y) = \left(g(x,y), h(x,y)\right)$$

Giả thiết T là ánh xạ 1 – 1 (đơn ánh) từ miền $R \subset \mathrm{R}^2$ lên miền $\Delta \subset \mathrm{R}^2$. Khi đó T có ánh xạ ngược T^{-1}: $\Delta \to R$ xác định như sau:

Với mỗi $(z,w) \in \Delta$ tồn tại duy nhất $(x,y) \in R$ sao cho $T(x,y) = (z,w)$, nghĩa là

$$\begin{cases} z = g(x,y) \\ w = h(x,y) \end{cases}$$

ta ký hiệu $T^{-1}(z,w) = (x,y)$.

Định thức Jacobi của T^{-1} được xác định theo công thức sau:

$$J(z,w) = \frac{D(x,y)}{D(z,w)} = \begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix}$$

Ta cũng có định thức Jacobi của T và công thức liên hệ với định thức Jacobi của T^{-1}:

$$J(x,y) = \frac{D(z,w)}{D(x,y)} = \begin{vmatrix} \dfrac{\partial z}{\partial x} & \dfrac{\partial z}{\partial y} \\ \dfrac{\partial w}{\partial x} & \dfrac{\partial w}{\partial y} \end{vmatrix} = \frac{1}{J(z,w)}.$$

Định lý 3.7: Giả sử X, Y là hai biến ngẫu nhiên liên tục với hàm mật độ xác suất đồng thời $f_{X,Y}(x,y)$. Ký hiệu $R_{X,Y}$ là miền giá trị của (X,Y):

$$R_{X,Y} = \left\{ (x,y) \in \mathrm{R}^2 \,\middle|\, f_{X,Y}(x,y) > 0 \right\}$$

Giả sử $T(x,y) = \big(z(x,y), w(x,y)\big)$ là ánh xạ 1 – 1 khả vi liên tục từ miền $R_{X,Y} \subset \mathrm{R}^2$ lên miền $\Delta \subset \mathrm{R}^2$. Khi đó $Z = z(X,Y),\ W = w(X,Y)$ là hai biến ngẫu nhiên có hàm mật độ xác suất đồng thời

$$f_{Z,W}(z,w) = \begin{cases} f_{X,Y}\big(x(z,w), y(z,w)\big)\left|\dfrac{D(x,y)}{D(z,w)}\right| & \text{nÕu } (z,w) \in \Delta \\ 0 & \text{nÕu ngîc l¹i} \end{cases} \qquad \textbf{(3.55)}$$

Ví dụ 3.21: Xét phép biến đổi T từ R^2 vào R^2 xác định bởi:

$$T(x,y) = \big(z(x,y), w(x,y)\big) \text{ trong đó } \begin{cases} z = x + y \\ w = x \end{cases} \Leftrightarrow \begin{cases} x = w \\ y = z - w \end{cases}$$

$$\frac{D(x,y)}{D(z,w)} = \begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ 1 & -1 \end{vmatrix} = -1.$$

Áp dụng công thức (3.55) ta nhận được:

$$f_{Z,W}(z,w) = f_{X,Y}(w, z-w).$$

Từ công thức (3.18) ta nhận được công thức (3.49) của biến ngẫu nhiên $Z = X + Y$:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,W}(z,w)dw = \int_{-\infty}^{\infty} f_{X,Y}(w, z-w)dw = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x)dx.$$

Ví dụ 3.22: Xét phép biến đổi T từ R^2 vào R^2 xác định bởi:

$$T(x,y) = \big(z(x,y), w(x,y)\big) \text{ trong đó } \begin{cases} z = xy \\ w = x \end{cases} \Leftrightarrow \begin{cases} x = w \\ y = \dfrac{z}{w} \end{cases}$$

$$\frac{D(x,y)}{D(z,w)} = \begin{vmatrix} \dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w} \\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w} \end{vmatrix} = \begin{vmatrix} 0 & 1 \\ \dfrac{1}{w} & -\dfrac{z}{w^2} \end{vmatrix} = -\frac{1}{w}.$$

Áp dụng công thức (3.55) ta nhận được:

$$f_{Z,W}(z,w) = \left|\frac{1}{w}\right| f_{X,Y}\left(w, \frac{z}{w}\right).$$

Từ công thức (3.18) ta nhận được công thức hàm mật độ của biến ngẫu nhiên $Z = XY$:

$$f_Z(z) = \int_{-\infty}^{\infty} f_{Z,W}(z,w)dw = \int_{-\infty}^{\infty} \left|\frac{1}{w}\right| f_{X,Y}\left(w, z/w\right)dw$$

$$f_Z(z) = \int_{-\infty}^{\infty} f_{X,Y}\left(x, z/x\right)\frac{dx}{|x|} \tag{3.56}$$

Ví dụ 3.23: Giả sử X, Y là hai biến ngẫu nhiên độc lập có phân bố đều trong khoảng $(0;1)$. Hàm mật độ xác suất đồng thời của X và Y:

$$f_{X,Y}(x,y) = \begin{cases} 1 & 0 < x < 1, 0 < y < 1 \\ 0 & \text{nÕu ng-îc l¹i} \end{cases}$$

Miền giá trị của $Z = XY$ cũng là khoảng $(0;1)$, do đó

$$f_{X,Y}\left(w, \frac{z}{w}\right) = \begin{cases} 1 & 0 < w < 1, 0 < z/w < 1 \\ 0 & \text{nÕu ng-îc l¹i} \end{cases}$$

$$f_{X,Y}\left(w, \frac{z}{w}\right) = \begin{cases} 1 & 0 < z < w < 1 \\ 0 & \text{nÕu ng-îc l¹i} \end{cases}$$

$$f_{Z,W}(z,w)=\left|\frac{1}{w}\right| f_{X,Y}\left(w,\frac{z}{w}\right)=\begin{cases}\dfrac{1}{w} & 0<z<w<1\\ 0 & \text{nÕu ng-îc l¹i}\end{cases}$$

$$f_Z(z)=\int_{-\infty}^{\infty} f_{Z,W}(z,w)dw=\int_z^1\frac{1}{w}dw=-\ln z\,,\quad 0<z<1$$

