Chương 3
CÁC ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Từ phân phối của đại lượng ngẫu nhiên người ta rút ra các số đặc trưng cho một mặt nào đó của đại lượng ngẫu nhiên. Các số này gọi là các đặc trưng của đại lượng ngẫu nhiên.

3.1. KỲ VỌNG

1- Định nghĩa kỳ vọng

Cho X là đại lượng ngẫu nhiên rời rạc có các bảng phân phối xác suất:

X	x_1	x_2	...	x_n	...
P	p_1	p_2	...	p_n	...

Khi đó ta gọi kỳ vọng của X là số:

$$E(X) = x_1p_1 + x_2p_2 + \ldots + x_np_n + \ldots = \sum_{n=1}^{\infty} x_np_n \qquad (3.1)$$

Trong trường hợp có vô hạn xn thì ta nói X có kỳ vọng và E(X) là kỳ vọng của nó nếu chuỗi (3.1) hội tụ tuyệt đối.

Vì $\sum_n p_n = 1$, nên:

$$\min x_i \le E(X) \le \max x_i$$

Do đó E(X) là một giá trị trung bình của các xi, mỗi xi được tính với tỷ trọng piA.

Vậy: kỳ vọng của đại lượng ngẫu nhiên là trung bình theo xác suất các giá trị có thể nhận của đại lượng ngẫu nhiên đó.

Trường hợp X là đại lượng ngẫu nhiên liên tục thì vai trò của hàm mật độ xác suất f(x) giống như bảng phân phối xác suất, tổng (1) tương ứng với tích phân $\int_{-\infty}^{+\infty} xf(x)dx$. Do đó:

Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ f(x) thì kỳ vọng của X là số:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx \qquad (3.2)$$

Ta nói X có kỳ vọng nếu tích phân (3.2) hội tụ tuyệt đối.

Ví dụ 3.1. Chùm chìa khoá có sáu chìa, trong đó có hai chìa mở được cửa. Thử từng chìa (thử xong bỏ ra ngoài) cho đến khi mở được cửa. Tìm số lần thử trung bình mở được cửa.

Giải. Gọi X là số lần để mở được cửa. Ta cần tìm E(X). Ta có:

$$X = \{1, 2, 3, 4, 5\}$$

Gọi Ai là biến cố lần thứ i mở được cửa ($i = \overline{1,5}$). Khi đó:

$$P(X=1) = P(A_1) = \frac{2}{6} = \frac{1}{3}$$

$$P(X=2) = P(\overline{A}_1 A_2) = P(\overline{A}_1).P(A_2/\overline{A}_1) = \frac{4}{6}.\frac{2}{5} = \frac{4}{15}$$

$$P(X=3) = P(\overline{A}_1 \overline{A}_2 A_3) = \frac{1}{5}$$

$$P(X=4) = P(\overline{A}_1 \overline{A}_2 \overline{A}_3 A_4) = \frac{2}{15}$$

$$P(X=5) = P(\overline{A}_1 \overline{A}_2 \overline{A}_3 \overline{A}_4 A_5) = \frac{1}{15}$$

Bảng phân phối xác suất của X là:

X	1	2	3	4	5
P	$\frac{1}{3}$	$\frac{4}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$

Từ đó:

$$E(X) = 1.\frac{1}{3} + 2.\frac{4}{15} + 3.\frac{1}{15} + 4.\frac{2}{15} + 5.\frac{1}{15} = \frac{7}{3}$$

Số lần thử trung bình là 2,33.

Ví dụ 3.2. Cho X là đại lượng ngẫu nhiên có hàm mật độ:

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

tìm kỳ vọng của X.

Giải.

$$E(X) = \int_{-\infty}^{+\infty} x f(x)dx = \int_0^1 2x^2 dx = \left.\frac{2x^3}{3}\right|_0^1 = \frac{2}{3}$$

2- Tính chất của kỳ vọng

Định lý 3.1. Với mọi đại lượng ngẫu nhiên X,Y và hằng số C ta có:

(i) $E(C) = C$

(ii) $E(X + Y) = E(X) + E(Y)$

(iii) $E(CX) = CE(X)$

(iv) $E(XY) = E(X)E(Y)$ nếu X và Y độc lập.

Chứng minh

(i) Bảng phân phối xác suất của X = C là:

X	C
P	1

Do đó E(X) = C.1 = C.

