Bài số 4

BIẾN NGẪU NHIÊN HAI CHIỀU VÀ PHÂN PHỐI XÁC SUẤT

Trong các chương trước, chúng ta đã nghiên cứu các biến ngẫu nhiên và phân phối xác suất của chúng mà không gian mẫu là không gian một chiều, trong đó các kết quả của một phép thử được coi như các giá trị của một biến ngẫu nhiên. Nhưng có những trường hợp ta cần xét các kết quả xảy ra đồng thời của một vài biến ngẫu nhiên.

I.BIẾN NGẪU NHIÊN HAI CHIỀU.

1. Định nghĩa. *Cho hai biến ngẫu nhiên X, Y. Cặp (có thứ tự) hai biến ngẫu nhiên (X,Y) được gọi là **một biến ngẫu nhiên hai chiều**..*

X, Y được gọi là các thành phần của biến ngẫu nhiên hai chiều (*X, Y*).

Chú ý: Ta thấy rằng có thể xảy ra các trường hợp sau:

+ X và Y cùng là biến ngẫu nhiên rời rạc
+ X và Y cùng là biến ngẫu nhiên liên tục
+ X là biến ngẫu nhiên rời rạc còn Y là biến ngẫu nhiên liên tục
+ X là biến ngãu nhiên liên tục còn Y là biến ngẫu nhiên rời rạc.

Tuy nhiên trong chương trình của chúng ta chỉ xét hai trường hợp

i. X và Y cùng là biến ngẫu nhiên rời rạc: khi đó (X,Y) được gọi là ***biến ngẫu nhiên hai chiều rời rạc***.

ii. X và Y cùng là biến ngẫu nhiên liên tục: khi đó (X,Y) được gọi là ***biến ngẫu nhiên hai chiều liên tục.***

2. Một số ví dụ.

Ví dụ 1. Đo lượng kết tủa P và thể tích V của khí bay ra từ một thí nghiệm hóa học, ta có một không gian mẫu 2 chiều gồm các kết quả (P, V): Khi đó (P,V) là biến ngẫu nhiên hai chiều liên tục.

Ví dụ 2. Xét độ rắn H và độ bền T của đồng cán nguội, ta có một không gian mẫu 2 chiều gồm các kết quả (H,T): Khi đó (H,T) là biến ngẫu nhiên hai chiều liên tục.

Biến ngẫu nhiên (*X,Y*) nhận giá trị (*x,y*), tức là đồng thời ta có X nhận giá trị là *x* và *Y* nhận giá trị *y*, tập giá trị của (*X,Y*) có thể được biểu diễn hình học bởi các điểm trên mặt phẳng toạ độ O*xy.*

Ví dụ 3. Tung đồng thời hai con xúc sắc và ta quan tâm đến X : số chấm xuất hiện trên xúc sắc thứ nhất, Y : số chấm xuất hiện trên mặt con xúc sắc thứ hai. Khi đó (X,Y) là biến ngẫu nhiên hai chiều rời rạc, trong trường hợp này tập giá trị của (X,Y) là

$$\left\{(i,j)\middle|\, i=1,2,...,6;\ j=1,2,3,...,6\right\}.$$

Chú ý: Trong một nghiên cứu để xác định khả năng thành công của các sinh viên ở bậc đại học dựa trên kết quả học tập ở bậc trung học ta có thể phải sử dụng không gian mẫu 3 chiều, gồm điểm kiểm tra năng khiếu, xếp loại ở bậc phổ thông và điểm trung bình cuối năm đầu ở bậc đại học. Và khi đó ta cũng có thể định nghĩa biến ngẫu nhiên n chiều ($n \geq 3$) một cách tương tự.

II.PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN HAI CHIỀU.

Điều ta quan tâm là quy luật phân bố xác suất của biến ngẫu nhiên. Nhớ lại rằng, khi xét các biến ngẫu nhiên một chiều: đối với biến ngẫu nhiên rời rạc, phân bố xác suất được xác định thông qua hàm xác suất của chúng; còn đối với biến ngẫu nhiên liên tục thì phân bố xác suất được xác định bởi hàm mật độ xác suất.

