Chương 3

Biến ngẫu nhiên nhiều chiều

Thông qua công cụ giải tích, chương này giới thiệu về

1. Khái niệm và phân loại biến ngẫu nhiên nhiều chiều.
2. Phân phối xác suất (đồng thời, biên) của biến ngẫu nhiên hai chiều rời rạc và liên tục.
3. Phân phối có điều kiện.
4. Tính độc lập của các biến ngẫu nhiên.
5. Khái niệm và tính chất của kỳ vọng, phương sai, hiệp phương sai, hệ số tương quan.
6. Luật số lớn và định lý giới hạn trung tâm.

BÀI 9 (2 tiết)

3.1 Khái niệm và phân loại biến ngẫu nhiên nhiều chiều

3.1.1 Khái niệm

Trong nhiều bài toán thực tế ta thường phải xét đồng thời nhiều biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ có quan hệ với nhau.

Một biến ngẫu nhiên n-chiều (véc-tơ ngẫu nhiên n-chiều) là một bộ có thứ tự $(X_1, X_2, \ldots, X_n)$ với các thành phần $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên xác định trong cùng một phép thử.

Ký hiệu biến 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.

3.1.2 Phân loại

Biến ngẫu nhiên n-chiều $(X_1, X_2, \ldots, X_n)$ là liên tục hay rời rạc nếu tất cả các biến ngẫu nhiên thành phần $X_1, X_2, \ldots, X_n$ là liên tục hay rời rạc.

Để cho đơn giản, ta nghiên cứu biến ngẫu nhiên hai chiều (X, Y), trong đó X và Y là các biến ngẫu nhiên một chiều. Hầu hết các kết quả thu được đều có thể mở rộng cho trường hợp biến ngẫu nhiên n-chiều.

Trong chương này ta không xét trường hợp biến ngẫu nhiên hai chiều có một biến ngẫu nhiên rời rạc và một biến ngẫu nhiên liên tục.

Ví dụ 3.1. Một nhà máy sản xuất ra một sản phẩm.

1. Gọi X, Y, Z lần lượt là các biến ngẫu nhiên chỉ chiều dài, chiều rộng và chiều cao của sản phẩm (đơn vị tính là cm).

   (a) (X, Y, Z) là biến ngẫu nhiên ba chiều mô tả kích thước của sản phẩm.

   (b) Nếu chỉ quan tâm đến chiều dài và chiều rộng của sản phẩm ta được biến ngẫu nhiên hai chiều (X, Y).

2. Nếu quan tâm đến trọng lượng G và thể tích V của sản phẩm thì ta được biến ngẫu nhiên hai chiều (G, V).

3.2 Bảng phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

3.2.1 Bảng phân phối xác suất đồng thời

Định nghĩa 3.1 (Bảng phân phối xác suất đồng thời)**.** Bảng phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc (X, Y) là

$X \backslash Y$	y_1	$\dots$	y_j	$\dots$	y_n	$\sum_j$
x_1	p_{11}	$\dots$	p_{1j}	$\dots$	p_{1n}	$P(X = x_1)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	p_{i1}	$\dots$	p_{ij}	$\dots$	p_{in}	$P(X = x_i)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	$\dots$	p_{mj}	$\dots$	p_{mn}	$P(X = x_m)$
$\sum_i$	$P(Y = y_1)$	$\dots$	$P(Y = y_j)$	$\dots$	$P(Y = y_n)$	$\sum_i \sum_j = 1$

trong đó $x_i, i = 1, \dots, m, y_j, j = 1, \dots, n$ là các giá trị có thể có của các biến thành phần X và Y tương ứng; các xác suất p_{ij} được xác định bởi

$$\boxed{p_{ij} = P(X = x_i, Y = y_j), \quad i = 1, \dots, m;\ j = 1, \dots, n} \tag{3.1}$$

Bảng này còn được gọi là bảng phân phối xác suất đồng thời của hai biến ngẫu nhiên rời rạc X và Y. Bảng này có thể ra vô hạn nếu m, n nhận giá trị ∞.

Tính chất 3.1. (a) $0 \le p_{ij} \le 1$ với mọi $i = 1, \dots, m, j = 1, \dots, n$.

(b) $\sum_{i=1}^{m} \sum_{j=1}^{n} p_{ij} = 1$.

Định nghĩa 3.2 (Biến ngẫu nhiên độc lập)**.** Hai biến ngẫu nhiên rời rạc X và Y độc lập với nhau nếu

$$\boxed{P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j) \quad \forall i = 1, \dots, m;\ j = 1, \dots, n} \tag{3.2}$$

Ví dụ 3.2. Lấy ngẫu nhiên 3 sản phẩm từ một lô hàng gồm 5 sản phẩm loại I, 4 sản phẩm loại II và 3 sản phẩm loại III. Gọi X, Y lần lượt là số sản phẩm loại I và loại II trong 3 sản phẩm lấy ra. (a) Lập bảng phân phối xác suất đồng thời của hai biến ngẫu nhiên X và Y. (b) Tính $P(X = 0)$.

Lời giải Ví dụ 3.2 (a) X có thể nhận các giá trị $0, 1, 2, 3$, Y có thể nhận các giá trị $0, 1, 2, 3$.

$$P(X = 0, Y = 0) = \frac{C_3^3}{C_{12}^3} = \frac{1}{220}.$$

$P(X=0, Y=1) = \dfrac{C_5^0 C_4^1 C_3^2}{C_{12}^3} = \dfrac{12}{220}$.

Tương tự ta có thể tính được $P(X=i, Y=j)$ với mọi $i, j = 0, 1, 2, 3$.

Bảng phân phối xác suất cần tìm là

X \ Y	0	1	2	3
0	1/220	12/220	18/220	4/220
1	15/220	60/220	30/220	0
2	30/220	40/220	0	0
3	10/220	0	0	0

(b) $P(X=0) = \sum_{j=0}^{3} P(X=0, Y=j) = \dfrac{1}{220} + \dfrac{12}{220} + \dfrac{18}{220} + \dfrac{4}{220} = \dfrac{35}{220}$.

3.2.2 Bảng phân phối xác suất thành phần (biên)

Từ Định nghĩa 3.1 ta suy ra:

Định nghĩa 3.3. **(a)** Bảng phân phối xác suất của biến ngẫu nhiên thành phần X:

X	x_1	x_2	...	x_m
P	$P(X=x_1)$	$P(X=x_2)$	...	$P(X=x_m)$

trong đó $P(X=x_i) = \sum_{j=1}^{n} p_{ij}$, $i = 1, \ldots, m$.

(b) Bảng phân phối xác suất của biến ngẫu nhiên thành phần Y:

Y	y_1	y_2	...	y_n
P	$P(Y=y_1)$	$P(Y=y_2)$	...	$P(Y=y_n)$

trong đó $P(Y=y_j) = \sum_{i=1}^{m} p_{ij}$, $j = 1, \ldots, n$.

Nhận xét 3.1. Từ các bảng phân phối thành phần ta có thể dễ dàng xác định các tham số đặc trưng của các biến ngẫu nhiên thành phần X và Y.

Ví dụ 3.3. Trong Ví dụ 3.2, bảng phân phối xác suất đồng thời của hai biến ngẫu nhiên X và Y được bổ sung thêm cột $P(X=i)$, $i = 0, 1, 2, 3$ và hàng $P(Y=j)$, $j = 0, 1, 2, 3$ là

X \ Y	0	1	2	3	$P(X=i)$
0	1/220	12/220	18/220	4/220	35/220
1	15/220	60/220	30/220	0	105/220
2	30/220	40/220	0	0	70/220
3	10/220	0	0	0	10/220
$P(Y=j)$	56/220	112/220	48/220	4/220	$\sum_i \sum_j = 1$

(a) Bảng phân phối xác suất của X và của Y là

X	0	1	2	3
P	35/220	105/220	70/220	10/220

Y	0	1	2	3
P	56/220	112?220	48/220	4/220

(b) $E(X) = \dfrac{(0)(35) + (1)(105) + (2)(70) + (3)(10)}{220} = 1.25.$

$E(X^2) = \dfrac{(0^2)(35) + (1^2)(105) + (2^2)(70) + (3^2)(10)}{220} \simeq 2.1591.$

$V(X) = E(X^2) - [E(X)]^2 = 2.1591 - (1.25)^2 = 0.5966.$

$\sigma(X) = \sqrt{V(X)} = \sqrt{0.5966} \simeq 0.7724.$

Ví dụ 3.4. Từ kết quả phân tích các số liệu thống kê trong tháng về doanh số bán hàng (X) và chi phí cho quảng cáo (Y) (đơn vị triệu đồng) của một công ty, thu được bảng phân phối xác suất đồng thời như sau:

Y \ X	100	200	300
1	$0,15$	$0,1$	$0,14$
$1,5$	$0,05$	$0,2$	$0,15$
2	$0,01$	$0,05$	$0,15$

(a) Tính giá trị trung bình và phương sai của doanh số bán hàng. (b) Tính giá trị trung bình và phương sai của chi phí cho quảng cáo.

Lời giải Ví dụ 3.4 Lập bảng phân phối xác suất của X và Y:

X	100	200	300
P	0,21	0,35	0,44

Y	1	1,5	2
P	0,39	0,4	0,21

(a) Trung bình và phương sai của doanh số bán hàng là $E(X)$ và $V(X)$ được tính dựa trên Định nghĩa 2.8(a) và công thức (2.18)

$$E(X) = 100 \times 0,21 + 200 \times 0,35 + 300 \times 0,44 = 223$$
$$E(X^2) = 100^2 \times 0,21 + 200^2 \times 0,35 + 300^2 \times 0,44 = 55700$$
$$V(X) = E(X^2) - [E(X)]^2 = 55700 - (233)^2 = 1411.$$

(b) Tương tự, trung bình và phương sai của chi phí cho quảng cáo là:

$$E(Y) = 1 \times 0,39 + 1,5 \times 0,4 + 2 \times 0,21 = 1,41$$
$$E(Y^2) = 1^2 \times 0,39 + 1,5^2 \times 0,4 + 2^2 \times 0,21 = 2,13$$
$$V(Y) = E(Y^2) - [E(Y)]^2 = 2,13 - (1,41)^2 = 0,2419.$$

Ví dụ 3.5. Cho X_1, X_2, X_3 là các biến ngẫu nhiên độc lập theo luật phân phối Poa-xông với tham số $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$. Tính xác suất của các sự kiện (a) Số lớn nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1. (b) Số nhỏ nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1.

Lời giải Ví dụ 3.5

(a) $P(\max\{X_1, X_2, X_3\} \geq 1) = 1 - P(\max\{X_1, X_2, X_3\} < 1)$. Sử dụng tính độc lập của các biến ngẫu nhiên X_1, X_2, X_3 suy ra

$$\begin{aligned} P(\max\{X_1, X_2, X_3\} < 1) &= P(X_1 < 1, X_2 < 1, X_3 < 1) = P(X_1 < 1)P(X_2 < 1)P(X_3 < 1) \\ &= P(X_1 = 0)P(X_2 = 0)P(X_3 = 0) = \left(e^{-1}\frac{1^0}{0!}\right)\left(e^{-2}\frac{2^0}{0!}\right)\left(e^{-3}\frac{3^0}{0!}\right) \\ &\simeq 0.0025. \end{aligned}$$

Suy ra $P(\max\{X_1, X_2, X_3\} \geq 1) \simeq 0.9975$.

(b) $P(\min\{X_1, X_2, X_3\} \geq 1) = P(X_1 \geq 1, X_2 \geq 1, X_3 \geq 1)$. Sử dụng tính độc lập của các biến ngẫu nhiên X_1, X_2, X_3 ta được:

$$\begin{aligned} P(\min\{X_1, X_2, X_3\} \geq 1) &= P(X_1 \geq 1)\; P(X_2 \geq 1)\; P(X_3 \geq 1) \\ &= \big(1 - P(X_1 < 1)\big)\big(1 - P(X_2 < 1)\big)\big(1 - P(X_3 < 1)\big) \\ &= \left(1 - \frac{(e^{-1})\;(1^0)}{0!}\right) \cdot \left(1 - \frac{(e^{-2})\;(2^0)}{0!}\right) \cdot \left(1 - \frac{(e^{-3})\;(3^0)}{0!}\right) \\ &\simeq 0.5194 \end{aligned}$$

3.2.3 Phân phối có điều kiện

Từ Định nghĩa 3.1 ta suy ra:

Định nghĩa 3.4. (a) Bảng phân phối xác suất của X với điều kiện $(Y = y_j)$:

$X\|(Y = y_j)$	x_1	x_2	$\ldots$	x_m
P	$p(x_1\|y_j)$	$p(x_2\|y_j)$	$\ldots$	$p(x_m\|y_j)$

trong đó $p(x_i|y_j) = P[(X = x_i)|(Y = y_j)]$, $i = 1, \ldots, m$; $j = 1, \ldots, n$.

(b) Bảng phân phối xác suất của Y với điều kiện $(X = x_i)$:

$Y\|(X = x_i)$	y_1	y_2	$\ldots$	y_n
P	$p(y_1\|x_i)$	$p(y_2\|x_i)$	$\ldots$	$p(y_n\|x_i)$

trong đó $p(y_j|x_i) = P[(Y = y_j)|(X = x_i)]$, $i = 1, \ldots, m$; $j = 1, \ldots, n$.

Nhận xét 3.2. **(a)** Từ bảng phân phối xác suất có điều kiện ta có thể tính được kỳ vọng (có điều kiện) của từng biến ngẫu nhiên.

(b) Các xác suất có điều kiện được tính như thông thường, tức là

$$P\left(X = x_i | Y = y_j\right) = \frac{P(X = x_i,\ Y = y_j)}{P(Y = y_j)} \quad \text{hoặc}$$
$$P\left(X = x_i | Y \in D\right) = \frac{P(X = x_i,\ Y \in D)}{P(Y \in D)}$$

và

$$P\left(Y = y_j | X = x_i\right) = \frac{P(Y = y_j,\ X = x_i)}{P(X = x_i)} \quad \text{hoặc}$$
$$P\left(Y = y_j | X \in D\right) = \frac{P(Y = y_j,\ X \in D)}{P(X \in D)}.$$

Ví dụ 3.6. Với giả thiết của Ví dụ 3.4,

(a) Nếu chi phí cho quảng cáo 1,5 triệu đồng thì doanh số trung bình là bao nhiêu?