Vậy
$$f_Z(z)=\begin{cases}-\ln z & 0<z<1\\ 0 & \text{nÕu ng-îc l¹i}\end{cases}$$

Ví dụ 3.24: Xét phép biến đổi T xác định bởi:

$$T(x,y)=\big(z(x,y),w(x,y)\big)\ \text{trong đó}\ \begin{cases}z=x/y\\ w=y\end{cases}$$

$$T^{-1}(z,w)=(x,y),\ \begin{cases}x=zw\\ y=w\end{cases}$$

$$\frac{D(x,y))}{D(z,w}=\begin{vmatrix}\dfrac{\partial x}{\partial z} & \dfrac{\partial x}{\partial w}\\ \dfrac{\partial y}{\partial z} & \dfrac{\partial y}{\partial w}\end{vmatrix}=\begin{vmatrix}w & z\\ 0 & 1\end{vmatrix}=w\,.$$

Áp dụng công thức (3.55) ta nhận được:

$$f_{Z,W}(z,w)=|w|\,f_{X,Y}\left(zw,w\right).$$

Từ công thức (3.18) ta nhận được công thức hàm mật độ của biến ngẫu nhiên $Z=\dfrac{X}{Y}$:

$$f_Z(z)=\int_{-\infty}^{\infty} f_{Z,W}(z,w)dw=\int_{-\infty}^{\infty}|w|\,f_{X,Y}\left(zw,w\right)dw=\int_{-\infty}^{\infty}|y|\,f_{X,Y}\left(zy,y\right)dy \qquad \textbf{(3.57)}$$

Ví dụ 3.25: Giả sử X , Y là hai biến ngẫu nhiên độc lập có phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Xét $Z=\dfrac{X}{Y}$. Áp dụng công thức (3.57) và từ tính chất độc lập của X , Y ta được.

$$f_Z(z)=\int_{-\infty}^{\infty}|y|\,f_{X,Y}\left(zy,y\right)dy=\int_{-\infty}^{\infty}|y|\,f_X\left(zy\right)f_Y\left(y\right)dy=\int_{-\infty}^{\infty}|y|\frac{1}{2\pi}e^{-y^2(1+z^2)/2}dy$$

$$=\frac{1}{\pi}\int_0^{\infty} ye^{-y^2(1+z^2)/2}dy=-\frac{1}{\pi(1+z^2)}\int_0^{\infty}e^{-y^2(1+z^2)/2}d\left(\frac{-y^2(1+z^2)}{2}\right)=\frac{1}{\pi(1+z^2)},\ \ -\infty<z<\infty.$$

Vậy $Z=\dfrac{X}{Y}$ là biến ngẫu nhiên có phân bố Cauchy tham số 1.

Ví dụ 3.26: Giả sử X , Y là hai biến ngẫu nhiên liên tục có hàm mật độ đồng thời $f_{X,Y}(x,y)$.

Xét phép biến đổi $\begin{cases} x = r\cos\theta \\ y = r\sin\theta;\ r \ge 0, 0 \le \theta < 2\pi \end{cases}$

Định thức Jacobi $$\frac{D(x,y)}{D(r,\theta)} = \begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r .$$

Đặt

$$\begin{cases} X = R\cos\Theta \\ Y = R\sin\Theta \end{cases} \qquad \textbf{(3.58)}$$

Áp dụng công thức (3.55) ta được

$$f_{R,\Theta}(r,\theta) = r f_{X,Y}\left(r\cos\theta, r\sin\theta\right) \qquad \textbf{(3.59)}$$

3.6.5 Kỳ vọng của hàm các biến ngẫu nhiên

A. Kỳ vọng của hàm một biến ngẫu nhiên

Định lý 3.8: Kỳ vọng của $Y = g(X)$ được tính theo công thức sau (xem công thức 2.49):

$$\mathrm{E}Y = \begin{cases} \sum\limits_i g(x_i) p_X(x_i) & \text{nÕu } X \text{ rêi r¹c cã hµm khèi lîng x¸c suÊt } p_X(x_i) \\ \int\limits_{-\infty}^{\infty} g(x) f_X(x)dx & \text{nÕu } X \text{ liªn tôc cã hµm mËt ®é } f_X(x) \end{cases} \qquad \textbf{(3.60)}$$

Chứng minh: Trường hợp X rời rạc có miền giá trị $R_X = \left\{x_i \middle| i \in I\right\}$ thì Y nhận các giá trị có dạng $g(x_i)$. Ký hiệu $p_X(x_i) = P\left\{X = x_i\right\}$ là hàm khối lượng xác suất của X.

Đặt $I = \bigcup\limits_k I_k$, trong đó các tập I_k thỏa mãn:

- $i, j \in I_k$ thì $g(x_i) = g(x_j)$,
- $i \in I_k$, $j \in I_{k'}$, $k \ne k'$ thì $g(x_i) \ne g(x_j)$.

Nói cách khác các tập I_k tạo thành một phân hoạch của I theo các giá trị bằng nhau của miền giá trị của $Y = g(X)$

Vậy $$\mathrm{E}Y = \sum_k g(x_k) \sum_{i \in I_k} p_X(x_i) = \sum_{i \in I} g(x_i) p_X(x_i) .$$

Trường hợp X liên tục, $g(x)$ đơn điệu, áp dụng công thức (3.43), (3.44a), (3.44b) ta có

$$\mathrm{E}Y = \int_{-\infty}^{\infty} y f_Y(y)dy = \int_{-\infty}^{\infty} g(x) f_Y(g(x))\left|g'(x)\right|dx = \int_{-\infty}^{\infty} g(x) f_X(x)dx .$$

Nếu $g(x)$ không đơn điệu, ta phân hoạch miền lấy tích phân thành các miền đơn điệu của $g(x)$ và nhận được công thức trên.

