XÁC SUẤT THỐNG KÊ VÀ ỨNG DỤNG

CÁC THAM SỐ ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Buổi học 4

CÁC THAM SỐ ĐẶC TRƯNG CỦA CÁC ĐẠI LƯỢNG NGẪU NHIÊN

1. KỲ VỌNG
2. HÀM CỦA CÁC ĐLNN
3. MỘT VÀI ĐỊNH LÝ VỀ KỲ VỌNG
4. PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN
5. MỘT VÀI ĐỊNH LÝ VỀ PHƯƠNG SAI
6. CHUẨN HÓA CÁC ĐLNN
7. HIỆP PHƯƠNG SAI–HỆ SỐ TƯƠNG QUAN
8. CÁC ĐỘ ĐO HƯỚNG TÂM KHÁC

KỲ VỌNG

Một khái niệm quan trọng trong xác suất và thống kê là *kỳ vọng toán, giá trị kỳ vọng* hay nói gọn là *kỳ vọng* của một ĐLNN.

1. TRƯỜNG HỢP RỜI RẠC. ĐLNN rời rạc X= $\{x_1,\ldots,x_n\}$ kỳ vọng của X được định nghĩa

$$EX = x_1P(X=x_1)+\cdots+x_nP(X=x_n)=\sum_{j=1}^{n}x_jP(X=x_j) \quad (1)$$

- **Nếu $P(X=x_j)=f(x_j)$,**

$$EX = x_1f(x_1)+\cdots+x_nf(x_n)=\sum_{j=1}^{n}x_jf(x_j)=\sum xf(x) \quad (2)$$

- **Nếu các xác suất trong (2) đều bằng nhau thì**

$$E(X)=\frac{x_1+x_2+\cdots+x_n}{n} \quad (3)$$

là *trung bình số học* của $x_1, x_2, \ldots, x_n$.

KỲ VỌNG

2. TRƯỜNG HỢP LIÊN TỤC. Với đại lượng ngẫu nhiên liên tục X có hàm mật độ *f(x)* thì kỳ vọng của X được định nghĩa

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \qquad (4)$$

Biểu thức (4) là tích phân hội tụ tuyệt đối.

- Kỳ vọng của X thường được gọi là *trung bình* của X, ký hiệu là μ_X, hay đơn giản là μ.
- Kỳ vọng của ĐLNN X là trung bình của các giá trị của ĐLNN X. Chính vì lý do đó nên kỳ vọng thường được gọi là *độ đo hướng tâm*.

KỲ VỌNG

- **VÍ DỤ 3.1.** Giả sử có trò chơi tung một con xúc xắc công bằng. Trong trò chơi này người chơi sẽ thắng \$20 nếu mặt 2 xuất hiện, thắng \$40 nếu mặt 4 xuất hiện và thua \$30 nếu mặt 6 xuất hiện; người chơi không thắng và cũng không thua nếu các mặt khác xuất hiện. Hãy tìm tổng số tiền mà người chơi hy vọng sẽ được trong trò chơi này.

KỲ VỌNG

- Gọi X là số tiền mà người chơi sẽ được trong một lần tung bất kỳ ứng với con xúc xắc xuất hiện các mặt 1, 2,..., 6

Mặt	1	2	3	4	5	6
X	0	+20	0	+40	0	-30
$f(x_j)$	1/6	1/6	1/6	1/6	1/6	1/6

- Ta có luật phân phối xác suất của X là

X	-30	0	+20	+40
$f(x_j)$	1/6	½	1/6	1/6

$$E(X)=(-30)\frac{1}{6}+(0)\frac{1}{2}+(20)\frac{1}{6}+(40)\frac{1}{6}=5$$

- Người chơi hy vọng thắng $5. Vậy, để trò chơi công bằng thì người chơi phải trả $5.

KỲ VỌNG

- **VÍ DỤ 3.2.** Hàm mật độ của ĐLNN X là

$$f(x)=\begin{cases}\frac{1}{2}x & 0<x<2\\ 0 & ngược\ lại\end{cases}$$

Tìm kỳ vọng của ĐLNN X

- Kỳ vọng của X được tính theo công thức

$$E(X)=\int_{-\infty}^{\infty} xf(x)dx=\int_0^2 x\left(\frac{1}{2}x\right)dx$$

$$=\int_0^2\left(\frac{x^2}{2}\right)dx=\frac{x^3}{6}\bigg|_0^2=\frac{4}{3}$$

HÀM CỦA CÁC ĐLNN

1. TRƯỜNG HỢP RỜI RẠC. X là ĐLNN rời rạc có hàm xác suất là *f(x)*, khi đó Y=g(X) cũng là ĐLNN rời rạc và hàm xác suất của Y là