(ii) Nếu X và Y rời rạc có bảng phân phối xác suất đồng thời như trong mục 2 phần 2.3 chương 2 và X + Y có bảng phân phối xác suất như trong mục 3 phần 2.4 chương 2 thì:

$$E(X+Y) = \sum_{k=1}^{s} z_k p_k^* = \sum_{k=1}^{s}\left(z_k \sum_{x_i + y_j = z_k} p_{ij} \right)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n}(x_i + y_j)p_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n}x_i p_{ij} + \sum_{i=1}^{m}\sum_{j=1}^{n}y_j p_{ij}$$

$$= \sum_{i=1}^{m}x_i\left(\sum_{j=1}^{n}p_{ij}\right) + \sum_{j=1}^{n}y_j\left(\sum_{j=1}^{n}p_{ij}\right)$$

$$= \sum_{i=1}^{n}x_i p_j + \sum_{j=1}^{n}y_j q_j = E(X) + E(Y)$$

Nếu X và Y liên tục có hàm mật độ xác suất đồng thời là f(x,y) thì:

$$E(X+Y) = \int_{-\infty}^{+\infty} z f_{X+Y}(z)dz = \int_{-\infty}^{+\infty} z\left(\iint_{x+y<z} f(x,y)dxdy\right)dz$$

$$= \iint_{R^2}(x+y)f(x,y)dxdy = \iint_{R^2} xf(x,y)dxdy + \iint_{R^2} yf(x,y)dxdy$$

$$= \int_{-\infty}^{+\infty} x\left(\int_{-\infty}^{+\infty} f(x,y)dy\right)dx + \int_{-\infty}^{+\infty} y\left(\int_{-\infty}^{+\infty} f(x,y)dx\right)dy$$

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

(iii) Nếu X rời rạc thì:

$$E(CX) = \sum_i (Cx_i)p_i = C\sum_i x_i p_i = CE(X)$$

Xét trường hợp X liên tục. Nếu C = 0 thì kết quả là hiển nhiên. Nếu $C \neq 0$ và X có hàm mật độ là f(x) thì dễ dàng thấy CX có hàm mật độ là $\frac{1}{|C|} f(\frac{x}{C})$ Từ đó:

$$C > 0 : E(CX) = \int_{-\infty}^{+\infty} \frac{x}{C} f\left(\frac{x}{C}\right) dx = C\int_{-\infty}^{+\infty} t f(t)dt = CE(X)$$

$$C < 0 : E(CX) = -\int_{-\infty}^{+\infty} \frac{x}{C} f\left(\frac{x}{C}\right) dx = -C\int_{+\infty}^{-\infty} t f(t)dt = CE(X)$$

(iv) Nếu X và Y rời rạc và độc lập thì:

$$E(X.Y) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_{ij} = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_i q_j = \left(\sum_{i=1}^{m} x_i p_i\right)\left(\sum_{i=1}^{n} y_j q_j\right)$$

$$= E(X)E(Y)$$

Nếu X và Y liên tục thì do chúng độc lập nên hàm mật độ đồng thời f(x,y) = fX(x).fY(y). Từ đó:

$$E(XY) = \iint_{R^2} xyf_X(x)f_y(y)dxdy = \int_{-\infty}^{+\infty} xf_X(x)dx.\int_{-\infty}^{+\infty} yf_Y(y)dy$$

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

3.2. PHƯƠNG SAI

1- Định nghĩa phương sai

Cho X là một đại lượng ngẫu nhiên có kỳ vọng E(X). Khi đó ta gọi phương sai của X là kỳ vọng của bình phương độ sai khác giữa X và E(X), ký hiệu là D(X). Vậy:

D(X) = E[(X – E(X))2]

Ký hiệu: a = E(X), thì:

D(X) = E[(X – a)2] = E(X – a)2

Nhận xét. Từ định nghĩa phương sai và ý nghĩa của kỳ vọng ta thấy phương sai là trung bình của bình phương sai số giữa X và E(X). Như vậy phương sai càng nhỏ thì các giá trị của X càng tập trung xung quanh E(X).

Nếu X rời rạc có bảng phân phối xác suất:

X	x_1 x_2 ... x_n ...
P	p_1 p_2 ... p_n ...

thì $(X - a)^2$ có bảng phân phối xác suất là:

$(X - a)^2$	$(x_1 - a)^2$	$(x_2 - a)^2$	...	$(x_n - a)^2$	...
P	p_1	p_2	...	p_n	...

Do đó ta có công thức:

$$D(X) = \sum_i (x_i - a)^2 p_i$$

Nếu X liên tục, có hàm mật độ xác suất f(x) thì ta có:

$$D(X) = \int_{-\infty}^{+\infty} (x - a)^2 f(x)dx$$

2- Tính chất của phương sai

Định lý 3.2. Với mọi đại lượng ngẫu nhiên X, Y và hằng số C ta có:

(i) $D(X) \geq 0,\ D(C) = 0$

(ii) $D(CX) = C^2 D(X)$

(iii) $D(X) = E(X^2) - (E(X))^2$

(iv) $D(X+Y) = D(X)+D(Y)$ nếu X và Y độc lập.

$D(X + C) = D(X)$

Chứng minh

(i) $D(C) = E(C - E(C))2 = E(C - C)2 = E(0) = 0.$

(ii) $D(CX) = E(CX - E(CX))2 = E(CX - CE(X))2$

$= C2E(X - E(X))2 = C2D(X)$

(iii) $D(X) = E(X - a)2 = E(X2 - 2aX + a2)$

$= E(X2) - 2aE(X) + a2$

$= E(X2) - a2$

(iv) $D(X + Y) = E(X + Y)2 - (E(X + Y))2$

$= E(X2 + 2XY + Y2) - (E(X) + E(Y))2$

$= E(X2) - (E(X))2 + E(Y2) - (E(Y))2$

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

(do X, Y độc lập nên $E(XY) = E(X)E(Y)$).