(b) Nếu muốn doanh số là 300 triệu đồng thì trung bình phải chi phí cho quảng cáo bao nhiêu?

Lời giải Ví dụ 3.6 Các bảng phân phối xác suất có điều kiện là:

$X\|(Y = 1,5)$	100	200	300
P	0,125	0,5	0,375

$Y\|(X = 300)$	1	1,5	2
P	$\frac{14}{44}$	$\frac{15}{44}$	$\frac{15}{44}$

(a) Nếu chỉ chi phí cho quảng cáo 1,5 triệu đồng thì doanh số trung bình là

$$E(X|Y = 1,5) = 100 \times 0,125 + 200 \times 0,5 + 300 \times 0,375 = 225 \quad \text{(triệu đồng)}.$$

(b) Nếu muốn doanh số là 300 triệu đồng thì trung bình phải chi phí cho quảng cáo là

$$E(Y|X = 300) = 1 \times \frac{14}{44} + 1,5 \times \frac{15}{44} + 2 \times \frac{15}{44} \simeq 1,5136 \quad \text{(triệu đồng)}.$$

3.3 Hàm phân phối xác suất

3.3.1 Hàm phân phối xác suất đồng thời

Định nghĩa 3.5 (Hàm phân phối đồng thời). Hàm hai biến $F_{XY}(x,y)$ xác định bởi:

$$\boxed{F_{XY}(x,y) = P(X < x, Y < y), \quad x, y \in \mathbb{R}} \tag{3.3}$$

được gọi là hàm phân phối xác suất của biến ngẫu nhiên hai chiều (X,Y) và còn được gọi là hàm phân phối xác suất đồng thời của các biến ngẫu nhiên X và Y.

Từ (3.3) và Định nghĩa 3.1, hàm phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc (X,Y) được xác định bởi

$$\boxed{F_{XY}(x,y) = \sum_{x_i<x}\sum_{y_j<y} P(X = x_i, Y = y_j)} \tag{3.4}$$

Ví dụ 3.7. Từ số liệu của Ví dụ 3.16, hãy tính $F_{X,Y}(2;3)$.

Lời giải Ví dụ 3.7 Sử dụng công thức (3.4),

$$F_{X,Y}(2,3) = \sum_{x_i<2}\sum_{y_j<3} P(X = x_i, Y = y_j) = p_{11} + p_{12} = 0,12 + 0,15 = 0,27.$$

Sau đây là một số tính chất của hàm phân phối xác suất của biến ngẫu nhiên hai chiều (X,Y).

Tính chất 3.2. **(a)** $0 \le F_{XY}(x,y) \le 1$.

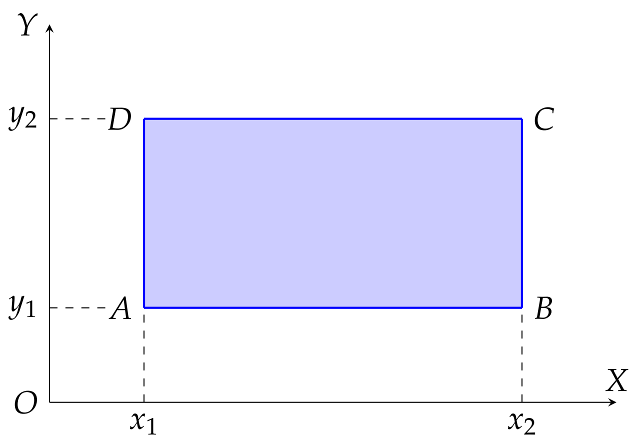
(b) Nếu $x < x_1, y < y_1$ thì $F_{XY}(x,y) \le F_{XY}(x_1,y_1)$.

(c) $F_{XY}(-\infty,y) = F_{XY}(x,-\infty) = 0$; $F_{XY}(+\infty,+\infty) = 1$ (giá trị ∞ hiểu theo nghĩa lấy giới hạn).

(d) Với $x_1 < x_2, y_1 < y_2$ ta luôn có

$$P(x_1 \le X < x_2, y_1 \le Y < y_2) = F_{XY}(x_2,y_2) - F_{XY}(x_2,y_1) - F_{XY}(x_1,y_2) + F_{XY}(x_1,y_1). \tag{3.5}$$

Nhận xét 3.3. Vế phải của (3.5) chính là xác xuất để điểm ngẫu nhiên (X,Y) rơi vào hình chữ nhật $ABCD$ trong Hình 3.1.



Hình 3.1: Miền $ABCD$ trong Nhận xét 3.3

3.3.2 Hàm phân phối xác suất thành phần (biên)

Các hàm

$$F_X(x) = P(X < x) = P(X < x, Y < +\infty) = \lim_{y\to+\infty} F_{XY}(x,y) = F_{XY}(x,+\infty),$$
$$F_Y(y) = P(Y < y) = P(X < +\infty, Y < y) = \lim_{x\to+\infty} F_{XY}(x,y) = F_{XY}(+\infty,y)$$

gọi là các phân phối biên của biến ngẫu nhiên hai chiều (X,Y). Đây cũng chính là các phân phối (một chiều) thông thường của X và Y tương ứng.

Định nghĩa 3.6 (Biến ngẫu nhiên độc lập)**.** Hai biến ngẫu nhiên X và Y độc lập nếu

$$\boxed{F_{XY}(x,y) = F_X(x)F_Y(y), \quad \forall x,y \in \mathbb{R}} \tag{3.6}$$

Nhận xét 3.4. Từ Định nghĩa 3.6, nếu X và Y độc lập ta có thể nghiên cứu từng biến ngẫu nhiên theo các phương pháp đã có ở Chương 2, đồng thời từ các phân phối biên của X và Y ta có thể xác định được hàm phân phối xác suất đồng thời $F_{X,Y}(x,y)$ của X và Y theo (3.6).

3.4 Hàm mật độ xác suất của biến ngẫu nhiên hai chiều liên tục

3.4.1 Hàm mật độ xác suất đồng thời

Định nghĩa 3.7 (Hàm mật độ đồng thời)**.** Giả sử hàm phân phối xác suất của biến ngẫu nhiên hai chiều liên tục (X,Y) là $F_{XY}(x,y)$. Khi đó, hàm mật độ xác suất của biến ngẫu nhiên hai chiều liên tục (X,Y) là hàm hai biến $f_{XY}(x,y) \geq 0$ thỏa mãn:

$$\boxed{F_{XY}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)dudv} \tag{3.7}$$

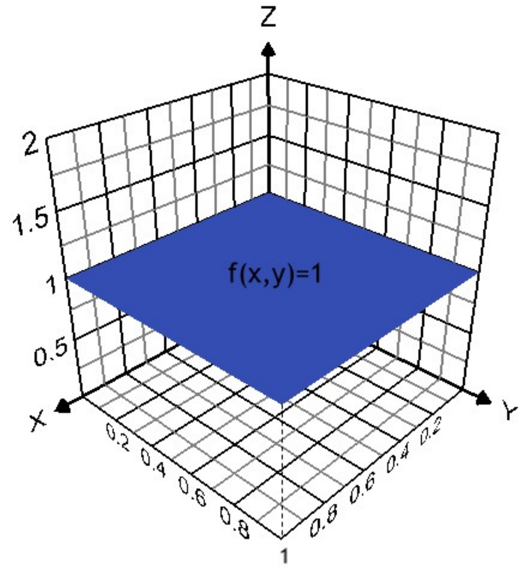
$f_{XY}(x,y)$ còn được gọi là hàm mật độ xác suất đồng thời của hai biến ngẫu nhiên liên tục X và Y.

Nhận xét 3.5. Về mặt hình học, hàm $f_{X,Y}(x,y)$ có thể xem như là một mặt cong trong $\mathbb{R}^3$ và được gọi là mặt phân phối xác suất.

Ví dụ 3.8. Hình 3.2 mô tả mặt phân phối xác suất của hàm $f_{X,Y}(x,y) = 1$ với $0 \leq x,y \leq 1$.

Hàm mật độ xác suất của biến ngẫu nhiên hai chiều (X,Y) có các tính chất sau.

Tính chất 3.3. **(a)** $f_{XY}(x,y) \geq 0$ với mọi $(x,y) \in \mathbb{R}^2$.



Hình 3.2: Mặt phân phối xác suất trong Ví dụ 3.8

(b) $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = 1.$

(c) $P\Big((X,Y) \in D\Big) = \int\int_{D\cap \mathbf{S}_{XY}} f_{XY}(x,y)dxdy$, với $D \subset \mathbb{R}^2$, $\mathbf{S}_{XY}$ là miền giá trị của biến ngẫu nhiên hai chiều (X,Y).

(d) Nếu $f_{XY}(x,y)$ liên tục theo cả hai biến thì $f_{XY}(x,y) = \dfrac{\partial^2 F_{XY}(x,y)}{\partial x \partial y}$.

Ví dụ 3.9. Biến ngẫu nhiên hai chiều liên tục (X,Y) có hàm mật độ xác suất là

$$f_{XY}(x,y) = \begin{cases} k, & 0 \le x \le 5,\ 0 \le y \le 3, \\ 0, & \text{trái lại.} \end{cases}$$

(a) Tìm hằng số k. (b) Tính $P(2 \le X < 3, 1 \le Y < 3)$.

Lời giải Ví dụ 3.9

(a) Sử dụng Tính chất 3.3(a),(b), ta có $k \ge 0$ và

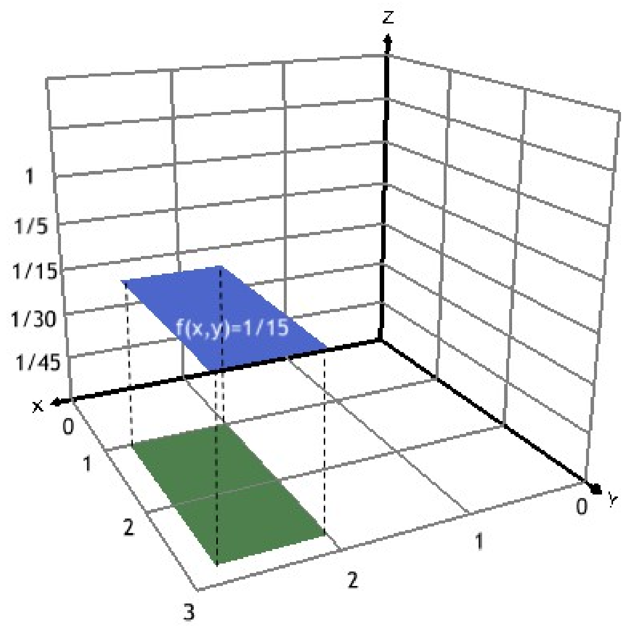
$$1 = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = \iint\limits_{D} kdxdy = \int\limits_{0}^{5} dx \int\limits_{0}^{3} kdy = 15k,$$

trong đó $D = \{(x,y) \in \mathbb{R}^2 : 0 \le x \le 5,\ 0 \le y \le 3\}$. Suy ra, $k = 1/15$.

(b) Sử dụng Tính chất 3.3(c),

$$P(2 \le X < 3, 1 \le Y < 3) = \int\limits_{2}^{3} dx \int\limits_{1}^{3} \frac{1}{15}dy = \frac{2}{15}.$$

Về mặt hình học đây là thể tích của hộp chữ nhật giới hạn bởi mặt phân phối $f_{X,Y}(x,y) = \dfrac{1}{15}$ có đáy là miền $D_1 = \{(x,y) \in \mathbb{R}^2 : 2 \le x < 3, 1 \le y < 3\}$, hình chiếu của mặt phân phối lên mặt Oxy (xem Hình 3.3).



Hình 3.3: Hình hộp chữ nhật trong Ví dụ 3.9(b)

3.4.2 Hàm mật độ xác suất biên

Định lý 3.1. Nếu (X,Y) là biến ngẫu nhiên hai chiều liên tục có hàm mật độ xác suất $f_{XY}(x,y)$ thì hàm mật độ xác suất của các biến ngẫu nhiên thành phần X và Y được xác định bởi

$$\boxed{f_X(x) = \int\limits_{-\infty}^{+\infty} f_{XY}(x,y)dy, \quad f_Y(y) = \int\limits_{-\infty}^{+\infty} f_{XY}(x,y)dx} \tag{3.8}$$

Chứng minh. Từ Tính chất 3.3(c) ta có thể viết

$$F_X(x) = P(X < x) = P(X < x, Y < \infty) = \int\limits_{-\infty}^{x} \left(\int\limits_{-\infty}^{+\infty} f_{XY}(u,y)dy \right) du.$$

Từ đây suy ra $f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$. Chứng minh tương tự cho $f_Y(y)$. □

Các hàm $f_X(x)$ và $f_Y(y)$ xác định trong công thức (3.8) còn gọi là các hàm mật độ xác suất biên của biến ngẫu nhiên hai chiều liên tục (X,Y).

Định nghĩa 3.8 (Biến ngẫu nhiên độc lập)**.** Giả sử biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất $f_{XY}(x, y)$. Khi đó X và Y độc lập nếu

$$\boxed{f_{XY}(x,y) = f_X(x)f_Y(y) \quad \forall x, y \in \mathbb{R}} \tag{3.9}$$

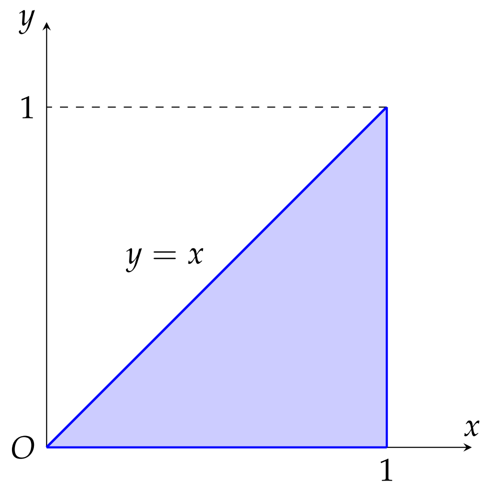
trong đó $f_X(x)$, $f_Y(y)$ lần lượt là hàm mật độ xác suất của các biến thành phần X và Y.

Ví dụ 3.10. Cho biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất là

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

(a) Tìm hằng số k. (b) Tìm các hàm mật độ xác suất của X và Y. (c) X và Y có độc lập không?

