HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

o0o

Chương 3: Véc tơ ngẫu nhiên

XÁC SUẤT THỐNG KÊ

Bộ môn Toán, Khoa Cơ bản 1

Hà Nội - 2023

Chương 3: Véc tơ ngẫu nhiên

3.1.1 Khái niệm véc tơ ngẫu nhiên

Định nghĩa

- Một véc tơ ngẫu nhiên n chiều hay một biến ngẫu nhiên n chiều là một bộ có thứ tự $(X_1, X_2, \ldots, X_n)$, trong đó các thành phần $X_1, X_2, \ldots, X_n$ là các biến ngẫu nhiên được xét một cách đồng thời.
- Véc tơ ngẫu nhiên n chiều được gọi là rời rạc nếu các thành phần của nó là các biến ngẫu nhiên rời rạc và được gọi là liên tục nếu các thành phần của nó là các biến ngẫu nhiên liên tục.

3.1.2 Hàm phân bố xác suất

Định nghĩa

Hàm phân bố xác suất của véc tơ ngẫu nhiên n chiều $(X_1, X_2, \ldots, X_n)$ được xác định như sau: $\forall x_1, x_2, \ldots, x_n \in \mathbb{R}$,

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots, X_n \leq x_n),$$

Ta cũng gọi $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n)$ là hàm phân bố xác suất đồng thời của các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$.

Tính chất của hàm phân bố xác suất đồng thời:

1) $0 \le F_{X_1,X_2,\dots,X_n}(x_1,x_2,\dots,x_n) \le 1,\ \forall x_1,x_2,\dots,x_n \in \mathbb{R}$.
2) $\lim\limits_{x_k \to -\infty} F_{X_1,X_2,\dots,X_n}(x_1,x_2,\dots,x_n) = 0,\ k = 1,2,\dots,n$.
3) $\lim\limits_{x_1,x_2,\dots,x_n \to \infty} F_{X_1,X_2,\dots,X_n}(x_1,x_2,\dots,x_n) = 1$.
4) Hàm phân bố xác suất đồng thời là hàm không giảm theo từng biến.
5) $P(x_1 < X \le x_2, y_1 < Y \le y_2) =$
$F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1)$
6) $\lim\limits_{y \to \infty} F_{X,Y}(x,y) = F_X(x);\ \lim\limits_{x \to \infty} F_{X,Y}(x,y) = F_Y(y)$
$F_X(x), F_Y(y)$ được gọi là hàm phân bố xác suất biên của thành phần X và Y tương ứng.
7) X, Y là hai biến ngẫu nhiên độc lập khi và chỉ khi

$$F_{X,Y}(x,y) = F_X(x)F_Y(y), \forall x, y \in \mathbb{R}.$$

Ví dụ 1: Cho véc tơ ngẫu nhiên (X, Y) có hàm phân bố xác suất

$$F_{X,Y}(x,y) = \begin{cases} (1-e^{-x})(1-e^{-y}) & \text{nếu } x \geq 0, y \geq 0 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a) Tìm hai hàm phân bố xác suất biên $F_X(x), F_Y(y)$.

b) Tính $P(X > 1, Y > 2)$.

Chương 3: Véc tơ ngẫu nhiên

3.2.1 Hàm khối lượng xác suất đồng thời và bảng phân bố xác suất đồng thời

Cho X, Y là hai biến ngẫu nhiên rời rạc có

$$R_X = \{x_1, \dots, x_n\}, R_Y = \{y_1, \dots, y_m\}.$$

Định nghĩa

Hàm khối lượng xác suất đồng thời của các biến ngẫu nhiên rời rạc X, Y được xác định bởi

$$p_{X,Y}(x, y) = P(X = x, Y = y),\ \forall x, y \in \mathbb{R}.$$

Tính chất của hàm khối lượng xác suất đồng thời:

1) $p_{X,Y}(x_i, y_j) \geq 0, \ \forall i = 1, \ldots, n; \forall j = 1, \ldots, m.$
2) $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{m} p_{X,Y}(x_i, y_j) = 1$
3) $\displaystyle F_{X,Y}(x, y) = \sum_{x_i \leq x}\sum_{y_j \leq y} p_{X,Y}(x_i, y_j)$

Bảng phân bố xác suất đồng thời của X, Y là bảng có dạng

$X \backslash Y$	y_1	y_2	$\dots$	y_m
x_1	$p_{X,Y}(x_1, y_1)$	$p_{X,Y}(x_1, y_2)$	$\dots$	$p_{X,Y}(x_1, y_m)$
x_2	$p_{X,Y}(x_2, y_1)$	$p_{X,Y}(x_2, y_2)$	$\dots$	$p_{X,Y}(x_2, y_m)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_n	$p_{X,Y}(x_n, y_1)$	$p_{X,Y}(x_n, y_2)$	$\dots$	$p_{X,Y}(x_n, y_m)$

- Bảng phân bố xác suất của X là

X	x_1	x_2	$\dots$	x_n
P	$p_X(x_1)$	$p_X(x_2)$	$\dots$	$p_X(x_n)$

trong đó

$$p_X(x_i) = P(X = x_i) = \sum_{j=1}^{m} p_{X,Y}(x_i, y_j), i = 1, \dots, n.$$

- Bảng phân bố xác suất của Y là

Y	y_1	y_2	$\dots$	y_m
P	$p_Y(y_1)$	$p_Y(y_2)$	$\dots$	$p_Y(y_n)$

trong đó

$$p_Y(y_j) = P(Y = y_j) = \sum_{i=1}^{n} p_{X,Y}(x_i, y_j), j = 1, \dots, m.$$

Nhận xét: X và Y độc lập khi và chỉ khi

$$p_{X,Y}(x_i, y_j) = p_X(x_i)p_Y(y_j),\ \forall i = 1, \ldots, n; \forall j = 1, \ldots, m.$$

Ví dụ 2: Cho X, Y là hai biến ngẫu nhiên rời rạc có bảng phân bố xác suất đồng thời

$X \backslash Y$	0	2	3	5
-2	0,1	0,15	0,1	0
1	$5k$	$3k$	0,05	0,07
4	0	$2k$	0	0,13

a) Tìm k và tính $F_{X,Y}(3, 2)$.

b) Tìm bảng phân bố xác suất của các biến ngẫu nhiên thành phần X và Y. Hai biến ngẫu nhiên X và Y có độc lập không?

3.2.2 Phân bố xác suất có điều kiện và kỳ vọng có điều kiện

Giả sử X là biến ngẫu nhiên rời rạc có $R_X = \{x_1, \dots, x_n\}$ và B là một biến cố có $P(B) > 0$. Khi đó hàm khối lượng xác suất của biến ngẫu nhiên X với điều kiện B được xác định bởi

$$p_{X|B}(x_i|B) = P(X = x_i|B) = \frac{P((X = x_i) \cap B)}{P(B)}$$

Bảng phân bố xác suất của X với điều kiện B là

$X\|B$	x_1	x_2	$\ldots$	x_n
P	$p_{X\|B}(x_1\|B)$	$p_{X\|B}(x_2\|B)$	$\ldots$	$p_{X\|B}(x_n\|B)$

Kỳ vọng của X với điều kiện B được xác định bởi

$$E(X|B) = \sum_{i=1}^{n} x_i P(X = x_i|B)$$

Cho X, Y là hai biến ngẫu nhiên rời rạc có

$$R_X = \{x_1, \dots, x_n\}, R_Y = \{y_1, \dots, y_m\}.$$

- Bảng phân bố xác suất của X với điều kiện $(Y = y_j)$ là

$X\|(Y = y_j)$	x_1	x_2	$\dots$	x_n
P	$p_{X\|Y}(x_1\|y_j)$	$p_{X\|Y}(x_2\|y_j)$	$\dots$	$p_{X\|Y}(x_n\|y_j)$

trong đó

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

- Kỳ vọng của X với điều kiện $(Y = y_j)$ là

$$E(X|Y = y_j) = \sum_{i=1}^{n} x_i P(X = x_i|Y = y_j).$$

- Bảng phân bố xác suất của Y với điều kiện $(X = x_i)$ là

$Y\|(X=x_i)$	y_1	y_2	$\dots$	y_m
P	$p_{Y\|X}(y_1\|x_i)$	$p_{Y\|X}(y_2\|x_i)$	$\dots$	$p_{Y\|X}(y_m\|x_i)$