Nhận xét. Do $D(X) \geq 0$ nên ta định nghĩa độ lệch của X là:

$$\sigma(X) = \sqrt{D(X)}$$

Độ lệch cũng có ý nghĩa như phương sai. Đôi lúc ta cũng ký hiệu phương sai là σ2.

Nếu X rời rạc thì:

$$E(X^2) = \sum_i x_i^2 p_i$$

Nếu X liên tục thì:

$$\mathbf{E(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)dx}$$

Do đó theo (iii), định lý 3.2 ta cũng có thể tính được phương sai.

Ví dụ 3.3. Cho X có bảng phân phối xác suất:

X	–1	0	3
P	0,2	0,5	0,3

Tính E(X) và D(X).

Giải. Ta có:

E(X) = –1.0,2 + 0.0,5 + 3.0,3 = 0,7

từ đó:D(X) = (–1 – 0,7)2.0,2 + (0 – 0,7)2.0,5 + (3 – 0,7)2.0,3 = 2,41

Có thể tính phương sai theo cách khác:

E(X2) = (–1)2.0,2 + 02.0,5 + 33.0,3 = 2,9

D(X) = E(X2) – [E(X)]2 = 2,9 – 0.72 = 2,41

Ví dụ 3.4. Cho X có hàm mật độ xác suất là:

$$f(x) = \begin{cases} 3x^2 & \text{nếu } x \in [0,1] \\ 0 & \text{nếu } x \notin [0,1] \end{cases}$$

Tính E(X) và D(X).

Giải. Ta có:

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 x.3x^2dx = \left.\frac{3x^4}{4}\right|_0^1 = \frac{3}{4}$$

$$D(X) = \int_{-\infty}^{+\infty} \left(x - \frac{3}{4}\right)^2 f(x)dx = \int_0^1 \left(x - \frac{3}{4}\right)^2 .3x^2dx$$

$$= \left.\left(\frac{3}{5}x^2 - \frac{9}{8}x^4 + \frac{9}{16}x^3\right)\right|_0^1 = \frac{3}{80}$$

Có thể tính D(X) theo cách khác:

$$E(X^2) = \int x^2 f(x)dx = \int 3xdx = \frac{3}{5}x^5\Big|_0^1 = \frac{3}{5}$$

từ đó:

$$D(X) = E(X^2) - (E(X))^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = \frac{3}{80}$$

3.3. MỘT SỐ ĐẶC TRƯNG KHÁC CỦA ĐẠI LƯỢNG NGẪU NHIÊN

1- Mốt của đại lượng ngẫu nhiên

Cho X là đại lượng ngẫu nhiên rời rạc có bảng phân phối xác suất:

X	x_1 x_2 ... x_n ...
P	p_1 p_2 ... p_n ...

Nếu $p_{k_o} = \max p_k$ thì ta gọi mốt của X là:

$$\mathrm{mod}(X) = x_{k_o}$$

Mốt của X gọi là số có khả năng nhất (xem mục 1.4 chương1).

Cho X là đại lượng ngẫu nhiên liên tục có hàm mật độ f(x). Nếu f(xo) = maxf(x) thì ta gọi mốt của X là:

$$\mathrm{mod}(X) = x_o$$

Ví dụ 3.5. Nếu X có bảng phân phối xác suất:

X	–1	0	2	3
P	0,1	0,4	0,3	0,2

thì: $$\mathrm{mod}(X) = 0.$$

Ví dụ 3.6. Nếu X có hàm mật độ:

$$f(x) = \begin{cases} 1 & \text{nếu } x \in [0,1] \\ 0 & \text{nếu } x \notin [0,1] \end{cases}$$

thì mod(X) là một giá trị bất kỳ thuộc đoạn [0,1].

2- Trung vị của đại lượng ngẫu nhiên

Cho X là một đại lượng ngẫu nhiên. Số m gọi là trung vị của X, ký hiệu là med(X) nếu:

$$P(X < m) \le \frac{1}{2} \text{ và } P(X > m) \le \frac{1}{2}$$

Ví dụ 3.7. Cho X, Y có bảng phân phối xác suất:

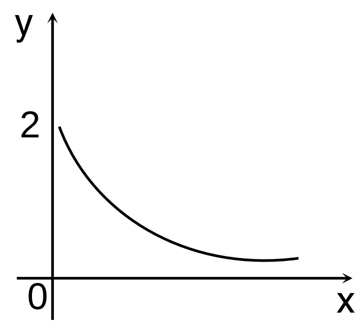
X	0	1	2	3
P	0,3	0,4	0,1	0,2

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

Khi đó med(X) = 1, med(Y) là một giá trị bất kỳ thuộc khoảng (1,2).