$$h(y)=P(Y=y)=\sum_{\{x|g(x)=y\}} P(X=x)=\sum_{\{x|g(x)=y\}} f(x)$$

- $X=\{x_1, x_2, ..., x_n\}$ và $Y=\{y_1, y_2, ..., y_m\}$, $(m \le n)$, $y_1h(y_1) + y_2h(y_2) + ... + y_mh(y_m) = g(x_1)f(x_1) + g(x_2)f(x_2) + ... + g(x_n)f(x_n)$.
- khi đó kỳ vọng của ĐLNN Y là

$$E[g(X)]=g(x_1)f(x_1)+g(x_2)f(x_2)+\dots+g(x_n)f(x_n)$$

$$=\sum_{j=1}^{n} g(x_j)f(x_j)=\sum g(x)f(x) \qquad (5)$$

HÀM CỦA CÁC ĐLNN

2. **TRƯỜNG HỢP LIÊN TỤC.** Nếu X là ĐLNN liên tục có hàm mật độ xác suất *f(x)* thì kỳ vọng của ĐLNN Y là

$$E\left[g(X)\right]=\int_{-\infty}^{\infty} g(x)f(x)dx \qquad (6)$$

❑ Tổng quát, ta dễ dàng xây dựng cho hàm hai biến hoặc nhiều biến ngẫu nhiên. Ví dụ, nếu X và Y là hai ĐLNN liên tục có hàm mật độ đồng thời *f(x,y)* thì kỳ vọng của *g(X,Y)* được cho bởi biểu thức

$$E[g(X,Y)]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \qquad (7)$$

KỲ VỌNG HÀM CỦA CÁC ĐLNN

- **VÍ DỤ 3.3.** Hàm mật độ của ĐLNN X là

$$f(x)=\begin{cases}\dfrac{1}{2}x & 0<x<2\\ 0 & \text{ngược lại}\end{cases}$$

Tìm kỳ vọng của ĐLNN Y = $3X^2 - 2X$.

- Kỳ vọng của Y được tính theo công thức

$$E(Y)=E\left(3X^2-2X\right)=\int_{-\infty}^{\infty}\left(3x^2-2x\right)f(x)\,dx$$

$$=\int_0^2\left(3x^2-2x\right)\left(\frac{1}{2}x\right)dx=\frac{10}{3}$$

MỘT VÀI ĐỊNH LÝ VỀ KỲ VỌNG

- ***Định lý 3-1:***

  Nếu c là một hằng số bất kỳ thì $E(c) = c$ và

  $E(cX) = c.E(X)$ (8)

- ***Định lý 3-2:***

  Nếu X và Y là hai ĐLNN bất kỳ, thì

  $E(X + Y) = E(X) + E(Y)$ (9)

- ***Định lý 3-3:***

  Nếu X và Y là hai ĐLNN độc lập, thì

  $E(XY) = E(X)E(Y)$ (10)

- Dễ dàng tổng quát cho các định lý này.

PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

Ngoài kỳ vọng của ĐLNN X, một tham số quan trọng khác trong xác suất - thống kê là *phương sai*, được định nghĩa bởi biểu thức

$$VarX = E(X - \mu)^2 \qquad (11)$$

- Phương sai là một giá trị không âm. Căn bậc hai của phương sai được gọi là độ lệch chuẩn, được cho bởi biểu thức

$$\sigma_X = \sqrt{Var(X)} = \sqrt{E[(X - \mu)^2]} \qquad (12)$$

- Độ lệch chuẩn, ký hiệu là σ thay vì σ_X và phương sai trong trường hợp như thế là σ^2.

PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

1. **TRƯỜNG HỢP RỜI RẠC.** ĐLNN rời rạc X = {x_1, x_2,..., x_n} và có hàm xác suất là *f(x)* thì phương sai được cho bởi biểu thức sau

$$\sigma_X^2 = E[(X-\mu)^2] = \sum_{j=1}^{n}(x_j-\mu)^2 f(x_j) = \sum (x-\mu)^2 f(x) \qquad (13)$$