1.Đối với biến ngẫu nhiên rời rạc

Định nghĩa. Cho X, Y là hai biến ngẫu nhiên rời rạc thì ***hàm phân phối xác suất đồng thời*** của chúng là một hàm hai biến $f(x, y)$ được xác định bởi

$$f(x, y) = P\{X = x; Y = y\}.$$

Ví dụ 4. Kểm tra một chiếc ti vi, gọi X là tuổi (làm tròn đến năm) và Y là số đèn điện tử bị hỏng của chiếc ti vi đó, thế thì $f(5,3) = P(X = 5; Y = 3)$ và đó là xác suất để chiếc ti vi 5 tuổi cần 3 đèn điện tử mới.

Nhận xét.
Hàm $f(x, y)$ là phân phối xác suất đồng thời của các biến ngẫu nhiên rời rạc X và Y nếu:

1. $f(x, y) \geq 0, \ \forall \ (x, y)$
2. $\sum_x \sum_y f(x, y) = 1$
3. Với miền A tùy ý trong mặt phẳng Oxy ta có $P[(X, Y) \in A] = \sum_A \sum f(x, y)$.

Tương tự như đối với biến ngẫu nhiên một chiều, khi (X, Y) là biến ngẫu nhiên hai chiều rời rạc (tức là X và Y là các biến ngẫu nhiên rời rạc), ta thường biểu diễn phân bố xác suất dưới dạng Bảng phân phối xác suất (đồng thời) như sau:

Y \ X	y_1	y_2	$\cdots$	y_k	$\cdots$
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\cdots$	$f(x_1, y_2)$	$\cdots$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\cdots$	$f(x_2, y_k)$	$\cdots$
$\cdots$			$\cdots$	$\cdots$	$\cdots$
x_k	$f(x_k, y_1)$	$f(x_k, y_2)$	$\cdots$	$f(x_k, y_k)$	$\cdots$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

Ví dụ 5. Chọn ngẫu nhiên hai chiếc ruột bút bi từ 1 hộp gồm 3 ruột bút xanh lơ, 2 ruột bút đỏ, 3 ruột bút màu vàng. Gọi *X* là số ruột bút xanh lơ được rút ra , *Y* là số ruột bút đỏ được chọn, tìm

a) Phân phối xác suất đồng thời $f(x, y)$.
b) $P[(X, Y) \in$ A] trong đó A là miền $\{(x, y) \mid x + y \leq 1\}$.

Giải: + Miền giá trị của (X, Y) là: $\{(0,0);(0,1);(1,0);(1,1);(0,2);(2,0)\}$

+ $f(0,1)$ là xác suất để chọn được một ruột bút đỏ và một ruột bút màu vàng.

+ Tổng số cách chọn 2 ruột bút từ 8 chiếc là $\mathrm{C}_8^2 = 28$.

+ Số cách chọn một ruột bút đỏ từ 2 ruột bút đỏ và một ruột bút xanh lá cây từ 3 ruột bút xanh lá cây là: $\mathrm{C}_2^1.\mathrm{C}_3^1 = 6$.

+ Do đó $f(0,1)\dfrac{6}{28} = \dfrac{3}{14}$.

a) Giả sử ta: rút được x ruột bút màu xanh từ 3 ruột bút: có C_3^x khả năng

rút được y ruột bút màu đỏ từ 2 ruột bút: có C_2^y khả năng

khi đó rút được $(2 - x - y)$ ruột bút màu vàng: có C_3^{2-x-y} khả năng

Do vậy hàm xác suất đồng thời là:

$$f(x, y) = \frac{C_3^x C_2^y C_3^{2-x-y}}{C_8^2} \text{ với } x = 0, 1, 2;\ y = 0, 1, 2;\ 0 \leq x + y \leq 2$$

Hay là:

$$f(x, y) = \begin{cases} \dfrac{C_3^x C_2^y C_3^{2-x-y}}{C_8^2}, (x, y) \in \{(i, j) \mid 0 \leq i, j \leq 2, 0 \leq i + j \leq 2\} \\ 0 \qquad\qquad ,(x, y) \in \{(i, j) \mid 0 \leq i, j \leq 2, i + j > 2\} \end{cases}$$

Ta có thể biểu diễn xác suất đồng thời như trong Bảng dưới đây. Chú ý rằng tổng các xác suất luôn bằng 1.

Y / X	0	1	2	**Tổng hàng**
0	$3/28$	$3/14$	$1/28$	$5/14$
1	$9/28$	$3/14$		$15/28$
2	$3/28$			$3/28$
Tổng cột	$15/28$	$3/7$	$1/28$	1

b) Dễ thấy: $P[(X, Y) \in$ A] $= P(X + Y \leq 1) = f(0, 0) + f(0,1) + f(1, 0) = \dfrac{3}{28} + \dfrac{3}{14} + \dfrac{9}{28} = \dfrac{9}{14}$.

