XÁC SUẤT THỐNG KÊ

Tôn Thất Tú

Đà Nẵng, 2019

Chương 3. Vectơ ngẫu nhiên

1. Định nghĩa

Giả sử $X_1, ..., X_n$ là n biến ngẫu nhiên liên quan đến thí nghiệm đang xét. Lúc đó ta gọi bộ gồm n biến ngẫu nhiên $(X_1, ..., X_n)$ là vectơ ngẫu nhiên n chiều.
Ở đây, ta hạn chế xét **trường hợp** $n = 2$.

2. Hàm phân phối đồng thời

Định nghĩa: Giả sử (X, Y) là vectơ ngẫu nhiên 2 chiều. Khi đó hàm hai biến $F(x, y)$ xác định như sau:

$$F_{X,Y}(x, y) = P(X < x, Y < y)$$

được gọi là hàm phân phối đồng thời của vectơ ngẫu nhiên (X, Y).

Tính chất:

i) $0 \leq F_{X,Y}(x, y) \leq 1, \forall x, y \in \mathbf{R}$ ii) $F_{X,Y}(x, y)$ không giảm theo từng biến

iii) $F_{X,Y}(x, y)$ liên tục trái

iv) $\lim\limits_{\substack{x \to +\infty \\ y \to +\infty}} F_{X,Y}(x, y) = 1, \ \lim\limits_{x \to -\infty} F_{X,Y}(x, y) = 0, \ \lim\limits_{y \to -\infty} F_{X,Y}(x, y) = 0$

2.1 Phân phối đồng thời rời rạc

a. Bảng phân phối đồng thời: Giả sử X nhận các giá trị trong tập $\{x_i, \ i \in I\}$ và Y nhận các giá trị trong tập $\{y_j, j \in J\}$.
Đặt $p_{ij} = P(X = x_i, Y = y_j), \ i \in I, \ j \in J$. Lúc đó ta có bảng phân phối hai chiều:

X \ Y	y_1	y_2	....	y_j
x_1	p_{11}	p_{12}	....	p_{1j}
x_2	p_{21}	p_{22}	....	p_{2j}
....			....	
....			....	
x_i	p_{i1}	p_{i2}	....	p_{ij}

Ví dụ 1

Một hộp có 2 bi trắng và 3 bi đen. Lấy ngẫu nhiên 1 viên không hoàn lại, rồi tiếp tục lấy ngẫu nhiên 2 viên không hoàn lại. Gọi X và Y là số bi đen lấy được ở lần 1 và lần 2.
a. Lập bảng phân phối đồng thời của X và Y.
b. Tính xác suất $P(X+Y<2)$.

Giải. a. Ta có X nhận các giá trị: 0, 1 và Y nhận các giá trị: 0, 1, 2.

$$P(X=0,Y=0)=P(X=0)P(Y=0|X=0)=2/5*0=0$$

$$P(X=0,Y=1)=P(X=0)P(Y=1|X=0)=2/5*\frac{C_3^1C_1^1}{C_4^2}=1/5$$

$$P(X=0,Y=2)=P(X=0)P(Y=2|X=0)=2/5*\frac{C_3^2}{C_4^2}=1/5$$

$$P(X=1,Y=0)=P(X=1)P(Y=0|X=1)=3/5*\frac{C_2^2}{C_4^2}=1/10$$

$$P(X=1,Y=1)=P(X=1)P(Y=1|X=1)=3/5*\frac{C_2^1C_2^1}{C_4^2}=2/5$$

$$P(X=1,Y=2)=P(X=1)P(Y=2|X=1)=3/5*\frac{C_2^2}{C_4^2}=1/10$$

Bảng phân phối đồng thời:

X \ Y	0	1	2
0	0	1/5	1/5
1	1/10	2/5	1/10

b. Xác suất:
$P(X+Y<2)=P(X=Y=0)+P(X=0,Y=1)+P(X=1,Y=0)$
$=0+1/5+1/10=3/10$

Ví dụ 2

Cho hai biến ngẫu nhiên X,Y độc lập có bảng phân phối:

X	1	2
P	0,4	0,6

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

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

Giải. Vì X,Y độc lập nên:
$P(X=1,Y=1)=0,4*0,3=0,12;\quad P(X=1,Y=2)=0,4*0,3=0,12$
$P(X=1,Y=3)=0,4*0,4=0,16;\quad P(X=2,Y=1)=0,6*0,3=0,18$
$P(X=2,Y=2)=0,6*0,3=0,18;\quad P(X=2,Y=3)=0,6*0,4=0,24$
Bảng phân phối đồng thời:

X \ Y	1	2	3
1	0,12	0,12	0,16
2	0,18	0,18	0,24

b. Phân phối biên: Ta đưa vào kí hiệu

$$p_{i.} = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij}$$

Từ đó ta được bảng phân phối của X và được gọi là phân phối biên X:

X	x_1	x_2	...	x_i	...
P	$p_{1.}$	$p_{2.}$	...	$p_{i.}$	...