B. Kỳ vọng của hàm n biến ngẫu nhiên

Định lý 3.9: Giả sử $X_1,...,X_n$ là n biến ngẫu nhiên. Xét hàm n biến liên tục $g(x_1,...,x_n)$, khi đó kỳ vọng của $Y = g(X_1,...,X_n)$ được tính theo công thức sau (xem công thức 2.49):

- Trường hợp $X_1,...,X_n$ rời rạc với hàm khối lượng xác suất $p_{X_1,...,X_n}(x_{i_1},....,x_{i_n})$ thì

$$\mathrm{E}Y = \sum_{i_1}...\sum_{i_n} g(x_{i_1},...,x_{i_n})\, p_{X_1,...,X_n}(x_{i_1},....,x_{i_n}) \qquad \textbf{(3.61a)}$$

- Trường hợp $X_1,...,X_n$ liên tục với hàm mật độ $f_{X_1,...,X_n}(x_1,...,x_n)$ thì

$$\mathrm{E}Y = \int_{-\infty}^{\infty}...\int_{-\infty}^{\infty} g(x_1,...,x_n)\, f_{X_1,...,X_n}(x_1,...,x_n)dx_1...dx_n \qquad \textbf{(3.61b)}$$

C. Tính chất tuyến tính của kỳ vọng, kỳ vọng của các biến ngẫu nhiên độc lập

a) Với mọi biến ngẫu nhiên $X_1,...,X_n$, với mọi hằng số $a_1,...,a_n$:

$$\mathrm{E}\left(\sum_{k=1}^{n} a_k X_k\right) = \sum_{k=1}^{n} a_k\,\mathrm{E}\left(X_k\right) \qquad \textbf{(3.62)}$$

b) Giả sử X,Y là hai biến ngẫu nhiên độc lập và g,h là hai hàm liên tục bất kỳ, khi đó ta có:

$$\mathrm{E}\left[g(X)h(Y)\right] = \mathrm{E}\left[g(X)\right]\mathrm{E}\left[h(Y)\right] \qquad \textbf{(3.63)}$$

Tính chất này có thể mở rộng cho n biến ngẫu nhiên độc lập $X_1,...,X_n$.

$$\mathrm{E}\left[\prod_{k=1}^{n} g_k(X_k)\right] = \prod_{k=1}^{n}\mathrm{E}\left[g_k(X_k)\right] \qquad \textbf{(3.64)}$$

Chứng minh: a) Ta chứng minh công thức đối với trường hợp hai biến ngẫu nhiên X,Y liên tục. Trường hợp khác được chứng minh tương tự.

Áp dụng công thức (3.61b) ta có:

$$\mathrm{E}\left(aX+bY\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by)\, f_{X,Y}(x,y)dxdy$$

$$= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f_{X,Y}(x,y)dxdy + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{X,Y}(x,y)dxdy$$

$$= a\int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy\right]dx + b\int_{-\infty}^{\infty} y\left[\int_{-\infty}^{\infty} f_{X,Y}(x,y)dx\right]dy$$

$$= a\int_{-\infty}^{\infty} x f_X(x)dx + b\int_{-\infty}^{\infty} y f_Y(y)dy = a\,\mathrm{E}(X) + b\,\mathrm{E}(Y).$$

b) X,Y độc lập do đó $f_{X,Y}(x,y) = f_X(x)f_Y(y)$.

$$\mathrm{E}\left[g(X)h(Y)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_{X,Y}(x,y)dxdy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dxdy$$

$$= \int_{-\infty}^{\infty} g(x)f_X(x)dx \int_{-\infty}^{\infty} h(y)f_Y(y)dy = \mathrm{E}\left[g(X)\right]\mathrm{E}\left[h(Y)\right].$$

3.7 PHÂN BỐ CÓ ĐIỀU KIỆN VÀ KỲ VỌNG CÓ ĐIỀU KIỆN

3.7.1 Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên rời rạc

Định nghĩa 3.7: *Giả sử X biến ngẫu nhiên rời rạc nhận các giá trị $\{x_1, x_2, ..., x_n\}$ và B là một biến cố trong cùng phép thử với X có xác suất $P(B) > 0$. Khi đó bảng phân bố xác suất của X với điều kiện B được xác định như sau*

$X\vert B$	x_1	x_2	$\dots$	x_i	$\dots$	x_n
P	$p_{X\vert B}(x_1\mid B)$	$p_{X\vert B}(x_2\mid B)$	$\dots$	$p_{X\vert B}(x_i\mid B)$	$\dots$	$p_{X\vert B}(x_n\mid B)$

trong đó

$$p_{X|B}(x_i|B) = \frac{P(\{X = x_i\}B)}{P(B)} \; ; \; i = 1,...,n \; . \qquad \textbf{(3.65)}$$

Hàm $p_{X|B}(x_i|B)$ xác định bởi công thức (3.65) được gọi là *hàm khối lượng xác suất của biến ngẫu nhiên X với điều kiện B*.

Định nghĩa 3.8: *Kỳ vọng của X với điều kiện B được định nghĩa:*

$$\mathrm{E}\left[X|B\right] = \sum_{i=1}^{n} x_i p_{X|B}(x_i|B) \qquad \textbf{(3.66)}$$

Định nghĩa 3.9: *Giả sử X, Y có tập các giá trị $\{x_1, x_2, ..., x_n\}$ và $\{y_1, y_2, ..., y_m\}$. Với mỗi y_j ta có bảng phân bố xác suất có điều kiện của X với điều kiện biến cố $\{Y = y_j\}$*:

$X\vert Y = y_j$	x_1	x_2	$\dots$	x_i	$\dots$	x_n
P	$p_{X\vert Y}(x_1\vert y_j)$	$p_{X\vert Y}(x_2\vert y_j)$	$\dots$	$p_{X\vert Y}(x_i\vert y_j)$	$\dots$	$p_{X\vert Y}(x_n\vert y_j)$

trong đó

$$p_{X|Y}(x_i \mid y_j) = \frac{p_{X,Y}(x_i, y_j)}{p_Y(y_j)} \; ; \; i = 1,\ldots,n, \; j = 1,\ldots,m \; . \qquad \textbf{(3.67)}$$

Tương tự với mỗi x_i ta có bảng phân bố xác suất có điều kiện của Y với điều kiện $\{X = x_i\}$:

$Y\big\vert X = x_i$	y_1	y_2	$\ldots$	y_j	$\ldots$	y_m
P	$p_{Y\vert X}(y_1\vert x_i)$	$p_{Y\vert X}(y_2\vert x_i)$	$\ldots$	$p_{Y\vert X}(y_j\vert x_i)$	$\ldots$	$p_{Y\vert X}(y_m\vert x_i)$

trong đó

$$p_{Y|X}(y_j | x_i) = \frac{p_{X,Y}(x_i, y_j)}{p_X(x_i)} \; ; \; i = 1,\ldots,n, \; j = 1,\ldots,m \; . \qquad \textbf{(3.68)}$$

Tính chất 3.5:

a) $0 \le p_{X|Y}(x_i \mid y_j) \le 1$; $\forall i = 1,\ldots,n, \; j = 1,\ldots,m$.

b) Với mỗi j : $\sum_{i=1}^{n} p_{X|Y}(x_i \mid y_j) = 1$.

c) Nếu X , Y độc lập thì $p_{X|Y}(x_i \mid y_j) = p_X(x_i)$ và $p_{Y|X}(y_j \mid x_i) = p_Y(y_j)$.