Lời giải Ví dụ 3.10 Ký hiệu $D = \left\{(x,y) \in \mathbb{R}^2 : 0 < x < 1; 0 < y < x\right\}$ (xem Hình 3.4).



Hình 3.4: Miền D của Ví dụ 3.10

(a) Theo Tính chất 3.3(a), $k \geq 0$ và theo Tính chất 3.3(b),

$$1 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = \iint_D kxdxdy = \int_0^1 dx \int_0^x kxdy = k\int_0^1 x^2dx = \frac{k}{3}.$$

Suy ra $k = 3$ và hàm mật độ xác suất $f_{X,Y}(x,y) = \begin{cases} 3x, & \text{nếu } (x,y) \in D, \\ 0, & \text{nếu trái lại.} \end{cases}$

(b) Tìm các hàm mật độ biên từ (3.8),

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \begin{cases} \int_0^x 3xdy, & 0 < x < 1, \\ 0, & \text{trái lại}, \end{cases} = \begin{cases} 3x^2, & 0 < x < 1, \\ 0, & \text{trái lại}, \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx = \begin{cases} \int_y^1 3xdx, & 0 < y < 1, \\ 0, & \text{trái lại}, \end{cases} = \begin{cases} \dfrac{3}{2} - \dfrac{3}{2}y^2, & 0 < y < 1, \\ 0, & \text{trái lại}. \end{cases}$$

(c) Vì $f_{X,Y}(x,y) \neq f_X(x) \times f_Y(y)$ với $(x,y) \in D$ nên X, Y không độc lập.

3.4.3 Hàm mật độ xác suất có điều kiện

Định nghĩa 3.9 (Hàm mật độ có điều kiện)**.** Hàm mật độ có điều kiện của thành phần X biết $Y = y$ là:

$$\boxed{f_X(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} = \frac{f_{XY}(x,y)}{\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx}} \tag{3.10}$$

Hàm mật độ có điều kiện của thành phần Y biết $X = x$ là:

$$\boxed{f_Y(y|x) = \frac{f_{XY}(x,y)}{f_X(x)} = \frac{f_{XY}(x,y)}{\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy}} \tag{3.11}$$

Chú ý 3.1. Các hàm mật độ có điều kiện cũng thỏa mãn các tính chất của một hàm mật độ thông thường.

Nhận xét 3.6. (a) Nếu $f_{X,Y}(x,y) \neq f_X(x)f_Y(y)$ thì X và Y không độc lập. Trong trường hợp này, nhờ (3.10) và (3.11) ta có thể xác định được hàm mật độ đồng thời $f_{X,Y}(x,y)$ bởi

$$\boxed{f_{X,Y}(x,y) = f_Y(y)f_X(x|y) = f_X(x)f_Y(y|x)} \tag{3.12}$$

(b) Nếu $f_X(x|y) = f_X(x)$ hoặc $f_Y(y|x) = f_Y(y)$ với mọi x, y thì ta lại có điều kiện độc lập (3.9).

(c) Từ (3.10) và (3.11) và (2.7) ta có các phân phối có điều kiện sau đây

$$\boxed{F_X(x|y) = \frac{\int_{-\infty}^{x} f_{X,Y}(u,y)du}{f_Y(y)}, \quad F_Y(y|x) = \frac{\int_{-\infty}^{y} f_{X,Y}(x,v)dv}{f_X(x)}} \tag{3.13}$$

Ngoài ra

$$\boxed{F_X(x|y_1 \le Y \le y_2) = P(X < x|y_1 \le Y \le y_2) = \frac{\int_{-\infty}^{x} du \int_{y_1}^{y_2} f_{X,Y}(u,v)dv}{\int_{-\infty}^{+\infty} dx \int_{y_1}^{y_2} f_{X,Y}(x,y)dy}} \tag{3.14}$$

(d) Từ (3.14) ta nhận được hàm mật độ có điều kiện của X với điều kiện $y_1 \le Y \le y_2$:

$$\boxed{f_X(x|y_1 \le Y \le y_2) = \frac{\int_{y_1}^{y_2} f_{X,Y}(x,v)dv}{\int_{-\infty}^{+\infty} dx \int_{y_1}^{y_2} f_{X,Y}(x,y)dy}} \tag{3.15}$$

Các công thức (3.10), (3.11), (3.13)–(3.15) cần có điều kiện biểu thức ở mẫu số phải khác không.

Ví dụ 3.11. Cho biến ngẫu nhiên hai chiều (X,Y) có hàm mật độ xác suất là

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

(a) Tìm các hàm mật độ xác suất của X và Y. (b) Tìm các hàm mật độ xác suất có điều kiện $f_X(x|y)$ và $f_Y(y|x)$.

Lời giải Ví dụ 3.11

(a) Các hàm mật độ xác suất biên là:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \begin{cases} \int_0^x \dfrac{1}{x}dy, & 0 < x < 1,\\ 0, & \text{trái lại,} \end{cases} = \begin{cases} 1, & 0 < x < 1,\\ 0, & \text{trái lại,} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx = \begin{cases} \int_y^1 \dfrac{1}{x}dx, & 0 < y < 1,\\ 0, & \text{trái lại,} \end{cases} = \begin{cases} -\ln y, & 0 < y < 1,\\ 0, & \text{trái lại.} \end{cases}$$

(b) Các hàm mật độ xác suất có điều kiện là:

$$f_X(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} -\dfrac{1}{x\ln y}, & 0 < y < x < 1,\\ 0, & \text{trái lại.} \end{cases}$$

$$f_Y(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \dfrac{1}{x}, & 0 < y < x < 1,\\ 0, & \text{trái lại.} \end{cases}$$

Ví dụ 3.12. Giả sử tại một trường đại học, một sinh viên đạt được điểm X trong bài kiểm tra năng khiếu toán học và điểm Y trong bài kiểm tra năng khiếu âm nhạc là một số trong khoảng từ 0 đến 1. Giả sử X và Y được phân phối theo hàm mật độ sau:

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{2}{5}(2x+3y), & 0 < x,y < 1,\\ 0, & \text{nếu ngược lại.} \end{cases}$$

(a) Tính tỷ lệ sinh viên đại học đạt điểm cao hơn 0,8 trong bài kiểm tra năng khiếu toán. (b) Giả sử điểm số của một sinh viên trong bài kiểm tra năng khiếu âm nhạc là 0,3. Tính xác suất để điểm của anh ấy trong bài kiểm tra năng khiếu toán học sẽ lớn hơn 0,8.

Lời giải Ví dụ 3.12

(a) $P(X > 0,8) = \int\limits_0^1 \int\limits_{0.8}^1 \frac{2}{5}(2x+3y)dxdy = \frac{2}{5}\int\limits_0^1 (0,36+0,6y)dy = 0,264.$

(b) Ta cần tính $P(X > 0,8|Y = 0,3) = \int\limits_{0,8}^1 f_X(x|y = 0,3)dx$, trong đó

$$f_X(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = \begin{cases} \dfrac{\frac{2}{5}(2x+3y)}{\frac{2}{5}\int_0^1 (2x+3y)dx}, & \text{nếu } 0 < x, y < 1, \\ 0, & \text{nếu trái lại} \end{cases}$$

$$= \begin{cases} \dfrac{2x+3y}{1+3y}, & \text{nếu} \quad 0 < x, y < 1, \\ 0, & \text{nếu trái lại,} \end{cases}$$

suy ra $f_X(x|y = 0,3) = \begin{cases} \dfrac{2x+0,9}{1,9}, & \text{nếu} \quad 0 < x < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$

Vậy,

$$P(X > 0,8|Y = 0,3) = \int\limits_{0,8}^1 \frac{2x+0,9}{1,9}dx = \frac{0,36+0,18}{1,9} = 0,2842.$$

3.5 Tính độc lập của các biến ngẫu nhiên

Một số dấu hiệu nhận biết tính độc lập của hai biến ngẫu nhiên X và Y dựa trên tính chất của bảng phân phối xác suất đồng thời, hàm phân phối xác suất đồng thời, hàm mật độ xác suất đồng thời như trong Định nghĩa 3.2, 3.6 và 3.8 với các công thức (3.2), (3.6) và (3.9) tương ứng.

BÀI 10 (2 tiết)

3.6 Hàm của hai biến ngẫu nhiên

3.6.1 Khái niệm

Cho hai biến ngẫu nhiên X và Y và định nghĩa $Z = g(X,Y)$ là một phép cho tương ứng mỗi cặp giá trị (x,y) của (X,Y) với duy nhất một giá trị $z = g(z,y)$ của Z. Z được gọi là hàm của hai biến ngẫu nhiên X và Y. Chẳng hạn $Z = X + Y$, $Z = XY$.

3.6.2 Phân phối xác suất

Xét biến ngẫu nhiên $Z = g(X,Y)$, trong đó (X,Y) là biến ngẫu nhiên hai chiều đã biết luật phân phối. Ta sẽ xét phân phối xác suất của Z trong một số trường hợp đơn giản theo cách sau.

Định lý 3.2.

$$\boxed{F_Z(z) = P(Z < z) = P\left(g(X,Y) < z\right) = P\left((X,Y) \in D\right)} \tag{3.16}$$

trong đó $D = \left\{(x,y) \in \mathbb{R}^2 : g(x,y) < z\right\}$.

Để tìm hàm mật độ xác suất $f_Z(z)$ của biến ngẫu nhiên Z, ta sử dụng tính chất

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

Ví dụ 3.13. Tính xác suất để hai người gặp được nhau trong Ví dụ 1.17.

Lời giải Ví dụ 3.13 Quy gốc thời gian về lúc 7h00. Gọi X và Y lần lượt là biến ngẫu nhiên chỉ thời điểm đến của hai người trong khoảng thời gian từ phút thứ 0 đến phút thứ 60. X và Y là các biến ngẫu nhiên liên tục có phân phối đều trên $[0;60]$. Do X và Y độc lập nên chúng có hàm mật độ xác suất đồng thời

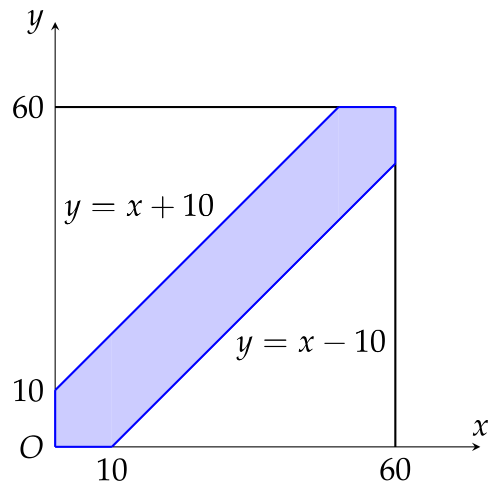
$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \begin{cases} \dfrac{1}{3600}, & (x,y) \in [0,60]^2, \\ 0, & \text{ngược lại.} \end{cases}$$

Đặt $Z = |X - Y|$. Khi đó, xác suất hai người gặp được nhau là

$$P(Z < 10) = P\left(|X - Y| < 10\right) = P\left((X,Y) \in D\right),$$

trong đó D là giao của miền $|x - y| < 10$ và hình vuông $[0;60]^2$ (xem Hình 3.5). Suy ra

$$P(Z < 10) = 1 - 2\int_{10}^{60} dx \int_{0}^{x-10} \frac{1}{3600} dy = 1 - \frac{2500}{3600} = \frac{11}{36}.$$



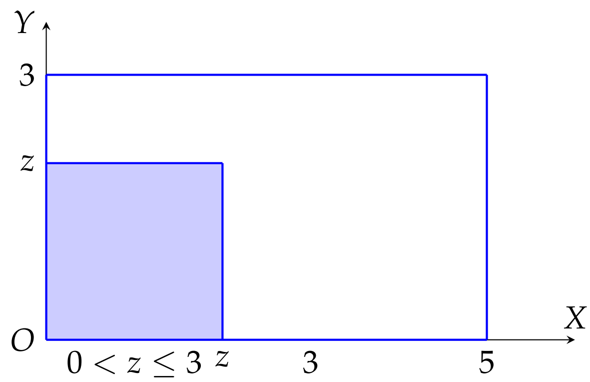
Hình 3.5: Miền D của Ví dụ 3.13

Ví dụ 3.14. Cho X và Y là hai biến ngẫu nhiên trong Ví dụ 3.9. Hãy tìm hàm mật độ của biến ngẫu nhiên $Z = \max(X, Y)$.

Lời giải Ví dụ 3.14 Trước hết ta tìm

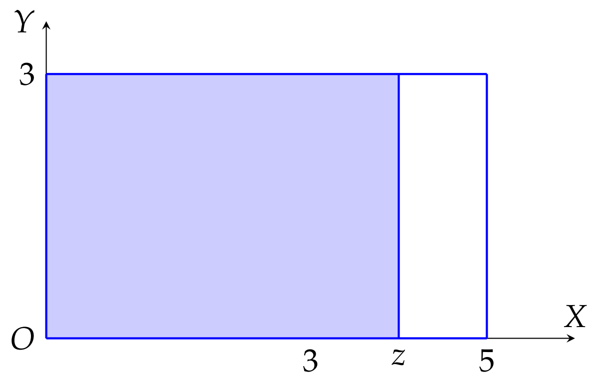
$$F_Z(z) = P(Z < z) = P(\max(X,Y) < z) = P(X < z, Y < z) = \iint\limits_D \frac{1}{15} dxdy,$$

trong đó $D = \{(x < z, y < z)\} \cap \{(0 \le x \le 3, 0 \le y \le 5)\}$.



Hình 3.6: Miền D với $0 < z \le 3$ của Ví dụ 3.14

Nếu $z \le 0$, $F_Z(z) = 0$.
Nếu $0 < z \le 3$, $F_Z(z) = \int_0^z dx \int_0^z \frac{1}{15} dy = \frac{z^2}{15}$.
Nếu $3 < z \le 5$, $F_Z(z) = \int_0^z dx \int_0^3 \frac{1}{15} dy = \frac{z}{5}$.
Nếu $z > 5$, $F_Z(z) = 1$.