2. Đối với biến ngẫu nhiên liên tục.

Khi X, Y là các biến ngẫu nhiên liên tục thì **hàm mật độ đồng thời** $f(x,y)$ là một hàm hai biến có đồ thị là một mặt cong nằm phía trên mặt phẳng Oxy. Khi đó với A là một miền nào đó trong Oxy thì $P[(X,Y)\in A]$ có giá trị bằng thể tích của khối trụ cong có đáy dưới là A và đáy trên là mặt cong $f(x,y)$.

Định nghĩa. Hàm $f(x, y)$ được gọi là **hàm mật độ xác suất đồng thời** của các ***<u>biến ngẫu nhiên liên tục</u>*** X và Y nếu:

1. $f(x, y) \geq 0,\ \forall (x, y)$
2. $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)dxdy = 1$
3. $P[(X,Y)\in A] = \iint\limits_A f(x,y)dxdy$ với A là miền tùy ý trong mặt phẳng Oxy.

Ví dụ 6. Một công ty kẹo phân phối các hộp kẹo sôcôla tổng hợp với các loại nhân kem, nhân bơ cứng và nhân quả hạnh nhân được phủ cả sôcôla đen và sôcôla trắng. Chọn ngẫu nhiên một hộp, gọi X, Y lần lượt là tỷ lệ sôcôla trắng nhân kem và sôcôla đen nhân kem. Giả sử hàm mật độ đồng thời của X, Y là:

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), (x,y)\in D: \{0\le x\le 1; 0\le y\le 1\} \\ 0,\ (x,y)\notin D \end{cases}$$

Tìm $P[(X, Y) \in A]$ trong đó $A = \left\{(x,y) \mid 0 < x < \frac{1}{2}; \frac{1}{4} < y < \frac{1}{2}\right\}$.

Giải: Ta có $P[(X, Y) \in A] = P(0 < X < \frac{1}{2},\ \frac{1}{4} < Y < \frac{1}{2}) = \iint\limits_A f(x,y)dxdy$

$$= \int\limits_{\frac{1}{4}}^{\frac{1}{2}}\int\limits_{0}^{\frac{1}{2}} \frac{2}{5}(2x+3y)dxdy \quad = \int\limits_{\frac{1}{4}}^{\frac{1}{2}} \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Bigg|_{x=0}^{x=\frac{1}{2}} dy$$

$$= \int\limits_{\frac{1}{4}}^{\frac{1}{2}} \left(\frac{1}{10} + \frac{3y}{5}\right)dy = \left(\frac{y}{10} + \frac{3y^2}{10}\right)\Bigg|_{\frac{1}{4}}^{\frac{1}{2}} = \frac{1}{10}\left[\left(\frac{1}{2}+\frac{3}{4}\right) - \left(\frac{1}{4}+\frac{3}{16}\right)\right] = \frac{13}{160}.$$

III. <u>PHÂN PHỐI BIÊN DUYÊN</u>

Bây giờ nếu đã biết phân phối xác suất đồng thời $f(x,y)$ của biến ngẫu nhiên hai chiều (X,Y), liệu ta có thể xác định được phân phối xác suất của các biến ngẫu nhiên thành phần hay không?

Định nghĩa. Biến ngẫu nhiên hai chiều (X,Y) rời rạc có hàm phân phối xác suất đồng thời là $f(x,y)$. Khi đó: **phân phối biên duyên** của X và Y được xác định bởi:

$$g(x)=\sum_{y}\mathrm{f}(\mathrm{x},\mathrm{y}) \text{ và } h(y)=\sum_{x}f(x,y).$$

Biến ngẫu nhiên hai chiều (X,Y) liên tục có hàm mật độ xác suất đồng thời là $f(x,y)$. Khi đó: **phân phối biên duyên** của X và Y được xác định bởi

$$g(x)=\int_{-\infty}^{\infty}f(x,y)dy \text{ và } h(y)=\int_{-\infty}^{\infty}f(x,y)dx.$$

Mô tả đối với biến ngẫu nhiên rời rạc:

Từ Bảng phân phối xác xuất đồng thời:

Y \ X	y_1	y_2	…	y_k	…	**Tổng theo hàng**
x_1	$f(x_1,y_1)$	$f(x_1,y_2)$	…	$f(x_1,y_2)$	…	p_1
x_2	$f(x_2,y_1)$	$f(x_2,y_2)$	…	$f(x_2,y_k)$	…	p_2
…			…	…	…	
x_k	$f(x_k,y_1)$	$f(x_k,y_2)$	…	$f(x_k,y_k)$	…	p_k
…	…	…	…	…	…	
Tổng theo cột	q_1	q_2		q_k		1

X	x_1	x_2	…	x_k	…	$g(x)=\sum_j f(x,y_j)$
$P(X=x_i)$	p_1	p_2		p_k		$\sum_i p_i$

Y	y_1	y_2	…	y_k	…	$h(y)=\sum_i f(x_i,y)$
$P(Y=y_j)$	q_1	q_2		q_k		$\sum_j q_j$

Ví dụ 7. Tìm phân phối biên duyên của X và Y biết phân phối xác suất đồng thời của chúng được cho trong Bảng sau:

Y \ X	0	1	2
0	$3/28$	$3/14$	$1/28$
1	$9/28$	$3/14$	
2	$3/28$		

Giải: Ta có bảng:

Y \ X	0	1	2	**Tổng hàng**
0	$3/28$	$3/14$	$1/28$	$5/14$
1	$9/28$	$3/14$		$15/28$
2	$3/28$			$3/28$
Tổng cột	$15/28$	$3/7$	$1/28$	1

Khi đó: ta có phân phối biên duyên của X là:

$$P(X=0)=g(0)=\sum_{y=0}^{2} f(0,y) \;=f(0,0)+f(0,1)+f(0,2)$$

$$=\frac{3}{28}+\frac{3}{14}+\frac{1}{28}=\frac{5}{14}.$$

$$P(X=1)=g(1)=\sum_{y=0}^{2} f(1,y) \;=f(1,0)+f(1,1)+f(1,2)$$

$$=\frac{9}{28}+\frac{3}{14}+0=\frac{15}{28}.$$

$$P(X=2)=g(2)=\sum_{y=0}^{2} f(2,y) \;=f(2,0)+f(2,1)+f(2,2)$$

$$=\frac{3}{28}+0+0=\frac{3}{28}.$$

X	0	1	2
$g(x)=P(X=x_i)$	$5/14$	$15/28$	$3/28$

Tương tự, phân phối biên duyên của Y là:

Y	0	1	2
$h(y)=P(Y=y_j)$	$15/28$	$3/7$	$1/28$

Ví dụ 8. Tìm phân phối biên duyên $g(x)$ và $h(y)$ với hàm mật độ đồng thời (trong Ví dụ 6) được cho bởi:

$$f(x,y)=\begin{cases}\frac{2}{5}(2x+3y), (x,y)\in D:\{0\le x\le 1; 0\le y\le 1\}\\ 0, (x,y)\notin D\end{cases}$$

Giải: Theo định nghĩa

$$g(x)=\begin{cases}\int_{-\infty}^{\infty} f(x,y)dy=\int_0^1 \frac{2}{5}(2x+3y)dy=\left(\frac{4xy}{5}+\frac{6y^2}{10}\right)\Bigg|\begin{matrix}y=1\\y=0\end{matrix}=\frac{4x+3}{5}, x\in[0;1]\\ 0, x\notin[0;1]\end{cases}$$

Tương tự,

$$h(y)=\begin{cases}\int_{-\infty}^{\infty} f(x,y)dx=\int_0^1 \frac{2}{5}(2x+3y)dx=\frac{2(1+3y)}{5}, y\in[0;1]\\ 0, y\notin[0;1]\end{cases}$$

<u>Chú ý:</u> Các phân phối biên duyên $g(x)$ và $h(y)$ thực sự là phân phối xác suất của các biến X và Y tương ứng vì nó thỏa mãn tất cả các điều kiện trong các Định nghĩa của Bài 3. Ví dụ, trong trường hợp liên tục:

$$\int_{-\infty}^{\infty} g(x)dx=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dydx=1$$

và $P(a < X < b) = P(a < X < b, -\infty < Y < \infty)$

$$=\int_a^b\int_{-\infty}^{\infty} f(x,y)dydx=\int_a^b g(x)dx.$$

IV.<u>PHÂN PHỐI XÁC SUẤT CÓ ĐIỀU KIỆN</u>

Với hai biến cố ngẫu nhiên một chiều A và B ta đã có công thức tính xác suất có điều kiện như sau:

$$P(B\,|\,A)=\frac{P(A\cap B)}{P(A)}, P(A)>0$$

- Nếu coi A là biến cố $X = x$, B là biến cố $Y = y$ trong đó X và Y là các biến ngẫu nhiên rời rạc thì ta có:

$$P(Y=y\,|\,X=x)=\frac{P(X=x,Y=y)}{P(X=x)}=\frac{f(x,y)}{g(x)}, g(x)>0$$

- Hàm $\frac{f(x,y)}{g(x)}$ được gọi là ***phân phối xác suất có điều kiện.*** Ta có thể dùng nó để tính các xác suất có điều kiện.

<u>Định nghĩa.</u> Giả sử (X,Y) biến ngẫu nhiên hai chiều rời rạc hoặc liên tục.

Phân phối có điều kiện của biến ngẫu nhiên Y với điều kiện $X = x$ được xác định bởi:

$$f\left(y|x\right)=\frac{f(x,y)}{g(x)},\quad g(x)>0.$$

Phân phối có điều kiện của biến ngẫu nhiên X với điều kiện $Y = y$ được xác định bởi:

$$f(x\,|\,y)=f\left(x|y\right)=\frac{f(x,y)}{h(y)}, h(y)>0.$$

- Xác suất để biến ngẫu nhiên rời rạc X lấy giá trị trong khoảng (a,b) khi đã biết biến ngẫu nhiên rời rạc $Y = y$ là:

$$P(a < X < b \mid Y = y) = \sum_{x} f(x \mid y),$$

trong đó tổng được lấy trên tất cả các giá trị của X nằm giữa a và b.

- Khi X và Y liên tục thì:

$$P(a < X < b \mid Y = y) = \int_a^b f(x \mid y)dx.$$

Ví dụ 9. Xét tiếp Ví dụ 7, tìm phân phối có điều kiện của X với điều kiện $Y = 1$ và dùng nó để xác định $P(X = 0 \mid Y = 1)$.

Giải: Chúng ta cần tìm $f(x \mid y)$, trong đó $y = 1$.

+ Ta có: $h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$

$$f(x \mid 1) = \frac{f(x,1)}{h(1)} = \frac{7}{3} f(x,1), \quad x = 0, 1, 2.$$

+ Do đó: $f(0 \mid 1) = \frac{7}{3} f(0,1) = \frac{7}{3}.\frac{3}{14} = \frac{1}{2}$, $\quad f(1 \mid 1) = \frac{7}{3} f(1,1) = \frac{7}{3}.\frac{3}{14} = \frac{1}{2}$

$$f(2 \mid 1) = \frac{7}{3} f(2,1) = \frac{7}{3}.0 = 0.$$

+ Vậy phân phối có điều kiện của X với điều kiện $Y = 1$ là:

x	0	1	2
$f(x \mid 1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

+ Từ đó: $P(X = 0 \mid Y = 1) = f(0 \mid 1) = \frac{1}{2}$.

Như vậy nếu biết một trong hai ruột bút được chọn có màu đỏ thì xác suất để chiếc ruột bút còn lại không có màu xanh lơ là $\frac{1}{2}$.

Ví dụ 10. Biến ngẫu nhiên hai chiều (X,Y), trong đó X là sự thay đổi nhiệt độ, Y là tỷ lệ thay đổi quang phổ mà một nguyên tử tạo ra, có hàm mật độ đồng thời như sau:

$$f(x,y) = \begin{cases} 10xy^2, \; 0 < x < y < 1 \\ \quad 0, \text{ tại các điểm khác} \end{cases}$$

a) Tìm hàm mật độ biên duyên $g(x)$, $h(y)$ và hàm mật độ có điều kiện $f(y|x)$.
b) Tìm xác suất để sự thay đổi quang phổ lớn hơn một nửa tổng số giá trị quan sát, biết nhiệt độ tăng 0,25 đơn vị.

