XÁC SUẤT THỐNG KÊ

Tôn Thất Tú

Đà Nẵng, 2019

Chương 2. Biến ngẫu nhiên

1. Định nghĩa

- Cho không gian xác suất $(\Omega, \mathcal{F}, P)$. Hàm X xác định trên không gian mẫu Ω và nhận giá trị trong $\mathbb{R}$ được gọi là *biến ngẫu nhiên* nếu với mọi $x \in \mathbb{R}$, tập hợp các kết quả $\{\omega : X(\omega) < x\}$ lập thành một biến cố ngẫu nhiên.
- Tập hợp các giá trị của X được gọi là miền giá trị của X, kí hiệu $X(\Omega)$.
- Nói một cách trực quan, biến ngẫu nhiên là một đại lượng có thể nhận giá trị này hay giá trị khác phụ thuộc vào kết quả của phép thử.

Ví dụ 1

- Gieo ngẫu nhiên 3 lần một đồng xu. Gọi X là số lần mặt sấp xuất hiện. Khi đó X là một biến ngẫu nhiên nhận các giá trị 0, 1, 2 và 3.
- Gọi Y là số người đến đổ xăng ở cửa hàng AB trong một ngày. Khi đó Y là biến ngẫu nhiên nhận các giá trị 0, 1, 2,. . . .
- Gọi Z là nhiệt độ trung bình trong một ngày (đơn vị: độ C). Khi đó Z là biến ngẫu nhiên nhận các giá trị trong khoảng $(-\infty, +\infty)$.

Phân loại

- BNN rời rạc: BNN có tập giá trị có số lượng hữu hạn hoặc vô hạn đếm được.
- BNN liên tục: BNN thỏa các điều kiện sau:
+ tập giá trị tạo thành 1 đoạn, khoảng hoặc hợp các đoạn, khoảng.
+ Với mọi c ta có $P(X = c) = 0$.
Chẳng hạn ở Ví dụ 1, X, Y là các BNN rời rạc, còn Z là BNN liên tục.

2. Hàm phân phối

Định nghĩa: Hàm số thực $F_X(x) = P(X < x),\ x \in \mathbf{R}$ được gọi là hàm phân phối của biến ngẫu nhiên X.
Nhận xét: Hàm phân phối $F_X(x)$ chính là xác suất X nhận giá trị trong khoảng $(-\infty, x)$.

Tính chất:
i) $0 \leq F_X(x) \leq 1, \forall x \in \mathbb{R}$
ii) $F_X(x)$ đơn điệu không giảm với mọi $x \in \mathbb{R}$
iii) $F_X(x)$ liên tục trái với mọi x, tức là

$$\lim_{x \to x_0^-} F_X(x) = F_X(x_0), \forall x_0 \in \mathbb{R}$$

iv) $\lim\limits_{x \to +\infty} F_X(x) = 1, \ \lim\limits_{x \to -\infty} F_X(x) = 0$

Nhận xét:
i) $P(X \geq a) = 1 - F(a)$
ii) $P(a \leq X < b) = F(b) - F(a)$

Ví dụ 2

Cho biến ngẫu nhiên X có hàm phân phối

$$F(x) = a + b.\arctan x, \, x \in \mathbf{R}$$

a) Tìm a và b.
b) Tìm x sao cho: $P(X \geq 1 - x) = 1/4$

Giải. a. Ta có:

$$\begin{cases} \lim\limits_{x \to +\infty} F(x) = 1 \\ \lim\limits_{x \to -\infty} F(x) = 0 \end{cases} \Leftrightarrow \begin{cases} a + \frac{b\pi}{2} = 1 \\ a - \frac{b\pi}{2} = 0 \end{cases} \Leftrightarrow \begin{cases} a = 1/2 \\ b = 1/\pi \end{cases}$$

b. $P(X \geq 1 - x) = 1 - P(X < 1 - x) = 1 - F(1 - x)$
$= 1 - \left[1/2 + 1/\pi * \arctan(1 - x)\right] = 1/2 - 1/\pi * \arctan(1 - x) = 1/4.$
Từ đó: $\arctan(1 - x) = \pi/4$ hay $x = 0$.

Ví dụ 3

Cho biến ngẫu nhiên X có hàm phân phối

$$F(x) = \begin{cases} 0, & x < 0 \\ ax^2 + b, & 0 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

a) Tìm a và b.
b) Tìm hàm phân phối của $Y = 2X + 1$.

Giải. a. Vì $F(x)$ liên tục trái nên

$$\begin{cases} \lim\limits_{x \to 0^-} F(x) = F(0) \\ \lim\limits_{x \to 2^-} F(x) = F(2) \end{cases} \Leftrightarrow \begin{cases} 0 = b \\ 4a + b = 1 \end{cases} \Leftrightarrow \begin{cases} a = 1/4 \\ b = 0 \end{cases}$$

b. Theo định nghĩa:

$F_Y(y) = P(Y < y) = P(2X + 1 < y)$

$= P(X < (y-1)/2) = F((y-1)/2)$

$$= \begin{cases} 0, & (y-1)/2 < 0 \\ 1/4 * [(y-1)/2]^2, & 0 \le (y-1)/2 < 2 \\ 1, & (y-1)/2 \ge 2 \end{cases}$$

$$= \begin{cases} 0, & y < 1 \\ 1/16 * (y-1)^2, & 1 \le y < 5 \\ 1, & y \ge 5 \end{cases}$$

3. Biến ngẫu nhiên rời rạc

Cho X là biến ngẫu nhiên rời rạc với tập giá trị $X(\Omega)$. Khi đó, hàm

$$p(x) = \begin{cases} P(X = x), & x \in X(\Omega), \\ 0, & x \notin X(\Omega), \end{cases}$$