Hình 3.7: Miền D với $3 < z \leq 5$ của Ví dụ 3.14

Vậy $F_z(z) = \begin{cases} 0 & \text{nếu } z \leq 0, \\ \dfrac{z^2}{15} & \text{nếu } 0 < z \leq 3, \\ \dfrac{z}{5} & \text{nếu } 3 < z \leq 5, \\ 1 & \text{nếu } z > 5, \end{cases}$ suy ra $f_z(z) = \begin{cases} \dfrac{2z}{15} & \text{nếu } 0 < z \leq 3, \\ \dfrac{1}{5} & \text{nếu } 3 < z \leq 5, \\ 0 & \text{nếu } z \leq 0,\ z > 5. \end{cases}$

Ví dụ 3.15. Cho X, Y là 2 biến ngẫu nhiên độc lập với nhau có cùng phân phối đều trên $[0, 2]$.

(a) Tìm hàm phân phối của các biến ngẫu nhiên $Z = X + Y$, $T = XY$, $U = X - Y$.

(b) Tính $P(-1 \leq Y - X \leq 1)$

Lời giải Ví dụ 3.15 Vì X và Y độc lập nên ta có hàm mật độ xác suất của (X, Y) là

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{1}{4}, & (x, y) \in \mathcal{D}, \\ 0, & \text{trái lại}, \end{cases}$$

trong đó $D := \{0 \leq x \leq 2;\ 0 \leq y \leq 2\}$.

(a1) Hàm phân phối xác suất của $Z = X + Y$ được xác định bởi

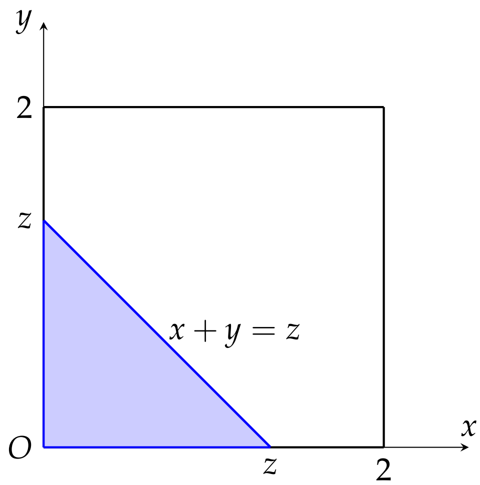
$$F_Z(z) = P(X + Y < z) = \iint\limits_{\{x+y<z\}} f_{X,Y}(x, y)dxdy = \frac{1}{4} \iint\limits_{\{x+y<z\} \cap D} dxdy.$$

Nếu $z \leq 0$ thì $F_Z(z) = 0$.

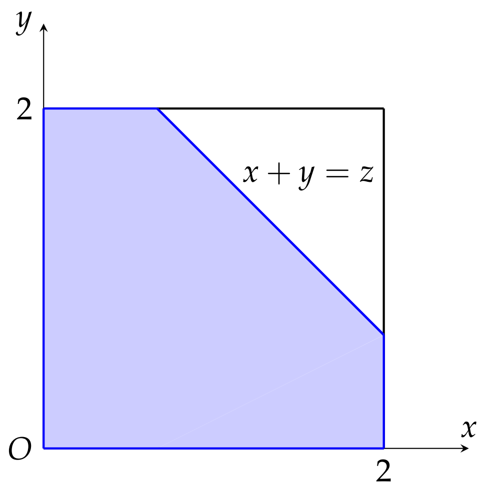
Nếu $0 < z \leq 2$ thì $F_Z(z) = \dfrac{1}{4} \displaystyle\int_0^z \left(\int_0^{z-x} dy \right) dx = \frac{1}{4} \times \frac{z^2}{2} = \frac{z^2}{8}$.

Nếu $2 < z \leq 4$ thì

$$F_Z(z) = 1 - \frac{1}{4} \int\limits_{z-2}^{2} \left(\int\limits_{z-x}^{2} dy \right) dx = 1 - \frac{1}{4} \int\limits_{z-2}^{2} (2 - z + x) dx = 1 - \frac{1}{8} \left(z^2 - 8z + 16 \right)$$
$$= \frac{1}{8}(-z^2 + 8z - 8).$$



Hình 3.8: Miền $\{x+y<z\}\cap D$ với $0<z\le 2$ trong Ví dụ 3.15(a1)



Hình 3.9: Miền $\{x+y<z\}\cap D$ với $2<z\le 4$ trong Ví dụ 3.15(a1)

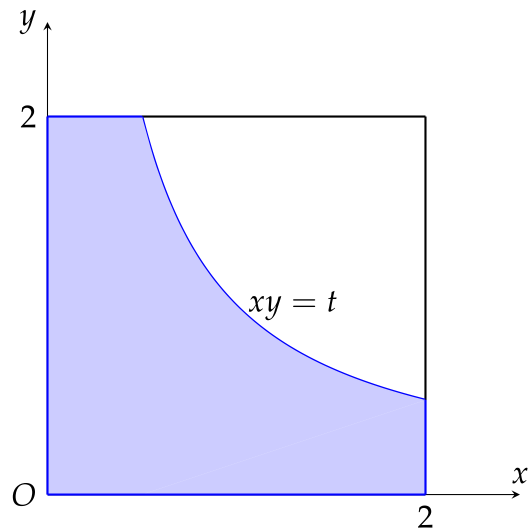
Nếu $z>4$ thì $F_Z(z)=\dfrac{1}{4}\displaystyle\iint_{\mathcal{D}} dxdy = 1$. Vậy

$$F_Z(z)=\begin{cases} 0, & \text{nếu } z\le 0,\\ \dfrac{z^2}{8}, & \text{nếu } 0<z\le 2,\\ \dfrac{1}{8}(-z^2+8z-8), & \text{nếu } 2<z\le 4,\\ 1, & \text{nếu } z>4. \end{cases}$$

(a2) Hàm phân phối xác suất của $T=XY$ được xác định như sau:

$$F_T(t)=P\left(XY<t\right)=\iint\limits_{\{xy<t\}} f_{X,Y}(x,y)dxdy=\frac{1}{4}\iint\limits_{\{xy<t\}\cap\mathcal{D}} dxdy.$$

Nếu $t\le 0$, $F_T(t)=0$.



Hình 3.10: Miền $\{xy < t\} \cap D$ với $0 < t \leq 4$ trong Ví dụ 3.15(a2)

Nếu $0 < t \leq 4$, $F_T(t) = \frac{1}{4}\left(\int_0^{t/2} dx \int_0^2 dy + \int_{t/2}^2 dx \int_0^{t/x} dy\right) = \frac{1}{4}\left(t + t\ln 2 - t\ln\frac{t}{2}\right)$.

Nếu $t > 4$, $F_T(t) = 1$. Vậy

$$F_T(t) = \begin{cases} 0, & \text{nếu } t \leq 0,\\ \frac{1}{4}\left(t + t\ln 2 - t\ln\frac{t}{2}\right), & \text{nếu } 0 < t \leq 4,\\ 1, & \text{nếu } t > 4. \end{cases}$$

(a3) Hàm phân phối xác suất của $U = X - Y$ là

$$F_U(u) = P(X - Y < u) = \iint\limits_{\{x-y<u\}} f_{X,Y}(x,y)dxdy = \frac{1}{4}\iint\limits_{\{x-y<u\}\cap\mathcal{D}} dxdy.$$

Nếu $u \leq -2$, $F_U(u) = 0$.

Nếu $-2 < u \leq 0$, $F_U(u) = \frac{1}{4}\int_0^{u+2} dx \int_{x-u}^2 dy = \frac{1}{4} \times \frac{1}{2}(2+u)^2 = \frac{1}{8}(2+u)^2$.

Nếu $0 < u \leq 2$ thì $F_U(u) = \frac{1}{4}\left[4 - \frac{1}{2}(2-u)^2\right] = \frac{1}{4}\left(-\frac{u^2}{2} + 2u + 2\right)$.

Nếu $u > 2$, $F_U(u) = 1$. Vậy

$$F_U(u) = \begin{cases} 0, & \text{nếu } u \leq -2,\\ \frac{1}{8}(2+u)^2, & \text{nếu } -2 < u \leq 0,\\ \frac{1}{8}(-u^2 + 4u + 4), & \text{nếu } 0 < u \leq 2,\\ 1, & \text{nếu } u > 2. \end{cases}$$

(b) $P(-1 \leq Y - X \leq 1) = P(X - 1 \leq Y \leq X + 1) = \dfrac{1}{4}\left(4 - 2 \times \dfrac{1}{2}\right) = \dfrac{3}{4}.$

Định lý 3.3. Nếu X_1 và X_2 là hai biến ngẫu nhiên độc lập có phân phối Poa-xông với tham số λ_1 và λ_2 tương ứng thì $X_1 + X_2$ cũng là biến ngẫu nhiên có phân phối Poa-xông với tham số $\lambda_1 + \lambda_2$.

Chứng minh. Đặt $Y_1 = X_1 + X_2$, $Y_2 = X_2$, khi đó $X_1 = Y_1 - Y_2$ và $X_2 = Y_2$. Vì X_1 và X_2 độc lập nên với $k = 0, 1, 2, \dots$, $l = 0, 1, 2, \dots, k$, $l \leq k$ (vì $X_1 \geq 0$),

$$\begin{aligned} P(Y_1 = k, Y_2 = l) &= P(X_1 = k - l, X_2 = l) = P(X_1 = k - l)P(X_2 = l) \\ &= \left(\frac{\lambda_1^{k-l}}{(k-l)!}e^{-\lambda_1}\right)\left(\frac{\lambda_2^l}{l!}e^{-\lambda_2}\right) = e^{-(\lambda_1+\lambda_2)}\frac{\lambda_1^{k-l}\lambda_2^l}{(k-l)!l!}. \end{aligned}$$

Từ đây suy ra

$$\begin{aligned} P(Y_1 = k) &= \sum_{l=0}^{k} P(Y_1 = k, Y_2 = l) = e^{-(\lambda_1+\lambda_2)}\sum_{l=0}^{k}\frac{\lambda_1^{k-l}\lambda_2^l}{(k-l)!l!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{k!}\sum_{l=0}^{k}\frac{k!}{(k-l)!l!}\lambda_1^{k-l}\lambda_2^l = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}, \quad k = 0, 1, 2, \dots. \end{aligned}$$

Vậy $X_1 + X_2$ là biến ngẫu nhiên có phân phối Poa-xông với tham số $\lambda_1 + \lambda_2$. □

Định lý 3.4. If $X_1, X_2, \dots, X_n$ là các biến ngẫu nhiên độc lập có phân phối chuẩn với kỳ vọng $\mu_1, \mu_2, \dots, \mu_n$ và phương sai $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ tương ứng, thì biến ngẫu nhiên

$$Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

có phân phối chuẩn với kỳ vọng

$$\mu_Y = a_1\mu_1 + a_2\mu_2 + \cdots + a_n\mu_n$$

và phương sai

$$\sigma_Y^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots + a_n^2\sigma_n^2.$$

3.7 Đặc trưng của biến ngẫu nhiên hai chiều

3.7.1 Kỳ vọng, phương sai của biến ngẫu nhiên thành phần

Định lý 3.5. (a) Nếu (X, Y) là biến ngẫu nhiên hai chiều rời rạc có bảng phân phối xác suất như trong Định nghĩa 3.1 thì

$$E(X) = \sum_i x_iP(X = x_i) = \sum_i\sum_j x_ip_{ij}; \quad E(Y) = \sum_j y_jP(Y = y_j) = \sum_j\sum_i y_jp_{ij} \tag{3.17}$$

$$V(X) = \sum_i\sum_j x_i^2p_{ij} - \big(E(X)\big)^2; \quad V(Y) = \sum_j\sum_i y_j^2p_{ij} - \big(E(Y)\big)^2. \tag{3.18}$$

(b) Nếu (X,Y) là biến ngẫu nhiên hai chiều liên tục có hàm mật độ xác suất $f_{XY}(x,y)$ thì

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_{XY}(x,y)dxdy; \qquad E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y f_{XY}(x,y)dxdy \tag{3.19}$$

$$V(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 f_{XY}(x,y)dxdy-\big(E(X)\big)^2; \quad V(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 f_{XY}(x,y)dxdy-\big(E(Y)\big)^2. \tag{3.20}$$

3.7.2 Kỳ vọng, phương sai của hàm của hai biến ngẫu nhiên

Định nghĩa 3.10. Cho biến ngẫu nhiên hai chiều (X,Y) có phân phối đã biết và ta xác định một biến mới $Z=g(X,Y)$ (g là hàm đo được).

$$E(Z)=E\left[g(X,Y)\right]=\sum_i\sum_j g(x_i,y_j)P(X=x_i,Y=y_j) \quad \text{(nếu } (X,Y) \text{ rời rạc)} \tag{3.21}$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f_{XY}(x,y)dxdy \quad \text{(nếu } (X,Y) \text{ liên tục).} \tag{3.22}$$

Đặc biệt, khi $g(X,Y)=X$ thay vào các công thức trên ta sẽ có $E(X)$.

Ví dụ 3.16. Cho biến ngẫu nhiên hai chiều rời rạc (X,Y) có bảng phân phối xác 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 và Y độc lập. (b) Tìm quy luật phân phối xác suất của biến ngẫu nhiên $Z=XY$. (c) Tính $E(Z)$ bằng hai cách và kiểm tra $E(Z)=E(X)E(Y)$.

Lời giải Ví dụ 3.16

(a) Bảng phân phối xác suất của X và Y là:

X	1	2
P	0,3	0,7

Y	1	2	3
P	0,4	0,5	0,1

Suy ra

$$P(X=x_i,Y=y_j)=P(X=x_i)P(Y=y_j), \quad \forall i=1,2;\ j=1,2,3$$

nên X,Y độc lập.

(b) Z nhận các giá trị $1, 2, 3, 4, 5, 6$ với

$$P(Z = 1) = P(X = 1, Y = 1) = 0,12$$
$$P(X = 2) = P(X = 1, Y = 2) + P(X = 2, Y = 1) = 0,15 + 0,28 = 0,43$$
$$P(X = 3) = P(X = 1, Y = 3) = 0,03$$
$$P(X = 4) = P(X = 2, Y = 2) = 0,35$$
$$P(Z = 6) = P(X = 2, Y = 3) = 0,07$$

Quy luật phân phối xác suất của $Z = XY$ là:

Z	1	2	3	4	6
P	0,12	0,43	0,03	0,35	0,07

(c) Cách 1: sử dụng Định nghĩa 2.8(a),

$$E(Z) = 1 \times 0,12 + 2 \times 0,43 + 3 \times 0,03 + 4 \times 0,35 + 6 \times 0,07 = 2,89.$$

Cách 2: sử dụng Định nghĩa 3.2(a),

$$\begin{aligned} E(Z) = E(XY) &= \sum_{i=1}^{2}\sum_{j=1}^{3} x_i y_j P(X = x_i, Y = y_j) \\ &= (1)(1)(0,12) + (1)(2)(0,15) + (1)(3)(0,03) + (2)(1)(0,28) \\ &\quad + (2)(2)(0,35) + (2)(3)(0,07) = 2,89. \end{aligned}$$

Ví dụ 3.17. Cho biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất

$$f_{X,Y}(x, y) = \begin{cases} \dfrac{x(1+3y^2)}{4}, & \text{nếu} \quad 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{trong trường hợp trái lại.} \end{cases}$$

Tính $E(Y/X)$.

