§2 CÁC SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN HAI CHIỀU

2.1 Các số đặc trưng biên

Ta đã biết các *đặc trưng cơ bản* của X và Y ở chương II. Chúng có thể tính được trực tiếp từ các khái niệm mới của PP đồng thời. Do tính tương tự ta viết chỉ viết các công thức cho biến X.

Nếu X rời rạc, *kỳ vọng biên* và *phương sai biên* là:

$$EX = \sum_i \sum_j x_i p\left(x_i, y_j\right) = \sum_i x_i\, p_1(x_i);$$

$$VX = \sum_i \sum_j (x_i - EX)^2 p\left(x_i, y_j\right) = \sum_i \sum_j {x_i}^2 p\left(x_i, y_j\right) - (EX)^2$$

$$= \sum_i (x_i - EX)^2\, p_1(x_i)$$

Nếu X liên tục:

$$EX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dx dy = \int_{-\infty}^{+\infty} x f_1(x) dx;$$

$$VX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - EX)^2 f(x, y) dx dy$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2 f(x, y) dx dy - (EX)^2$$

$$= \int_{-\infty}^{+\infty} (x - EX)^2 f_1(x) dx.$$

Ta có thể viết công thức tổng quát hơn, chẳng hạn nếu (X, Y) có

PP đã biết và ta cũng có một hàm $Z = g(X,Y)$, khi đó

$$E[g(X,Y)] = \sum_i \sum_j g(x_i, y_j) p\left(x_i, x_j\right) \text{ (biến rời rạc);}$$

hoặc $$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f(x,y) dxdy \text{ (biến liên tục).}$$

Dễ thấy VX chính là $E[g(X,Y)]$ với $g(X,Y) = (x - EX)^2$.

Ngoài ra ta còn có một số đặc trưng quan trọng là các *độ lệch chuẩn biên*:

$$\sigma_X = \sqrt{VX}, \sigma_Y = \sqrt{VY}.$$

2.2 Hiệp phương sai và hệ số tương quan

*** Định nghĩa 1**. *Hiệp phương sai* (*covariance*) *của hai biến* X *và* Y, ký hiệu là μ_{XY}, được xác định như sau

$$\mu_{XY} = E[(X - EX)(Y - EY)] = E(XY) - EX.EY.$$

Phụ thuộc vào (X,Y) rời rạc hay liên tục, ta có các công thức tính:

$$\mu_{XY} = \sum_i \sum_j x_i y_j p\left(x_i, y_j\right) - EX.EY;$$

$$\mu_{XY} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf(x,y) dxdy - EX.EY.$$

Dễ thấy $VX = \mu_{XX}$. Trong chừng mực nào đó, hiệp phương sai được dùng là độ đo quan hệ giữa hai biến: *nếu hiệp phương sai dương* ⇔ *hai biến có khuynh hướng đồng biến* và *nếu chúng có* *khuynh hướng nghịch biến* $\Leftrightarrow$ *hiệp phương sai âm*.

*** Định nghĩa 2**. Nếu $\mu_{XY} = 0$ ta nói rằng hai biến X và Y *không tương quan*.

Rõ ràng nếu X, Y độc lập $\Rightarrow$ chúng không tương quan; điều ngược lại nói chung không đúng.

Nhiều khi để đơn giản các ký hiệu, người ta tập hợp các hiệp phương sai và phương sai của một véc tơ ngẫu nhiên vào một ma trận được gọi là *ma trận hiệp phương sai* (*covariance matrix*); trong trường hợp biến 2 chiều $\boldsymbol{X} = (X, Y)$ đó là ma trận

$$\Gamma = \begin{bmatrix} VX & \mu_{XY} \\ \mu_{XY} & VY \end{bmatrix}.$$

Do hiệp phương sai có nhiều hạn chế (khó xác định miền biến thiên để so sánh quan hệ hai biến, thứ nguyên phức tạp,...), người ta đưa ra khái niệm dưới đây:

*** Định nghĩa 3**. *Hệ số tương quan của hai biến X, Y* (*correlation coefficient*), ký hiệu là ρ_{XY}, được xác định như sau

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y}.$$

Hệ số tương quan có tính chất quan trọng $|\rho_{XY}| \leq 1$ và:

- Nếu $\rho_{XY} = \pm 1 \Rightarrow \exists\, a, b: Y = aX + b$ (quan hệ tuyến tính);
- Nếu $\rho_{XY} = 0 \Leftrightarrow X, Y$ không tương quan;
- Nếu $0 < |\rho_{XY}| \leq 1 \Rightarrow X, Y$ phụ thuộc;
- Nếu $\rho_{XY} > 0$, ta có tương quan dương (khuynh đồng biến);
- Nếu $\rho_{XY} < 0$, tương quan âm (khuynh nghịch biến)…

*** <u>Thí dụ 1</u>**. Từ bảng PPXS đồng thời của X và Y

x \ y	1	2	3
1	0,10	0,25	0,15
2	0,15	0,00	0,35

hãy tính hiệp phương sai và hệ số tương quan của chúng.

Giải. Đầu tiên ta tính các đặc số biên: $EX = 1{,}5$; $VX = 0{,}25$;

$EY = 2{,}25$; $VY = 0{,}6845$; sau đó tính $E(XY) = 3{,}45$. Từ đó:

$$\mu_{XY} = E(XY) - EX.EY = 3{,}45 - 1{,}5.2{,}25 = 0{,}075;$$

$$\rho_{XY} = \frac{\mu_{XY}}{\sigma_X \sigma_Y} = \frac{0{,}075}{\sqrt{0{,}25.0{,}6845}} \approx 0{,}1813.$$

*** <u>Thí dụ 2</u>**. Cho biến ngẫu nhiên 2 chiều (X, Y) có mật độ đồng thời

$$f(x, y) = \begin{cases} \dfrac{1}{2\pi}, & 4x^2 + y^2 \leq 4; \\ 0, & 4x^2 + y^2 > 4. \end{cases}$$

Chứng tỏ X, Y phụ thuộc và tính hiệp phương sai μ_{XY}.

Giải. Trước hết ta tính các mật độ biên:

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, |x| \le 1; \\ 0, \quad |x| > 1. \end{cases}$$

$$f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \begin{cases} \frac{1}{2\pi}\sqrt{4-y^2}, |y| \le 2; \\ 0, \quad |y| > 2. \end{cases}$$

Vì $f(x,y) \neq f_1(x)\, f_2(y)$ nên X và Y phụ thuộc. Do $f_1(x), f_2(y)$ là các hàm chẵn nên $EX = EY = 0$, từ đó

$$\mu_{XY} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy$$

$$= \frac{1}{2\pi}\int_{-1}^{1} xdx \int_{-2\sqrt{1-x^2}}^{2\sqrt{1-x^2}} ydy = 0.$$

Rõ ràng X và Y không tương quan, nhưng vẫn phụ thuộc.

2.3 Các số đặc trưng có điều kiện

Dùng các khái niệm xác suất có điều kiện và mật độ có điều kiện ở tiết trước ta có thể định nghĩa các đặc trưng có điều kiện, chẳng hạn *kỳ vọng có điều kiện* của X với điều kiện $Y = y$:

$$E(X|y_k) = \sum_i x_i P(X = x_i | Y = y_k) \text{ (X rời rạc)},$$

$$E(X|y) = \int_{-\infty}^{+\infty} x\varphi(x|y)\, dx \text{ (X liên tục)}.$$

Tương tự có thể định nghĩa $E(Y|x)$ và các *phương sai có điều kiện* tương ứng, chẳng hạn:

$V(X|y_k) = \sum_i {x_i}^2 P(X = x_i|Y = y_k)$ – $[E(X|y_k)]^2$ (X rời rạc),

$V(X|y) = \int_{-\infty}^{+\infty} x^2 \varphi(x|y)\, dx$ – $[E(X|y)]^2$ (X liên tục).

Kỳ vọng có điều kiện $E(Y|x) = E(Y|X = x)$ là hàm của x, trong thống kê người ta gọi là *hàm hồi quy* của Y đối với X và đồ thị của hàm trên mặt phẳng toạ độ Đề-các được gọi là đường hồi quy.

Để ý kỳ vọng có điều kiện $E(Y|X)$, cũng như các đặc trưng có điều kiện khác, là biến ngẫu nhiên và đến lượt mình có thể có các đặc số tương ứng.