Giải:

a) Theo định nghĩa ta có:

$$g(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_{x}^{1}10xy^2dy=\frac{10}{3}xy^3\Big|_{y=x}^{y=1}=\frac{10}{3}x(1-x^3),\ 0<x<1$$

$$h(y)=\int_{-\infty}^{\infty} f(x,y)dx=\int_{0}^{y}10xy^2dx=5x^2y^2\Big|_{x=0}^{x=y}=5y^4,\ 0<y<1$$

$$f(y\,|\,x)=\frac{f(x,y)}{g(x)}=\frac{10xy^2}{\frac{10}{3}x(1-x^3)}=\frac{3y^2}{1-x^3},\ 0<x<y<1.$$

b) Do đó

$$P(Y>\frac{1}{2}\,|\,X=0{,}25)=\int_{\frac{1}{2}}^{1} f(y\,|\,x=0{,}25)dy=\int_{\frac{1}{2}}^{1}\frac{3y^2}{1-0{,}25^3}dy=\frac{8}{9}.$$

Ví dụ 11

Cho hàm mật độ đồng thời

$$f(x,y)=\begin{cases}\dfrac{x(1+3y^2)}{4}, & 0<x<2,\ 0<y<1\\ 0 & ,\ (x,y)\notin(0,2)\times(0,1)\end{cases}$$

Tìm $g(x)$, $h(y)$, $f(x\,|\,y)$ và $P(\frac{1}{4}<X<\frac{1}{2}\,|\,Y=\frac{1}{3})$.

Giải:

Theo định nghĩa ta có:

$$g(x)=\int_{-\infty}^{\infty} f(x,y)dy=\int_{0}^{1}\frac{x(1+3y^2)}{4}dy=\left(\frac{xy}{4}+\frac{xy^3}{4}\right)\Bigg|_{y=0}^{y=1}=\frac{x}{2},\ 0<x<2$$

$$h(y)=\int_{-\infty}^{\infty} f(x,y)dx=\int_{0}^{2}\frac{x(1+3y^2)}{4}dx=\left(\frac{x^2}{8}+\frac{3x^2y^2}{8}\right)\Bigg|_{x=0}^{x=2}=\frac{1+3y^2}{2},\ 0<y<1.$$

Do đó:

$$f(x\,|\,y)=\frac{f(x,y)}{h(y)}=\frac{\frac{x(1+3y^2)}{4}}{\frac{(1+3y^2)}{2}}=\frac{x}{2},\ 0<x<2$$

và $$P\left(\frac{1}{4}<X<\frac{1}{2}\,|\,Y=\frac{1}{3}\right)=\int_{\frac{1}{4}}^{\frac{1}{2}}\frac{x}{2}dx=\frac{3}{64}.$$

V.ĐỘC LẬP THỐNG KÊ

Dễ thấy rằng: Nếu $f(x\,|\,y)$ không phụ thuộc vào y thì kết quả của biến ngẫu nhiên Y không ảnh hưởng đến kết quả của biến ngẫu nhiên X. Nói cách khác X và Y là các biến ngẫu nhiên độc lập.

Định nghĩa. Giả sử (X,Y) biến ngẫu nhiên hai chiều **rời rạc** hoặc ***liên tục,*** có phân phối xác suất đồng thời $f(x,y)$ và các phân phối biên duyên tương ứng $g(x)$, $h(y)$. Các biến ngẫu nhiên X và Y được gọi là **độc lập thống kê** khi và chỉ khi $f(x,y) = g(x).h(y)$ với $\forall\,(x,y)$ nằm trong miền giá trị của (X, Y).

Chú ý: Kiểm tra tính độc lập thống kê của các biến ngẫu nhiên rời rạc đòi hỏi sự cẩn thận hơn vì tích của các phân phối biên duyên có thể chỉ bằng phân phối xác suất đồng thời tại một vài điểm chứ không phải tất cả các điểm (x,y). Nếu chúng ta tìm được bất cứ điểm (x,y) nào nằm trong tập xác định của $f(x,y)$ mà $f(x,y) \neq g(x).h(y)$ thì ta kết luận ngay được các biến ngẫu nhiên rời rạc X và Y không độc lập thống kê.

Ví dụ 12. Chỉ ra rằng các biến ngẫu nhiên trong Ví dụ 7 không độc lập thống kê.

Giải: + Xét bảng sau:

Y / X	0	1	2	**Tổng hàng**
0	$3/28$	$3/14$	$1/28$	$5/14$
1	$9/28$	$3/14$		$15/28$
2	$3/28$			$3/28$
Tổng cột	$15/28$	$3/7$	$1/28$	1

+ Để ý điểm (0,1). Từ Bảng trên ta tìm được ba xác suất $f(0,1)$, $g(0)$ và $h(1)$ là:

$$f(0,1) = \frac{3}{14}$$

$$g(0) = \sum_{y=0}^{2} f(0,y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$$

$$h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$$

Rõ ràng: $f(0,1) \neq g(0)h(1)$. Do đó X và Y không độc lập thống kê.