Nhận xét: Trong thực hành, để tìm phân phối biên ta có thể xuất phát từ bảng phân phối đồng thời tính tổng các xác suất theo hàng hoặc theo cột như bảng minh họa sau đây:

X \ Y	y_1	y_2	....	y_j	$P(X = x_i)$
x_1	p_{11}	p_{12}	....	p_{1j}	$p_{1.}$
x_2	p_{21}	p_{22}	....	p_{2j}	$p_{2.}$
....	....				....
....	....				...
x_i	p_{i1}	p_{i2}	....	p_{ij}	$p_{i.}$
$P(Y = y_j)$	$p_{.1}$	$p_{.2}$	....	$p_{.j}$	

Phân phối biên Y được tìm hoàn toàn tương tự.

c. Phân phối có điều kiện: Kí hiệu

$$P(x_i|y_j) = P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{p_{.j}}$$

được gọi là xác suất có điều kiện để X nhận giá trị x_i khi biết $Y = y_j$.
Ứng với mỗi y_j cố định ta có bảng phân phối xác suất có điều kiện của X với điều kiện $Y = y_j$:

$X|Y = y_j$	x_1	x_2	...	...
P	$P(x_1|y_j)$	$P(x_2|y_j)$	...	...

Tương tự ứng với mỗi x_i cố định ta có bảng phân phối xác suất có điều kiện của Y với điều kiện $X = x_i$.

Ví dụ 3

Vecto ngẫu nhiên (X, Y) có phân phối được cho ở bảng sau:

Y \ X	1	2	3
5	0.1	0.2	0.3
6	0.08	0.16	0.16

Tìm phân phối biên của X, Y và tính $P(Y = 6|X = 1)$.

Giải. Ta có:

Y \ X	1	2	3	$P(Y = y_j)$
5	0.1	0.2	0.3	0,6
6	0.08	0.16	0.16	0,4
$P(X = x_i)$	0.18	0.36	0.46	

Phân phối biên của X:

X	1	2	3
P	0,18	0,36	0,46

Phân phối biên của Y:

Y	5	6
P	0,6	0,4

Xác suất:

$$P(Y = 6|X = 1) = \frac{P(X = 1, Y = 6)}{P(X = 1)} = \frac{0,08}{0,18} = \frac{4}{9}$$

2.2 Phân phối đồng thời liên tục

a. Định nghĩa: Vectơ ngẫu nhiên (X, Y) được gọi là có phân phối đồng thời liên tục nếu hàm phân phối đồng thời của (X, Y) có thể biểu diễn ở dạng

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) du dv$$

Hàm $f_{X,Y}(u, v)$ được gọi là hàm mật độ đồng thời của X, Y.

Tính chất:

$$i) f_{X,Y}(u, v) \geq 0, \quad ii) f_{X,Y}(u, v) = \frac{\partial^2 F_{X,Y}(u, v)}{\partial u \partial v}, \quad iii) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(u, v) du dv = 1$$

b. Phân phối biên:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$$

c. Mật độ có điều kiện: Hàm

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

là mật độ điều kiện của X với điều kiện $Y = y$.

Ví dụ 4

Vectơ ngẫu nhiên (X,Y) có hàm mật độ

$$f_{X,Y}(x,y) = \begin{cases} C(x+xy), & (x,y) \in [0,1]\times[0,1] \\ 0, & (x,y) \notin [0,1]\times[0,1] \end{cases}$$

Tìm $C, f_X(x)$ và $f_{Y|X}(y|x)$.