*** <u>Thí dụ 3</u>**. Cho bảng PP đồng thời của X và Y

$x \backslash y$	1	2	3
2	0,10	0,25	0,15
4	0,15	0,00	0,35

Tính các kỳ vọng và phương sai có điều kiện $E(X|1)$, $E(Y|4)$, $V(X|1)$, $V(Y|4)$.

Giải. Dễ thấy:

$$P(X = 2|Y = 1) = \frac{p_{11}}{p_2(1)} = \frac{0,10}{0,25} = 0,4;$$

$$P(X = 4|Y = 1) = \frac{p_{21}}{p_2(1)} = \frac{0,15}{0,25} = 0,6.$$

Từ đó $\quad E(X|1) = 2.0,4 + 4.0,6 = 3,2;$

$$V(X|1) = 2^2.0,4 + 4^2.0,6 - 3,2^2 = 0,96.$$

Tiếp theo

$$P(Y = 1|X = 4) = P(Y = 1|4) = \frac{p_{21}}{p_1(2)} = \frac{0,15}{0,5} = 0,3;$$

$$P(Y = 2|4) = \frac{p_{22}}{p_1(2)} = \frac{0}{0,5} = 0;$$

$$P(Y = 3|4) = \frac{p_{32}}{p_1(2)} = \frac{0,35}{0,5} = 0,7.$$

Và ta có: $\quad E(Y|4) = 1.0,3 + 3.0,7 = 2,4;$

$$V(Y|4) = 1.0,3 + 3^2.0,7 - 2,4^2 = 0,84.$$

*** <u>Thí dụ 4</u>**. Cho X và Y có hàm mật độ đồng thời

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

Tìm hàm mật độ có điều kiện $\psi(y|x)$ và tính $P(X^2 + Y^2 \le 1)$.

Giải. Đầu tiên ta tìm hàm mật độ biên

$$f_1(x) = \begin{cases} 1, & 0 < x \le 1; \\ 0, & \text{nếu trái lại.} \end{cases}$$

Từ đó

$$\psi(y|x) = \frac{f(x,y)}{f_1(x)} = \begin{cases} \frac{1}{x}, & 0 < y \le x \le 1; \\ 0, & \text{nếu trái lại;} \end{cases}$$

(PP có điều kiện là phân phối đều trên $(0, x)$).

Theo tính chất (iii) của hàm mật độ

$$P(X^2 + Y^2 \le 1) = \iint_{\mathcal{D}} \frac{dx}{x},$$

với $\mathcal{D}$ là miền $\{(x, y): 0 < y \le x \le 1,\ x^2 + y^2 \le 1\}$. Dùng toạ độ cực ta có

$$P(X^2 + Y^2 \le 1) = \int_0^1 dr \int_0^{\pi/4} \frac{r d\varphi}{r\, cos\varphi} = \ln\left(\left|tan\frac{3\pi}{8}\right|\right) = \ln\left(1 + \sqrt{2}\right).$$

*** <u>Thí dụ 5</u>**. Điểm ngẫu nhiên (X, Y) rơi đồng khả năng vào miền e-líp có trục đối xứng nằm trên trục toạ độ và độ dài các bán trục là a và b.

a) Xác định mật độ XS của từng toạ độ và mật độ XS có điều kiện.

b) Tính hiệp phương sai của X và Y.

Giải. a) Dễ thấy hàm mật độ đồng thời của X và Y

$$f(x, y) = \begin{cases} \frac{1}{\pi ab}, & \frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1; \\ 0, & \frac{x^2}{a^2} + \frac{y^2}{b^2} > 1. \end{cases}$$

Rõ ràng $f_1(x) \neq 0$ chỉ nếu $|x| \le a$, khi đó

$$f_1(x) = \frac{2\sqrt{a^2-x^2}}{\pi a^2}, |x| \leq a; \text{tương tự } f_2(y) = \frac{2\sqrt{b^2-y^2}}{\pi b^2}, |y| \leq b.$$

