Chương III: VECTƠ NGẪU NHIÊN

(ĐẠI LƯỢNG NGẪU NHIÊN NHIỀU CHIỀU)

III.1. Khái niệm.

Nếu các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$ cùng xác định trên các kết quả của một phép thử thì ta nói $Z = (X_1, X_2, \ldots, X_n)$ là một vectơ ngẫu nhiên *n* chiều.

III.2. Vectơ ngẫu nhiên rời rạc 2 chiều (X,Y).

III.2.1 Bảng phân phối XS đồng thời.

III.2.2 Phân phối XS theo các BNN thành phần X, Y (PP lề).

III.2.3 PP XS có điều kiện.

III.2.4 Điều kiện độc lập của X và Y.

III.2.5 Hàm phân phối XS của (X,Y).

III.3 Một số tham số đặc trưng của vectơ ngẫu nhiên.

* Kỳ vọng toán
* Kỳ vọng có điều kiện
* Covarian (Hiệp phương sai)
* Hệ số tương quan & ý nghĩa.
* Ma trận tương quan
* Sử dụng máy tính bỏ túi để tính một số tham số đặc trưng.

III.4. Hàm của vectơ ngẫu nhiên (X,Y).

III.2 PHÂN PHỐI XÁC SUẤT của VTNN RỜI RẠC 2 CHIỀU

III.2.1 Bảng phân phối XS đồng thời:

Cho $X = \{x_1, x_2, \ldots, x_m\}$; $Y = \{y_1, y_2, \ldots, y_n\}$.

Đặt $p_{ij} = P(X = x_i, Y = y_j)$; $i = \overline{1,m},\ j = \overline{1,n},$. Dưới đây là **bảng phân phối xác suất đồng thời** của (X, Y):

X \ Y	y_1	y_2	...	Y_n
x_1	p_{11}	p_{12}	...	p_{1n}
x_2	P_{21}	P_{22}	...	P_{2n}
...	...	...	...	...
x_m	p_{m1}	p_{m2}	...	p_{mn}

Khi đó $0 \le p_{ij} \le 1$ và $\sum_i \sum_j p_{ij} = 1$.

III.2.2 Phân phối XS theo các BNN thành phần X, Y (PP lề).

Đặt: $$p_i = \sum_{j=1}^{n} p_{ij} = P(X = x_i), i = \overline{1, m}$$

Ta được bảng phân phối xác suất của X:

X	x_1	x_2	...	x_m
$\mathbf{P^X}$	p_1	p_2	...	p_m

Đặt: $$q_j = \sum_{i=1}^{m} p_{ij} = P(Y = y_j), j = \overline{1, n}$$

Ta được bảng phân phối xác suất của Y:

X	y_1	y_2	...	y_n
$\mathbf{P^Y}$	q_1	q_2	...	q_n

III.2.3 Phân phối xác suất có điều kiện:

Bảng PPXS của X với điều kiện $Y = y_j (j = \overline{1, n})$ là:

X	x_1	x_2	...	x_m
P^{X/y_j}	$\frac{p_{1j}}{q_j}$	$\frac{p_{2j}}{q_j}$	...	$\frac{p_{mj}}{q_j}$

tức là $P(X=x_i | Y=y_j) = \frac{p_{ij}}{q_j}$

Bảng PPXS của Y đối với điều kiện $X = x_i (i = \overline{1, m})$ là:

Y	y_1	y_2	...	y_n
P^{Y/x_i}	$\frac{p_{i1}}{p_i}$	$\frac{p_{i2}}{p_i}$	...	$\frac{p_{in}}{p_i}$

III.2.4 Điều kiện độc lập của X và Y.

Hai biến ngẫu nhiên X và Y độc lập với nhau nếu quy luật phân phối xác suất của X không phụ thuộc vào việc biến Y nhận giá trị nào, và ngược lại.

Các tính chất: X và Y độc lập

$\Leftrightarrow P(X=x_i, Y=y_j) = P(X=x_i).P(Y=y_j)\ \forall i,j$ hay $p_{ij} = p_i q_j\ \forall i, j$.

