VÉC-TƠ NGẪU NHIÊN

BỘ MÔN THỐNG KÊ TOÁN HỌC
KHOA TOÁN - TIN HỌC
ĐẠI HỌC KHOA HỌC TỰ NHIÊN TP.HCM

Tháng 9 năm 2021

Outline

Outline

1. Giới thiệu
   - Khái niệm véc-tơ ngẫu nhiên
   - Phân phối xác suất của véc-tơ ngẫu nhiên 2 chiều
2. Véc-tơ ngẫu nhiên rời rạc 2 chiều
   - Phân phối đồng thời
   - Phân phối lề
   - Phân phối có điều kiện và sự độc lập
3. Hiệp phương sai và hệ số tương quan
   - Hiệp phương sai
   - Hệ số tương quan

Véc-tơ ngẫu nhiên

Một bộ gồm n biến ngẫu nhiên $(X_1, \ldots, X_n)$ gọi là một véc-tơ ngẫu nhiên n chiều.
Nếu $X_1, \ldots, X_n$ là các biến ngẫu nhiên rời rạc thì $(X_1, \ldots, X_n)$ là véc-tơ ngẫu nhiên rời rạc.
Nếu $X_1, \ldots, X_n$ là các biến ngẫu nhiên liên tục thì $(X_1, \ldots, X_n)$ là véc-tơ ngẫu nhiên liên tục.

Ví dụ 1

Một nhà máy sản xuất một loại sản phẩm, nếu kích thước của sản phẩm được đo bằng chiều dài X và chiều rộng Y thì ta có véctơ ngẫu nhiên hai chiều, còn nếu xét thêm cả chiều cao Z nữa thì ta có véctơ ngẫu nhiên ba chiều. Nếu ta chỉ quan tâm đến trọng lượng và thể tích của sản phẩm ta cũng được biến ngẫu nhiên hai chiều.

Hàm phân phối của véc-tơ ngẫu nhiên 2 chiều

Định nghĩa 2 (Joint probability distribution function)

Hàm phân phối xác suất đồng thời của véc-tơ ngẫu nhiên (X, Y) là hàm $F(x, y)$ được định nghĩa

$$F(x, y) = \mathbb{P}(X \leq x, Y \leq y) \quad \forall x, y \in \mathbb{R} \tag{1}$$

Định nghĩa 3 (Marginal probability distribution function)

Nếu véc-tơ ngẫu nhiên (X, Y) có hàm phân phối xác suất đồng thời $F(x, y)$ thì **hàm phân phối xác suất lề** cho X và Y được định nghĩa

$$F_X(x) = \mathbb{P}(X \leq x) = F(x, +\infty) = \lim_{y \to +\infty} F(x, y) \tag{2}$$

$$F_Y(y) = \mathbb{P}(Y \leq y) = F(+\infty, y) = \lim_{x \to +\infty} F(x, y) \tag{3}$$

Hàm phân phối của véc-tơ ngẫu nhiên 2 chiều

Tính chất

1. $F(x, y)$ là hàm không giảm theo từng biến số

$$F(x_1, y) \leq F(x_2, y) \text{ khi } x_1 \leq x_2$$
$$F(x, y_1) \leq F(x, y_2) \text{ khi } y_1 \leq y_2$$

2. $$\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = \lim_{\substack{x \to -\infty \\ y \to -\infty}} F(x, y)) = 0$$

3. $$\lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1$$

Outline

Hàm mật độ đồng thời

TH rời rạc

Định nghĩa 4 (Joint probability mass function)

Hàm mật độ xác suất đồng thời (hay ngắn gọn là **hàm mật độ đồng thời**) của véc-tơ ngẫu nhiên rời rạc (X, Y), ký hiệu là $f_{X,Y}(x, y)$, là một hàm thực thỏa

(1) $f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$
(2) $f_{X,Y}(x, y) \geq 0$
(3) $\sum_x \sum_y f_{X,Y}(x, y) = 1$

Hàm mật độ đồng thời của (X, Y) được biểu diễn bằng bảng phân phối xác suất đồng thời.