Ví dụ 3.8. Cho X là đại lượng ngẫu nhiên có hàm mật độ:

$$f(x) = \begin{cases} 2e^{-2x} & \text{nếu } x \ge 0 \\ 0 & \text{nếu } x < 0 \end{cases}$$



Tìm mod(X), med(X).

Giải. Từ đồ thị của hàm f(x) ta thấy:

max f(x) = f(0) = 2

nên: mod(X) = 0

Gọi m là med(X), ta cần tìm m để:

$$P(X < m) = \frac{1}{2}$$

tức là:

$$\int_{-\infty}^{m} f(x)dx = \frac{1}{2} \Rightarrow \int_0^m 2e^{-2x} = \frac{1}{2} \Rightarrow -e^{-2x}\Big|_0^m = \frac{1}{2}$$

$$\Rightarrow e^{2m} = 2 \qquad \Rightarrow m = \frac{\ln 2}{2}$$

vậy: $\mathbf{med(X)} = \frac{\ln 2}{2}$

3- Mômen trung tâm

Cho X là một đại lượng ngẫu nhiên có kỳ vọng E(X) = a. Ta gọi mômen trung tâm cấp k của X là:

$$\mu_k = \mu_k(X) = E(X - a)^k$$

Ta gọi mômen gốc cấp k là: $\gamma_k = E(X^k)$. Ta có : $\gamma_1 = a$

Theo công thức nhị thức Newton:

$$\mu_n = E(X-a)^n = E\left(\sum_{k=0}^{n} C_n^k (-a)^k X^{n-k}\right)$$

$$= \sum_{k=0}^{n} C_n^k (-1)^k a^k E(X^{n-k})$$

hay: $$\mu_n = \sum_{k=0}^{n} (-1)^k \gamma_{n-k} \gamma_1^k$$

Theo định nghĩa, ta có ngay:

$$\mu_o = 1,\ \mu_1 = 0,\ \mu_2 = D(X)$$

với mọi $n \geq 2$, do $\gamma_o = 1$ nên:

$$\mu_n = \sum_{k=0}^{n-2} (-1)^k \gamma_{n-k} \gamma_1^k + (-1)^{n-1}(n-1)\gamma_1^n$$

từ công thức này ta có:

$$\mu_2 = \gamma_2 - \gamma_1^2$$

$$\gamma_3 = \gamma_3 - 3\gamma_2\gamma_1 + 2\gamma_1^3$$

$$\mu_4 = \gamma_4 - 4\gamma_3\gamma_1 + 6\gamma_2\gamma_1^2 - 3\gamma_1^4$$

...

Ví dụ 3.9. Cho X có bảng phân phối xác suất:

X	–1	0	1	2
P	0,1	0,2	0,4	0,3

Tìm các mômen để cấp 4 của X.

Giải: $\gamma_o = 1;\ \gamma_1 = a = 0{,}9;\quad \gamma_2 = 1{,}7;\quad \gamma_3 = 2{,}7;\quad \gamma_4 = 6{,}5$

Từ đó: $\mu_2 = 1{,}7 - 0{,}9^2 = 0{,}89$

$$\mu_3 = 2{,}7 - 3.1{,}7.0{,}9 + 2.0{,}9^3 = -0{,}432$$

$$\mu_4 = 6{,}5 - 4.2{,}7.0{,}9 + 6.1{,}7.0{,}9^2 - 3.0{,}9^4 = 3{,}0737$$

Ví dụ 3.10. Cho X có hàm mật độ là

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

Tìm các mômen đến cấp 4 của X.

Giải. Có thể chứng minh được:

$$E(X^n)=\int_{-\infty}^{+\infty} x^n f(x)dx$$

do đó: $\gamma_o=1,\ \gamma_1=\frac{2}{3},\ \gamma_2=\frac{1}{2},\ \gamma_3=\frac{2}{5},\ \gamma_4=\frac{1}{3}$

Từ đó:

$$\mu_2=\frac{1}{2}-\left(\frac{2}{3}\right)^2=\frac{1}{18}$$

$$\mu_3=\frac{2}{5}-3.\frac{1}{2}.\frac{2}{3}+2.\left(\frac{2}{3}\right)^2=\frac{1}{135}$$

$$\mu_4=\frac{1}{3}-4.\frac{2}{5}.\frac{2}{3}+6.\frac{1}{2}.\left(\frac{2}{3}\right)^2-3.\left(\frac{2}{3}\right)^4=\frac{1}{135}$$