Lời giải Ví dụ 3.17 Sử dụng công thức (3.22),

$$E\left(\frac{Y}{X}\right) = \int_0^1 \int_0^2 \frac{y(1+3y^2)}{4} dxdy = \int_0^1 \frac{y+3y^3}{2} dy = \frac{5}{8}.$$

Sau đây là một mở rộng của Định lý 2.6.

Định lý 3.6. Cho $h(X, Y)$, $g(X, Y)$ là các hàm của biến ngẫu nhiên hai chiều (X, Y). Khi đó

$$E[g(X, Y) \pm h(X, Y)] = E[g(X, Y)] \pm E[h(X, Y)].$$

Hệ quả 3.1. (a) Nếu $g(X, Y) = g(X)$ và $h(X, Y) = h(Y)$ thì

$$E[g(X) \pm h(Y)] = E[g(X)] \pm E[h(Y)].$$

(b) Nếu $g(X,Y) = X$ và $h(X,Y) = Y$ thì

$$E(X \pm Y) = E(X) \pm E(Y).$$

Định lý 3.7. Nếu X và Y là hai biến ngẫu nhiên độc lập thì

$$E(XY) = E(X)E(Y).$$

Chứng minh. Giả sử X và Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất lần lượt là $f_X(x)$ và $f_Y(y)$. Vì X và Y độc lập nên $f_{X,Y}(x,y) = f_X(x)f_Y(y)$, ở đây $f_{X,Y}(x,y)$ là hàm mật độ xác suất đồng thời của hai biến ngẫu nhiên X và Y. Theo Định nghĩa 3.10,

$$\begin{aligned} E(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy f_{X,Y}(x,y)dxdy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xy f_X(x)f_Y(y)dxdy \\ &= \int_{-\infty}^{+\infty} x f_X(x)dx \int_{-\infty}^{+\infty} y f_Y(y)dy = E(X)E(Y). \end{aligned}$$

□

3.7.3 Hiệp phương sai

Nếu $g(X,Y) = (X - E(X))(Y - E(Y))$ thì Định nghĩa 3.10 cho ta một giá trị kỳ vọng, gọi là hiệp phương sai của hai biến ngẫu nhiên X và Y.

Định nghĩa 3.11 (Hiệp phương sai). Cho biến ngẫu nhiên hai chiều (X,Y), hiệp phương sai của hai thành phần X và Y, ký hiệu là $cov(X,Y)$ được xác định bởi

$$\boxed{cov(X,Y) = E\left[(X - E(X))(Y - E(Y))\right]} \tag{3.23}$$

Một công thức khác để tính hiệp phương sai, tương đương với công thức (3.23) được nêu trong định lý sau đây.

Định lý 3.8.

$$\boxed{cov(X,Y) = E(XY) - E(X).E(Y)} \tag{3.24}$$

Chứng minh. Ta chứng minh trong trường hợp biến ngẫu nhiên rời rạc. Giả sử biến ngẫu nhiên hai chiều (X,Y) có bảng phân phối xác suất đồng thời cho trong Định nghĩa 3.1. Theo Định nghĩa 3.2 và 3.11,

$$\begin{aligned} cov(X,Y) &= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_i - E(X))(y_j - E(Y))p_{ij} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_{ij} - E(X)\sum_{i=1}^{m}\sum_{j=1}^{n} y_j p_{ij} - E(Y)\sum_{i=1}^{m}\sum_{j=1}^{n} x_i p_{ij} + E(X)E(Y)\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} \\ &= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) = E(XY) - E(X)E(Y), \end{aligned}$$

vì $\sum_{i=1}^{m}\sum_{j=1}^{n} y_j p_{ij} = E(Y)$, $\sum_{i=1}^{m}\sum_{j=1}^{n} x_i p_{ij} = E(X)$, $\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1$.

Đối với biến ngẫu nhiên liên tục được chứng minh tương tự. □

Từ Định lý 3.7 và 3.8 ta thu được các tính chất sau đây của hiệp phương sai.

Định lý 3.9. (a) $cov(X,Y) = cov(Y,X)$.

(b) $V(X) = cov(X,X)$, $V(Y) = cov(Y,Y)$.

(c) Nếu X,Y độc lập thì $cov(Y,X) = 0$, điều ngược lại chưa chắc đã đúng.

(d) $cov(aX,Y) = acov(X,Y)$ với a là hằng số.

(e) $cov(X+Z,Y) = cov(X,Y) + cov(Z,Y)$.

(f) $cov\Big(\sum_{i=1}^n X_i, Y\Big) = \sum_{i=1}^n cov(X_i,Y)$.

Ví dụ 3.18. Cho biến ngẫu nhiên hai chiều (X,Y) có bảng phân phối xác suất là

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

(a) Tìm $E(X)$, $E(Y)$, $cov(X,Y)$. (b) X và Y có độc lập không?

Lời giải Ví dụ 3.18

(a) Ta có

$$E(X) = (-1)\times\frac{9}{15} + 0\times\frac{4}{15} + 1\times\frac{2}{15} = -\frac{7}{15}.$$
$$E(Y) = (-1)\times\frac{5}{15} + 0\times\frac{5}{15} + 1\times\frac{5}{15} = 0.$$
$$E(XY) = (-1)\times(-1)\times\frac{4}{15} + (-1)\times(1)\times\frac{4}{15} + 1\times(-1)\times 0 + 1\times 1\times 0 = 0.$$

Suy ra $cov(X,Y) = E(XY) - E(X)\times E(Y) = 0$.

(b) Dễ kiểm tra được $P(X=-1,Y=-1) \neq P(X=-1)\times P(Y=-1)$ nên X,Y không độc lập.

Trong Ví dụ 3.18 ta có $cov(X,Y) = 0$ nhưng hai biến ngẫu nhiên X và Y không độc lập.

Nhận xét 3.7. Hiệp phương sai được dùng làm độ đo quan hệ giữa hai biến X và Y:

(a) $cov(X,Y) > 0$ cho thấy xu thế Y tăng khi X tăng.

(b) $cov(X,Y) < 0$ cho thấy xu thế Y giảm khi X tăng.

Sau đây là một số tính chất nhằm cung cấp thêm công cụ để tính phương sai và độ lệch chuẩn.

Định lý 3.10. Cho (X, Y) là biến ngẫu nhiên hai chiều, a, b, c là các hằng số. Khi đó

$$V(aX + bY + c) = a^2V(X) + b^2V(Y) + 2ab\, cov(X, Y).$$

Chứng minh. Sử dụng Định nghĩa 2.9,

$$V(aX + bY + c) = E[(aX + bY + c) - E(aX + bY + c)]^2.$$

Theo Hệ quả 3.1 và Hệ quả 2.1,

$$E(aX + bY + c) = aE(X) + bE(Y) + c.$$

Suy ra

$$\begin{aligned} V(aX + bY + c) &= E\{a[X - E(X)] + b[Y - E(Y)]\}^2 \\ &= a^2E[X - E(X)]^2 + b^2E[Y - E(Y)]^2 + 2abE\{[X - E(X)][Y - E(Y)]\} \\ &= a^2V(X) + b^2V(Y) + 2ab\, cov(X, Y). \end{aligned}$$

□

Từ Định lý 3.10, nếu cho $b = 0$, $c = 0$ ta nhận được Định lý 2.9(a), nếu cho $a = 0$, $b = 0$ ta nhận được Định lý 2.9(b).

Hệ quả 3.2. (a) Nếu $a = 1$, $b = 1$ và $c = 0$ ta có

$$V(X + Y) = V(X) + V(Y) + 2cov(X, Y).$$

(b) Nếu $a = 1$, $b = -1$ và $c = 0$ ta có

$$V(X - Y) = V(X) + V(Y) - 2cov(X, Y).$$

Từ Định lý 3.10 và Định lý 3.9(c) ta có hệ quả sau.

Hệ quả 3.3. Nếu X và Y là hai biến ngẫu nhiên độc lập thì

$$V(aX + bY) = a^2V(X) + b^2V(Y).$$
$$V(aX - bY) = a^2V(X) + b^2V(Y).$$

Định nghĩa 3.12 (Ma trận hiệp phương sai)**.** Ma trận hiệp phương sai của biến ngẫu nhiên hai chiều (X, Y) được xác định bởi

$$\Gamma = \begin{bmatrix} cov(X, X) & cov(X, Y) \\ cov(Y, X) & cov(Y, Y) \end{bmatrix} = \begin{bmatrix} V(X) & cov(X, Y) \\ cov(X, Y) & V(Y) \end{bmatrix}$$

Tính chất 3.4. (a) Ma trận hiệp phương sai là ma trận đối xứng.

(b) Ma trận hiệp phương sai là ma trận của dạng toàn phương không âm.

Nhận xét 3.8. Hiệp phương sai có hạn chế cơ bản là khó xác định được miền biến thiên, nó thay đổi từ cặp biến thiên này sang cặp biến thiên khác. Chưa kể về mặt vật lý nó có đơn vị đo bằng bình phương đơn vị đo của biến ngẫu nhiên X, Y (nếu chúng cùng đơn vị đo). Vì thế cần đưa ra một số đặc trưng khác để khắc phục hạn chế này, đó là "hệ số tương quan".

3.7.4 Hệ số tương quan

Định nghĩa 3.13 (Hệ số tương quan)**.** Hệ số tương quan của hai biến ngẫu nhiên X và Y, ký hiệu là ρ_{XY}, được xác định như sau:

$$\boxed{\rho_{XY} = \frac{cov(X,Y)}{\sqrt{V(X).V(Y)}} = \frac{cov(X,Y)}{\sigma(X)\sigma(Y)}} \tag{3.25}$$

Tính chất 3.5. **(a)** $|\rho_{XY}| \le 1$.

(b) Nếu $\rho_{XY} = \pm 1$ ta nói hai biến ngẫu nhiên X và Y có quan hệ tuyến tính (tức là tồn tại a và b sao cho $Y = aX + b$).

(c) Nếu $\rho_{XY} = 0$ ta nói hai biến ngẫu nhiên X và Y là không tương quan.

Nói chung $0 < |\rho_{XY}| < 1$, trong trường hợp này ta nói hai biến X và Y tương quan với nhau. Chú ý rằng, hai biến tương quan thì phụ thuộc (không độc lập), nhưng không tương quan thì chưa chắc độc lập.

Nhận xét 3.9. Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa X và Y. Khi $|\rho_{XY}|$ càng gần 1 thì tính chất tương quan tuyến tính càng chặt. Khi $|\rho_{XY}|$ càng gần 0 thì sự phụ thuộc tuyến tính càng ít, càng lỏng lẻo. Khi $\rho_{XY} = 0$ 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.

Ví dụ 3.19. Cho biến ngẫu nhiên hai chiều rời rạc (X,Y) có bảng phân phối xác suất là

X \ Y	1	2	3
1	0,17	0,13	0,25
2	0,10	0,30	0,05

(a) Tìm ma trận hiệp phương sai của (X,Y). (b) Tìm hệ số tương quan $\rho_{X,Y}$.

Lời giải Ví dụ 3.19

(a) Tính

$$E(X) = 1 \times 0,55 + 2 \times 0,45 = 1,45;\ V(X) = 1 \times 0,55 + 4 \times 0,45 - (1,45)^2 = 0,2475.$$
$$E(Y) = 2,03;\ V(Y) = 0,5691.$$
$$E(XY) = 2,88 \Longrightarrow cov(X,Y) = E(XY) - E(X) \times E(Y) = -0,0635.$$

Vậy ma trận hiệp phương sai

$$\Gamma = \begin{pmatrix} V(X) & cov(X,Y) \\ cov(Y,X) & V(Y) \end{pmatrix} = \begin{pmatrix} 0,2475 & -0,0635 \\ -0,0635 & 0,5691 \end{pmatrix}$$

(b) Hệ số tương quan $\rho_{XY} = \dfrac{cov(X,Y)}{\sqrt{V(X)\times V(Y)}} = -0,1692.$

Ví dụ 3.20. Trọng lượng của những người chồng tuân theo luật phân phối chuẩn với trọng lượng trung bình là 70kg và độ lệch chuẩn 9kg, còn trọng lượng của những người vợ tuân theo luật phân phối chuẩn với trọng lượng trung bình là 55kg và độ lệch chuẩn 4kg. Hệ số tương quan trọng lượng giữa vợ và chồng là $\frac{2}{3}$. Tính xác suất vợ nặng hơn chồng.

Lời giải Ví dụ 3.20 Gọi X và Y lần lượt là các biến ngẫu nhiên chỉ "trọng lượng của chồng" và "trọng lượng của vợ". Ta có $X \sim \mathcal{N}(70;9^2)$, $Y \sim \mathcal{N}(55;4^2)$. Ta cần tính $P(X<Y)$.

Vì $X-Y$ là biến ngẫu nhiên có phân phối chuẩn với

$$E(X-Y) = E(X) - E(Y) = 70 - 55 = 15$$

và

$$V(X-Y) = V(X) + V(Y) - 2\text{cov}(X,Y) = 9^2 + 4^2 - (2)(24) = 49,$$

trong đó $\text{cov}(X,Y) = \rho_{XY}\sigma(X)\sigma(Y) = \dfrac{2}{3}(9)(4) = 24$. Vậy $X-Y \sim \mathcal{N}(15;49)$. Suy ra

$$P(X<Y) = P(X-Y<0) = 0,5 + \phi\left(\frac{0-15}{\sqrt{49}}\right) = 0,5 - \phi(2,14) \simeq 0,5 - 0,48382 = 0.01618,$$

trong đó $\phi(2,14) = 0,48382$ được tra từ bảng hàm số Láp-la-xơ (Phụ lục 2).