Bảng phân phối xác suất đồng thời
TH rời rạc

X \ Y	y_1	y_2	$\cdots$	y_j	$\cdots$	y_n	Tổng dòng
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	$\cdots$	$f(x_1, y_j)$	$\cdots$	$f(x_1, y_n)$	$f(x_1, \bullet)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	$\cdots$	$f(x_2, y_j)$	$\cdots$	$f(x_2, y_n)$	$f(x_2, \bullet)$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
x_i	$f(x_i, y_1)$	$f(x_i, y_2)$	$\cdots$	$f(x_i, y_j)$	$\cdots$	$f(x_i, y_n)$	$f(x_i, \bullet)$
$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$
x_m	$f(x_m, y_1)$	$f(x_m, y_2)$	$\cdots$	$f(x_m, y_j)$	$\cdots$	$f(x_m, y_n)$	$f(x_m, \bullet)$
Tổng cột	$f(\bullet, y_1)$	$f(\bullet, y_2)$	$\cdots$	$f(\bullet, y_j)$	$\cdots$	$f(\bullet, y_n)$	1

Bảng 1: Phân phối xác suất đồng thời của (X, Y)

Hàm mật độ đồng thời
Ví dụ

Ví dụ 5

Cho (X, Y) là véc-tơ ngẫu nhiên rời rạc 2 chiều có hàm mật độ đồng thời $f(x, y)$ cho bởi bảng sau

X \ Y	-1	0	1
1	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{6}$
0	$\frac{1}{9}$	0	$\frac{1}{6}$
-1	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{9}$

Tính:

(a) $\mathbb{P}(X + Y = 1)$

(b) $\mathbb{P}(X = 0)$

(c) $\mathbb{P}(X < Y)$

Hàm mật độ lề
TH rời rạc

Định nghĩa 6 (Marginal probability mass function)

Nếu véc-tơ ngẫu nhiên rời rạc (X, Y) có hàm mật độ đồng thời là $f_{X,Y}(x, y)$ thì hàm mật độ lề cho biến ngẫu nhiên X và Y được xác định như sau

$$f_X(x) = \mathbb{P}(X = x) = \sum_y f_{X,Y}(x, y) \tag{4}$$

$$f_Y(y) = \mathbb{P}(Y = y) = \sum_x f_{X,Y}(x, y) \tag{5}$$

Hàm mật độ lề
TH rời rạc

Bảng phân phối lề của biến ngẫu nhiên X

X	x_1	x_2	$\cdots$	x_m
$\mathbb{P}_X$	$f_X(x_1)$	$f_X(x_2)$	$\cdots$	$f_X(x_m)$

với $f_X(x_i) = f(x_i, \bullet) = \sum_{j=1}^{n} f(x_i, y_j) \quad (i = 1, \ldots, m)$

Bảng phân phối lề của biến ngẫu nhiên Y

Y	y_1	y_2	$\cdots$	y_n
$\mathbb{P}_Y$	$f_Y(y_1)$	$f_Y(y_2)$	$\cdots$	$f_Y(y_n)$

với $f_Y(y_j) = f(\bullet, y_j) = \sum_{i=1}^{m} f(x_i, y_j) \quad (j = 1, \ldots, n)$

Hàm mật độ lề
TH rời rạc

Ví dụ 7

(X, Y) là véc-tơ ngẫu nhiên rời rạc 2 chiều có hàm mật độ đồng thời $f_{X,Y}(x, y)$ cho bởi bảng sau

X \ Y	-1	0	1
1	$\frac{1}{18}$	$\frac{1}{9}$	$\frac{1}{6}$
0	$\frac{1}{9}$	0	$\frac{1}{6}$
-1	$\frac{1}{6}$	$\frac{1}{9}$	$\frac{1}{9}$

Tìm hàm xác suất lề cho X và Y.