4- Hệ số bất đối xứng. Hệ số nhọn

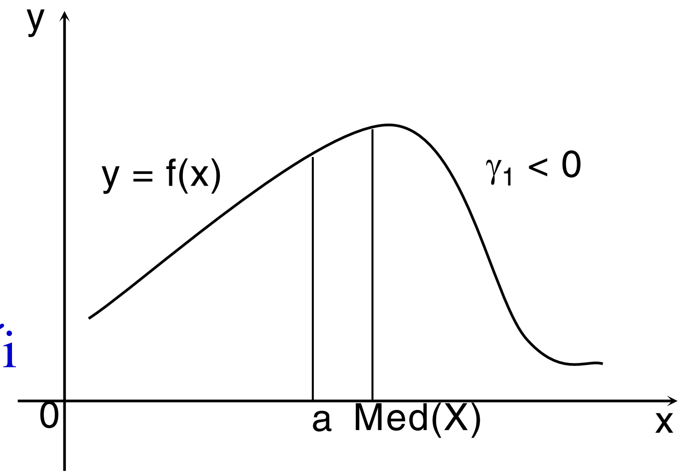
Với các ký hiệu như trong mục 3, ta gọi hệ số đối xứng của X là số

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\mu_3}{(\mu_2)^{3/2}}$$

Khi: $\gamma_1 = 0$ thì phân phối đối xứng

$\gamma_1 > 0$ thì phân phối lệch bên phải

$\gamma_1 < 0$ thì phân phối lệch bên trái



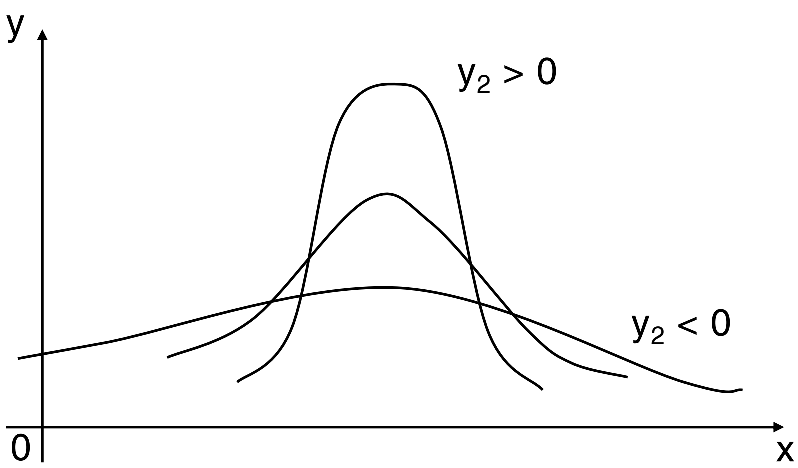
Trên hình vẽ: , a < med(X), phân phối lệch bên trái.

Ta gọi hệ số nhọn của X là:

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\mu_4}{\mu_2^2} - 3$$

Xét các phân phối có kỳ vọng a, phương sai σ^2. Phân phối có hàm mật độ:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$$



gọi là phân phối chuẩn. Với phân phối chuẩn ta có $\gamma_2 = 0$. So sánh đồ thị hàm mật độ các phân phối có kỳ vọng a, phương sai σ^2 ta thấy:

$\gamma_2 > 0$ có độ nhọn cao hơn phân phối chuẩn

$\gamma_2 < 0$ có độ nhọn thấp hơn phân phối chuẩn

3.4. ĐẶC TRƯNG CỦA VECTƠ NGẪU NHIÊN HAI CHIỀU VÀ NHIỀU CHIỀU

1. Kỳ vọng

Cho vectơ ngẫu nhiên hai chiều Z = (X,Y); ta gọi kỳ vọng của Z là vectơ:

E(Z) = (E(X), E(Y)) R2

2. Kỳ vọng của hàm một vectơ ngẫu nhiên

1- Trường hợp rời rạc

Giả sử (X,Y) có phân phối đồng thời phân phối (X = xi,Y = yj) = pij và $Z = \varphi(X, Y)$, khi đó ta có:

$$E(Z) = \sum_{i} z_i P(\varphi(X, Y) = z_i)$$

$$= \sum_{i=i}^{m} \sum_{j=i}^{n} \varphi(x_i, y_j) p_{ij}$$

2- Trường hợp liên tục

Giả sử (X,Y) có hàm mật độ đồng thời là f(x,y) và Z = . Khi đó ta có:

$$E(Z) = \iint_{(R^2)} \varphi(x, y) f(x, y) dxdy$$

3. Kỳ vọng có điều kiện

1- Trường hợp rời rạc

Với các ký hiệu như trong mục 2.3, ta có các kỳ vọng điều kiện:

$$E(X/Y=y_j) = \sum_{i=1}^{m} x_i \frac{p_{ij}}{m_j}, \ j=\overline{1,n}$$

$$E(Y/X=x_i) = \sum_{j=1}^{n} y_j \frac{p_{ij}}{p_i}, \ i=\overline{1,m}$$

2- Trường hợp liên tục

Với các ký hiệu như trong mục 2.3, ta có các kỳ vọng có điều kiện:

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

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

Ta cũng ký hiệu E(X/Y) là đại lượng ngẫu nhiên nhận giá trị E(X/y) khi Y = y và E(Y/X) là đại lượng ngẫu nhiên nhận giá trị E(Y/x) khi X = x.

