Chương 3: Biến ngẫu nhiên nhiều chiều

Vũ Thị Huệ [1]

Viện Toán ứng dụng và Tin học, ĐHBK Hà Nội

Hà Nội, tháng 9 năm 2020

[1] Email: hue.hnue@gmail.com

Nội dung

Các khái niệm cơ sở

- Ở chương trước chúng ta quan tâm đến xác suất của biến ngẫu nhiên riêng rẽ. Nhưng trong thực tế nhiều khi ta phải xét đồng thời nhiều biến khác nhau có quan hệ tương hỗ (ví dụ khi nghiên cứu về sinh viên một trường đại học thì cần quan tâm đến chiều cao, cân nặng, tuổi, ...). Do đó dẫn đến khái niệm biến ngẫu nhiên nhiều chiều hay véctơ ngẫu nhiên.
- Để cho đơn giản, ta nghiên cứu biến ngẫu nhiên hai chiều (X, Y), trong đó X, Y là các biến ngẫu nhiên một chiều. Hầu hết các kết quả thu được đều có thể mở rộng khá dễ dàng cho trường hợp biến ngẫu nhiên n chiều.
- Biến ngẫu nhiên hai chiều được gọi là rời rạc (liên tục) nếu các thành phần của nó là các biến ngẫu nhiên rời rạc (liên tục).

Các khái niệm cơ sở

Định nghĩa 1.1

Hàm phân phối xác suất của biến ngẫu nhiên hai chiều (X, Y) được xác định như sau

$$\boxed{F(x, y) = P(X < x,\ Y < y), \quad x, y \in \mathbb{R}.} \tag{1.1}$$

Nhiều tài liệu gọi hàm trên là hàm phân phối xác suất đồng thời của hai biến X và Y.

Tính chất

- $0 \leq F(x, y) \leq 1,\ \forall x, y \in \mathbb{R}$;
- $F(x, y)$ là hàm không giảm theo từng đối số;
- $F(-\infty, y) = F(x, -\infty) = 0,\ \forall x, y \in \mathbb{R}$ và $F(+\infty, +\infty) = 1$;
- Với $x_1 < x_2,\ y_1 < y_2$ ta luôn có

$$P(x_1 \leq X \leq x_2,\ y_1 \leq y \leq y_2) = F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1).$$

Các khái niệm cơ sở

Tính chất (tiếp)

- Các hàm

$$\begin{aligned} F(x,+\infty) &= P(X<x,\ Y<+\infty) = P(X<x) =: F_X(x) \\ F(+\infty,y) &= P(X<+\infty,\ Y<y) = P(Y<y) =: F_Y(x) \end{aligned}$$

là các hàm phân phối riêng của các biến ngẫu nhiên X và Y và còn được gọi là các *phân phối biên* của biến ngẫu nhiên hai chiều (X,Y).

Định nghĩa 1.2

Hai biến ngẫu nhiên X,Y được gọi là *độc lập* nếu

$$F(x,y) = F_X(x).F_Y(y), \quad \forall x,y \in \mathbb{R}.$$

PPXS của biến ngẫu nhiên hai chiều rời rạc

Định nghĩa 1.3

Bảng phân phối xác suất của biến ngẫu nhiên hai chiều (X, Y) rời rạc được xác định như sau

[.3cm]XY	y_1	$\dots$	y_j	$\dots$	y_n	$\sum_j$
x_1	p_{11}	$\dots$	p_{1j}	$\dots$	p_{1n}	$P(X = x_1)$
x_2	p_{21}	$\dots$	p_{2j}	$\dots$	p_{2n}	$P(X = x_2)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	p_{i1}	$\dots$	p_{ij}	$\dots$	p_{in}	$P(X = x_i)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_m	p_{m1}	$\dots$	p_{mj}	$\dots$	p_{mn}	$P(X = x_m)$
$\sum_i$	$P(Y = y_1)$	$\dots$	$P(Y = y_j)$	$\dots$	$P(Y = y_n)$	1

PPXS của biến ngẫu nhiên hai chiều rời rạc

Trong đó

$$\boxed{p_{ij} = P\left(X = x_i,\ Y = y_j\right) \quad \forall i = \overline{1,m},\ j = \overline{1,n}.}$$

Kích thước bảng này có thể chạy ra vô hạn khi m, n chạy ra vô hạn.