$\Leftrightarrow F(x,y) = F_X(x).F_Y(y);$

III.2.5 Hàm phân phối xác suất đồng thời của (X,Y) .

$$F(x,y) = P(X < x, Y < y) \overset{RR}{=} \sum_{x_i<x} \sum_{y_j<y} p_{ij}$$

Lưu ý:

- *F(x,y) chính là xác suất để điểm ngẫu nhiên M(X,Y) rơi vào hình chữ nhật vô hạn có đỉnh phía trên, bên phải là (x,y).*

III.3 MỘT SỐ ĐẶC TRƯNG của BNN hai chiều rời rạc:

*** Kỳ vọng toán:** $E(X,Y) = (E(X),E(Y))$

*** Hiệp phương sai (Covarian, mômen tương quan):**

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

ở đây: $E(X.Y) \overset{RR}{=} \sum_j \sum_i x_i y_j p_{ij}$

Nhận xét:

$$\text{cov}(X,X) = E[(X-E(X))^2] = E(X^2) - E^2(X) \equiv D(X)$$

$$\text{cov}(Y,Y) = \quad \dots \quad = E(Y^2) - E^2(Y) \equiv D(Y)$$

* **Ma trận tương quan** (ma trận hiệp phương sai) của (X,Y):

$$D(X,Y)=\begin{bmatrix} \mathrm{cov}(X,X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \mathrm{cov}(Y,Y) \end{bmatrix}$$

$$=\begin{bmatrix} D(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & D(Y) \end{bmatrix}$$

* **Hệ số tương quan của X và Y:**

$$R_{XY}=\frac{\mathrm{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{E(XY)-E(X).E(Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

Hệ số tương quan và covarian dùng để đặc trưng cho mức độ chặt chẽ của mối liên hệ phụ thuộc giữa các BNN X và Y.

Hệ số tương quan không có đơn vị đo và **$|R_{XY}| \leq 1$**.

Nếu $R_{XY} = 0$ thì ta nói X, Y không tương quan, ngược lại khi $R_{XY} \neq 0$ ta nói X, Y có tương quan.

Nếu X, Y độc lập thì cov(X,Y)= $R_{XY} = 0$.

Điều ngược lại không đúng, tức là nếu cov(X,Y)= 0 thì hoặc X, Y độc lập, hoặc X, Y phụ thuộc ở một dạng thức nào đó.

Nếu $R_{XY} = \pm 1$ thì X, Y có tương quan tuyến tính (thuận /nghịch).

Khi $R_{XY} \approx \pm 1$ thì X, Y có tương quan "gần" tuyến tính.

III.4 HÀM CỦA BIẾN NGẪU NHIÊN 2 CHIỀU

Nếu ứng với mỗi cặp giá trị có thể có của biến ngẫu nhiên 2 chiều (X, Y) có một gía trị có thể có của Z thì ta nói Z là hàm của 2 biến ngẫu nhiên , ký hiệu Z = φ(X,Y).

Khi biến ngẫu nhiên 2 chiều (X, Y) có bảng phân phối xác suất đồng thời trong III.2.1, ta có thể tìm phân phối xác suất của Z theo công thức: $P(Z = z_k) = \sum_{\varphi(x_i, y_j) = z_k} p_{ij}$

Một số tính chất:

Nếu $Z = \varphi(X, Y) \Rightarrow E(Z) = \sum_{i;j} \varphi(x,y) \times p_{ij}$

$E(aX \pm bY) = a\,E(X) \pm b\,E(Y)$; a,b là các hằng số

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

* Khi X, Y độc lập : $E(XY) = E(X) \times E(Y)$

$D(X \pm Y) = D(X) + D(Y)$

* Giả sử $X_1, X_2, \ldots, X_n$ là các BNN độc lập, có cùng phân phối xác suất.

Ký hiệu $E(X_i) = a$; $D(X_i) = \sigma^2$; $\forall i$. Ta có :

- BNN $U = X_1 + X_2 + \ldots + X_n$ có $E(U) = n.a$ và $D(U) = n.\sigma^2$;
- BNN $\overline{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ co' $E\left(\overline{X}\right) = a$; $D\left(\overline{X}\right) = \dfrac{\sigma^2}{n} \ll \sigma^2$

$\Rightarrow$ Khi đo một đại lượng vật lý, người ta thường đo nhiều lần rồi lấy trung bình cộng các kết quả làm giá trị của đại lượng đó.

Ví dụ 1

Một hộp đựng 5 sản phẩm, trong đó có 3 phế phẩm mà không kiểm tra thì không biết. Các sản phẩm được lấy ra kiểm tra cho đến khi phát hiện thấy 2 phế phẩm thì dừng lại.

Kí hiệu X là BNN chỉ số lần kiểm tra cho tới khi phế phẩm đầu tiên được phát hiện. Y là BNN chỉ số lần kiểm tra tiếp theo cho tới khi phế phẩm thứ hai được phát hiện.