Định lý 3.3. (Công thức kỳ vọng toàn phần):

$$E(E(X/Y)) = E(X)$$

Chứng minh. Trường hợp rời rạc:

$$E(E(X/Y)) = \sum_{j=1}^{n} E(X/Y=y_j)q_j$$

$$= \sum_{j=1}^{n}\sum_{i=1}^{m}\left(x_i.\frac{p_{ij}}{q_j}\right)q_j = \sum_{j=1}^{n}\sum_{i=1}^{m} x_i p_{ij} =$$

$$= \sum_{i=1}^{m} x_i\left(\sum_{j=1}^{n} p_{ij}\right) = \sum_{i=1}^{m} x_i p_i = E(X)$$

Trường hợp liên tục:

$$E(E(X/Y)) = \int_{-\infty}^{+\infty} E(X/y)f_Y(y)dy$$

$$= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} x\, f_{X/y}(x)dx\right).f_Y(y)dy$$

$$= \int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty} x.\frac{f(x,y)}{f_Y(y)}dx\right)f_Y(y)dy$$

$$= \int_{-\infty}^{+\infty} x\left(\int_{-\infty}^{+\infty} f(x,y)dy\right)dx$$

$$= \int_{-\infty}^{+\infty} xf_X(x)dx = E(X)$$

4. Covarian. Ma trận tương quan

Cho vectơ ngẫu nhiên Z = (X,Y). Khi đó ta gọi covarian của Z là:

$$\mathrm{cov}(X,Y) = E[(X - E(X))\,(Y - E(Y))]$$

Bằng biến đổi đơn giản ta có:

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

Khi tính covarian, cần chú ý rằng trong trường hợp rời rạc thì:

$$E(XY) = \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_{ij}$$

và trong trường hợp liên tục thì:

$$E(XY) = \iint_{(R^2)} xyf(x,y)\,dxdy$$

Từ định nghĩa ta có:

$$D(X) = \mathrm{cov}(X,X)$$

Ta gọi ma trận tương quan (hay ma trận hiệp phương sai) của (X,Y) là:

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

Ví dụ 3.11. Cho vectơ ngẫu nhiên Z = (X,Y) có hàm mật độ:

$$f(x,y) = \begin{cases} 8xy & \text{nếu } 0 \le x < y \le 1 \\ 0 & \text{nếu trái lại} \end{cases}$$

tìm cov(X,Y) và D(X,y).

Giải. Ta có:

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \begin{cases} 4x(1-x^2) & \text{nếu } x \in [0,1] \\ 0 & \text{nếu } x \notin [0,1] \end{cases}$$

$$f_Y(y) = \begin{cases} 4y^3 & \text{nếu } y \in [0,1] \\ 0 & \text{nếu } y \notin [0,1] \end{cases}$$

từ đó:

$$E(X) = \int_o^1 4x^2(1-x^2)dx = \frac{8}{15}$$

$$E(Y) = \int_o^1 4y^2 dy = \frac{4}{5}$$

$$E(X^2) = \int_o^1 4x^3(1-x^2)dx = \frac{1}{3}$$

$$E(Y^2) = \int_o^1 4y^5 dy = \frac{2}{3}$$

$$E(XY) = \iint\limits_{(R^2)} xyf(x,y)dxdy = \int_0^1 dy \int_x^1 8x^2y^2 dx = \frac{4}{9}$$

$$D(X) = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{125}$$

$$D(Y) = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75}$$

Vậy:

$$\text{cov}(X,Y) = \frac{4}{9} - \frac{8}{15}.\frac{4}{5} = \frac{4}{225}$$

và:

$$D(X,Y) = \begin{pmatrix} \frac{11}{225} & \frac{4}{225} \\ \frac{4}{225} & \frac{2}{75} \end{pmatrix}$$

5. Hệ số tương quan

Ta gọi số:

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

là hệ số tương quan giữa X và Y.

Định lý 3.4. Với mọi (X,Y), ta có:

(i) $|R_{XY}| \le 1$

(ii) RXY = ±1 nếu và chỉ nếu X và Y tương quan tuyến tính, tức là tồn tại các số A, B, C sao cho: AX + BY = C (h.k.n)

Chứng minh. Ta có:

$$D\left(\frac{X}{\sigma(X)} \pm \frac{Y}{\sigma(Y)}\right) = E\left(\frac{X}{\sigma(X)} \pm \frac{Y}{\sigma(Y)}\right)^2 - E^2\left(\frac{X}{\sigma(X)} \pm \frac{Y}{\sigma(Y)}\right)$$

$$= \frac{E(X^2)}{\sigma^2(X)} \pm \frac{2E(XY)}{\sigma(X)\sigma(Y)} + \frac{E(Y^2)}{\sigma^2(Y)} - \frac{E^2(X)}{\sigma^2(X)} \mp \frac{2E(X)E(Y)}{\sigma(X)\sigma(Y)} - \frac{E^2(Y)}{\sigma^2(Y)}$$

$$= 2 \pm 2\frac{E(XY) - E(X)E(Y)}{\sigma(X)\sigma(Y)}$$

Vì vậy với mọi X và Y khác hằng số ta có:

$$D\left(\frac{X}{\sigma(X)} \pm \frac{Y}{\sigma(Y)}\right) = 2(1 \pm R_{XY}) \qquad (*)$$