Tính chất

- $p_{ij} \geq 0 \quad \forall i, j$;
- $\sum\limits_{i,j} p_{ij} = 1$;
- Hàm phân phối xác suất được xác định theo công thức $F(x, y) = \sum\limits_{i,j:\ x_i < x,\ y_j < y} p_{ij}$;
- Các phân phối biên được xác định như sau:

$$\begin{aligned} P(X = x_i) &= \sum_j P(X = x_i,\ Y = y_j) = \sum_j p_{ij} \\ P(Y = y_j) &= \sum_i P(X = x_i,\ Y = y_j) = \sum_i p_{ij}. \end{aligned}$$

PPXS của biến ngẫu nhiên hai chiều rời rạc

Ví dụ 1

Cho bảng phân phối xác suất đồng thời của (X, Y) như sau:

[.3cm]XY	1	2	3
1	0.10	0.25	0.10
2	0.15	0.05	0.35

Tìm bảng phân phối xác suất của X và Y, sau đó tính $F(2;3)$.

PPXS của biến ngẫu nhiên hai chiều rời rạc

Giải

Lấy tổng của hàng, cột tương ứng ta thu được

X	1	2
$P(X = x)$	0.45	0.55

Y	1	2	3
$P(Y = x)$	0.25	0.30	0.45

Ta có

$$F(2,3) = \sum_{x_i<2} \sum_{y_j<3} p_{ij} = p_{11} + p_{12} = 0.35.$$

PPXS của biến ngẫu nhiên hai chiều rời rạc

Giải

Lấy tổng của hàng, cột tương ứng ta thu được

X	1	2
$P(X=x)$	0.45	0.55

Y	1	2	3
$P(Y=x)$	0.25	0.30	0.45

Ta có

$$F(2,3)=\sum_{x_i<2}\sum_{y_j<3} p_{ij}=p_{11}+p_{12}=0.35.$$

PPXS của biến ngẫu nhiên hai chiều rời rạc

Ví dụ 2

Ta lấy ngẫu nhiên 3 pin từ một nhóm gồm 3 pin mới, 4 pin đã qua sử dụng nhưng vẫn dùng được và 5 pin hỏng. Nếu ký hiệu X, Y *tương ứng là số pin mới và số pin đã qua sử dụng nhưng vẫn dùng được trong 3 pin lấy ra. Lập bảng phân phối xác suất đồng thời cho* (X, Y).

PPXS của biến ngẫu nhiên hai chiều rời rạc

Bài làm

$P(X=0,Y=0)=C_5^3/C_{12}^3=10/220$
$P(X=0,Y=1)=C_4^1.C_5^2/C_{12}^3=40/220$
$P(X=0,Y=2)=C_4^2.C_5^1/C_{12}^3=30/220$
$P(X=0,Y=3)=C_4^3/C_{12}^3=4/220$
$P(X=1,Y=0)=C_3^1.C_5^2/C_{12}^3=30/220$
$P(X=1,Y=1)=C_3^1.C_4^1.C_5^1/C_{12}^3=60/220$
$P(X=1,Y=2)=C_3^1.C_4^2/C_{12}^3=18/220$
$P(X=2,Y=0)=C_3^2.C_5^1/C_{12}^3=15/220$
$P(X=2,Y=1)=C_3^2.C_4^1/C_{12}^3=12/220$, $P(X=3,Y=0)=C_3^3/C_{12}^3=1/220$

[.3cm]XY	0	1	2	3	$P(X=i)$
0	$10/220$	$40/220$	$30/220$	$4/220$	$84/220$
1	$30/220$	$60/220$	$18/220$	0	$108/220$
2	$15/220$	$12/220$	0	0	$27/220$
3	$1/220$	0	0	0	$1/220$
$P(Y=j)$	$56/220$	$112/220$	$48/220$	$4/220$	

PPXS của biến ngẫu nhiên hai chiều rời rạc

Ví dụ 3

15% các gia đình trong một cộng đồng nào đó không có con, 20% có 1, 35% có 2, và 30% có 3 con. Giả sử rằng các con được sinh ra là độc lập với nhau và khả năng là trai hay gái đều là 0,5. Một gia đình được lựa chọn ngẫu nhiên từ cộng đồng này, sau đó gọi B là số con trai và G là số con gái. Lập bảng phân phối xác suất đồng thời cho (B, G)