3.8 Luật số lớn và định lý giới hạn trung tâm

Trong mục này ta nghiên cứu luật số lớn đó là sự hội tụ theo xác suất của dãy các biến ngẫu nhiên và định lý giới hạn trung tâm: khảo sát sự hội tụ theo phân phối xác suất của dãy các biến ngẫu nhiên.

3.8.1 Sự hội tụ của dãy các biến ngẫu nhiên

Định nghĩa 3.14 (Hội tụ theo xác suất)**.** Xét dãy biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ và biến ngẫu nhiên X trong cùng một phép thử. Ta nói rằng dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ hội tụ theo xác suất về biến ngẫu nhiên X nếu với mọi $\varepsilon > 0$

$$\boxed{\lim_{n\to\infty} P\Big(|X_n - X| > \varepsilon\Big) = 0} \tag{3.26}$$

Như vậy dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ hội tụ theo xác suất về biến ngẫu nhiên X thì với n đủ lớn, thực tế gần như chắc chắn ta có thể coi rằng X_n không khác mấy so với X.

Định nghĩa 3.15 (Hội tụ theo phân phối)**.** Dãy biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ được gọi là hội tụ theo phân phối về biến ngẫu nhiên X nếu dãy các hàm phân phối xác suất $\{F_{X_n}(x)\}_{n=1}^{\infty}$ hội tụ về hàm phân phối $F_X(x)$. Tức là với mọi $x \in \mathbb{R}$

$$\boxed{\lim_{n\to\infty} F_{X_n}(x) = F_X(x)} \tag{3.27}$$

3.8.2 Luật số lớn Trê-bư-sep

Định lý 3.11 (Bất đẳng thức Markov)**.** Cho Y là biến ngẫu nhiên không âm có kỳ vọng hữu hạn. Khi đó với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{P(Y \geq \epsilon) < \frac{E(Y^2)}{\epsilon^2}} \tag{3.28}$$

Chứng minh. Ta chứng minh cho trường hợp Y là biến ngẫu nhiên liên tục có hàm mật độ xác suất là $f_Y(y)$.

$$P(Y \geq \varepsilon) = \int\limits_{\varepsilon}^{+\infty} f_Y(y)dy = \frac{1}{\varepsilon^2}\int\limits_{\varepsilon}^{+\infty} \varepsilon^2 f_Y(y)dy \leq \frac{1}{\varepsilon^2}\int\limits_{\varepsilon}^{+\infty} y^2 f_Y(y)dy \leq \frac{1}{\varepsilon^2}\int\limits_{0}^{+\infty} y^2 f_Y(y)dy = \frac{E(Y^2)}{\varepsilon^2}.$$

Dấu bằng không thể đồng thời xảy ra ở cả 2 dấu "=" và "$\leq$" trong biểu thức trên. □

Định lý 3.12 (Bất đẳng thức Trê-bư-sep)**.** Cho X là biến ngẫu nhiên có kỳ vọng $E(X) = \mu$ và phương sai $V(X) = \sigma^2$ hữu hạn. Khi đó với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{P(|X-\mu| \geq \varepsilon) < \frac{\sigma^2}{\varepsilon^2}} \tag{3.29}$$

hay tương đương

$$\boxed{P(|X-\mu| < \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}} \tag{3.30}$$

Chứng minh. Ta chứng minh cho trường hợp X là biến ngẫu nhiên liên tục. Đặt $Y = |X - \mu| \geq 0$ và áp dụng Định lý 3.11. □

Các bất đẳng thức (3.29) và (3.30) được gọi là bất đẳng thức Trê-bư-sep.

Nhận xét 3.10. Bất đẳng thức Trê-bư-sep có nhiều ứng dụng. Trước hết nó cho phép ta đánh giá cận trên hoặc cận dưới của xác suất để biến ngẫu nhiên X nhận giá trị sai lệch so với kỳ vọng $E(X)$ không quá ϵ. Bất đẳng thức Trê-bư-sep có ý nghĩa to lớn về mặt lý thuyết, nó được sử dụng để chứng minh các định lý của luật số lớn.

Ví dụ 3.21. Để ước lượng nhanh chóng sai số của số vải bán ra trong một tháng của một cửa hàng, ta tiến hành như sau:

1. Giả sử có n khách hàng trong một tháng và số vải của mỗi khách hàng được làm tròn bởi số nguyên gần nhất (ví dụ trong sổ ghi 195,6m thì được làm tròn là 196m).

2. Ký hiệu X_i là sai số giữa số mét vải thực bán và số mét vải đã làm tròn của khách hàng thứ i trong tháng, $i = 1, 2, \ldots, n$. Các sai số $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập cùng có phân bố đều trên đoạn $[-0,5; 0,5]$. Khi đó $E(X_i) = 0$ và $V(X_i) = \dfrac{1}{12}$,

$i = 1, 2, \ldots, n$. Sai số tổng cộng trong cả tháng là $X = X_1 + X_2 + \cdots + X_n$. Ta tính được $E(X) = 0$ và $V(X) = \dfrac{n}{12}$. Theo bất đẳng thức Trê-bư-sep, xác suất để sai số vượt quá ε mét sẽ được đánh giá bởi:

$$P(|X - 0| \geq \varepsilon) = P(|X| \geq \varepsilon) < \frac{n}{12\varepsilon^2}.$$

3. Bây giờ giả sử $n = 10000$. Để xác suất $P(|X| \geq \varepsilon)$ bé hơn 0,01 ta phải có $\dfrac{n}{12\varepsilon^2} \leq 0,01$ hay $\varepsilon \geq 288,6746$.

4. Vậy ta có thể kết luận: Với xác suất 0,99 sai số giữa số vải thực bán với số vải đã tính tròn không vượt quá 289m, nếu số khách hàng là 100000.

Định lý 3.13 (Định lý Trê-bư-sep)**.** Nếu dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n, \ldots$ độc lập, có kỳ vọng hữu hạn và phương sai bị chặn đều ($V(X_i) \leq C, \forall i = 1, 2, \ldots, C$ là hằng số dương), thì với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| < \epsilon\right) = 1} \tag{3.31}$$

Chứng minh. Áp dụng Định lý (3.12) với $X = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ ta nhận được kết quả của Định lý Trê-bư-sep. □

Hệ quả 3.4. Nếu dãy các biến ngẫu nhiên $X_1, X_2, \ldots, X_n, \ldots$ độc lập, có cùng kỳ vọng hữu hạn ($E(X_i) = \mu, i = 1, 2, \ldots$) và phương sai bị chặn đều ($V(X_i) \leq C\ \forall i = 1, 2, \ldots, C$ là hằng số dương), thì với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{\lim_{n\to+\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \epsilon\right) = 1} \tag{3.32}$$

Nhận xét 3.11. (a) Định lý Trê-bư-sep chứng tỏ rằng trung bình số học của các biến ngẫu nhiên độc lập hội tụ theo xác suất về trung bình số học của kỳ vọng tương ứng của nó. Nói cách khác có sự ổn định của trung bình số học của một số lớn các biến ngẫu nhiên xung quanh trung bình số học của các kỳ vọng của các biến ngẫu nhiên ấy. Như vậy mặc dù từng biến ngẫu nhiên độc lập có thể nhận giá trị khác nhiều so với kỳ vọng của chúng, song trung bình số học của một số lớn các biến ngẫu nhiên lại nhận giá trị gần bằng trung bình số học các kỳ vọng của chúng với xác suất rất lớn. Điều đó cho phép dự đoán giá trị trung bình số học của các biến ngẫu nhiên. Chẳng hạn, gieo một con xúc xắc cân đối đồng chất. Gọi X là số chấm xuất hiện ở mặt trên con xúc xắc thì $E(X) = 3,5$. Một nhà thống kê đã gieo một con xúc xắc cân đối đồng chất 1 triệu lần (nhờ sự trợ giúp của máy vi tính) và ghi lại số chấm xuất hiện ở mặt trên con xúc xắc. Số chấm trung bình của 1 triệu lần gieo được tìm thấy là $\dfrac{x_1 + x_2 + \cdots + x_{10^6}}{10^6} \simeq 3,500867$

(b) Định lý Trê-bư-sep có ứng dụng rộng rãi trong nhiều lĩnh vực, chẳng hạn nó chính là cơ sở cho phương pháp đo lường trong vật lý. Để xác định giá trị của một đại lượng vật lý nào đó người ta thường tiến hành đo n lần độc lập và lấy trung bình số học của các kết quả đo làm giá trị thực của đại lượng cần đo. Thật vậy, giả sử xem kết quả của n lần đo là các biến ngẫu nhiên $X_1, X_2, \dots, X_n$. Ta thấy rằng các biến ngẫu nhiên này độc lập, có cùng kỳ vọng bằng chính giá trị thực của đại lượng vật lý (giả sử không có sai số hệ thống), các phương sai của chúng đều bị chặn trên bởi bình phương của độ chính xác của thiết bị đo. Do đó theo định lý Trê-bư-sep ta có thể cho rằng trung bình số học của các kết quả đo sẽ sai lệch rất ít so với giá trị thực của đại lượng vật lý với xác suất gần như bằng một.

(c) Định lý Trê-bư-sep còn là cơ sở cho lý thuyết mẫu, ứng dụng trong thống kê.

3.8.3 Luật số lớn Béc-nu-li

Áp dụng luật số lớn Trê-bư-sep với trường hợp $X_i \sim \mathcal{B}(1, p)$ chính là số lần xảy ra A trong phép thử thứ i, $i = 1, 2, \dots, n$ ta có luật số lớn Béc-nu-li.

Định lý 3.14 (Định lý Béc-nu-li)**.** Giả sử ta có n phép thử Béc–nu–li với $P(A) = p$ và m là số lần xảy ra A trong n phép thử đó. Khi đó với $\varepsilon > 0$ tùy ý cho trước ta có:

$$\boxed{\lim_{n \to +\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1} \tag{3.33}$$

Nhận xét 3.12. Định lý Béc-nu-li chỉ ra rằng tần suất xuất hiện của sự kiện A trong n phép thử độc lập sẽ hội tụ theo xác suất về xác suất của sự kiện đó khi số lần thử tăng lên vô hạn. Chính vì vậy định lý Béc-nu-li là cơ sở lý thuyết của định nghĩa thống kê về xác suất ở Chương 1: Khi $n \to +\infty$ thì $\dfrac{m}{n} \to p$.

Ví dụ 3.22. Giả sử p là tỉ lệ cử tri sẽ bầu cho ứng cử viên A. Để ước lượng trước tỉ lệ này người ta phỏng vấn ngẫu nhiên n cử tri. Có thể coi kết quả bầu của cử tri thứ i là biến ngẫu nhiên X_i có phân phối Béc-nu-li tham số p (X_i nhận giá trị 1 nếu cử tri thứ i bầu cho ứng cử viên A và nhận giá trị 0 trong trường hợp ngược lại). Các biến ngẫu nhiên X_i, $i = 1, 2, \dots, n$ độc lập và có cùng phân phối Béc-nu-li tham số p với kỳ vọng $E(X_i) = p$ và phương sai $V(X_i) = p(1-p)$. Đặt $K_n = X_1 + X_2 + \cdots + X_n$, áp dụng bất đẳng thức Trê-bư-sep ta có

$$P(|K_n - p| \geq \varepsilon) < \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \quad \text{hay} \quad P(|K_n - p| < \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}.$$

Chẳng hạn, nếu $\varepsilon = 0,1$ và $n = 100$ thì $P(|K_{100} - p| \geq 0,1) < \dfrac{1}{4 \times 100 \times (0,1)^2} = 0,25$. Nói cách khác, nếu phỏng vấn 100 cử tri và lấy kết quả này để ước lượng cho tỉ lệ cử tri sẽ bỏ phiếu cho ứng cử viên A thì sai số vượt quá 0,1 có xác suất nhỏ hơn 0,25.

Nếu muốn ước lượng tin cậy hơn (chẳng hạn xác suất lớn hơn 95%) và chính xác hơn (với sai số 0,01) thì $P(|K_n - p| < 0,01) \geq 1 - \dfrac{1}{4n(0,01)^2}$. Vậy số cử tri phải phỏng vấn thỏa mãn

$$1 - \frac{1}{4n(0,01)^2} \geq 0,95 \quad \text{suy ra} \quad n \geq 50000.$$

Như vậy để ước lượng với độ tin cậy cao và độ chính xác cao thì cần phải lấy mẫu với số lượng lớn. Tuy nhiên ở đây ta chỉ dựa vào bất đẳng thức Trê-bư-sep để giải quyết bài toán. Trong Chương 4 ta sẽ nghiên cứu về bài toán ước lượng này và bằng phương pháp khác ta sẽ chỉ ra số cử tri phải phỏng vấn nhỏ hơn kết quả trên.

3.8.4 Định lý giới hạn trung tâm

Định lý 3.15 (Định lý giới hạn trung tâm). Giả sử $\{X_n\}$ là dãy biến ngẫu nhiên độc lập cùng phân phối với $E(X_n) = \mu$, $V(X_n) = \sigma^2$ với mọi n. Đặt $\overline{X}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$. Khi đó dãy biến ngẫu nhiên $U_n = \dfrac{\overline{X}_n - \mu}{\sigma}\sqrt{n}$ hội tụ theo phân phối về phân phối chuẩn tắc $\mathcal{N}(0;1)$, tức là

$$\lim_{n\to\infty} F_{U_n}(x) = \Phi(x) \quad x \in \mathbb{R}, \tag{3.34}$$

ở đây $\Phi(x)$ là hàm phân phối xác suất của biến ngẫu nhiên phân phối chuẩn tắc được xác định bởi (2.42).

Nhận xét 3.13. **(a)** Ý nghĩa của Định lý giới hạn trung tâm là khi có nhiều nhân tố ngẫu nhiên tác động (sao cho không có nhân tố nào vượt trội lấn át các nhân tố khác) thì kết quả của chúng có dạng phân phối tiệm cận chuẩn.

(b) Trong thực hành, nếu n đủ lớn ta có:

$$\boxed{U_n = \frac{\overline{X}_n - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0;1)} \tag{3.35}$$

3.8.5 Xấp xỉ phân phối nhị thức bằng phân phối chuẩn

Ta sử dụng định lý giới hạn để trình bày lại việc xấp xỉ phân phối nhị thức bằng phân phối Poa-xông và phân phối chuẩn đã đề cập ở Chương 2.