Tổng quát: Giả sử $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên, rời rạc hoặc liên tục, có phân phối xác suất đồng thời $f(x_1, x_2, \ldots, x_n)$ và các phân phối biên duyên tương ứng là $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$. Các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ được gọi là **độc lập thống kê** với nhau khi và chỉ khi:

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\ldots f(x_n)$$

với mọi $(x_1, x_2, \ldots, x_n)$ nằm trong miền giá trị của $(X_1, X_2, \ldots, X_n)$.

VI. HÀM CỦA CÁC BIẾN NGẪU NHIÊN.

Vấn đề: Nếu ta đã biết phân phối xác suất của biến ngẫu nhiên một chiều (hay hai chiều) nào đó, một biến ngẫu nhiên một chiều (hoặc hai chiều) khác có mối liên hệ với nó. Ta có thể tìm được phân phối xác suất của các biến ngẫu nhiên sau hay không?

Định lý 1. Giả sử rằng X là biến ngẫu nhiên **rời rạc** với phân phối xác suất là $f(x)$. Giải sử $Y = f(X)$ xác định phép biến đổi một một giữa các giá trị của X và Y sao cho phương trình $y = u(x)$ có thể giải được **duy nhất nghiệm** x tính theo y, gọi là $x = w(y)$. Khi đó phân phối xác suất của Y được xác định bởi:

$$g(y) = f[w(y)].$$

Ví dụ 13. Cho X là biến ngẫu nhiên hình học với phân phối xác suất là

$$f(x) = \frac{3}{4}\left(\frac{1}{4}\right)^{x-1},\ x = 1,2,3,....$$

Tìm phân phối xác suất của biến $Y = X^2$.

Giải: Do biến ngẫu nhiên X chỉ nhận các giá trị dương, nên phép tương ứng một một giữa x và y là $y = x^2$ và $x = \sqrt{y}$. Do đó

$$g(y) = \begin{cases} f(\sqrt{y}) = \frac{3}{4}\left(\frac{1}{4}\right)^{\sqrt{y}-1}, & y = 1,4,9,... \\ 0, & \text{trái lại.} \end{cases}$$

Định lý 2. Giả sử X_1, X_2 là các biến ngẫu nhiên **rời rạc** với phân phối xác suất đồng thời là $f(x_1, x_2)$. Giả sử $Y_1 = u_1(X_1, X_2)$ và $Y_2 = u_2(X_1, X_2)$ xác định phép biến đổi một một giữa tập các điểm (x_1, x_2) và tập các điểm (y_1, y_2), vì thế hệ phương trình $y_1 = u_1(x_1, x_2)$ và $y_2 = u_2(x_1, x_2)$ sẽ có ***nghiệm duy nhất*** x_1 và x_2 tính qua y_1 và y_2, ta gọi chúng là $x_1 = w_1(y_1, y_2)$ và $x_2 = w_2(y_1, y_2)$. Khi đó phân phối xác suất đồng thời của Y_1, Y_2 được xác định bởi:

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)].$$

Định lý 3. Giả sử X là biến ngẫu nhiên **liên tục** với phân phối xác suất là $f(x)$. Giả sử $Y = u(X)$ xác định phép biến đổi một một giữa các giá trị của X và các giá trị của Y, vì thế phương trình $y = u(x)$ có **nghiệm duy nhất** $x = w(y)$. Khi đó phân phối xác suất của Y được xác định bởi:

$$g(y) = f[w(y)]|J|,$$

ở đó $J = w'(y)$ được gọi là **Jacobian** của phép đổi biến.

Định lý 4. Giả sử X_1, X_2 là các biến ngẫu nhiên **liên tục** với phân phối xác suất đồng thời là $f(x_1, x_2)$. Giả sử $Y_1 = u_1(X_1, X_2)$ và $Y_2 = u_2(X_1, X_2)$ là phép biến đổi một một giữa tập các điểm (x_1, x_2) và (y_1, y_2) vì thế ***có thể giải được duy nhất*** x_1, x_2 theo y_1, y_2, gọi chúng là $x_1 = w_1(y_1, y_2), x_2 = w_2(y_1, y_2)$. Khi đó phân phối xác suất đồng thời của Y_1 và Y_2 được xác định bởi:

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)]|J|,$$

ở đó Jacobi là định thức cấp 2×2

$$J = \begin{vmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 \end{vmatrix}$$

và $\partial x_1/\partial y_1$ là đạo hàm của $x_1 = w_1(y_1, y_2)$ theo biến y_1 với y_2 được giữ không đổi, hay theo giải tích thìđó chính là đạo hàm riêng của x_1 theo biến y_1. Các đạo hàm riêng khác được định nghĩa tương tự.