PPXS của biến ngẫu nhiên hai chiều rời rạc

Bài làm

[.3cm]BG	0	1	2	3	$P(B=i)$
0	0,15	0,10	0,0875	0,0375	0,3750
1	0,10	0,175	0,1125	0	0,3875
2	0,0875	0,1125	0	0	0,2000
3	0,0375	0	0	0	0,0375
$P(G=j)$	0,3750	0,3875	0,2000	0,0375	

$P(B=2, G=1) = P(\text{có 3 con và có đúng 1 gái})$
$= P(\text{có 3 con}).P(\text{có đúng 1 gái}|\text{có 3 con}) = 0,3.C_3^1.0,5.0,5^2 = 0,1125$

Phân phối xác suất của biến ngẫu nhiên hai chiều rời rạc

Chú ý 1.1

- Hai biến ngẫu nhiên X, Y được gọi là độc lập với nhau nếu ta có

$$\boxed{P(X = x_i,\ Y = y_j) = P(X = x_i).P(Y = y_j),\quad \forall i = \overline{1, m},\ j = \overline{1, n}}$$

- Các xác suất có điều kiện vẫn được tính như thông thường, tức là

$$P\left(X = x_i | Y = y_j\right) = \frac{P(X = x_i,\ Y = y_j)}{P(Y = y_j)} \quad \text{hoặc}$$

$$P\left(X = x_i | Y \in D\right) = \frac{P(X = x_i,\ Y \in D)}{P(Y \in D)}$$

Công thức cũng tương tự với $P\left(Y = y_j | X = x_i\right),\ P\left(Y = y_j | X \in D\right)$.

Phân phối xác suất của biến ngẫu nhiên hai chiều liên tục

Định nghĩa 1.4

Hàm hai biến không âm, liên tục $f(x,y)$ được gọi là *hàm mật độ xác suất đồng thời* của biến ngẫu nhiên hai chiều liên tục $(X<Y)$ nếu nó thỏa mãn

$$\boxed{P((X,Y)\in) = \iint f(x,y)dxdy \quad \forall \subset \mathbb{R}^2.} \tag{1.2}$$

Tính chất

- $F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(u,v)dudv;$
- $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)dxdy.$

Phân phối xác suất của biến ngẫu nhiên hai chiều liên tục

Tính chất (tiếp)

- $f(x,y) = \dfrac{\partial^2 F(x,y)}{\partial x \partial y}$;
- Các hàm mật độ biên
  - theo x : $\quad f_X(x) = \displaystyle\int_{-\infty}^{+\infty} f(x,y)dy$;
  - theo y : $\quad f_Y(y) = \displaystyle\int_{-\infty}^{+\infty} f(x,y)dx$.
- Hai biến ngẫu nhiên X và Y được gọi là độc lập nếu $f(x,y) = f_X(x).f_Y(y) \;\; \forall x, y$.
- Hàm mật độ có điều kiện của X khi đã biết $Y = y$:

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

PPXS của biến ngẫu nhiên hai chiều liên tục

Ví dụ 4

Hàm mật độ đồng thời của X, Y được cho bởi:

$$f(x,y) = \begin{cases} 2.e^{-x}.e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \textit{trường hợp khác} \end{cases}$$

Tính $P(X > 1, Y < 1)$, $P(X < Y)$, $P(X < a)$

Bài làm

$$P(X > 1, Y < 1) = \int_0^1 \int_1^\infty 2.e^{-x}.e^{-2y} dxdy = e^{-1}(1 - e^{-2})$$

$$P(X < Y) = \int_0^\infty \int_0^y 2.e^{-x}.e^{-2y} dxdy = 1/3$$

$$P(X < a) = \int_0^a \int_0^\infty 2.e^{-x}.e^{-2y} dydx = 1 - e^{-a}$$

PPXS của biến ngẫu nhiên hai chiều liên tục

Ví dụ 4

Hàm mật độ đồng thời của X, Y được cho bởi:

$$f(x,y) = \begin{cases} 2.e^{-x}.e^{-2y} & 0 < x < \infty, 0 < y < \infty \\ 0 & \textit{trường hợp khác} \end{cases}$$