Giả sử $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên độc lập có cùng phân phối Béc-nu-li tham số p. Theo công thức (2.30) và (2.33) ta có $X = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n,p)$. Công thức (1.19) cho phép tính xác suất $P(X = k) = P_n(k) = C_n^k p^k (1-p)^{n-k}$. Tuy nhiên khi n khá lớn ta không thể áp dụng công thức này để tính mà cần đến công thức xấp xỉ (vì sẽ tràn bộ nhớ khi tính toán có sử dụng máy tính).

Định lý 3.16 (Định lý giới hạn địa phương)**.** Giả sử X là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n,p)$. Đặt $x_k = \dfrac{k-np}{\sqrt{np(1-p)}}$, khi đó

$$\boxed{P(X=k) = \frac{\varphi(x_k)}{\sqrt{np(1-p)}}(1+\varepsilon_{n,k})} \tag{3.36}$$

trong đó $|\varepsilon_{n,k}| < \dfrac{C}{\sqrt{n}}$ với C là hằng số.

Khi n đủ lớn ta có thể xấp xỉ xác suất $P_n(k)$ trong (1.19) bởi

$$\boxed{P_n(k) = P(X=k) \simeq \frac{\varphi(x_k)}{\sqrt{np(1-p)}}} \tag{3.37}$$

trong đó $\varphi(x) = \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ là hàm Gao–xơ với các giá trị được tính trong bảng giá trị hàm Gao–xơ (Phụ lục 1) đối với các giá trị x dương. Hàm $\varphi(x)$ là hàm chẵn, tức là $\varphi(-x) = \varphi(x)$. Khi $x > 4$ ta có thể lấy $\varphi(x) \simeq 0$.

Để tính xấp xỉ giá trị của hàm phân phối xác suất nhị thức ta có thể áp dụng Định lý Moa-vrơ–Láp-la-xơ.

Định lý 3.17 (Moa-vrơ–Láp-la-xơ)**.** Đối với các các biến ngẫu nhiên độc lập $X_1, X_2, \ldots, X_n$ có cùng phân phối Béc-nu-li tham số p thì

$$\lim_{n\to\infty} P\left(\frac{X_1+X_2+\cdots+X_n-np}{\sqrt{npq}} < x\right) = \Phi(x) \quad x \in \mathbb{R}. \tag{3.38}$$

Chứng minh. Áp dụng định lý giới hạn trung tâm (Định lý 3.15) cho dãy các biến ngẫu nhiên độc lập $X_1, X_2, \ldots, X_n$ có cùng phân bố Béc-nu-li tham số p ta được kết quả của Định lý 3.17. □

Nếu X là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n,p)$, thì khi n khá lớn ta có thể sử dụng công thức xấp xỉ:

$$\boxed{P(X<x) = P\left(\frac{X-np}{\sqrt{np(1-p)}} < \frac{x-np}{\sqrt{np(1-p)}}\right) \simeq \Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)} \tag{3.39}$$

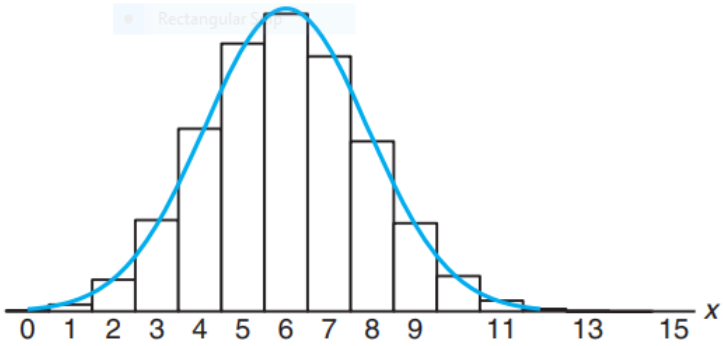
và

$$\boxed{P(k_1 \le X \le k_2) \simeq \Phi\left(\frac{k_2-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1-np}{\sqrt{np(1-p)}}\right)} \tag{3.40}$$

Công thức xấp xỉ trên là tốt nếu $np \ge 5$ và $n(1-p) \ge 5$.

Nhận xét 3.14. Khi $k_1 = k_2 = k$, $0 \le k \le n$, vế trái của công thức (3.40) sẽ là $P(X=k) \ne 0$, trong khi đó vế phải bằng 0. Điều này xảy ra vì ta đã dùng hàm phân phối liên tục để xấp xỉ phân phối rời rạc.

Ví dụ 3.23. Để minh họa việc xấp xỉ phân phối chuẩn cho phân phối nhị thức, ta vẽ biểu đồ của $\mathcal{B}(15; 0,4)$ và vẽ đường cong chuẩn có cùng kỳ vọng $\mu = np = 15 \times 0,4 = 6$ và phương sai $\sigma^2 = npq = 15 \times 0,4 \times 0,6 = 3,6$ với biến ngẫu nhiên tuân theo luật phân phối nhị thức X (xem Hình 3.11).



Hình 3.11: Xấp xỉ phân phối chuẩn cho phân phối nhị thức $\mathcal{B}(15; 0,4)$

Trong hình minh họa về xấp xỉ phân phối nhị thức bởi phân phối chuẩn, vì ta xấp xỉ một phân phối rời rạc bằng một phân phối liên tục, nên cần một sự hiệu chỉnh để giảm sai số.

Định nghĩa 3.16. Cho X là biến ngẫu nhiên tuân theo luật phân phối nhị thức $\mathcal{B}(n; p)$. Phân phối xác suất của X được xấp xỉ bởi phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với $\mu = np$ và $\sigma^2 = np(1-p)$ và

$$\boxed{P(X = k) = C_n^k p^k (1-p)^{n-k} \simeq \Phi\left(\frac{k + 0,5 - \mu}{\sigma}\right) - \Phi\left(\frac{k - 0,5 - \mu}{\sigma}\right)} \tag{3.41}$$

$$\boxed{P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k (1-p)^{n-k} \simeq \Phi\left(\frac{k_2 + 0,5 - \mu}{\sigma}\right) - \Phi\left(\frac{k_1 - 0,5 - \mu}{\sigma}\right)} \tag{3.42}$$

Xấp xỉ là khá tốt nếu $np \ge 5$ và $n(1-p) \ge 5$.

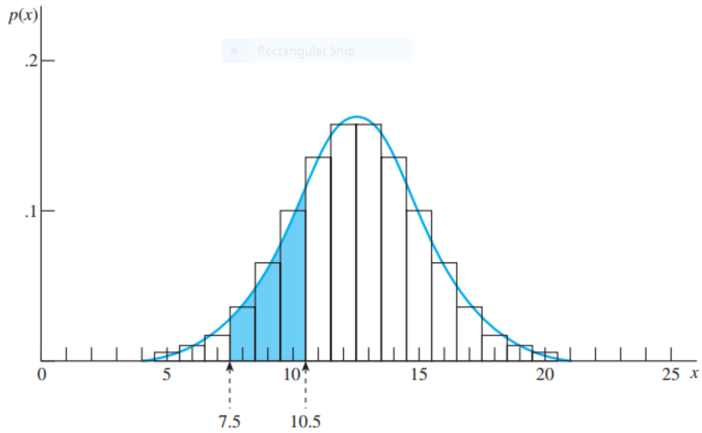
Nhận xét 3.15. Hình 3.12 và 3.13 biểu thị biểu đồ xác suất nhị thức với $n = 25$ và $p = 0,5$, $p = 0,1$ tương ứng. Phân phối trong Hình 3.12 là hoàn toàn đối xứng.

Việc thêm $+0,5$ và $-0,5$ chính là yếu tố hiệu chỉnh và gọi là hiệu chỉnh liên tục.

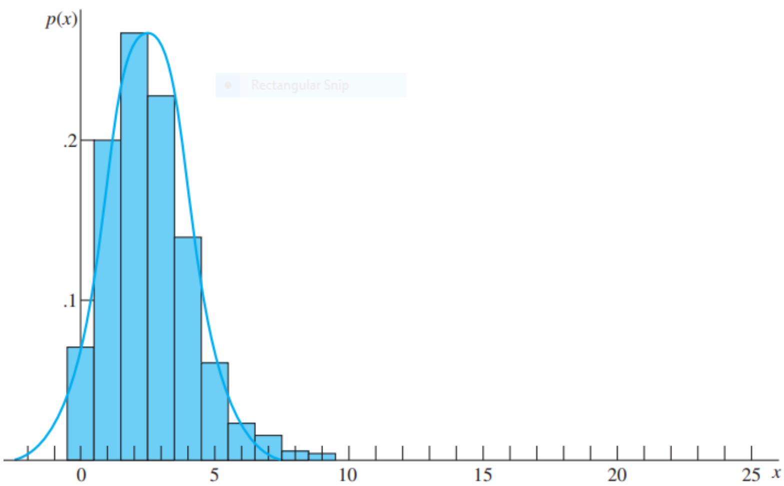
3.8.6 Xấp xỉ phân phối nhị thức bằng phân phối Poa-xông

Khi điều kiện xấp xỉ phân phối nhị thức bằng phân phối chuẩn không thỏa mãn (tức là điều kiện $np > 5$ và $n(1-p) > 5$ không thỏa mãn), ta có thể xấp xỉ phân phối nhị thức bằng phân phối Poa-xông.

Định lý 3.18. Cho dãy $\{X_n\}_{n=1}^{\infty}$ là dãy các biến ngẫu nhiên có phân bố nhị thức, với mỗi X_n có phân phối nhị thức $X_n \sim \mathcal{B}(n, p_n)$. Nếu tồn tại giới hạn $\lim_{n\to\infty} np_n = \lambda > 0$ thì X_n hội tụ theo phân phối về biến ngẫu nhiên X có phân phối Poa-xông tham số λ.



Hình 3.12: Phân phối nhị thức với $n = 25$ và $p = 0,5$ xấp xỉ bởi phân phối chuẩn với $\mu = 12,5$ và $\sigma = 2,5$



Hình 3.13: Phân phối nhị thức và xấp xỉ phân phối chuẩn với $n = 25$ và $p = 0,1$

3.9 Tổng hợp một số đề thi

Ví dụ 3.24 (Đề thi cuối kỳ 20183)**.** Cho biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất là

$$f_{X,Y}(x,y) = \begin{cases} kx^2, & \text{nếu} \quad -1 \le x \le 1,\ 0 \le y \le x^2, \\ 0, & \text{nếu trái lại.} \end{cases}$$

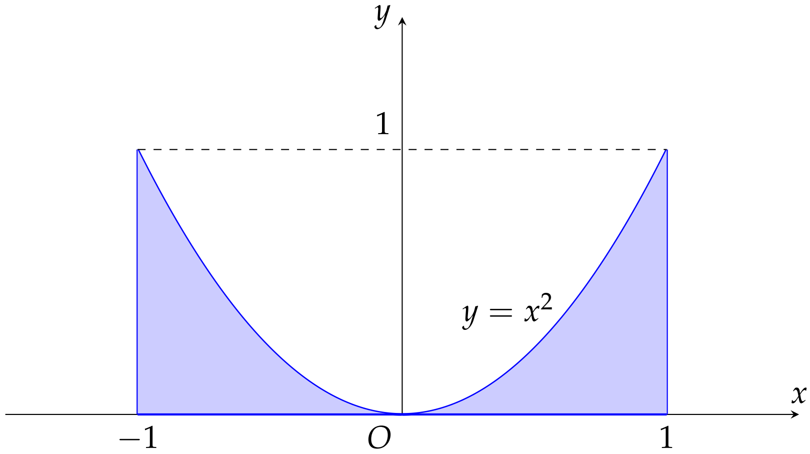
(a) Tìm k.

(b) Tính $P\left(Y \le \dfrac{1}{4}\right)$.

Lời giải Ví dụ 3.24 (a) Sử dụng Tính chất 3.3(a), $k \ge 0$. Sử dụng Tính chất 3.3(b),

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x,y)dxdy = \iint_D kx^2dxdy = \int_{-1}^{1}\int_{0}^{x^2} kx^2dxdy = k\frac{x^5}{5}\Big|_{-1}^{1} = \frac{2k}{5},$$

trong đó $D = \{(x,y) \in \mathbb{R}^2 : -1 \le x \le 1,\ 0 \le y \le x^2\}$ (xem Hình 3.14). Suy ra $k = 5/2$.



Hình 3.14: Miền D của Ví dụ 3.24(a)

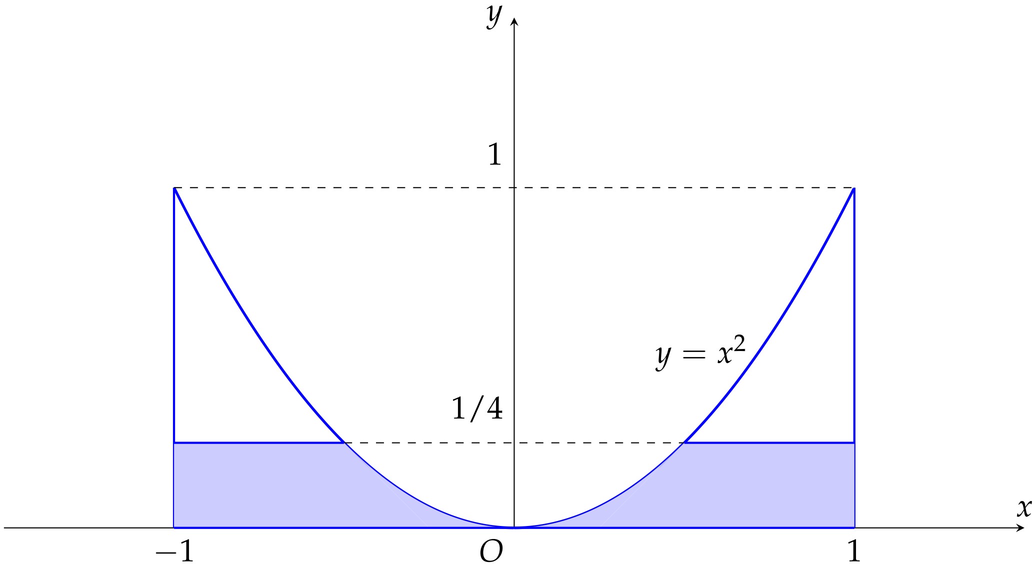
(b) Sử dụng Tính chất 3.3(c) ta tính

$$P\left(Y \le \frac{1}{4}\right) = \iint_{D_1} \frac{5}{2}dxdy = 2\left[\int_0^{1/2} dx \int_0^{x^2} \frac{5}{2}x^2dy + \int_{1/2}^{1} dx \int_0^{1/4} \frac{5}{2}x^2dy\right] = \frac{19}{48} \simeq 0,3958.$$

Hoặc

$$P\left(Y \le \frac{1}{4}\right) = \iint_{D_1} \frac{5}{2}dxdy = 2\int_0^{1/4} dy \int_{\sqrt{y}}^{1} \frac{5}{2}x^2dx = \frac{19}{48} \simeq 0,3958.$$

Ở đây $D_1 = \{(x,y) \in \mathbb{R}^2 : -1 \le x \le 1,\ 0 \le y \le x^2,\ y \le 1/4\}$ (xem Hình 3.15)



Hình 3.15: Miền D_1 của Ví dụ 3.24(b)

Ví dụ 3.25 (Đề thi cuối kỳ 20191)**.** Cho U và V là hai biến ngẫu nhiên liên tục, độc lập với nhau và có cùng phân phối đều trên $[10; 30]$.