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

- Kỳ vọng của Y với điều kiện $(X = x_i)$ là
$$E(Y|X = x_i) = \sum_{j=1}^{m} y_j P(Y = y_j|X = x_i).$$

Nhận xét: Nếu X, Y độc lập thì $\forall i = 1, \dots, n; \forall j = 1, \dots, m,$
$$p_{X|Y}(x_i|y_j) = p_X(x_i),\ p_{Y|X}(y_j|x_i) = p_Y(y_j).$$

Ví dụ 3: Từ kết quả phân tích các số liệu thống kê trong tháng về doanh thu (X) và chi phí cho quảng cáo (Y) (đơn vị triệu đồng) của một công ty, thu được bảng phân bố xác suất đồng thời như sau:

Y \ X	100	200	300
1	0,15	0,2	0,04
1,5	0,05	0,2	0,15
2	0,01	0,05	0,15

(a) Tìm bảng phân bố xác suất của X với điều kiện $Y = 1,5$.

(b) Nếu chi phí cho quảng cáo là 1,5 triệu đồng thì doanh thu trung bình là bao nhiêu?

3.3.1 Hàm mật độ xác suất đồng thời

Định nghĩa

Hàm mật độ xác suất của véc tơ ngẫu nhiên hai chiều (X, Y) là hàm hai biến $f_{X,Y}(x, y) \geq 0$ thỏa mãn

$$F_{X,Y}(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f_{X,Y}(u, v)dudv$$

$f_{X,Y}(x, y)$ còn được gọi là hàm mật độ xác suất đồng thời của hai biến ngẫu nhiên X, Y.

Tính chất của hàm mật độ xác suất đồng thời:

1) $f_{X,Y}(x, y) \geq 0 \; \forall x, y \in \mathbb{R}$
2) Nếu hàm $f_{X,Y}(x, y)$ liên tục theo cả hai biến trên miền $D \subset \mathbb{R}^2$ thì

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}, \; \forall (x, y) \in D$$

3) $P((X, Y) \in D) = \iint\limits_D f_{X,Y}(x, y) dxdy$ với $D \subset \mathbb{R}^2$
4) $\int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} f_{X,Y}(x, y) dxdy = 1$

Ví dụ 4: Cho véc tơ ngẫu nhiên hai chiều (X, Y) có hàm mật độ xác suất

$$f_{X,Y}(x,y) = \frac{c}{(1+x^2)(1+y^2)};\; x, y \in \mathbb{R}$$

a) Tìm hằng số c.

b) Tìm hàm phân bố xác suất của (X, Y).

c) Tính $P(1 < X \leq \sqrt{3}, 0 < Y \leq 1)$.

3.3.2 Hàm mật độ xác suất biên

- Hàm mật độ xác suất của biến ngẫu nhiên thành phần X là

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

- Hàm mật độ xác suất của biến ngẫu nhiên thành phần Y là

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

Nhận xét: Hai biến ngẫu nhiên liên tục X và Y độc lập khi và chỉ khi

$$f_{X,Y}(x,y) = f_X(x)f_Y(y), \ \forall x, y \in \mathbb{R}.$$

Ví dụ 5: Cho véc tơ ngẫu nhiên hai chiều (X, Y) có hàm mật độ xác suất

$$f_{X,Y}(x,y) = \begin{cases} \dfrac{1}{6\pi} & \text{nếu } \dfrac{x^2}{9} + \dfrac{y^2}{4} \le 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a) Tìm các hàm mật độ xác suất của X và Y.

b) Hai biến ngẫu nhiên X và Y có độc lập không?