Tính $P(X > 1, Y < 1)$, $P(X < Y)$, $P(X < a)$

Bài làm

$$P(X > 1, Y < 1) = \int_0^1 \int_1^\infty 2.e^{-x}.e^{-2y} dxdy = e^{-1}(1 - e^{-2})$$

$$P(X < Y) = \int_0^\infty \int_0^y 2.e^{-x}.e^{-2y} dxdy = 1/3$$

$$P(X < a) = \int_0^a \int_0^\infty 2.e^{-x}.e^{-2y} dydx = 1 - e^{-a}$$

Nội dung

Kỳ vọng và phương sai của các thành phần

Trường hợp (X, Y) rời rạc

$$EX = \sum_i P(X = x_i) = \sum_i \sum_j x_i p_{ij}; \quad EY = \sum_j y_j P(Y = y_j) = \sum_i \sum_j y_j p_{ij}$$

$$VX = \sum_i \sum_j x_i^2 p_{ij} - (EX)^2; \quad VY = \sum_i \sum_j y_j^2 p_{ij} - (EY)^2.$$

Trường hợp (X, Y) liên tục

$$EX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x.f(x, y)dxdy; \quad EY = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y.f(x, y)dxdy$$

$$VX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2.f(x, y)dxdy - (EX)^2; \quad VY = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y^2.f(x, y)dxdy - (EY)^2.$$

Kỳ vọng và phương sai của các thành phần

Trường hợp (X, Y) rời rạc

$$EX = \sum_i P(X = x_i) = \sum_i \sum_j x_i p_{ij}; \quad EY = \sum_j y_j P(Y = y_j) = \sum_i \sum_j y_j p_{ij}$$

$$VX = \sum_i \sum_j x_i^2 p_{ij} - (EX)^2; \quad VY = \sum_i \sum_j y_j^2 p_{ij} - (EY)^2.$$

Trường hợp (X, Y) liên tục

$$EX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x.f(x, y)dxdy; \quad EY = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y.f(x, y)dxdy$$

$$VX = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^2.f(x, y)dxdy - (EX)^2; \quad VY = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y^2.f(x, y)dxdy - (EY)^2.$$

Kỳ vọng và phương sai của các thành phần

Chú ý 2.1

Đối với biến ngẫu nhiên $Z = g(X, Y)$ ta có

$$EZ = E\left[g(X,Y)\right] = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} g(x,y).f(x,y)dxdy$$

Hiệp phương sai và hệ số tương quan

Định nghĩa 2.1

Cho biến ngẫu nhiên hai chiều (X, Y), hiệp phương sai của hai thành phần X và Y, kí hiệu là $cov(X, Y)$, được xác định bởi

$$\boxed{cov(X,Y) = E\left[(X - EX)(Y - EY)\right] = E(XY) - EX.EY,} \tag{2.3}$$

trong đó $E(XY)$ được xác định theo công thức

$$E(XY) = \begin{cases} \sum\limits_i \sum\limits_j x_i y_j p_{ij}, & \text{đối với biến ngẫu nhiên rời rạc} \\ \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} xy.f(x,y), & \text{đối với biến ngẫu nhiên liên tục} \end{cases}$$

Ý nghĩa: Hiệp phương sai là một chỉ báo quan hệ của X, Y:

- $cov(X, Y) > 0$ cho thấy xu thế Y tăng khi X tăng
- $cov(X, Y) < 0$ cho thấy xu thế Y giảm khi X tăng

Hiệp phương sai và hệ số tương quan

Định nghĩa 2.2

Ta nói rằng X và Y *không tương quan* nếu $cov(X, Y) = 0$.