(a) Tìm hàm mật độ xác suất $f_{U,V}(u, v)$ của biến ngẫu nhiên hai chiều (U, V).

(b) Tính $P(|U - V| < 10)$.

Lời giải Ví dụ 3.25

(a) Vì U, V là hai biến ngẫu nhiên liên tục, có phân phối đều trên $[10; 30]$ nên

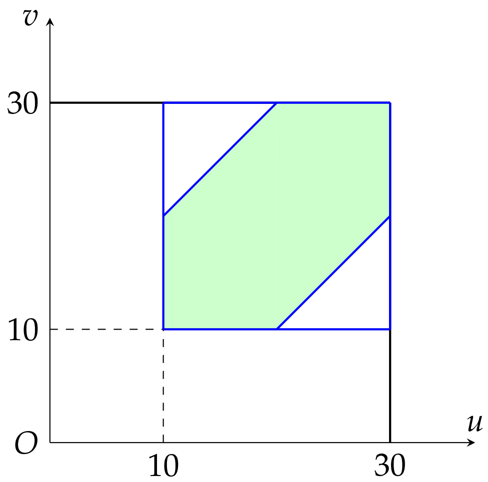
$$f_U(u) = \begin{cases} \dfrac{1}{20}, & u \in [10;30], \\ 0, & u \notin [10;30], \end{cases} \quad f_V(v) = \begin{cases} \dfrac{1}{20}, & v \in [10;30], \\ 0, & v \notin [10;30]. \end{cases}$$

Mặt khác vì U và V độc lập nên $f_{U,V}(u, v) = \begin{cases} \dfrac{1}{400}, & (u,v) \in [10;30]^2, \\ 0, & (u,v) \notin [10;30]^2. \end{cases}$

(b) $P(|U - V| < 10) = \displaystyle\int\int_{D \cap S_{U,V}} f_{U,V}(u, v)dudv$ với $D = \{(u, v) \in \mathbb{R}^2 : |u - v| < 10\}$. Sử dụng tính chất của tích phân hai lớp suy ra

$$P(|U - V| < 10) = \frac{1}{400}(20^2 - 10^2) = \frac{3}{4} = 0,75.$$

Ví dụ 3.26 (Đề thi cuối kỳ 20192)**.** Thời gian hoạt động X_i, $i = 1, 2, 3$, của linh kiện điện tử I, II, III là các biến ngẫu nhiên độc lập, tuân theo luật phân phối mũ với hàm mật độ xác suất tương ứng là $f_{X_i}(x) = \lambda_i e^{-\lambda_i x}$, $x > 0$, $\lambda_i > 0$, $i = 1, 2, 3$. (a) Tìm hàm phân phối xác suất của biến ngẫu nhiên chỉ thời gian hoạt động của một hệ thống gồm 3 linh kiện trên mắc nối tiếp. (b) Tìm kỳ vọng và phương sai của biến ngẫu nhiên chỉ thời gian hoạt động của hệ thống đó.



Hình 3.16: Miền $D \cap S_{U,V}$ trong Ví dụ 3.25(b)

Lời giải Ví dụ 3.26 (a) Gọi X : "thời gian hoạt động của hệ thống", $X = \min\{X_1, X_2, X_3\}$.

$$F_X(x) = P(X < x) = 1 - P(\min\{X_1, X_2, X_3\} \geq x) = 1 - \Pi_{i=1}^{3} P(X_i \geq x)$$
$$= 1 - \Pi_{i=1}^{3}\Big[1 - P(X_i < x)\Big] = 1 - e^{-(\lambda_1+\lambda_2+\lambda_3)x}, \ x > 0.$$

(b) Sử dụng công thức tính kỳ vọng, phương sai của biến ngẫu nhiên phân phối mũ,

$$E(X) = \frac{1}{\lambda_1 + \lambda_2 + \lambda_3}, \quad V(X) = \frac{1}{(\lambda_1 + \lambda_2 + \lambda_3)^2}.$$

Bài tập Chương 3

Biến ngẫu nhiên rời rạc

Bài tập 3.1. Cho biến ngẫu nhiên X và Y 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 và Y độc lập.

(b) Lập bảng phân phối xác suất của X và Y.

(c) Tìm quy luật phân phối của biến ngẫu nhiên $Z = XY$.

(d) Tính $E(Z)$ bằng 2 cách và kiểm tra $E(Z) = E(X).E(Y)$.

Bài tập 3.2. Cho biến ngẫu nhiên X và Y có bảng phân bố xác suất đồng thời là

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

(a) Tìm $E(X)$, $E(Y)$, $cov(X,Y)$.

(b) X và Y có độc lập không?

(c) Tìm bảng phân phối xác suất của X, của Y.

Bài tập 3.3. Cho biến ngẫu nhiên X và Y có bảng phân bố xác suất đồng thời là

$X \backslash Y$	1	2	3
1	0,17	0,13	0,25
2	0,10	0,30	0,05

(a) Lập bảng phân phối xác suất của X và của Y.

(b) Lập ma trận Covarian của (X,Y).

(c) Tìm hệ số tương quan.

(d) X và Y có độc lập không?

Bài tập 3.4. Thống kê về giá thành sản phẩm Y(triệu đồng) và sản lượng X(tấn) của một ngành sản xuất thu được bảng phân phối xác suất sau:

$X \backslash Y$	30	50	80	100
6	0,05	0,06	0,08	0,11
7	0,06	0,15	0,04	0,08
8	0,07	0,09	0,10	0,11

(a) Tìm giá thành sản phẩm trung bình và mức độ phân tán của nó.

(b) Tìm sản lượng trung bình khi giá thành bằng 8.

(c) X và Y có độc lập không?

(d) X và Y có tương quan không?

Bài tập 3.5. Cho X_1, X_2, X_3 là các biến ngẫu nhiên độc lập theo luật phân phối Poa-xông với tham số $\lambda_1 = 1$, $\lambda_2 = 2$, $\lambda_3 = 3$. Tính xác suất của các sự kiện sau:

(a) Số lớn nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1.

(b) Số lớn nhất trong các số X_1, X_2, X_3 bằng 1.

(c) Số nhỏ nhất trong các số X_1, X_2, X_3 không nhỏ hơn 1.

(d) Số nhỏ nhất trong các số X_1, X_2, X_3 bằng 1.

Bài tập 3.6. Cho X và Y là hai biến ngẫu nhiên có bảng phân phối xác suất là:

X	0	1	2	3	4	5
P	0,15	0,3	0,25	0,2	0,08	0,02

Y	0	1	2	3	4	5
P	0,3	0,2	0,2	0,15	0,1	0,05

(a) Tính $E(X)$, $E(Y)$, $V(X)$, $V(Y)$.

(b) Nếu X và Y độc lập, tính $P(X + Y \leq 2)$ và lập bảng phân phối xác suất của $X + Y$.

Bài tập 3.7. Lấy ngẫu nhiên 3 viên bi từ một hộp gồm 3 bi đỏ, 5 bi xanh và 4 bi vàng. Gọi X, Y lần lượt là số bi xanh, bi vàng trong 3 bi lấy ra. Lập bảng phân phối xác suất đồng thời cho biến ngẫu nhiên hai chiều (X, Y).

Biến ngẫu nhiên liên tục

Bài tập 3.8. Cho X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

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

(a) Tìm hằng số k

(b) X và Y có độc lập không?

Bài tập 3.9. Cho X, Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

$$f_{X,Y}(x,y) = \begin{cases} k\left(x^2 + \dfrac{xy}{2}\right), & \text{nếu } 0 < x < 1,\ 0 < y < 2, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm hằng số k.

(b) Tìm hàm phân phối đồng thời của X và Y.

Bài tập 3.10. Cho X, Y là hai biến ngẫu nhiên có hàm mật độ xác suất đồng thời

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{6\pi}, & \text{nếu } \dfrac{x^2}{9} + \dfrac{y^2}{4} < 1, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm hàm mật độ xác suất biên của X, của Y.

(b) Tìm xác suất để (X, Y) nằm trong hình chữ nhật $O(0,0); A(0,1); B(1,2); D(2,0)$.

Bài tập 3.11. Cho biến ngẫu nhiên hai chiều liên tục (X, Y) có hàm mật độ xác suất là

$$f_{X,Y}(x,y) = \begin{cases} kx^2, & \text{nếu } \quad -1 \le x \le 1,\ 0 \le y \le x^2, \\ 0, & \text{nếu trái lại.} \end{cases}$$

(a) Tìm k.

(b) Tìm hàm mật độ xác suất biên $f_X(x)$, $f_Y(y)$.

(c) Tính $P\left(Y \le \dfrac{1}{4}\right)$.

Bài tập 3.12. Cho X và Y là hai biến ngẫu nhiên liên tục có hàm mật độ xác suất đồng thời là

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

(a) Tìm hàm mật độ xác suất biên của X, của Y.

(b) Tìm hàm mật độ xác suất có điều kiện $f_1(x|y)$, $f_2(y|x)$.

Bài tập 3.13. Một linh kiện điện tử có thời gian hoạt động X là biến ngẫu nhiên có phân phối mũ với hàm mật độ xác suất là $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$, $\lambda > 0$.

(a) Tìm hàm phân phối xác suất của biến ngẫu nhiên chỉ thời gian hoạt động của một mạng gồm 2 linh kiện loại trên được mắc song song/mắc nối tiếp.

(b) Tính kỳ vọng, phương sai của thời gian hoạt động của mạng đó.

Bài tập 3.14. Cho X và Y là hai biến ngẫu nhiên độc lập với nhau có cùng phân phối đều trên $[0, 2]$.

(a) Tìm hàm phân phối của các biến ngẫu nhiên $Z = X + Y$; $T = XY$; $U = X - Y$.

(b) Tính $P(-1 \le Y - X \le 1)$.

Bài tập 3.15. Hai người A và B hẹn gặp nhau tại cổng trường trong khoảng từ 7h00 đến 8h00. Gọi X và Y lần lượt là thời gian đến điểm hẹn của người A và B trong khoảng thời gian trên. Giả sử X và Y độc lập và có cùng phân phối đều trên [7;8].

(a) Tìm hàm phân phối xác suất đồng thời của X và Y.

(b) Với quy ước chỉ đợi nhau trong vòng 10 phút, tìm xác suất để 2 người được gặp nhau.

Bài tập 3.16. Cho X và Y là hai biến ngẫu nhiên độc lập, $X \sim \mathcal{N}(5; 1^2)$, $Y \sim \mathcal{N}(3; (0,2)^2)$.

(a) Tìm $P(X + Y < 5,5)$.

(b) Tìm $P(X < Y)$; $P(X > 2Y)$.

(c) Tìm $P(X < 1; Y < 1)$.

Bài tập 3.17. Trọng lượng của những người chồng tuân theo luật phân phối chuẩn với kỳ vọng 70kg và độ lệch chuẩn 9kg, còn trọng lượng của những người vợ tuân theo luật phân phối chuẩn với kỳ vọng 55kg và độ lệch chuẩn 4kg. Hệ số tương quan trọng lượng giữa vợ và chồng là $^2/_3$. Tính xác suất vợ nặng hơn chồng.

Bài tập 3.18. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất $f_X(x)$. Tìm hàm mật độ xác suất $g_Y(y)$ của biến ngẫu nhiên Y nếu:

(a) $Y = X + 1$, $-\infty < x < +\infty$.

(b) $Y = 2X$, $-a < x < a$.

Bài tập 3.19. Giả sử tại một trường đại học, một sinh viên đạt được điểm X trong bài kiểm tra năng khiếu toán học và điểm Y trong bài kiểm tra năng khiếu âm nhạc là một số trong khoảng từ 0 đến 1. Giả sử X và Y được phân phối theo hàm mật độ sau

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{2}{5}(2x+3y) & 0 < x, y < 1, \\ 0 & \text{nếu ngược lại.} \end{cases}$$

(a) Tính tỷ lệ sinh viên đại học đạt điểm cao hơn 0,8 trong bài kiểm tra năng khiếu toán.

(b) Giả sử điểm số của một sinh viên trong bài kiểm tra năng khiếu âm nhạc là 0,3. Tính xác suất để điểm của anh ấy trong bài kiểm tra năng khiếu toán học sẽ lớn hơn 0,8.

(c) Giả sử điểm số của một sinh viên trong bài kiểm tra năng khiếu toán là 0,3. Tính xác suất để điểm của anh ấy trong bài kiểm tra năng khiếu âm nhạc sẽ lớn hơn 0,8.

Bài tập 3.20. Một mảnh đất bằng phẳng có hình tam giác vuông với một bờ phía nam dài 200m, bờ phía đông dài 100m. Ta quan tâm đến điểm mà một hạt giống rơi từ trên cao xuống tiếp đất. Giả sử rằng hạt giống nằm trong ranh giới của mảnh đất với tọa độ X và Y của nó được phân bố đều trên bề mặt của tam giác vuông.

(a) Tìm c với c là giá trị của hàm mật độ xác suất của điểm nằm trong ranh giới mảnh đất.

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

(c) Tìm hàm mật độ xác suất của Y biết $X = x$ và tính $P(0,1 \le Y \le 0,7 \mid X = 0,5)$.