- Trường hợp đặc biệt, nếu các xác suất đều bằng nhau thì biểu thức (13) trở thành

$$\sigma^2 = [(x_1-\mu)^2 + (x_2-\mu)^2 + \ldots + (x_n-\mu)^2]/n \qquad (14)$$

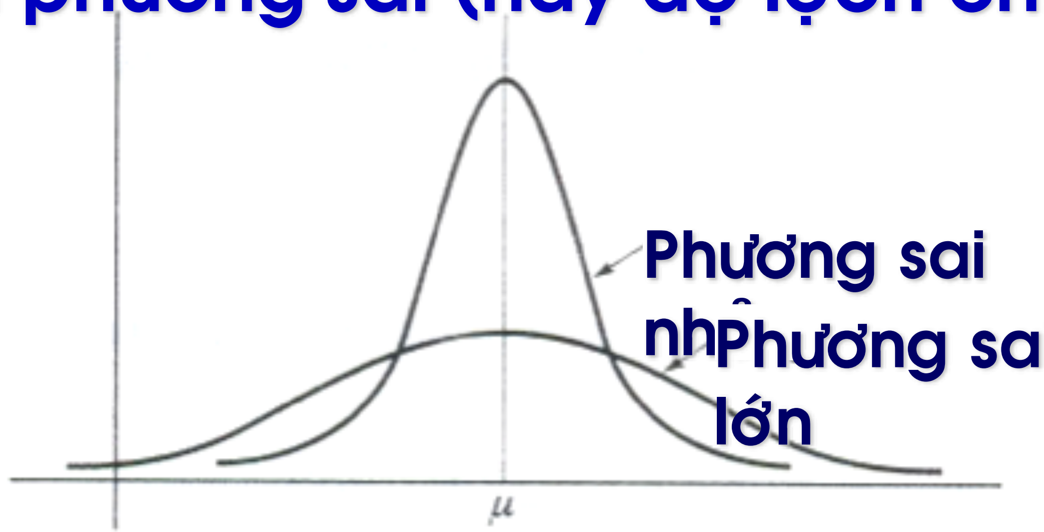
- Nếu X nhận một dãy các giá trị vô hạn x_1, x_2, ... thì $\sigma_X^2 = \sum_{j=1}^{\infty}(x_j-\mu)^2 f(x_j)$ là một chuỗi hội tụ.

PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

2. TRƯỜNG HỢP LIÊN TỤC. Nếu X là một ĐLNN liên tục có hàm mật độ *f(x)* thì phương sai được cho bởi biểu thức sau

$$\sigma_X^2 = E[(X-\mu)^2] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \qquad (15)$$

- Ý nghĩa phương sai (hay độ lệch chuẩn).



PHƯƠNG SAI VÀ ĐỘ LỆCH CHUẨN

- **VÍ DỤ 3.4.** Hàm mật độ của ĐLNN X là

$$f(x) = \begin{cases} \frac{1}{2}x & 0 < x < 2 \\ 0 & ngủợc\ lại \end{cases}$$

Tìm phương sai và độ lệch chuẩn của X

- Theo ví dụ 3.2., ta có μ = E(X) = 4/3. Khi đó phương sai được cho bởi biểu thức

$$\sigma^2 = E[(X - \frac{4}{3})^2] = \int_{-\infty}^{\infty} (x - \frac{4}{3})^2 f(x)dx = \int_0^2 (x - \frac{4}{3})^2 (\frac{1}{2}x)dx = \frac{2}{9}$$

và độ lệch chuẩn là $\sigma = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3}$

- Chú ý, nếu X có *đơn vị đo lường* là cm thì VarX có đơn vị là cm² và độ lệch chuẩn có cùng đơn vị với X là cm.

MỘT VÀI ĐỊNH LÝ VỀ PHƯƠNG SAI

- ***Định lý 3-4:***

$$\sigma^2 = E(X-\mu)^2 = E(X^2) - \mu^2 = E(X^2) - (E(X))^2 \quad (16)$$

- ***Định lý 3-5:*** Nếu c là một hằng số bất kỳ

$$Var(cX) = c^2 Var(X) \quad (17)$$

- ***Định lý 3-6:*** $E(X-a)^2$ nhỏ nhất khi $a = E(X)$.
- ***Định lý 3-7:*** X và Y là hai ĐLNN độc lập

$$Var(X+Y) = Var(X) + Var(Y) \quad (18)$$

$$Var(X-Y) = Var(X) + Var(Y) \quad (19)$$

Dễ dàng tổng quát định lý 3-7 cho trường hợp nhiều hơn hai biến ngẫu nhiên độc lập.

CHUẨN HÓA MỘT ĐLNN

❑ X là một ĐLNN có trung bình μ và độ lệch chuẩn σ ($\sigma > 0$). Ta có thể xác định *ĐLNN được chuẩn hóa* bằng biểu thức

$$X^* = \frac{X - \mu}{\sigma} \qquad (20)$$

❑ Một tính chất quan trọng của X^* là nó có trung bình 0 và phương sai 1, nghĩa là

$$E(X^*) = 0, \qquad Var(X^*) = 1 \qquad (21)$$

khi đó ĐLNN X được biểu diễn *chuẩn tắc*.