Hãy :

a) Lập bảng phân phối xác suất đồng thời của (X, Y).

b) Tính cov(X,Y) và hệ số tương quan của X, Y.

c) X,Y có độc lập hay không ?

d) Tìm phân phối XS và kỳ vọng có điều kiện của X khi Y=2.

Hướng dẫn:

Gọi A_i là biến cố lần thứ i lấy được chính phẩm; i = 1,2,3…

$\overline{A_i}$ là biến cố lần thứ i lấy được phế phẩm; i = 1,2,3…

VD: $$\underbrace{A_1.\overline{A_2}}_{X=2}.\underbrace{\overline{A_3}}_{Y=1} \Rightarrow P(X=2;Y=1)=\frac{2}{5}\times\frac{3}{4}\times\frac{2}{3}=0,2$$

$$\underbrace{\overline{A_1}}_{X=1}.\underbrace{A_2.A_3\overline{A_4}}_{Y=3} \Rightarrow P(X=1;Y=3)=\frac{3}{5}\times\frac{2}{4}\times\frac{1}{3}\times\frac{2}{2}=0,1$$

X \ Y	1	2	3
1	?	?	0,1
2	0,2	?	?
3	?	?	?

Y / X	1	2	3
1	0,3	0,2	0,1
2	0,2	0,1	0
3	0,1	0	0

$$p_{11} = P(X=1;Y=1) = P(\overline{A_1}.\overline{A_2}) = \frac{3}{5}.\frac{2}{4} = \frac{3}{10}$$

$$p_{12} = P(X=1;Y=2) = P(\overline{A_1}.A_2.\overline{A_3}) = \frac{3}{5}.\frac{2}{4}.\frac{2}{3} = \frac{1}{5}$$

$$p_{13} = P(X=1;Y=3) = P(\overline{A_1}.A_2.A_3.\overline{A_4}) = \frac{3}{5}.\frac{2}{4}.\frac{1}{3}.1 = \frac{1}{10}$$

$$p_{21} = P(X=2;Y=1) = P(A_1.\overline{A_2}.\overline{A_3}) = \frac{2}{5}.\frac{3}{4}.\frac{2}{3} = \frac{1}{5}$$

$$p_{22} = P(X=2;Y=2) = P(A_1.\overline{A_2}.A_3.\overline{A_4}) = \frac{2}{5}.\frac{3}{4}.\frac{1}{3}.1 = \frac{1}{10}$$

$$p_{31} = P(X=3;Y=1) = P(A_1.A_2.\overline{A_3}.\overline{A_4}) = \frac{2}{5}.\frac{1}{4}.1.1 = \frac{1}{10}$$

$$p_{23} = P(X=2;Y=3) = 0$$

$$p_{32} = P(X=3;Y=2) = 0$$

$$p_{33} = P(X=3;Y=3) = 0$$

b) Tính Cov(X,Y) và R_{XY}:

X \ Y	1	2	3	P^X
1	0,3	0,2	0,1	0,6
2	0,2	0,1	0	0,3
3	0,1	0	0	0,1
P^Y	0,6	0,3	0,1	

Viết lại các bảng PPXS thành phần của X và Y (phân phối lề):

X	1	2	3
P^X	0,6	0,3	0,1

Y	1	2	3
P^Y	0,6	0,3	0,1

E(X)= E(Y)=1,5 D(X)= D(Y) = 0,45

$$E\left(XY\right)=1\times1\times0,3+1\times2\times0,2+1\times3\times0,1+2\times1\times0,2+2\times2\times0,1+3\times1\times0,1=2,1$$

$$\mathrm{cov}(X,Y)=E(XY)-E(X).E(Y)=-0,15. \quad R_{XY}=\frac{E(XY)-E(X).E(Y)}{\sqrt{D(X)}\sqrt{D(Y)}}=\frac{-1}{3}$$

HD Sử dụng MTBT tìm 1 số đặc trưng của VTNN rời rạc:

Các bước thực hiện	Máy CASIO fx 570 ES (PLUS)...	Máy CASIO fx 580 VNX....
Mở cột tần số (nếu máy chưa mở)	**SHIFT -- MODE (SETUP) -- ▼ --** **-- 4 (STAT) -- 1 (ON)**	
Vào chế độ thống kê hai biến.	**MODE -- 3 (STAT) -- 2 (A+BX)**	
Nhập dữ liệu	**X Y FREQ** **1 x1 y1 p11** **2 x1 y2 p12** **... ... ...** **... xn ym pnm** **AC**	
Đọc kết quả **E(X); E(Y)**	**SHIFT – 1 (STAT)- 4 (VAR) –** **--- 2 ($\overline{x}$) -- =** ***Muốn có kq E(Y) thì chọn*** $\overline{y}$	
Đọc kết quả $\sqrt{D(X)}$ $\sqrt{D(Y)}$	**SHIFT – 1 (STAT)- 4 (VAR) –** **--- 3 (σX) -- =** ***Muốn có kq*** $\sqrt{D(Y)}$ ***thì chọn*** σY	
Đọc kết quả R_{XY}	**SHIFT – 1 (STAT)-6(REG)–3 (r) --=**	
Tham khảo các KQ trung gian	**SHIFT – 1 (STAT)- 3 (SUM) - ....**	