3.3.3 Hàm mật độ xác suất có điều kiện và kỳ vọng có điều kiện

- Hàm mật độ xác suất có điều kiện của Y với điều kiện $X = x$ là

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} \text{ với điều kiện } f_X(x) > 0.$$

- Kỳ vọng của Y với điều kiện $X = x$ được xác định bởi

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

- Hàm mật độ xác suất có điều kiện của X với điều kiện $Y = y$ là

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} \text{ với điều kiện } f_Y(y) > 0.$$

- Kỳ vọng của X với điều kiện $Y = y$ được xác định bởi

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

Ví dụ 6: Cho véc tơ ngẫu nhiên hai chiều (X, Y) có hàm mật độ xác suất

$$f_{X,Y}(x,y) = \begin{cases} x+y & \text{nếu } 0 \le x; y \le 1 \\ 0 & \text{nếu ngược lại} \end{cases}$$

a) Tìm hàm mật độ xác suất có điều kiện $f_{Y|X}(y|x)$ và tính $E(Y|X = x)$.

b) Tìm hàm mật độ xác suất có điều kiện $f_{X|Y}(x|y)$ và tính $E(X|Y = y)$.

Chương 3: Véc tơ ngẫu nhiên

3.4.1 Hiệp phương sai (Covariance)

Định nghĩa

Hiệp phương sai của hai biến ngẫu nhiên X và Y, ký hiệu $\operatorname{cov}(X, Y)$, được xác định như sau

$$\operatorname{cov}(X, Y) = E[(X - E(X))(Y - E(Y))] \tag{1}$$

Kỳ vọng của hàm của hai biến ngẫu nhiên

- Nếu X, Y là các biến ngẫu nhiên rời rạc có

$$R_X = \{x_1, \dots, x_n\}, R_Y = \{y_1, \dots, y_m\}$$

thì

$$E(g(X,Y)) = \sum_{i=1}^{n}\sum_{j=1}^{m} g(x_i, y_j) p_{X,Y}(x_i, y_j)$$

- Nếu X, Y là các biến ngẫu nhiên liên tục thì

$$E(g(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y) dxdy$$

Tính chất của hiệp phương sai:

1) $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$
2) Nếu X và Y độc lập thì $\text{cov}(X, Y) = 0$.
3) Nếu a, b, c, d là các hằng số thì

$$\text{cov}(aX + c, bY + d) = ab\text{cov}(X, Y).$$

4) Nếu a, b là các hằng số thì

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

Ma trận hiệp phương sai

Cho véc tơ ngẫu nhiên $X = (X_1, X_2, \ldots, X_n)$. Ma trận

$$M = [C_{ij}]_{n\times n} \text{ với } C_{ij} = \operatorname{cov}(X_i, X_j)$$

được gọi là ma trận hiệp phương sai của véc tơ ngẫu nhiên X.

Nhận xét: Ma trận hiệp phương sai là ma trận đối xứng.

3.4.2 Hệ số tương quan

Định nghĩa

Hệ số tương quan của hai biến ngẫu nhiên X và Y, ký hiệu $\rho_{X,Y}$, được xác định như sau

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)D(Y)}}. \tag{2}$$

Tính chất của hệ số tương quan:

1) Nếu X và Y độc lập thì $\rho_{X,Y} = 0$.
   Khi $\rho_{X,Y} = 0$ ta nói X, Y không tương quan.
2) $-1 \leq \rho_{X,Y} \leq 1$
3) $\rho_{X,Y} = \pm 1 \Leftrightarrow X$ và Y tương quan tuyến tính, tức là tồn tại $a \neq 0$ và b sao cho $Y = aX + b$.