Giải. Theo tính chất của hàm mật độ:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = \int_0^1\int_0^1 C(x+xy)dxdy = \frac{C}{2}\int_0^1 (y+1)dy = \frac{3C}{4} = 1$$

Suy ra: $C = 4/3$. Mật độ biên của X:

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy = \int_0^1 \frac{4}{3}(x+xy)dy = 2x, x \in [0,1].$$

Do đó:

$$f_X(x) = \begin{cases} 2x, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$

Mật độ có điều kiện:

$$f_{Y|X}(y|x) = \frac{f_{X,Y(x,y)}}{f_X(x)} = \frac{4/3(x+xy)}{2x} = \frac{2}{3}(y+1), (x,y) \in [0,1]\times[0,1]$$

2.3 Sự độc lập của các biến ngẫu nhiên

Hai biến ngẫu nhiên X và Y độc lập khi và chỉ khi:

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

- Nếu X và Y có phân phối đồng thời rời rạc thì điều kiện trên trở thành:

$$P(X = x_i, Y = y_j) = P(X = x_i).P(Y = y_j), \forall i, j$$

- Nếu X và Y có phân phối đồng thời liên tục thì điều kiện trên trở thành:

$$f_{X,Y}(x,y) = f_X(x).f_Y(y)$$

3. Kỳ vọng có điều kiện

Kỳ vọng có điều kiện của biến ngẫu nhiên Y với điều kiện $X = x$, kí hiệu $E(Y|X = x)$ được xác định:

- Nếu X và Y có phân phối đồng thời rời rạc thì:

$$E(Y|X = x) = \sum_j y_j P(Y = y_j|X = x)$$

- Nếu X và Y có phân phối đồng thời liên tục thì:

$$E(Y|X = x) = \int_{-\infty}^{+\infty} y f_{Y|X}(y|x)dy$$

4. Hiệp phương sai và hệ số tương quan

a. Hiệp phương sai:

- Hiệp phương sai của hai biến ngẫu nhiên X và Y, kí hiệu $Cov(X,Y)$ được xác định theo biểu thức:

$$Cov(X,Y) = E\left[(X - EX)(Y - EY)\right]$$

- Hai biến ngẫu nhiên X và Y được gọi là tương quan với nhau nếu $Cov(X,Y) \neq 0$ và không tương quan nếu $Cov(X,Y) = 0$.

Tính chất:

i) $Cov(X,Y) = Cov(Y,X)$

ii) $Cov(aX + b, cY + d) = acCov(X,Y),\ a,b,c,d = const$

iii) $Cov(X,Y) = E(XY) - E(X).E(Y)$

iv) $D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X,Y)$

Nhận xét:

i) Nếu X, Y độc lập thì $Cov(X,Y) = 0$. Do đó, từ "tính chất độc lập" suy ra "tính chất không tương quan", tuy nhiên điều ngược lại nói chung không đúng.

ii) Trong thực hành, ta tính $Cov(X,Y) = E(XY) - E(X).E(Y)$, trong đó:

$$E(XY) = \begin{cases} \sum x_i y_j p_{ij}, & X,Y \text{ có phân phối đồng thời rời rạc} \\ \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} \mathrm{xyf}_{XY}(x,y)dxdy, & X,Y \text{ có phân phối đồng thời liên tục} \end{cases}$$

b. Hệ số tương quan

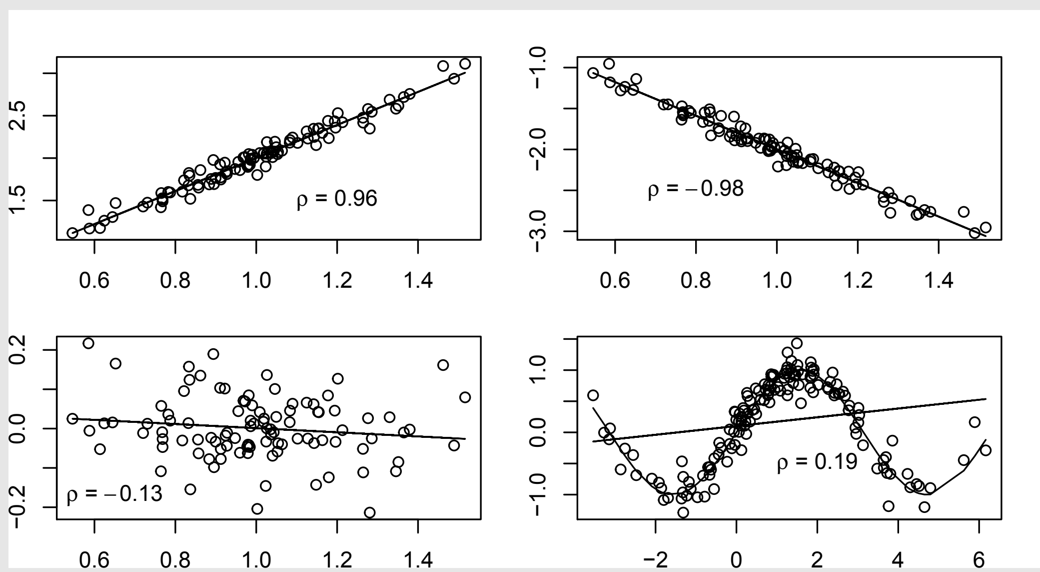
Hệ số tương quan của hai biến ngẫu nhiên X và Y, kí hiệu $\rho(X,Y)$ được xác định bởi biểu thức:

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{D(X)D(Y)}}$$

Tính chất:

i) $\rho(X,Y) = \rho(Y,X)$ ii) $|\rho(X,Y)| \leq 1$

Ý nghĩa hệ số tương quan: Hệ số tương quan đo mức độ phụ thuộc tuyến tính giữa X và Y. Khi $|\rho(X,Y)|$ càng gần 1 thì quan hệ tuyến tính giữa X và Y càng mạnh.



Ví dụ 5

Vecto ngẫu nhiên (X,Y) có phân phối được cho ở bảng sau:

Y \ X	1	2	3
1	0.1	0.1	0.2
2	0.2	0.1	0.3

Tính $E(Y|X=1), Cov(X,Y), \rho(X,Y)$.

Giải. Phân phối biên:

X	1	2	3
P	0,3	0,2	0,5

Y	1	2
P	0,4	0,6

Xác suất có điều kiện:

$$P(Y=1|X=1) = \frac{P(X=1,Y=1)}{P(X=1)} = \frac{0,1}{0,3} = \frac{1}{3}$$

$$P(Y=2|X=1) = \frac{P(X=1,Y=2)}{P(X=1)} = \frac{0,2}{0,3} = \frac{2}{3}$$

Phân phối có điều kiện:

$Y\|X=1$	1	2
P	1/3	2/3

Do đó: $E(Y|X=1) = \sum y_i P(Y=y_i|X=1) = 5/3$.
Hiệp phương sai: $Cov(X,Y) = E(XY) - E(X)E(Y)$, trong đó:
$E(X) = \sum p_i x_i = 2,2; \quad E(Y) = \sum p_i y_i = 1,6; \quad E(XY) = \sum p_{ij} x_i y_j = 3,5$
$\Rightarrow Cov(X,Y) = 3,5 - 2,2 * 1,6 = -0,02$
Phương sai:
$D(X) = E(X^2) - (EX)^2 = \sum p_i x_i^2 - 2,2^2 = 0,76$
$D(Y) = E(Y^2) - (EY)^2 = \sum p_i y_i^2 - 1,6^2 = 0,24$

$$\Rightarrow \rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{D(X)D(Y)}} = \frac{-0,02}{\sqrt{0,76*0,24}} = -0,0468$$

Ví dụ 6

Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ đồng thời:

$$f(x,y) = \begin{cases} x+y, & (x,y) \in [0,1]^2 \\ 0, & (x,y) \notin [0,1]^2 \end{cases}$$

Tính $P(0 < X < 1/2 < Y < 1), Cov(X,Y), D(X+2Y)$.

Giải.Xác suất:

$$P(0 < X < 1/2 < Y < 1) = \int_0^{1/2} \int_{1/2}^1 f(x,y)dydx = ...$$

Phân phối biên:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \int_0^1 (x+y)dy = x + 1/2, x \in [0,1]$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \int_0^1 (x+y)dx = y + 1/2, y \in [0,1]$$

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x)dx = \int_0^1 x(x+1/2)dx = 7/12; E(Y) = 7/12$$

$$E(XY) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy = \int_0^1\int_0^1 xy(x+y)dxdy = 1/3$$

Suy ra: $Cov(X,Y) = E(XY) - E(X)E(Y) = 1/3 - (7/12)^2 = -1/144$.
Mặt khác,

$$D(X+2Y) = D(X) + D(2Y) + 2Cov(X,2Y) = D(X) + 4D(Y) + 4Cov(X,Y)$$

$$E(X^2) = \int_{-\infty}^{+\infty} x^2 f_X(x)dx = \int_0^1 x^2(x+1/2)dx = 5/12; E(Y^2) = 5/12$$

$$D(X) = E(X^2) - (EX)^2 = 5/12 - (7/12)^2 = 11/144; D(Y) = 11/144$$

Vậy, $D(X+2Y) = 11/144 + 4*11/144 - 4*1/144 = 17/48$