(i) Do $D\left(\frac{X}{\sigma(X)} \pm \frac{Y}{\sigma(Y)}\right) \ge 0$ nên theo (*)

$$1 \pm R_{XY} \geq 0 \Leftrightarrow -1 \leq R_{XY} \leq 1$$

(ii) $$R_{XY} = \pm 1 \Leftrightarrow D\left(\frac{X}{\sigma(X)} \mp \frac{Y}{\sigma(Y)}\right) = 0$$

$$\Leftrightarrow \frac{X}{\sigma(X)} \mp \frac{Y}{\sigma(Y)} = C\,(h.k.n)$$

Vậy $R_{XY} = \pm 1$ nếu và chỉ nếu X và Y tương quan tuyến tính.

Nhận xét. Nếu X và Y độc lập thì cov(X,Y) = 0, do đó Rxy = 0. Khi Rxy = 0 thì chưa chắc X và Y đã độc lập, trong trường hợp này ta nói Xvà Y không tương quan với nhau.

6. Vài đặc trưng của vectơ ngẫu nhiên nhiều chiều

Cho vectơ ngẫu nhiên n chiều:

Z = (X1, X2, ..., Xn)

1- Kỳ vọng

Ta gọi kỳ vọng của Z là vectơ n chiều:

E(Z) = (E(X1), E(X2), ..., E(Xn))

2- Mômen bậc k

Ta gọi mômen hỗn hợp bật k của (X1, X2, ..., E(xn)) đối với (a1,a2, ...,an) là số:

$$E[(X_1 - a_1)^{k_1}(x_2 - a_2)^{k_2} \ldots (x_n - a_n)^{k_n}]$$

trong đó: k1 + k2 + ... + kn = k.

nếu: (a1,a2, ...,an) = (0,0,...,0)

thì mômen gọi là mômen gốc.

nếu: (a1,a2, ...,an) = (E(X1),E(X2), ..., E(xn))

thì mômen gọi là mômen trung tâm.

Covarian của Xi, Yj chính là mômen hỗn hợp bậc

$$\alpha_{ij} = \mathrm{cov}(X_i, X_j) = E(X_i, X_j) - E(X_i), E(X_j)$$

2:*3- Ma trận tương quan*

Ta gọi ma trận tương quan hay ma trận hiệp phương sai của Z là:

$$D(X_1, X_2, \ldots, X_n) = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \ldots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \ldots & \alpha_{2n} \\ \alpha_{n1} & \alpha_{n2} & \ldots & \alpha_{nn} \end{pmatrix}$$

Ma trận tương quan là một ma trận đối xứng.

BÀI TẬP

3.1. Tính kỳ vọng E(Z1), E(Z2), phương sai D(Z1), D(Z2) trong bài 2.1.

3.2. Tính kỳ vọng E(X), phương sai D(X) của biến ngẫu nhiên X trong các bài 2.2, 2.3, 2.4, 2.5, 2.10.

3.3. Hàm mật độ phân phối của biến ngẫu nhiên X có dạng:

$$\varphi(x) = \begin{cases} \dfrac{1}{2e} & \text{nếu } |x - a| \le e \\ 0 & \text{nếu } |x - a| > e \end{cases}$$

Tìm kỳ vọng và phương sai của biến ngẫu nhiên X.

3.4. Tìm kỳ vọng và phương sai của biến ngẫu nhiên X tuân theo biến đổi Laplace, có hàm mật độ:

$$\varphi(x) = \frac{1}{2} e^{-|x|}$$

3.5. Một dụng cụ đo có sai số hệ thống 3m với độ lệch chuẩn $\delta = 20$ m. Tính sai số sao cho sai số của phép đo không vượt quá 5m về giá trị tuyệt đối.

3.6. Người ta tiện một loại chi tiết có độ dài quy định là a = 20 cm, biết độ lệch chuẩn là δ = 0,2 cm. Tính xác suất để kích thước của chi tiết sản xuất ra lệch so với kích thước quy định không quá ±0,3 cm.

3.7. Cho biến ngẫu nhiên X với hàm mật độ có dạng:

$$\varphi(x) = \begin{cases} ae^{-x^2} & \text{khi } x \geq 0 \\ 0 & \text{khi } x < 0 \end{cases}$$

a) Tính số a

b) Tính $P(-1<X<1)$

c) Tính E(X).

3.8. Cho biến ngẫu nhiên X có hàm mật độ:a)

$$\varphi(x) = \begin{cases} 6x(1-x) & \text{nếu } x \in [0,1] \\ 0 & \text{nếu } x \notin [0,1] \end{cases}$$

a) Tìm hàm mật độ của biến ngẫu nhiên Y = 2X–3

b) Tìm hàm mật độ của biến ngẫu nhiên Z = X2

c) Tính E(Y), E(Z), D(Y).