Ví dụ 14 Cho X_1, X_2 là các biến ngẫu nhiên liên tục có phân phối xác suất đồng thời là

$$f(x_1, x_2) = \begin{cases} 4x_1x_2, & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0, & \text{trái lại} \end{cases}$$

Tìm phân phối xác suất đồng thời của $Y_1 = X_1^2$ và $Y_2 = X_1X_2$.

Giải:

+ Các nghiệm ngược của $y_1 = x_1^2$ và $y_2 = x_1x_2$ là $\begin{cases} x_1 = \sqrt{y_1} \\ x_2 = y_2/\sqrt{y_1} \end{cases}$, từ đó chúng ta có được

$$J = \begin{vmatrix} 1/(2\sqrt{y_1}) & 0 \\ -y_2/2y_1^{3/2} & 1/\sqrt{y_1} \end{vmatrix} = \frac{1}{2y_1}.$$

+ Để xác định tập các điểm *B* trong mặt phẳng y_1y_2 là ảnh của tập các điểm *A* trong mặt phẳng x_1x_2, chúng ta viết

$$\begin{cases} x_1 = \sqrt{y_1} \\ x_2 = y_2/\sqrt{y_1} \end{cases}$$

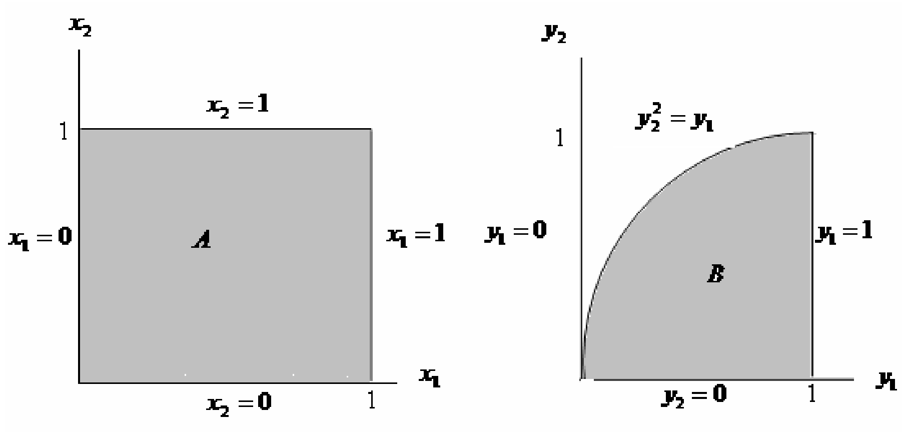
các đường $x_1 = 0, x_2 = 0, x_1 = 1, x_2 = 1$ bao tất cả các điểm *A* sẽ được chuyển thành các đường $y_1 = 0, y_2 = 0, y_1 = 1, y_2 = \sqrt{y_1}$ hay $y_1 = y_2^2$.

+ Hai miền này được minh họa trong Hình vẽ.

+ Rõ ràng phép biến đổi trên là một một, biến tập $A = \{(x_1, x_2) \mid 0 < x_1 < 1, 0 < x_2 < 1\}$ thành tập $B = \{(y_1, y_2) \mid y_2^2 < y_1 < 1, 0 < y_2 < 1\}$

+ Theo Định lý 4. phân phối xác suất đồng thời của Y_1, Y_2 là:

$$g(y_1,y_2)=4(\sqrt{y_1})\frac{y_2}{\sqrt{y_1}}\frac{1}{2y_1}=\begin{cases}\dfrac{2y_2}{y_1}, & {y_2}^2<y_1<1, 0<y_2<1\\ 0, & \text{trái lại.}\end{cases}.$$



Hình : Ánh xạ tập A vào tập B

Về nhà:

Bài tập: Tr. **94**

Đọc trước các Mục 4.1 và **4.2** chuẩn bị cho **Bài số :**

Kì vọng và phương sai