c) Theo đn, X,Y độc lập $\Leftrightarrow$ $P(X=x_i; Y=y_j) = P(X=x_i).P(Y=y_j)$; $\forall$ i,j.

Trong bảng PPXS đồng thời: $P(X=1;Y=1) = 0{,}3$

nhưng $P(X=1)\times P(Y=1) = (6/10)\times(6/10) = 0{,}36$

$\left.\begin{matrix} P(X=1;Y=1) = 0{,}3 \\ P(X=1)\times P(Y=1) = 0{,}36 \end{matrix}\right\} \neq$

nên ta kết luận X,Y không độc lập.

(Cách khác: Do $R_{XY} \neq 0$ nên suy ra X,Y không độc lập)

d) Từ bảng PPXS đồng thời, suy ra bảng phân phối xác suất của X với điều kiện Y=2:

$X^{\mid Y=2}$	1	2
$P^{X\mid Y=2}$	$\frac{0,2}{0,3}=\frac{2}{3}$	$\frac{0,1}{0,3}=\frac{1}{3}$

và $E(X|Y=2) = 4/3$.

Ví dụ 2

Cho hai đại lượng ngẫu nhiên X, Y độc lập có các bảng phân phối xác suất:

Y	0	1
P	$\frac{1}{2}$	$\frac{1}{2}$

X	-1	1	2
P	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

a) Lập bảng phân phối xác suất của $Z = 3X^2 + 2Y$; Tính E(Z),D(Z).

b) Tính E(U),D(U) với $U = 5X - 3Y + 10$.

Hướng dẫn: X,Y độc lập nên $P(X=a, Y=b) = P(X=a) \times P(Y=b), \forall a,b$.

$\Rightarrow P(X=-1; Y=0) = P(X=-1) \times P(Y=0) = \frac{1}{4} \times \frac{1}{2} = 1/8$.

Lập bảng PPXS đồng thời của (X,Y) rồi tính giá trị hàm $Z=3X^2+2Y$.

$Z=3X^2+2Y$

X \ Y	0	1
-1	1/8 $Z = 3$	1/8 $Z = 5$
1	2/8 $Z = 3$	2/8 $Z = 5$
2	1/8 $Z = 12$	1/8 $Z = 14$

Suy ra bảng phân phối xác suất của Z:

$$\begin{array}{c|cccc} Z & 3 & 5 & 12 & 14 \\ \hline P & \frac{3}{8} & \frac{3}{8} & \frac{1}{8} & \frac{1}{8} \end{array}$$

Vậy E(Z) = 6,25 và D(Z) = 16,1875.

b) HD: E(5X - 3Y +10) = 5E(X) - 3E(Y) + 10.

D(5X -3Y + 10) = 25D(X) + 9D(Y).