3.9. Cho biến ngẫu nhiên X có hàm mật độ:

$$\varphi(x) = \begin{cases} \dfrac{1}{2}\sin x & \text{nếu } x \in [0,\pi] \\ 0 & \text{nếu } x \notin [0,\pi] \end{cases}$$

Tìm kỳ vọng và phương sai của biến ngẫu nhiên Y = X2.

3.10. Cho véctơ ngẫu nhiên (X,Y) có hàm mật độ:

$$\varphi(x,y) = \begin{cases} a(x+y^2) & \text{khi } (x,y) \in [0,1]\times[0,1] \\ 0 & \text{ở những nơi khác} \end{cases}$$

a) Xác định a

b) Tính xác suất P(0,2 < X < 0,5; 0,4 < Y < 0,6)

c) Tính hệ số tương quan giữa X và Y.

3.11. Cho véctơ ngẫu nhiên (X,Y) có hàm mật độ đồng thời:

$$\varphi(x,y) = \begin{cases} axy & \text{khi } 0 < y < \dfrac{x}{2} < 1 \\ 0 & \text{ở những nơi khác} \end{cases}$$

a) Xác định a

b) Tính hệ số tương quan giữa X và Y

c) Tìm ma trận hiệp phương sai của (X,Y).

3.12. Cho véctơ ngẫu nhiên (X,Y) có hàm mật độ đồng thời:

$$\varphi(x,y) = \begin{cases} e^{-y} & \text{với} \quad 0 \le x < +\infty \quad x < y < +\infty \\ 0 & \text{vớ i} \quad \text{mọi} \quad (x,y) \quad \text{khác} \end{cases}$$

a) Tìm hàm phân phối đồng thời F(x,y) của (X,Y)

b) Tìm phương sai và ma trận hiệp phương sai của (X,Y).

3.13. Cho các biến ngẫu nhiên X, Y với:

$$E(X) = -2,\ E(Y) = 4$$

$$D(X) = 4,\ \ D(Y) = 9$$

Hệ số tương quan $r_{xy} = -\frac{1}{2}$

Tìm kỳ vọng của biến ngẫu nhiên Z = 3X2 – 2XY + Y2 – 3.

3.14. Giả sử X1, X2, ..., Xp; Y1, Y2, ..., Yq; Z1, Z2, ..., Zr là các biến ngẫu nhiên độc lập và có phương sai bằng 1. Chứng minh rằng hệ số tương quan của hai biến ngẫu nhiên

u = X1 + X2 +...+ Xp + Y1 + Y2 +...+ Yq :

và: $v = X1 + X2 + ... + Xp + Z1 + Z2 + ... + Zr$

là: $$p = \frac{p}{\sqrt{(p+q)(p+r)}}$$

3.15. Cho các biến ngẫu nhiên X1, X2, ..., Xm; Y1, Y2, ..., Yn có phương sai đều bằng 1:

$$cov(Xi,Xj) = p1; \quad cov(Yi,Yj) = p2; \quad cov(Xi,Yj) = p3$$

Tìm hệ số tương quan của hai biến ngẫu nhiên:

$$u = (X1 + X2 + ... + Xm) \text{ và } v = (Y1 + Y2 + ... + Yn)$$

3.16. Một hộp đựng ba bi đỏ, hai bi xanh và một bi vàng. Lấy ngẫu nhiên ra từng bi cho đến khi gặp bi đỏ thì dừng lại. Gọi X là biến ngẫu nhiên chỉ số bi xanh, Y là biến chỉ số bi vàng đã lấy ra.

a) Lập bảng phân phối xác suất đồng thời của X và Y

b) Tính hệ số tương quan giữa X và Y

c) Tính cov(X,Y)

d) Tìm D(X,Y).

3.17. Có ba đồng tiền gồm hai đồng công bằng, một đồng thiên vị (cả hai mặt đều ngửa). Tung cả ba đồng tiền đó, gọi X là số mặt sấp, Y là số mặt ngửa xuất hiện.

a) Lập bảng phân phối xác suất đồng thời của X và Y; X và Y có độc lập không?

b) Tính hệ số tương quan giữa X và Y

c) Tính cov(X,Y)

d) Tìm D(X,Y).

3.18. Cho (X,Y) có hàm mật độ đồng thời là:

$$f(x,y) = \begin{cases} 8xy & \text{nếu } 0 \le y \le x \le 1 \\ 0 & \text{nơi khác} \end{cases}$$

a) Tìm hàm mật độ có điều kiện $f_{x/_{y=1/2}}(x)$

b) Tìm kỳ vọng có điều kiện $E\left(X/Y = \frac{1}{2}\right)$

c) Tìm các xác suất có điều kiện

$$E\left(X/Y = \frac{1}{2}\right);\ P\left(X < \frac{1}{2}\Big/ Y < \frac{1}{2}\right);\ P\left(X < \frac{1}{2}\Big/ Y = \frac{1}{2}\right)$$