Từ đó:

$$\varphi(x|y) = \frac{f(x,y)}{f_2(y)} = \begin{cases} \dfrac{b}{2a\sqrt{b^2-y^2}}, & \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \leq 1; \\ 0, & \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} > 1. \end{cases}$$

$$\psi(y|x) = \frac{f(x,y)}{f_1(x)} = \begin{cases} \dfrac{a}{2b\sqrt{a^2-x^2}}, & \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \leq 1; \\ 0, & \dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} > 1. \end{cases}$$

b) Do miền lấy tích phân đối xứng nên $EX = EY = 0$, nên

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

với để ý hàm $f(x,y)$ chỉ khác 0 trong e-líp $\dfrac{x^2}{a^2} + \dfrac{y^2}{b^2} \leq 1$. Để tính tích phân ta đổi biến sang toạ độ cực suy rộng $\dfrac{1}{\pi ab}$

$$x = arcos\varphi, y = brsin\varphi,$$

$$\mu_{XY} = \int_0^{2\pi}\int_0^1 abr^2\,cos\varphi sin\varphi\,\frac{1}{\pi ab}\,abr\,drd\varphi = 0.$$

Như vậy hai biến X và Y không tương quan, nhưng chúng phụ thuộc vì $f(x,y) \neq f_1(x)\,f_2(y)$.

*** Tính chất của kỳ vọng có điều kiện**

(i) Với mọi hàm $g(.)$ liên tục $E[g(X)Y|X] \ = \ g(X)E(Y|X)$;

(ii) $E(X_1 + X_2|X) = E(X_1|X) + E(X_2|X)$;

(iii) Nếu X, Y độc lập $E(Y|X) = EY$;

(iv) $E[E(Y|X)] = EY$.

2.4 Phân phối chuẩn hai chiều

Ta dùng các ký hiệu rút gọn:

$$\mu_X = EX, \mu_Y = EY, \sigma_X^2 = VX, \sigma_Y^2 = VY, \rho = \rho_{XY}.$$

Từ đó có thể xác định *hàm mật độ chuẩn đồng thời* của hai biến X và Y, ký hiệu là $\mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$, có dạng

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \times$$

$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho\frac{x-\mu_X}{\sigma_X}.\frac{y-\mu_Y}{\sigma_Y}\right]\right\}.$$

Để ý $X \sim\mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim\mathcal{N}(\mu_Y, \sigma_Y^2)$ và nếu X và Y không tương quan ($\rho = 0$), thì giả thiết chuẩn cho phép kết luận chúng độc lập (vì có thể chứng minh dễ dàng $f(x, y) = f_1(x)f_2(y)$).

Sử dụng khái niệm véc tơ – ma trận

$$\boldsymbol{x} = \binom{x}{y}, \boldsymbol{\mu} = \binom{\mu_X}{\mu_Y}, \boldsymbol{\Gamma} = \begin{pmatrix} \sigma_X^2 & \rho \\ \rho & \sigma_Y^2 \end{pmatrix},$$

ta có biểu diễn

$$f(\boldsymbol{x}) = f(x,y) = \frac{1}{2\pi\sqrt{det\,\boldsymbol{\Gamma}}} \exp\left\{-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

Ở đây det là ký hiệu định thức ma trận, T – phép chuyển vị. Biến chuẩn 2 chiều trong trường hợp này ký hiệu $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$.

Trường hợp tổng quát nếu $\boldsymbol{X}$ là véc tơ n chiều và $\sim\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Gamma})$, khi đó hàm mật độ có dạng

$$f(\boldsymbol{x}) = (2\pi)^{-n/2}\,(det\,\boldsymbol{\Gamma})^{-1/2} \exp\left\{-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Gamma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

*** <u>Thí dụ 6</u>**. Cho biến 2 chiều $(X,Y) \sim \mathcal{N}(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho)$. Tính các kỳ vọng và phương sai có điều kiện.

Giải. Dễ dàng tìm được $f_2(y)$ từ mật độ đồng thời ở trên

$$f_2(y) = \frac{1}{\sigma_Y\sqrt{2\pi}}\, e^{-\frac{1}{2\sigma_Y^2}(x-\mu_Y)^2},$$

từ đó

$$\varphi(x|y) = \frac{f(x,y)}{f_2(y)}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_X\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_X^2(1-\rho^2)}\left[x-\mu_X-\rho\frac{\sigma_X}{\sigma_Y}(y-\mu_Y)\right]^2\right\}.$$