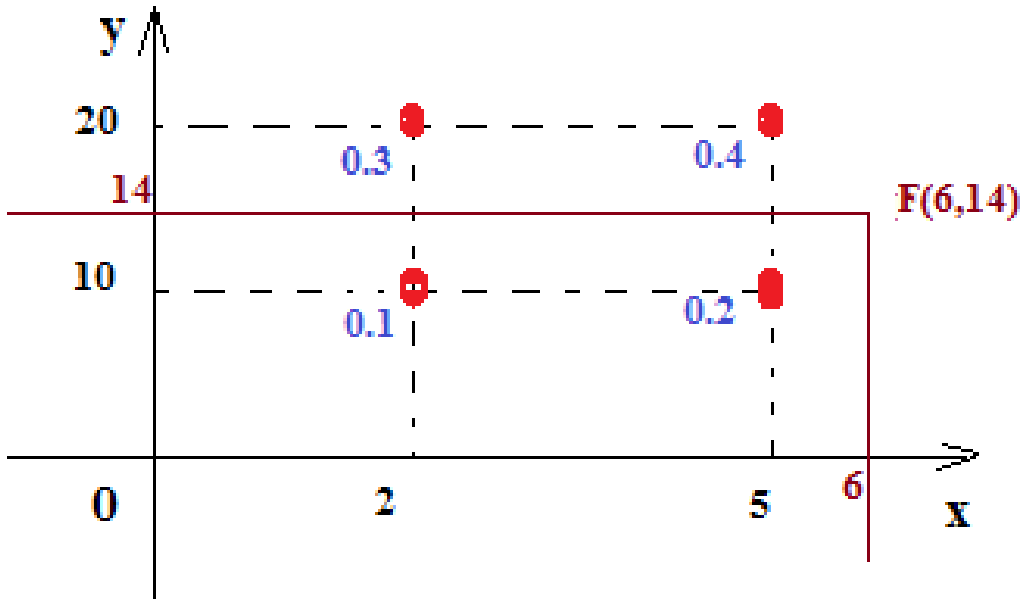
Ví dụ 3

Dưới đây là bảng PPXS đồng thời của 2 biến ngẫu nhiên X,Y. Tìm hàm phân phối XS của (X,Y).

X \ Y	10	20
2	0.1	0.3
5	0.2	0.4

Hướng dẫn :

$$F(x,y) = P(X<x; Y<y) = \sum_{i,j} p_{ij} \quad ; x_i < x \ \& y_j < y.$$



Ví dụ:

+ F(x,y) = P(X<x; Y<y).

+ F(6; 14)=P(X< 6; Y< 14) = 0,3

+ F(3; 20)=P(X<3,Y<20)= 0,1.

+F(4;25)=P(X<4; Y<25) =0,4.

Đáp số:

$$F(x,y)=\begin{cases} 0,1 & (x,y)\in(2,\ 5]\times(10,\ 20] \\ 0,1+0,3 & (x,y)\in(2,\ 5]\times(20,+\infty) \\ 0,1+0,2 & (x,y)\in(5,+\infty)\times(10,\ 20] \\ 1 & (x,y)\in(5,+\infty)\times(20,+\infty) \\ 0 & (x,y)\neq \end{cases}$$

Ví dụ 4

Biến ngẫu nhiên 2 chiều (X,Y) có bảng phân phối xác suất đồng thời như sau:

X \ Y	2	3	4
0	0,12	0,15	0,03
1	0,28	0,35	0,07

a) Chứng minh X ,Y là độc lập.

b) Tìm hệ số tương quan Rxy.

c) Tìm phân phối xác suất của biến ngẫu nhiên $Z = X^2Y + 5$

d) Tính E(Z) bằng 2 cách khác nhau.

Hướng dẫn:

b) Do X,Y độc lập nên $R_{XY} = 0$.

c)

X \ Y	2	3	4
0	0,12 Z=5	0,15 Z=5	0,03 Z=5
1	0,28 Z=7	0,35 Z=8	0,07 Z=9

suy ra:

$Z = X^2Y+5$	5	7	8	9
p_i	0,3	0,28	0,35	0,07

d) Tính E(Z) bằng 2 cách khác nhau.

Cách 1: Dùng trực tiếp bảng trong câu c): $E(Z) = 6,89$

Cách 2: Dùng tính chất kỳ vọng của các biến ngẫu nhiên độc lập, với $E(X) = 0,7$; $E(X^2) = 0,7$; $E(Y) = 0,41$.

$E(Z) = E(X^2Y) + E(5) = E(X^2)\times E(Y) + 5 = 6,89$.

Ví dụ 5

Một sinh viên có xác suất nghỉ một buổi học bất kỳ là 5%; xác suất đi học trễ một buổi là 20%. Giả thiết trong 1 tuần, sinh viên đó có 5 buổi học trên trường.

a) Lập bảng phân phối xác suất đồng thời giữa biến X là số buổi sinh viên đó nghỉ trong 1 tuần và Y là số buổi sinh viên đó đi học trễ trong cùng tuần đó.

b) Lập bảng phân phối xác suất của Y với điều kiện trong tuần có 1 buổi sinh viên nghỉ học.

Hướng dẫn:

a)

Y \ X	0	1	2	3	4	5
0	0.2373	0.0791	0.0105	0.0007	2.34E-05	3.125E-07
1	0.3164	0.0844	0.0084	0.0004	6.25E-06	0
2	0.1688	0.0338	0.0023	5E-05	0	0
3	0.045	0.006	0.0002	0	0	0
4	0.006	0.0004	0	0	0	0
5	0.0003	0	0	0	0	0

b)

Y|X=1	0	1	2	3	4
Pi	0.38846	0.4144	0.1657	0.02947	0.002