Định nghĩa 3.10: *Kỳ vọng có điều kiện của X với điều kiện $Y = y_j$ và kỳ vọng có điều kiện của Y với điều kiện $X = x_i$ tương ứng* được tính theo công thức sau

$$\mathrm{E}\left[X \middle| Y = y_j\right] = \sum_{i=1}^{n} x_i p_{X|Y}(x_i \mid y_j) ; \quad \mathrm{E}\left[Y \middle| X = x_i\right] = \sum_{j=1}^{m} y_j p_{Y|X}(y_j \mid x_i) . \qquad \textbf{(3.69)}$$

Ví dụ 3.27: Thực hiện lặp lại nhiều lần phép thử Bernoulli, xác suất xuất hiện của biến cố A trong mỗi lần thử là p , $0 < p < 1$. Gọi Y là biến ngẫu nhiên chỉ lần thử đầu tiên xuất hiện biến cố A. Gọi B là biến cố: "Trong n lần thử đầu tiên có duy nhất một lần xuất hiện biến cố A".

a) Tìm bảng phân bố xác suất của Y .

b) Tìm phân bố của Y với điều kiện B.

Giải: a) Ta có bảng phân bố xác suất của Y (xem ví dụ 2.9)

Y	1	2	$\ldots$	k	$\ldots$
P	p	qp	$\ldots$	$q^{k-1}p$	$\ldots$

với $q = 1 - p$.

b) Phân bố của Y với điều kiện B:

Khi $P(B)>0$ thì $P\left(\{Y=k\}\middle|B\right)=\dfrac{P\left(\{Y=k\}B\right)}{P(B)}$.

- Rõ ràng khi $k>n$ thì biến cố $\{Y=k\}$ kéo theo trong n phép thử đầu tiên biến cố A không xuất hiện. Do đó $\{Y=k\}B=\varnothing$, vậy $P\left(\{Y=k\}B\right)=0$.

- Khi $k\le n$, áp dụng công thức Bernoulli ta có: $P(B)=C_n^1pq^{n-1}=npq^{n-1}$.

Mặt khác $P\left(\{Y=k\}B\right)=P\{\text{chØxuÊt hiÖn biÕn cè } A \text{ ë lÇn thö thø } k\}=pq^{n-1}$

Vậy $$P\left(\{Y=k\}\middle|B\right)=\begin{cases}\dfrac{1}{n} & \text{nÕu } \ k\le n\\ 0 & \text{nÕu } \ k>n\end{cases}.$$

Ví dụ 3.28: Thống kê dân cư của một thành phố nọ ở độ tuổi trưởng thành về thu nhập hàng tháng X và lứa tuổi Y thu được kết quả trong bảng sau.

X \ Y	30	45	70
1	0,02	0,01	0,05
2	0,06	0,03	0,10
3	0,21	0,18	0,15
4	0,08	0,07	0,04

trong đó $X=1,2,3,4$ tương ứng chỉ thu nhập triệu đồng /tháng.

$Y=30,\ 45,\ 70$ chỉ độ tuổi của người dân trong khoảng: 25-35, 35-55, 55-85.

Tìm thu nhập trung bình theo lứa tuổi.

Giải: Thu nhập trung bình theo lứa tuổi là kỳ vọng có điều kiện của X theo Y.

Với $Y=30$ bảng phân bố xác suất điều kiện tương ứng:

$X\|Y=30$	1	2	3	4
P	$\dfrac{0,02}{0,37}$	$\dfrac{0,06}{0,37}$	$\dfrac{0,21}{0,37}$	$\dfrac{0,08}{0,37}$

Từ đó $\mathrm{E}\left[X\middle|Y=30\right]=1\cdot\dfrac{2}{37}+2\cdot\dfrac{6}{37}+3\cdot\dfrac{21}{37}+4\cdot\dfrac{8}{37}=\dfrac{109}{37}=2,9459$.

Tương tự $\mathrm{E}\left[X \middle| Y=45\right]=\dfrac{89}{29}=3,069$. $\mathrm{E}\left[X \middle| Y=70\right]=\dfrac{86}{34}=2,5294$.

Vậy thu nhập trung bình ở độ tuổi 30 là 2.945.900đ/tháng, độ tuổi 45 là 3.069.000đ/tháng và độ tuổi 70 là 2.529.400 đ/tháng.

3.7.2 Phân bố có điều kiện và kỳ vọng có điều kiện của biến ngẫu nhiên liên tục

Định nghĩa 3.11: *Xét biến ngẫu nhiên X và biến cố B trong cùng một phép thử, thỏa mãn điều kiện $P(B)>0$. Hàm phân bố xác suất của X với điều kiện B được định nghĩa và ký hiệu như sau*

$$F_{X|B}(x\,|\,B)=P\left\{X\le x \middle| B\right\}=\frac{P\left(\{X\le x\}B\right)}{P(B)}. \qquad \textbf{(3.70)}$$

Hàm mật độ xác suất của X với điều kiện B được định nghĩa và ký hiệu như sau

$$f_{X|B}(x\,|\,B)=\frac{dF_{X|B}(x\,|\,B)}{dx}. \qquad \textbf{(3.71)}$$

Định nghĩa 3.12: *Giả sử $f_{X,Y}(x,y)$ là hàm mật độ xác suất đồng thời của hai biến ngẫu nhiên liên tục X, Y và $f_X(x)$ là hàm mật độ của biến ngẫu nhiên thành phần X. Hàm phân bố xác suất của Y với điều kiện $\{X=x\}$ được định nghĩa và ký hiệu như sau:*

$$F_{Y|X}(y\,|\,x)=\int_{-\infty}^{y}\frac{f_{X,Y}(x,v)}{f_X(x)}dv,\ \text{nếu}\ \ f_X(x)>0. \qquad \textbf{(3.72)}$$