Kỳ vọng và phương sai từ phân phối đồng thời
TH rời rạc

Định nghĩa 8

Xét véc-tơ ngẫu nhiên (X, Y), nếu X có hàm mật độ lề $f_X(x)$ thì

$$\mathbb{E}(X) = \mu_X = \sum_x x f_X(x) = \sum_x \sum_y x f_{X,Y}(x, y) \tag{6}$$

và

$$\mathbb{V}ar(X) = \sigma_X^2 = \sum_x (x - \mu_X)^2 f_X(x) = \sum_x \sum_y (x - \mu_X)^2 f_{X,Y}(x, y) \tag{7}$$

Ta cũng có định nghĩa tương tự cho Y.

Hàm mật độ có điều kiện
TH rời rạc

Xét véc-tơ ngẫu nhiên rời rạc (X, Y), khi biết trước $X = x$ thì hàm mật độ có điều kiện của Y cho bởi

$$f_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x)$$

Áp dụng công thức xác suất có điều kiện ta có

$$\begin{aligned} f_{Y|X}(y|x) = \mathbb{P}(Y = y|X = x) &= \frac{\mathbb{P}\left[(X = x) \cap (Y = y)\right]}{\mathbb{P}(X = x)} \\ &= \frac{\mathbb{P}(X = x, Y = y)}{P(X = x)} \end{aligned}$$

trong đó $P(X = x, Y = y) = f_{X,Y}(x, y)$ và $P(X = x) = f_X(x)$.

Hàm mật độ có điều kiện
TH rời rạc

Định nghĩa 9 (Conditional probability mass function)

Xét véc-tơ ngẫu nhiên rời rạc (X, Y), **hàm mật độ có điều kiện** của Y cho trước X nhận giá trị x được định nghĩa

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{với } f_X(x) > 0 \tag{8}$$

Tương tự, hàm mật độ có điều kiện của X cho trước $Y = y$ được định nghĩa

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} \quad \text{với } f_Y(y) > 0 \tag{9}$$

Hàm mật độ có điều kiện
TH rời rạc

Hệ quả 10

Hàm mật độ đồng thời $f_{XY}(x,y)$ của véc-tơ ngẫu nhiên (X,Y) có thể được viết dưới dạng sau

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

Hàm phân phối có điều kiện
TH rời rạc

Định nghĩa 11 (Conditional probability distribution function)

Hàm phân phối xác suất của biến ngẫu nhiên X khi biết Y nhận giá trị y được định nghĩa:

$$F_{X|Y}(x|y) = \mathbb{P}(X \leq x|Y = y) = \sum_{x_i \leq x} f_{X|Y}(x_i|y) \tag{10}$$

Tương tự, hàm phân phối xác suất của biến ngẫu nhiên Y khi biết $X = x$

$$F_{Y|X}(y|x) = \mathbb{P}(Y \leq y|X = x) = \sum_{y_j \leq y} f_{Y|X}(y_j|x) \tag{11}$$

Kỳ vọng có điều kiện
TH rời rạc

Định nghĩa 12 (Conditional mean)

Kỳ vọng có điều kiện của biến ngẫu nhiên Y cho trước $X = x$, ký hiệu $\mathbb{E}(Y|X = x)$ hay $\mu_{Y|x}$ được định nghĩa

$$\mathbb{E}(Y|X = x) = \sum_y y f_{Y|X}(y|x) \tag{12}$$

Tương tự, kỳ vọng có điều kiện của X cho trước $Y = y$

$$\mathbb{E}(X|Y = y) = \sum_x x f_{X|Y}(x|y) \tag{13}$$

Kỳ vọng có điều kiện
TH rời rạc

Tính chất của kỳ vọng có điều kiện

Nếu X và Y có phân phối đồng thời, ta có

1. $$\mathbb{E}\left[\mathbb{E}(X|Y)\right] = \mathbb{E}(X)$$
2. $$\mathbb{V}ar(X) = \mathbb{E}\left[\mathbb{V}ar(X|Y)\right] + \mathbb{V}ar\left[\mathbb{E}(X|Y)\right]$$

Phân phối có điều kiện
TH rời rạc

Ví dụ 13

(X, Y) là véc-tơ ngẫu nhiên rời rạc 2 chiều có hàm xác suất đồng thời $f(x, y)$ cho bởi bảng như ví dụ 5.