(nghĩa là σ là thành phần đơn vị trong $X - \mu$).

❑ ĐLNN được chuẩn hóa rất hữu ích trong việc so sánh các luật phân phối khác nhau.

HIỆP PHƯƠNG SAI

- **TRƯỜNG HỢP LIÊN TỤC.** Cho X và Y là hai ĐLNN liên tục có hàm mật độ đồng thời *f(x,y)*, kỳ vọng của X và Y là

$$\mu_X = E(X) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)dxdy;\ \mu_Y = E(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)dxdy \quad (22)$$

và phương sai là

$$\sigma_X^2 = E[(X-\mu_X)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)^2 f(x,y)dxdy \quad (23)$$

$$\sigma_Y^2 = E[(Y-\mu_Y)^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y-\mu_Y)^2 f(x,y)dxdy$$

- Hai ĐLNN X và Y có *hiệp phương sai* được định nghĩa như sau

$$\sigma_{XY} = Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \quad (24)$$

HIỆP PHƯƠNG SAI

- Dưới dạng hàm mật độ đồng thời f(x,y), ta có

$$\sigma_{XY} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f(x,y)dxdy \quad (25)$$

2. TRƯỜNG HỢP RỜI RẠC. Trong trường hợp này biểu thức (43) và (46) được thay bởi

$$\mu_X = \sum_x\sum_y xf(x,y) \qquad \mu_Y = \sum_x\sum_y yf(x,y) \quad (26)$$

- *Hiệp phương sai* của X và Y như sau

$$\sigma_{XY} = \sum_x\sum_y (x-\mu_X)(y-\mu_Y)f(x,y) \quad (27)$$

CÁC ĐỊNH LÝ VỀ HIỆP PHƯƠNG SAI

- ***Định lý 3-14:***

$$\sigma_{XY} = E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y \quad (28)$$

- ***Định lý 3-15:*** **X và Y là hai ĐLNN độc lập,**

$$\sigma_{XY} = Cov(X,Y) = 0 \quad (29)$$

- ***Định lý 3-16:***

$$Var(X \pm Y) = Var(X) + Var(Y) \pm 2Cov(X,Y) \quad (30)$$

hay

$$\sigma^2_{X \pm Y} = \sigma^2_X + \sigma^2_Y \pm 2\sigma_{XY} \quad (31)$$

- ***Định lý 3-17:*** $\left|\sigma_{XY}\right| \le \sigma_X \sigma_Y$ **(32)**

- **Điều ngược lại của định lý 3-15 không đúng. Nếu X và Y độc lập thì định lý 3-16 trở thành định lý 3-7.**

HỆ SỐ TƯƠNG QUAN

- Nếu X và Y độc lập thì cov(X,Y) = σ_{XY} = 0. Nếu X và Y phụ thuộc hoàn toàn, khi đó X = Y, thì cov(X,Y) = σ_{XY} = $\sigma_X\sigma_Y$. Như vậy, ta có *độ đo phụ thuộc* của ĐLNN X và Y như sau

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \qquad (33)$$

- Ta gọi ρ là *hệ số tương quan*. Từ định lý 3-17 ta thấy rằng –1 ≤ ρ ≤ 1. Trong trường hợp *ρ* = 0 (nghĩa là hiệp phương sai bằng 0), ta nói 2 ĐLNN X và Y *không liên quan*. Tuy nhiên, trong trường hợp này các ĐLNN có thể độc lập mà cũng có thể không độc lập.

CÁC ĐỘ ĐO HƯỚNG TÂM KHÁC

1. **MODE.** *mode* của ĐLNN rời rạc X là giá trị $x \in X$ mà tại đó có xác suất xảy ra lớn nhất. Mode của ĐLNN liên tục X là giá trị $x \in X$ mà tại đó hàm mật độ xác suất có giá trị lớn nhất. Đôi khi ta có hai, ba, hay nhiều giá trị *mode*.

2. **TRUNG VỊ.** Số *trung vị* là giá trị x mà tại đó $P(X< x) \leq ½$ và $P(X> x) \leq ½$. Trường hợp luật phân phối liên tục thì $P(X< x) = ½ = P(X> x)$ số trung vị tách đường cong hàm mật độ thành hai phần riêng biệt có cùng diện tích là ½. Trường hợp luật phân phối rời rạc, số trung vị có thể không tồn tại.

THE END