Đạo hàm của hàm phân bố xác suất có điều kiện $\dfrac{d}{dy}F_{Y|X}(y\,|\,x)$ được gọi là hàm mật độ xác suất có điều kiện của Y với điều kiện $\{X=x\}$ và ký hiệu là $f_{Y|X}(y\,|\,x)$. (Đây là hàm của biến y còn x đóng vai trò là tham số).

$$f_{Y|X}(y\,|\,x)=\frac{f_{X,Y}(x,y)}{f_X(x)}\ \text{với mọi}\ (x,y)\in R_{X,Y}\ \text{và thoả mãn}\ f_X(x)>0. \qquad \textbf{(3.73)}$$

Tương tự ta có hàm phân bố xác suất và hàm mật độ xác suất có điều kiện của X với $\{Y=y\}$

$$F_{X|Y}(x\,|\,y)=\int_{-\infty}^{x}\frac{f_{X,Y}(u,y)}{f_Y(y)}du,\ \text{nếu}\ \ f_Y(y)>0. \qquad \textbf{(3.74)}$$

$$f_{X|Y}(x\,|\,y)=\frac{f_{X,Y}(x,y)}{f_Y(y)}\ \text{với mọi}\ (x,y)\in R_{X,Y}\ \text{và thoả mãn}\ f_Y(y)>0. \qquad \textbf{(3.75)}$$

Tính chất 3.6:

a) $f_{X|Y}(x\,|\,y)\ge 0$.

b) Với mỗi y thỏa mãn $f_Y(y) > 0$ thì có $\int\limits_{-\infty}^{\infty} f_{X|Y}(x\,|\,y)dx = 1$.

c) Trường hợp X, Y độc lập ta có $f_{X|Y}(x\,|\,y) = f_X(x)$ và $f_{Y|X}(y\,|\,x) = f_Y(y)$.

Ví dụ 3.29: Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời đã cho trong ví dụ 3.8 thì

$$f_{X|Y}(x\,|\,y) = \begin{cases} \dfrac{1}{2(1-|y|)} & \text{nÕu } |y| < 1 \text{ vµ } |x| + |y| \le 1 \\ 0 & \text{nÕu ngîc l¹i} \end{cases}.$$

Định nghĩa 3.13: *Giả sử hai biến ngẫu nhiên X, Y có $f_{Y|X}(y\,|\,x)$ là hàm mật độ xác suất có điều kiện của Y với điều kiện $\{X = x\}$. Khi đó kỳ vọng của Y với điều kiện $\{X = x\}$ được ký hiệu và định nghĩa theo công thức sau*

$$\mathrm{E}\left[Y \middle| X = x\right] = \int\limits_{-\infty}^{\infty} y f_{Y|X}(y\,|\,x)dy \qquad \textbf{(3.76)}$$

$\mathrm{E}\left[Y \middle| X = x\right]$ xác định ở công thức (3.69), (3.76) *là một hàm của x, được gọi là hàm hồi qui của Y đối với X.*

$\mathrm{E}\left[X \middle| Y = y\right]$ *là một hàm của y, được gọi là hàm hồi qui của X đối với Y.*

Ví dụ 3.30: Giả sử X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời đã cho trong ví dụ 3.7

$$f_{X,Y}(x,y) = \begin{cases} 2 & \text{nÕu } 0 < y \le x < 1 \\ 0 & \text{nÕu ngîc l¹i} \end{cases}$$

Tìm hàm mật độ có điều kiện $f_{X|Y}(x\,|\,y)$, $f_{Y|X}(y\,|\,x)$ và kỳ vọng điều kiện.

Giải: Theo kết quả của ví dụ 3.7 ta có

$$f_X(x) = 2x \qquad \text{nÕu } 0 < x < 1$$

$$f_Y(y) = 2(1-y) \quad \text{nÕu } 0 < y < 1$$

Áp dụng công thức (3.73), (3.75) ta được

$$f_{Y|X}(y\,|\,x) = \frac{1}{x} \qquad 0 < y \le x < 1,\ 0 < x < 1$$

$$f_{X|Y}(x\,|\,y) = \frac{1}{1-y} \qquad 0 < y \le x < 1,\ 0 < y < 1.$$

Kỳ vọng của Y với điều kiện $\{X = x\}$

$$\mathrm{E}\left[Y\middle|X=x\right]=\int_{-\infty}^{\infty}yf_{Y|X}(y\,|\,x)dy=\int_{0}^{x}y\left(\frac{1}{x}\right)dy=\frac{y^2}{2x}\Bigg|_0^x=\frac{x}{2}\quad \text{nếu } 0<x<1.$$

Kỳ vọng của X với điều kiện $\{Y=y\}$

$$\mathrm{E}\left[X\middle|Y=y\right]=\int_{-\infty}^{\infty}xf_{X|Y}(x\,|\,y)dx=\int_{y}^{1}x\left(\frac{1}{1-y}\right)dx=\frac{x^2}{2(1-y)}\Bigg|_{x=y}^{x=1}=\frac{1+y}{2}\quad \text{nếu } 0<y<1.$$

3.7.3 Biến ngẫu nhiên kỳ vọng có điều kiện

Hàm hồi qui của Y đối với X là một hàm phụ thuộc biến x: $H(x)=\mathrm{E}\left[Y\middle|X=x\right]$.