Nhận xét

- $cov(X, Y) = cov(Y, X)$;
- $VX = cov(X, X),\ VY = cov(Y, Y)$;
- Nếu X, Y độc lập, ta có $E(XY) = EX.EY$ tức là X và Y không tương quan. Điều ngược lại chưa chắc đã đúng.
- $cov(aX, Y) = a.cov(X, Y)$
- $cov(X + Z, Y) = cov(X, Y) + cov(Z, Y)$
- $cov\left(\left(\sum_{i=1}^{n} X_i, Y\right) = \sum_{i=1}^{n} cov(X_i, Y)\right.$
- $X_1, X_2, ..., X_n$ độc lập: $Var(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} Var(X_i)$

Hiệp phương sai và hệ số tương quan

Định nghĩa 2.3

Ma trận hiệp phương sai của biến ngẫu nhiên hai chiều (X, Y) được xác định bởi

$$\Gamma = \begin{bmatrix} cov(X,X) & cov(X,Y) \\ cov(Y,X) & cov(Y,Y) \end{bmatrix} = \begin{bmatrix} VX & cov(X,Y) \\ cov(X,Y) & VY \end{bmatrix}$$

Định nghĩa 2.4

Hệ số tương quan của hai biến ngẫu nhiên X và Y, ký hiệu là ρ_{XY} và được xác định theo công thức

$$\boxed{\rho_{XY} = \frac{cov(X,Y)}{\sqrt{VX.VY}}} \qquad (2.4)$$

Chú ý 2.2

- $|\rho_{XY}| \leq 1$.
- Nếu $\rho_{XY} = \pm 1$ ta nói hai biến ngẫu nhiên X và Y có quan hệ tuyến tính.
- Nếu $\rho_{XY} = 0$ ta nói hai biến ngẫu nhiên X và Y là không tương quan.

Hiệp phương sai và hệ số tương quan

Định nghĩa 2.3

Ma trận hiệp phương sai của biến ngẫu nhiên hai chiều (X, Y) được xác định bởi

$$\Gamma = \begin{bmatrix} cov(X,X) & cov(X,Y) \\ cov(Y,X) & cov(Y,Y) \end{bmatrix} = \begin{bmatrix} VX & cov(X,Y) \\ cov(X,Y) & VY \end{bmatrix}$$

Định nghĩa 2.4

Hệ số tương quan của hai biến ngẫu nhiên X và Y, ký hiệu là ρ_{XY} và được xác định theo công thức

$$\boxed{\rho_{XY} = \frac{cov(X,Y)}{\sqrt{VX.VY}}} \quad (2.4)$$

Chú ý 2.2

- $|\rho_{XY}| \leq 1$.
- Nếu $\rho_{XY} = \pm 1$ ta nói hai biến ngẫu nhiên X và Y có quan hệ tuyến tính.
- Nếu $\rho_{XY} = 0$ ta nói hai biến ngẫu nhiên X và Y là không tương quan.

Nội dung

Hàm của một biến ngẫu nhiên

Nếu ta xác định là một hàm của biến ngẫu nhiên X thì Z trở thành một biến ngẫu nhiên mới. Ta sẽ tìm hàm phân phối xác suất cho Z trong một số trường hợp đơn giản.

Định nghĩa 3.1

Cho biến ngẫu nhiên X có hàm phân phối xác suất. Khi đó hàm phân phối xác suất của Z được xác định theo cách sau:

$$F_Z(z) = P(Z < z) = P\left(g(X) < z\right) = P(X \in D), \tag{3.5}$$

trong đó $D = \{x|g(x) < z\}$.
Tuy nhiên tùy vào từng bài có thể có các cách giải ngắn hơn.

Hàm của một biến ngẫu nhiên

Ví dụ 5

Cho biến ngẫu nhiên X có bảng phân phối xác suất

X	-1	0	1	2	3
$P(X=x)$	0.1	0.2	0.3	0.2	0.2

Xác định luật phân phối xác suất của $Z = X^2$ và tìm kỳ vọng của Z.

Giải

Ta có $X \in \{-1, 0, 1, 2, 3\}$, suy ra $Z \in \{0, 1, 4, 9\}$ với các xác suất tương ứng:

$$P(Z=0) = P(X=0) = 0.2; \qquad P(Z=1) = P(X=1) + P(X=-1) = 0.4;$$
$$P(X=4) = P(X=2) = 0.2; \qquad P(Z=9) = P(X=3) = 0.2.$$

Z	0	1	4	9
$P(Z=z)$	0.2	0.4	0.2	0.2

Kỳ vọng $EZ = \sum_i z_i p_i = 3.$

Hàm của một biến ngẫu nhiên

Ví dụ 5

Cho biến ngẫu nhiên X có bảng phân phối xác suất

X	-1	0	1	2	3
$P(X=x)$	0.1	0.2	0.3	0.2	0.2

Xác định luật phân phối xác suất của $Z = X^2$ và tìm kỳ vọng của Z.