(a) Lập bảng phân phối có điều kiện của X cho trước $Y = 1$ và tính $f_{X|Y}(-1|Y = 1)$.
(b) Tính $\mathbb{E}(X|Y = 1)$ và $\mathbb{V}ar(X|Y = 1)$.
(c) Lập bảng phân phối có điều kiện của Y cho trước $X = -1$ và tính $f_{Y|X}(1|X = -1)$.
(d) Tính $\mathbb{E}(Y|X = -1)$ và $\mathbb{V}ar(Y|X = -1)$.

Sự độc lập
TH rời rạc

Sự độc lập của hai biến ngẫu nhiên rời rạc

Hai biến ngẫu nhiên rời rạc X và Y gọi là độc lập với nhau nếu thỏa một trong các tính chất sau

(1) $f_{X,Y}(x, y) = f_X(x).f_Y(y)\ \forall x, y$.
(2) $f_{Y|X}(y|x) = f_Y(y)\ \forall x, y$ và $f_X(x) > 0$.
(3) $f_{X|Y}(x|y) = f_X(x)\ \forall x, y$ và $f_Y(y) > 0$.
(4) $\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A).\mathbb{P}(Y \in B)$ với tập A, B bất kỳ trên miền giá trị tương ứng của X và Y.

Ví dụ 14

Kiểm tra tính độc lập của hai biến ngẫu nhiên trong ví dụ 5.

Ví dụ 15

Cho véc-tơ ngẫu nhiên (X, Y) có hàm mật độ đồng thời

$$f(x, y) = c(x + y) \quad x = 1, 2, 3 \text{ và } y = 1, 2, 3$$

(a) Tìm c.
(b) Tính $\mathbb{P}(X = 1, Y \leq 4)$, $\mathbb{P}(X = 1)$, $\mathbb{P}(Y = 2)$, $\mathbb{P}(X \leq 2, Y \leq 2)$.
(d) Tìm phân phối lề cho X, phân phối lề cho Y.
(e) Tìm phân phối của Y cho biết $X = 1$; phân phối của X cho biết $Y = 2$.
(f) Tính $\mathbb{E}(Y|X = 1)$ và $\mathbb{E}(X|Y = 2)$.
(g) X và Y có độc lập?

Outline

Hiệp phương sai

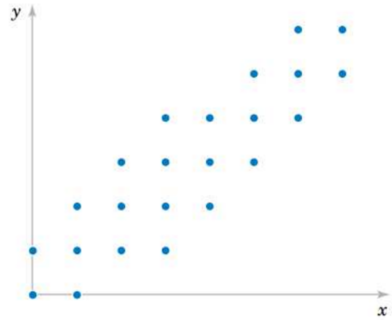
Định nghĩa 16 (Covariance)

Cho X và Y là hai biến ngẫu nhiên, **hiệp phương sai** giữa X và Y, ký hiệu $\mathbb{C}ov\,(X, Y)$ (hay $\sigma_{X,Y}$) được định nghĩa như sau

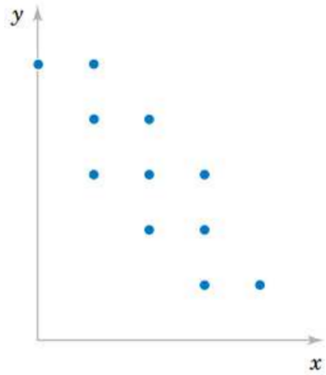
$$\begin{aligned} \mathbb{C}ov\,(X, Y) &= \mathbb{E}\,(X - \mathbb{E}\,[X])\,(Y - \mathbb{E}\,[Y]) \qquad (14) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

Hiệp phương sai là đại lượng dùng để đo mối liên hệ tuyến tính giữa hai biến ngẫu nhiên X và Y.