Ví dụ 7: Cho X, Y là hai biến ngẫu nhiên rời rạc có bảng phân bố xác suất đồng thời

X \ Y	0	2	3	5
-2	0,1	0,15	0,1	0
1	0,2	0,12	0,05	0,07
4	0	0,08	0	0,13

a) Tìm hệ số tương quan của X và Y.

b) Tính $D(2X - 3Y)$.

Chương 3: Véc tơ ngẫu nhiên

3.5.1 Hội tụ theo xác suất và hội tụ theo phân bố của dãy biến ngẫu nhiên

Hội tụ theo xác suất

Dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ được gọi là hội tụ theo xác suất về biến ngẫu nhiên X, ký hiệu $X_n \xrightarrow[n\to\infty]{P} X$, nếu $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P(|X_n - X| \leq \varepsilon) = 1.$$

Hội tụ theo phân bố

Dãy các biến ngẫu nhiên $\{X_n\}_{n=1}^{\infty}$ được gọi là hội tụ theo phân bố về biến ngẫu nhiên X nếu $\forall x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_{X_n}(x) = F(x).$$

Bất đẳng thức Chebyshev

Nếu X là biến ngẫu nhiên có kỳ vọng và phương sai hữu hạn thì $\forall \varepsilon > 0$,

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}.$$

Luật số lớn Chebyshev

Nếu $\{X_n\}_{n=1}^{\infty}$ là dãy các biến ngẫu nhiên độc lập, có các kỳ vọng hữu hạn và tất cả các phương sai bị chặn trên bởi hằng số C thì $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\ldots+X_n}{n} - \frac{E(X_1)+\ldots+E(X_n)}{n}\right| \leq \varepsilon\right) = 1.$$

Hệ quả. Nếu $\{X_n\}_{n=1}^{\infty}$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố với kỳ vọng μ và phương sai σ^2 thì

$$\frac{X_1+\ldots+X_n}{n} \xrightarrow[n\to\infty]{P} \mu.$$

Luật số lớn Bernoulli

Tiến hành một dãy n phép thử độc lập, xác suất xuất hiện biến cố A ở mỗi phép thử là p. Gọi $f_n(A)$ là tần suất xuất hiện biến cố A. Khi đó

$$f_n(A) \xrightarrow[n\to\infty]{P} p.$$

3.7.3 Định lý giới hạn trung tâm

Định lý giới hạn trung tâm

Giả sử $\{X_n\}_{n=1}^{\infty}$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố với kỳ vọng μ và phương sai σ^2. Đặt

$$S_n = \frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Khi đó $\{S_n\}_{n=1}^{\infty}$ hội tụ theo phân bố về phân bố chuẩn tắc, tức là $\forall x \in \mathbb{R}$,

$$\lim_{n\to\infty} F_{S_n}(x) = \Phi(x).$$

Định lý Moivre-Laplace

Nếu $\{X_n\}_{n=1}^{\infty}$ là dãy các biến ngẫu nhiên độc lập có cùng phân bố Bernoulli với tham số p thì $\forall x \in \mathbb{R}$,

$$\lim_{n\to\infty} P\left(\frac{X_1 + X_2 + \ldots + X_n - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x).$$

Xấp xỉ phân bố nhị thức bởi phân bố chuẩn

Nếu $X \sim B(n;p)$ và $np > 5, n(1-p) > 5$ thì $\dfrac{X - np}{\sqrt{np(1-p)}}$ có phân bố xấp xỉ $N(0;1)$.

$$P(X \leq k) = \Phi\left(\frac{k + 0,5 - np}{\sqrt{np(1-p)}}\right),$$

$$P(k_1 \leq X \leq k_2) = \Phi\left(\frac{k_2 + 0,5 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - 0,5 - np}{\sqrt{np(1-p)}}\right).$$

Ví dụ 8: Xác suất để sản phẩm sau khi sản xuất không được kiểm tra chất lượng là 0,2. Tìm xác suất để trong 400 sản phẩm được sản xuất ra,

a) có nhiều nhất 80 sản phẩm không được kiểm tra chất lượng.

b) có từ 70 đến 100 sản phẩm không được kiểm tra chất lượng.