Biểu thức trên chính là hàm mật độ của phân phối chuẩn

$$\mathcal{N}\left(\mu_X + \rho\frac{\sigma_X}{\sigma_Y}(y - \mu_Y);\ \sigma_X^2(1 - \rho^2)\right),$$

từ đó:

$$E(X|y) = \mu_X + \rho\ \frac{\sigma_X}{\sigma_Y}\ (y - \mu_Y); V(X|y) = \sigma_X^2(1 - \rho^2).$$

Tương tự:

$$E(Y|x) = \mu_Y + \rho\ \frac{\sigma_Y}{\sigma_X}\ (x - \mu_X); V(Y|x) = \sigma_Y^2(1 - \rho^2).$$

BÀI TẬP

1. Cho X và Y là toạ độ của một điểm ngẫu nhiên có PP đều trong hình chữ nhật giới hạn bởi các đường $x = a, x = b, y = c, y = d$ $(b > a, d > c)$. Tìm hàm mật độ đồng thời của X và Y, sau đó tính kỳ vọng của các toạ độ của điểm ngẫu nhiên đó.

2. Cho X và Y có bảng PPXS

x \ y	1	2	3	4
1	0,16	0,08	0,10	0,20
2	0,08	0,10	0,14	0,14

   Tìm các kỳ vọng biên, kỳ vọng có điều kiện và các phương sai tương ứng.

3. Cho X và Y có hàm mật độ đồng thời

$$f(x,y) = \frac{A}{\pi^2(16+x^2)(25+y^2)}.$$

   a) Xác định A.
   b) X và Y có độc lập không?
   c) Tìm các đặc trưng biên của X và Y.

4. Xác định XS rơi của điểm (X, Y) vào miền $\{1 \leq x \leq 2, 1 \leq y \leq 2\}$, nếu hàm PP của (X, Y) có dạng $(a > 0)$

$$F(x,y) = \begin{cases} 1 - a^{-x^2} - a^{-2y^2} + a^{-(x^2+2y^2)}, & x \geq 0, y \geq 0; \\ 0, & x < 0 \text{ hay } y < 0. \end{cases}$$

   Sau đó tính các đặc trưng có điều kiện.

5. Bảng PPXS đồng thời của số lỗi vẽ màu X và số lỗi đúc Y của một loại sản phẩm nhựa ở một công ty được cho bởi

$x \backslash y$	0	1	2
0	0,59	0,06	0,03
1	0,10	0,05	0,01
2	0,06	0,05	0,01
3	0,02	0,02	0,00

Tính các kỳ vọng biên và ma trận tương quan của (X, Y).

6. Cho luật PP của một biến 2 chiều như sau (X, Y)

$x \backslash y$	2	3	5
1	0,1	0	0,1
4	0,2	0,5	0,1

Tìm luật PPXS của các hàm $X+Y$ và XY, sau đó tính các kỳ vọng và phương sai.

7. Biến ngẫu nhiên 2 chiều (X, Y) có PP chuẩn trong một hình vuông có cạnh a. Các đường chéo của hình vuông trùng với các trục toạ độ.
a) Xác định mật độ của (X, Y).
b) Tính các mật độ biên và mật độ có điều kiện.
c) Tính ma trận tương quan của (X, Y).
d) X và Y có phụ thuộc không?

8. Biến ngẫu nhiên 2 chiều có hàm mật độ đồng thời

$$f(x, y) = a(x^2 + y^2), \text{ nếu } x^2 + y^2 \leq 4.$$

Tìm hệ số a, sau đó tìm các đặc trưng biên và đặc trưng có điều kiện.

9. Cho hàm mật độ đồng thời của X và Y

$$f(x, y) = a\, e^{-\frac{1}{2}(x^2+2xy+5y^2)}.$$

Xác định hằng số a, sau đó tìm các đặc trưng biên và đặc trưng có điều kiện.

10. Cho X và Y là hai biến ngẫu nhiên có hàm mật độ đồng thời $f(x,y) = 1/2x^2y$, nếu $x \geq 1$ và $x \geq y \geq 1/x$. Tìm các kỳ vọng có điều kiện.

11. Cho X và Y là hai biến ngẫu nhiên: Y có phân phối đều trong (0;10); còn hàm mật độ có điều kiện $\varphi(x|y) = 1/y$, với $0 < x < y < 10$. Tính
    a) $E(X|Y = y)$; b) EX.