Xét biến ngẫu nhiên của hàm $H(x)$ xác định bởi công thức (3.42):

$$\mathrm{E}\left[Y\middle|X\right]=H(X),\ \text{trong đó } H(x)=\mathrm{E}\left[Y\middle|X=x\right] \tag{3.77}$$

Tương tự nếu đặt hàm hồi qui của X đối với Y là $G(y)=\mathrm{E}\left[X\middle|Y=y\right]$, ta định nghĩa

$$\mathrm{E}\left[X\,|\,Y\right]=G(Y),\ \text{trong đó } G(y)=\mathrm{E}\left[X\middle|Y=y\right] \tag{3.78}$$

Biến ngẫu nhiên xác định bởi công thức (3.76), (3.77) *gọi là biến ngẫu nhiên kỳ vọng có điều kiện.*

Áp dụng công thức (3.60) ta được $\mathrm{E}(Y)=\mathrm{E}\left(\mathrm{E}\left[Y\,|\,X\right]\right)$, cụ thể:

- Nếu X rời rạc với hàm khối lượng xác suất $p_X(x_i)=P\{X=x_i\}$ thì

$$\mathrm{E}(Y)=\sum_{x_i\in R_X}\mathrm{E}\left[Y\middle|X=x_i\right]p_X(x_i) \tag{3.79}$$

- Nếu X liên tục với hàm mật độ xác suất $f_X(x)$ thì

$$\mathrm{E}(Y)=\int_{-\infty}^{\infty}\mathrm{E}\left[Y\middle|X=x\right]f_X(x)dx, \tag{3.80}$$

3.8 PHÂN BỐ CHUẨN NHIỀU CHIỀU

3.8.1 Khái niệm phân bố chuẩn n chiều

Định nghĩa 3.14: *Véc tơ ngẫu nhiên* $(X_1,X_2,...,X_n)$ *được gọi là có phân bố chuẩn* n *chiều nếu mật độ đồng thời có dạng*

$$f(x)=\frac{1}{\sqrt{(2\pi)^n\det M}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})M^{-1}(\boldsymbol{x}-\boldsymbol{a})^t\right\} \tag{3.81}$$

trong đó $\boldsymbol{x}=(x_1,x_2,...,x_n)\in \mathrm{R}^n$; $\boldsymbol{a}=(\mu_1,\mu_2,...,\mu_n)\in \mathrm{R}^n$; $M=\left[C_{ij}\right]_{n\times n}$ là ma trận đối xứng xác định dương; ma trận cột $(\boldsymbol{x}-\boldsymbol{a})^t$ là chuyển vị của ma trận hàng $(\boldsymbol{x}-\boldsymbol{a})$.

Ký hiệu $\left(X_1,X_2,...,X_n\right)\sim \mathbf{N}(\mathbf{a},M)$

Định lý 3.10: Nếu véc tơ ngẫu nhiên $X=\left(X_1,X_2,...,X_n\right)$ có phân bố chuẩn n chiều với mật độ đồng thời xác định như trong định nghĩa trên thì

a) Với mọi $i=1,...,n$: $\mu_i=\mathrm{E}X_i$.

b) M là ma trận hiệp phương sai của véc tơ ngẫu nhiên $X=\left(X_1,X_2,...,X_n\right)$, nghĩa là:

$$C_{ij}=\operatorname{cov}(X_i,X_j). \tag{3.82}$$

c) Các biến ngẫu nhiên $X_1,X_2,...,X_n$ độc lập khi và chỉ khi chúng không tương quan, nói cách khác M là ma trận chéo.

d) Biến ngẫu nhiên thành phần X_i có phân bố chuẩn $\mathbf{N}(\mu_i;C_{ii})$.

e) $\left(Y_1,Y_2,...,Y_n\right)=H\left(X_1,X_2,...,X_n\right)^t$ cũng có phân bố chuẩn, trong đó H là ma trận vuông cấp n bất kỳ có $\det H\neq 0$.

Ví dụ 3.31: Tìm hàm phân bố của biến ngẫu nhiên $T=2X-3Y+7Z$ biết rằng:

$$\left(X,Y,Z\right)\sim \mathbf{N}(\mathbf{a},M),\ \mathbf{a}=(8,3,-2),\ M=\begin{bmatrix}3 & -0,5 & 0,4\\ -0,5 & 2 & 0,7\\ 0,4 & 0,7 & 5\end{bmatrix}.$$

Giải: Theo định lý trên T có phân bố chuẩn, ta chỉ cần tìm kỳ vọng và phương sai.

$$\mathrm{E}T=2\mathrm{E}X-3\mathrm{E}Y+7\mathrm{E}Z=(2)(8)-(3)(3)+(7)(-2)=-7.$$

$$\mathrm{D}T=\mathrm{E}\left[T-\mathrm{E}T\right]^2=\mathrm{E}\left[2(X-8)-3(Y-3)+7(Z+2)\right]^2$$

$$=4\mathrm{E}\left[(X-8)\right]^2+9\mathrm{E}\left[(Y-3)\right]^2+49\mathrm{E}\left[(Z+2)\right]^2$$
$$-12\mathrm{E}\left[(X-8)(Y-3)\right]+28\mathrm{E}\left[(X-8)(Z+2)\right]-42\mathrm{E}\left[(Y-3)(Z+2)\right]$$

$$=(4)(3)+(9)(2)+(49)(5)-(12)(-0,5)+(28)(0,4)-(42)(0,7)=262,8.$$

Vậy $T\sim \mathbf{N}(-7;262,8)$.

3.8.2 Phân bố chuẩn hai chiều

Định nghĩa 3.15: *Véc tơ ngẫu nhiên (X,Y) được gọi là có phân bố chuẩn hai chiều nếu hàm mật độ xác suất đồng thời có dạng:*

$$f_{X,Y}(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\exp\left(-\frac{1}{2}Q(x,y)\right), \tag{3.83}$$

trong đó $\sigma_1 > 0$, $\sigma_2 > 0$, $|\rho| < 1$ và

$$Q(x,y) = \frac{1}{(1-\rho^2)}\left\{\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right\}.$$

Định lý 3.11: Giả sử véc tơ ngẫu nhiên (X,Y) có phân bố chuẩn hai chiều với hàm mật độ xác suất đồng thời xác định như trong định nghĩa 3.13. Khi đó

a) X có phân bố chuẩn $\mathbf{N}(\mu_1;\sigma_1^2)$, Y có phân bố chuẩn $\mathbf{N}(\mu_2;\sigma_2^2)$.

b) Hệ số tương quan $\rho(X,Y) = \rho$.

c) X, Y không tương quan khi và chỉ khi X, Y độc lập.

d) Hàm mật độ xác suất của X với điều kiện $Y = y$:

$$f_{X|Y}(x\,|\,y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \frac{1}{\sigma_1\sqrt{1-\rho^2}}\varphi\left(\frac{x-\mu_1-(\sigma_1/\sigma_2)\rho(y-\mu_2)}{\sigma_1\sqrt{1-\rho^2}}\right) \quad \textbf{(3.84)}$$

trong đó $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$ là hàm mật độ xác suất của biến ngẫu nhiên chuẩn tắc $\mathbf{N}(0;1)$.