Giải

Ta có $X \in \{-1, 0, 1, 2, 3\}$, suy ra $Z \in \{0, 1, 4, 9\}$ với các xác suất tương ứng:

$$P(Z=0) = P(X=0) = 0.2; \qquad P(Z=1) = P(X=1) + P(X=-1) = 0.4;$$
$$P(X=4) = P(X=2) = 0.2; \qquad P(Z=9) = P(X=3) = 0.2.$$

Z	0	1	4	9
$P(Z=z)$	0.2	0.4	0.2	0.2

Kỳ vọng $EZ = \sum_i z_i p_i = 3$.

Hàm của một biến ngẫu nhiên

Ví dụ 6

Thanh AB dài 10cm bỗng nhiên bị gãy ở một điểm C bất kỳ. Hai đoạn AC và BC được dùng làm hai cạnh của một hình chữ nhật. Tìm hàm phân phối xác suất của biến ngẫu nhiên chỉ diện tích hình chữ nhật đó.

Giải

Gọi X là biến ngẫu nhiên chỉ độ dài đoạn AC, ta có $X \sim U(0;10)$. Gọi Y là biến ngẫu nhiên chỉ diện tích hình chữ nhật, ta có $Y = X(10-X)$. Do
$X \in (0;10) \Rightarrow Y = X(10-X) \in (0;25)$. Vậy ta có hàm phân phối xác suất của Y là
$F_Y(y) = \begin{cases} 0, & y \le 0 \\ 1, & y > 25 \end{cases}$.
Với $0 < y \le 25$ ta có

$$\begin{aligned} F_Y(y) &= P(Y<y) = P(X(10-X)<y) = P\left(X^2 - 10X + y > 0\right) \\ &= P\left(X < 5 - \sqrt{25-y}\right) + P\left(X > 5 + \sqrt{25-y}\right) \\ &= P\left(0 < X < 5 - \sqrt{25-y}\right) + P\left(10 > X > 5 + \sqrt{25-y}\right) = \frac{5-\sqrt{25-y}}{5}. \end{aligned}$$

Hàm của hai biến ngẫu nhiên

Xét biến ngẫu nhiên $Z = g(X, Y)$, trong đó (X, Y) là biến ngẫu nhiên hai chiều đã biết luật phân phối. Ta sẽ xét luật phân phối xác suất của Z trong một số trường hợp đơn giản theo cách sau:

$$F_Z(z) = P(Z < z) = P\left(g(X, Y) < z\right) = P\left((X, Y) \in D\right),$$

trong đó $D\left\{(x, y)|g(x, y) < z\right\}$.
Đối với biến ngẫu nhiên hai chiều liên tục (X, Y) với hàm mật độ đồng thời $f(x, y)$ ta có

$$P\left((X, Y) \in D\right) = \iint\limits_D f(x, y)dxdx,$$

đồng thời kỳ vọng

$$EZ = E\left(g(X, Y)\right) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} g(x, y).f(x, y)dxdy.$$

Hàm của hai biến ngẫu nhiên

Ví dụ 7

Hai người bạn hẹn gặp nhau ở công viên trong khoảng thời gian từ 17h đến 18h. Họ hẹn nhau nếu người nào đến trước thì sẽ đợi người kia trong vòng 10 phút. Sau 10 phút đợi nếu không gặp sẽ về. Thời điểm đến của hai người là ngẫu nhiên và độc lập với nhau trong khoảng thời gian trên. Tính xác suất hai người gặp được nhau.

Giải

Quy gốc thời gian về lúc 17h. Gọi X, Y là biến ngẫu nhiên chỉ thời điểm người A, B đến, ta có $X, Y \sim U(0;60)$. Do X, Y độc lập nên chúng có hàm mật độ đồng thời $f(x,y) = \begin{cases} \dfrac{1}{3600}, & (x,y) \in [0;60]^2 \\ 0, & \text{ngược lại} \end{cases}$. Gọi Z là biến ngẫu nhiên chỉ khoảng thời gian giữa thời điểm hai người đến. Ta có $Z = |X - Y|$. Khi đó, xác suất hai người gặp nhau là

$$P(Z < 10) = P\left(|X - Y| < 10\right) = P\left((X,Y) \in D\right),$$

trong đó D là giao miền $|X - Y| < 10$ và hình vuông $[0;60]^2$. Vậy

$$P(Z < 10) = \frac{S_D}{\ldots} = \frac{1100}{\ldots} = \frac{11}{\ldots}$$

Nội dung

Luật số lớn

Bất đẳng thức Trebyshev

Định lý 1: Cho Y là biến ngẫu nhiên không âm. Khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$P(Y \geq \epsilon) < \frac{E(Y^2)}{\epsilon^2}$$