Hiệp phương sai

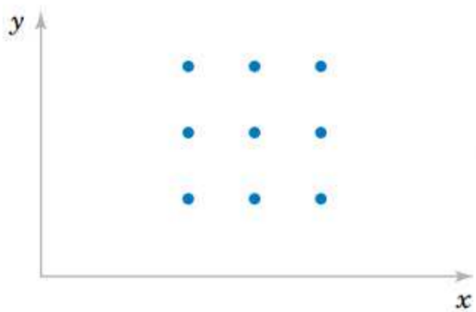


Tương quan dương

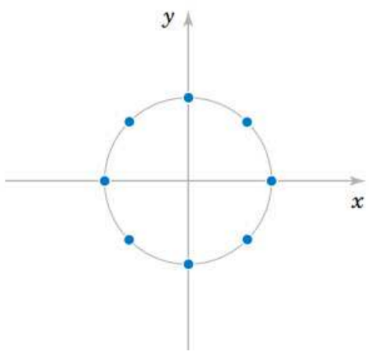


Tương quan âm

Hiệp phương sai



Không tương quan



Không tương quan

Hiệp phương sai

Tính chất

Nếu hai biến ngẫu nhiên X và Y độc lập và có phương sai hữu hạn thì

$$\mathbb{C}ov\,(X, Y) = 0 \qquad (15)$$

và phương sai của $X + Y$

$$\mathbb{V}ar\,(X + Y) = \mathbb{V}ar\,(X) + \mathbb{V}ar\,(Y) \qquad (16)$$

Chú ý

Nếu hai biến ngẫu nhiên X và Y có $\mathbb{C}ov\,(X, Y) = 0$ thì ta nói X và Y không tương quan, nhưng không thể suy ra được X và Y là độc lập.

Hiệp phương sai

Định lí 17 (Phương sai của tổng n biến ngẫu nhiên)

Nếu $X_1, \ldots, X_n$ là n biến ngẫu nhiên sao cho $\mathbb{V}ar\,(X_i) < +\infty$ với mọi $i = 1, \ldots, n$ thì

$$\mathbb{V}ar\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}ar\,(X_i) + 2\sum_{i<j}\sum \mathbb{C}ov\,(X_i, X_j) \qquad (17)$$

Trường hợp hai biến

Với a, b và c là hằng số, ta có

$$\mathbb{V}ar(aX + bY + c) = a^2\mathbb{V}ar(X) + b^2\mathbb{V}ar(Y) + 2ab\mathbb{C}ov(X, Y)$$

Hệ số tương quan

Định nghĩa 18 (Coefficient of Correlation)

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

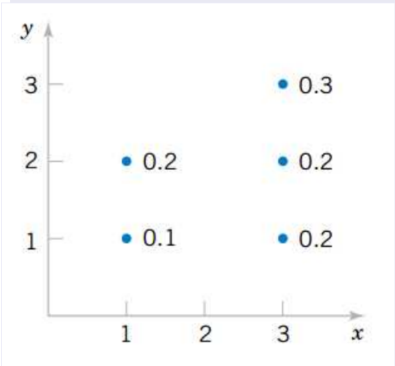
$$\rho_{X,Y} = \frac{\mathbb{C}ov(X, Y)}{\sqrt{\mathbb{V}ar(X)\mathbb{V}ar(Y)}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \qquad (18)$$

Tính chất

$$-1 \le \rho_{X,Y} \le +1$$

Hệ số tương quan

Ví dụ 19



Cho véc-tơ ngẫu nhiên rời rạc (X, Y) có phân phối xác suất đồng thời như hình bên. Tính $\mathbb{C}ov(X, Y)$ và $\rho_{X,Y}$.

Hệ số tương quan

Ví dụ 20

Cho véc-tơ ngẫu nhiên (X, Y) có $\rho_{X,Y} = \frac{1}{3}$, và $\sigma_X^2 = a$, $\sigma_Y^2 = 4a$. Biến ngẫu nhiên $Z = 3X - 4Y$ có $\sigma_Z^2 = 11$. Tìm a.

Ôn tập

- Phân phối đồng thời
- Phân phối lề
- Phân phối có điều kiện
- Sự độc lập
- Hiệp phương sai
- Hệ số tương quan