Nhận xét 3.3:

a) Ma trận tương quan của (X,Y) là

$$M = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \Rightarrow M^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1/\sigma_1^2 & -\rho/(\sigma_1\sigma_2) \\ -\rho/(\sigma_1\sigma_2) & 1/\sigma_2^2 \end{bmatrix}. \quad \textbf{(3.85)}$$

b) Từ định lý 3.11 suy ra rằng phân bố của X với điều kiện $Y = y$ có phân bố chuẩn với kỳ vọng $\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2)$ và phương sai $\sigma_1^2(1-\rho^2)$, do đó kỳ vọng có điều kiện:

$$\mathrm{E}\left[X \middle| Y = y\right] = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y-\mu_2). \quad \textbf{(3.86)}$$

là một hàm tuyến tính theo y. Đây là một tính chất rất quan trọng của phân bố chuẩn 2 chiều.

TÓM TẮT

Tương tự biến ngẫu nhiên, véc tơ ngẫu nhiên cũng được nghiên cứu theo hàm phân bố xác suất đồng thời (3.1).

Trường hợp véc tơ ngẫu nhiên rời rạc hai chiều (X,Y), hàm phân bố xác suất đồng thời được xác định từ hàm khối lượng xác suất đồng thời (3.8)-(3.10). Hàm phân bố xác suất đồng thời của véc tơ ngẫu nhiên liên tục được xác định từ hàm mật độ xác suất đồng thời (3.14).

Từ bảng phân bố xác suất đồng thời của véc tơ ngẫu nhiên rời rạc hai chiều (X,Y) ta có thể xác định được bảng phân bố xác suất thành phần X và thành phần Y (Nhận xét 3.1).

Công thức (3.18)-(3.19) cho phép tìm hàm mật độ thành phần X và thành phần Y từ hàm mật độ xác suất đồng thời của véc tơ ngẫu nhiên liên tục (X,Y).

Các tiêu chuẩn để nhận biết hai biến ngẫu nhiên X,Y độc lập (công thức 3.20−3.23).

Ngoài hai đặc trưng kỳ vọng và phương sai đã được xét ở chương 2, trong chương này còn xét các đặc trưng hiệp phương sai $\text{cov}(X,Y)$ (công thức 3.32a–3.34), hệ số tương quan ρ_{XY} (công thức 3.39) và ma trận hiệp phương sai (công thức 3.38).

Hàm khối lượng xác suất có điều kiện (công thức 3.65, 3.67, 3.68). Hàm mật độ xác suất có điều kiện (công thức 3.70 − 3.75).

Kỳ vọng có điều kiện (công thức 3.66, 3.69, 3.76).

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

3.1 Bảng phân bố xác suất của X và Y cho phép xác định bảng phân bố xác suất đồng thời của (X,Y).

Đúng ☐ Sai ☐.

3.2 Bảng phân bố xác suất đồng thời của (X,Y) xác định bảng phân bố xác suất của hai biến ngẫu nhiên thành phần X và Y.

Đúng ☐ Sai ☐.

3.3 Nếu hai biến ngẫu nhiên X, Y độc lập thì bảng phân bố xác suất của X và Y cho phép xác định bảng phân bố xác suất đồng thời của (X,Y).

Đúng ☐ Sai ☐.

3.4 Hai biến ngẫu nhiên độc lập có hiệp phương sai bằng 0.

Đúng ☐ Sai ☐.

3.5 Hai biến ngẫu nhiên có hiệp phương sai bằng 0 thì độc lập.

Đúng ☐ Sai ☐.

3.6 Hiệp phương sai luôn nhận giá trị dương.

Đúng ☐ Sai ☐.

3.7 Nếu $Y = aX + b$, $a \neq 0$ thì hệ số tương quan $\rho_{X,Y}$ luôn luôn bằng 1.

Đúng ☐ Sai ☐.

3.8 Nếu $\{x_1, x_2, ..., x_n\}$ là tập giá trị của X thì $\{f(x_1), f(x_2), ..., f(x_n)\}$ là tập giá trị của hàm hồi quy $f(x) = \text{E}(Y|X = x)$ của Y đối với X.

Đúng ☐ Sai ☐.

3.9 Nếu hai biến ngẫu nhiên X, Y độc lập thì hàm hồi quy $f(x)=\mathrm{E}\left(Y\middle|X=x\right)$ của Y đối với X và hàm hồi quy $g(y)=\mathrm{E}\left(X\middle|Y=y\right)$ của X đối với Y là hai hàm hằng.

Đúng ☐ Sai ☐.

3.10 Nếu hiệp phương sai của hai biến ngẫu nhiên bằng 0 thì hai kỳ vọng của chúng bằng nhau ($\mathrm{cov}(X,Y)=0$ thì $\mathrm{E}X=\mathrm{E}Y$).

Đúng ☐ Sai ☐.

3.11 Hàm mật độ xác suất đồng thời của véc tơ ngẫu nhiên liên tục (X,Y) bằng tích của hai hàm mật độ xác suất thành phần X và Y.

Đúng ☐ Sai ☐.

3.12 Giả sử (X,Y) là véc tơ ngẫu nhiên phân bố chuẩn, khi đó X, Y độc lập khi và chỉ khi X, Y không tương quan.

Đúng ☐ Sai ☐.

3.13 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời như sau

X \ Y	y_1	y_2	y_3
x_1	0,18	0,22	0,16
x_2	0,08	0,16	0,20

Tìm bảng phân bố xác suất của hai biến ngẫu nhiên thành phần X, Y.

3.14 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời như sau

X \ Y	-1	1
-1	1/6	1/4
0	1/6	1/8
1	1/6	1/8

Hãy xác định $\mathrm{E}X, \mathrm{E}Y$, $\mathrm{cov}(X,Y)$ và $\rho_{X,Y}$.