Chứng minh

Ta chứng minh cho trường hợp Y là biến ngẫu nhiên liên tục.

$$P(Y \geq \epsilon) = \int_{\epsilon}^{+\infty} f(y)dy = \frac{1}{\epsilon^2}\int_{\epsilon}^{+\infty} \epsilon^2.f(y)dy \leq \frac{1}{\epsilon^2}\int_{\epsilon}^{+\infty} y^2.f(y)dy$$

$$\leq \frac{1}{\epsilon^2}\int_{0}^{+\infty} y^2.f(y)dy = \frac{E(Y^2)}{\epsilon^2}$$

Tuy nhiên dấu bằng không thể đồng thời xảy ra ở cả 2 dấu $\leq$ nên ta có ĐPCM.

Luật số lớn

Bất đẳng thức Trebyshev

Định lý 2: Cho X là biến ngẫu nhiên có $EX = \mu, VX = \sigma^2$ hữu hạn. Khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$P(|X - \mu| \geq \epsilon) < \frac{\sigma^2}{\epsilon^2}$$

hay tương đương

$$P(|X - \mu| \leq \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

Chứng minh

Ta chứng minh cho trường hợp X là biến ngẫu nhiên liên tục.
Ta chỉ cần đặt $Y = |X - \mu|$, lập tức áp dụng định lý 1 ta có ĐPCM.

Luật số lớn

Áp dụng định lý 2 với $X = \frac{1}{n}\sum\limits_{i=1}^{n} X_i$ ta có luật số lớn Trebyshev

Luật số lớn Trebyshev

Nếu dãy các biến ngẫu nhiên $X_1, X_2, ...X_n, ...$ độc lập, có kỳ vọng hữu hạn và phương sai bị chặn ($VX_i \leq C$ với C là hằng số), khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$\lim_{n\to+\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i| < \epsilon) = 1$$

Hệ quả

Nếu dãy các biến ngẫu nhiên $X_1, X_2, ...X_n, ...$ độc lập, có cùng kỳ vọng ($EX_i = \mu$) và phương sai bị chặn ($VX_i \leq C$ với C là hằng số), khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$\lim_{n\to+\infty} P(|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \epsilon) = 1$$

Luật số lớn Bernoulli

Áp dụng luật số lớn Trebyshev với trường hợp $X_i \sim B(1, p)$ chính là số lần xảy ra A trong phép thử thứ i ta có luật số lớn Bernoulli.

Luật số lớn Bernoulli

Xét n phép thử độc lập, cùng điều kiện.
Trong mỗi phép thử, xác suất xảy ra A luôn là p.
m là số lần xảy ra A trong n phép thử.
khi đó với $\epsilon > 0$ tuỳ ý cho trước ta có:

$$\lim_{n\to+\infty} P(|\frac{m}{n} - p| < \epsilon) = 1$$

Với luật số lớn Bernoulli ta đã chứng minh được điều thừa nhận trong phần ĐỊNH NGHĨA XÁC SUẤT THEO THỐNG KÊ, đó là với $n \to +\infty$ thì $\frac{m}{n} \to p$

Định lý giới hạn trung tâm

Định lý giới hạn trung tâm

Giả sử $\{X_n\}$ là dãy biến ngẫu nhiên độc lập cùng phân phối với $EX_i = \mu, VX_i = \sigma^2$.
Đặt $\overline{X_n} = \sum_{i=1}^{n} X_i$. Khi đó với n đủ lớn ta có:

$$\overline{X_n} \sim N(\mu, \frac{\sigma^2}{n})$$

hay là

$$\frac{\overline{X_n} - \mu}{\sigma}\sqrt{n} \sim N(0;1)$$