3.15 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời như sau

X \ Y	1	2	3
1	0,12	0,15	0,03
2	0,28	0,35	0,07

a) Chứng minh rằng X, Y có độc lập.

b) Tìm quy luật phân bố của biến ngẫu nhiên $Z = XY$.

c) Tính các kỳ vọng $\mathrm{E}X, \mathrm{E}Y, \mathrm{E}Z$.

3.16 Cho X, Y là hai biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X	0	1	2	3
P	0,4	0,3	0,2	0,1

Y	0	1	2	3	4
P	0,1	0,3	0,4	0,15	0,05

Tìm bảng phân bố xác suất đồng thời của X, Y. Tính xác suất $P\{X > Y\}$.

3.17 Gieo đồng thời một con xúc xắc và một đồng tiền. Gọi X là biến ngẫu nhiên chỉ số chấm của con xúc xắc và Y là biến ngẫu nhiên chỉ mặt sấp (1) hay mặt ngửa (0) của đồng tiền. Lập bảng phân bố xác suất đồng thời của X và Y.

3.18 Cho bảng phân bố xác suất đồng thời của X, Y

Y \ X	26	30	41	50
23	0,05	0,08	0,12	0,04
27	0,09	0,30	0,11	0,21

Tìm bảng phân bố xác suất điều kiện của Y khi $X = 26$ và của X khi $Y = 27$.

3.19 Cho bảng phân bố xác suất đồng thời của X, Y

Y \ X	1	3	4	8
3	0,15	0,06	0,25	0,04
6	0,30	0,10	0,03	0,07

a) Tìm kỳ vọng có điều kiện của Y khi $X = 1$.

b) Tìm các kỳ vọng $\mathrm{E}X, \mathrm{E}Y$ và phương sai $\mathrm{D}X, \mathrm{D}Y$.

3.20 Cho X, Y là hai biến ngẫu nhiên có bảng phân bố xác suất đồng thời

X \ Y	-1	0	1
-1	4α	α	4α
0	α	2α	α

1	0	2α	0

a) Tìm α. Tính $\mathrm{E}X$, $\mathrm{E}Y$.

b) Tính $\mathrm{cov}(X,Y)$, $\rho_{X,Y}$.

c) X và Y có độc lập không.

3.21 Giả sử X là biến ngẫu nhiên rời rạc có phân bố Poisson tham số λ. Tìm hàm khối lượng xác suất với điều kiện $B=\{X \text{ chẵn}\}$.

3.22 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{X,Y}(x,y)=\begin{cases} kx & \text{nếu } 0<y<x<1 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

a) Tìm k.

b) Tìm các hàm mật độ xác suất của X và của Y.

c) X và Y có độc lập không?

3.23 Hàm phân bố xác suất của véc tơ ngẫu nhiên (X,Y) có dạng

$$F_{X,Y}(x,y)=\begin{cases} 1-e^{-x}-e^{-y}+e^{-x-y} & \text{nếu } x>0, y>0 \\ 0 & \text{nếu ngược lại} \end{cases}.$$

Tìm hàm mật độ xác suất đồng thời $f_{XY}(x,y)$ và hàm mật độ có điều kiện $f(x|y)$.

3.24 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{X,Y}(x,y)=\frac{k}{(1+x^2)(1+y^2)}.$$

a) Tìm k. Tìm hàm phân bố xác suất đồng thời mật độ của X, Y.

b) X và Y có độc lập không?

c) Tính xác suất để véc tơ ngẫu nhiên (X,Y) nhận giá trị nằm trong hình chữ nhật với các đỉnh là $A(1,1); B(\sqrt{3},1); C(1,0)$ và $D(\sqrt{3},0)$.

3.25 Chứng minh rằng hàm phân bố điều kiện và hàm mật độ điều kiện của biến ngẫu nhiên X với điều kiện $B=\{a<X\le b\}$ có dạng

$$F_X(x\,|\,a<X\le b)=\begin{cases} 0 & x\le a \\ \dfrac{F_X(x)-F_X(a)}{F_X(b)-F_X(a)} & a<x\le b; \\ 1 & x>b \end{cases}$$

$$f_X(x \mid a < X \le b) = \begin{cases} 0 & x \le a \\ f_X(x) / \int_a^b f_X(y)dy & a < x \le b \\ 0 & x > b \end{cases}$$

3.26 Giả sử biến ngẫu nhiên $X \sim \mathbf{N}(0;\sigma^2)$. Tìm hàm mật độ điều kiện, hàm phân bố điều kiện và kỳ vọng điều kiện với điều kiện $B = \{X > 0\}$.

3.27 Cho véc tơ ngẫu nhiên (X,Y) có hàm mật độ xác suất xác định như sau:

$$f_{X,Y}(x,y) = \begin{cases} k(x+y) & \text{nÕu } 0 < x < 2,\ 0 < y < 2 \\ 0 & \text{nÕu ng-îc l¹i} \end{cases}$$

a) Tìm k. Tìm hàm mật độ xác suất của X và của Y. X, Y có độc lập không.

b) Tìm hàm mật độ xác suất có điều kiện $f_{X|Y}(x \mid y)$ và $f_{Y|X}(y \mid x)$.

c) Tính $P\{0 < Y < 1/2 \mid X = 1\}$.

3.28 Cho X và Y là hai biến ngẫu nhiên có phân bố chuẩn đồng thời với

$\mathrm{E}X = 35,\ \mathrm{E}Y = 20,\ \mathrm{D}X = 36,\ \mathrm{D}Y = 16$ và $\rho_{X,Y} = 0{,}8$.

Tìm kỳ vọng và phương sai của $2X - 3Y$.

3.29 Cho hai biến ngẫu nhiên X, Y

a) Chứng minh bất đẳng thức Cauchy-Schwarz: $\left(\mathrm{E}[XY]\right)^2 \le \left(\mathrm{E}\,X^2\right)\left(\mathrm{E}\,Y^2\right)$.

b) Chứng minh hệ số tương quan thỏa mãn: $\left|\rho_{X,Y}\right| \le 1$.

3.30 Cho X là biến ngẫu nhiên với hàm phân bố $F_X(x)$ và hàm mật độ $f_X(x)$. Đặt $Y = aX + b$, $a \ne 0$ và b là hai hằng số thực. Tìm hàm phân bố và hàm mật độ của Y.