được gọi là hàm khối xác suất (probability mass function).
Trong trường hợp $X(\Omega) = \{x_1, ..., x_n\}$ hữu hạn và $p_i = P(X = x_i)$, ta có bảng phân phối xác suất:

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Nhận xét: i) $\sum p_i = 1$
ii) Hàm phân phối của X sẽ là

$$F_X(x) = P(X < x) = \sum_{x_i < x} P(X = x_i)$$

Ví dụ 4

Một lô sản phẩm có 12 sản phẩm, trong đó có 8 chính phẩm và 4 phế phẩm. Lấy ngẫu nhiên 2 sản phẩm. Gọi X là số chính phẩm trong 2 sản phẩm lấy ra. Tìm phân phối của X, xác định hàm phân phối và tính xác suất $P(1 \le X < 3)$.

Giải. Ta có X là biến ngẫu nhiên rời rạc nhận các giá trị: 0, 1, 2.

$$P(X = 0) = C_4^2/C_{12}^2 = 1/11$$

$$P(X = 1) = C_8^1 C_4^1/C_{12}^2 = 16/33$$

$$P(X = 2) = C_8^2/C_{12}^2 = 14/33$$

Bảng phân phối xác suất:

X	0	1	2
P	1/11	16/33	14/33

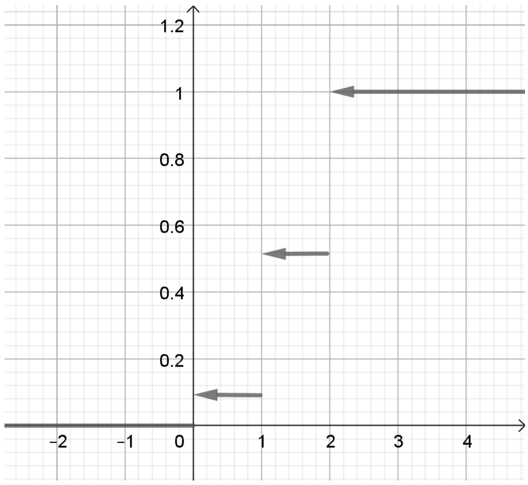
Bảng phân phối xác suất:

X	0	1	2
P	1/11	16/33	14/33

Hàm phân phối:

$$F_X(x) = P(X < x) = \sum_{x_i < x} P(X = x_i)$$

$$= \begin{cases} 0, & x \le 0 \\ 1/11, & 0 < x \le 1 \\ 1/11 + 16/33, & 1 < x \le 2 \\ 1/11 + 16/33 + 14/33, & x > 2 \end{cases} = \begin{cases} 0, & x \le 0 \\ 1/11, & 0 < x \le 1 \\ 19/33, & 1 < x \le 2 \\ 1, & x > 2 \end{cases}$$



Hình 1. Đồ thị hàm phân phối

Xác suất: $P(1 \le X < 3) = P(X = 1) + P(X = 2) = 16/33 + 14/33 = 10/11.$

Ví dụ 5

Một xạ thủ có 4 viên đạn, xạ thủ đó bắn từng phát độc lập cho đến khi có phát trúng hoặc hết đạn thì thôi. Gọi X là số viên đạn đã bắn. Lập bảng phân phối xác suất và tìm hàm phân phối của X, biết rằng xác suất bắn trúng mỗi lần đều bằng 0,7.

Giải. Ta có X là biến ngẫu nhiên nhận các giá trị:1,2,3,4.
Gọi A_i là biến cố viên đạn thứ i trúng đích, $i = 1, 2, 3, 4$. Ta có các biến cố A_i độc lập.
$P(X = 1) = P(A_1) = 0,7$; $P(X = 2) = P(\bar{A}_1 A_2) = 0,3 * 0,7 = 0,21$
$P(X = 3) = P(\bar{A}_1 \bar{A}_2 A_3) = 0,3 * 0,3 * 0,7 = 0,063$
$P(X = 4) = P(\bar{A}_1 \bar{A}_2 \bar{A}_3) = 0,3 * 0,3 * 0,3 = 0,027$
Bảng phân phối xác suất và hàm phân phối:

X	1	2	3	4
P	0,7	0,21	0,063	0,027

$$F_X(x) = \sum_{x_i < x} P(X = x_i) = \begin{cases} 0, & x \le 1 \\ 0,7, & 1 < x \le 2 \\ 0,91, & 2 < x \le 3 \\ 0,973, & 3 < x \le 4 \\ 1, & x > 4 \end{cases}$$

Ví dụ 6

Cho 2 hộp chứa bi. Hộp 1 chứa 2T và 8Đ. Hộp 2 chứa 3T và 6Đ. Lấy ngẫu nhiên 1 viên từ hộp 1 chuyển sang hộp 2, sau đó từ hộp 2 lấy ngẫu nhiên 1 viên. Gọi X là số bi trắng lần 2 lấy được. Lập bảng phân phối xác suất của X.

Giải. Gọi H_1, H_2 là biến cố lần 1 lấy bi trắng, đen. Ta có $\{H_1, H_2\}$ là nhóm đầy đủ.
Từ giả thiết, X là biến ngẫu nhiên rời rạc nhận các giá trị: 0, 1.
$P(X=0) = P(H_1)P(X=0|H_1) + P(H_2)P(X=0|H_2)$
$= 2/10 * 6/10 + 8/10 * 7/10 = 0,68$
$P(X=1) = P(H_1)P(X=1|H_1) + P(H_2)P(X=1|H_2)$
$= 2/10 * 4/10 + 8/10 * 3/10 = 0,32$
Bảng phân phối xác suất:

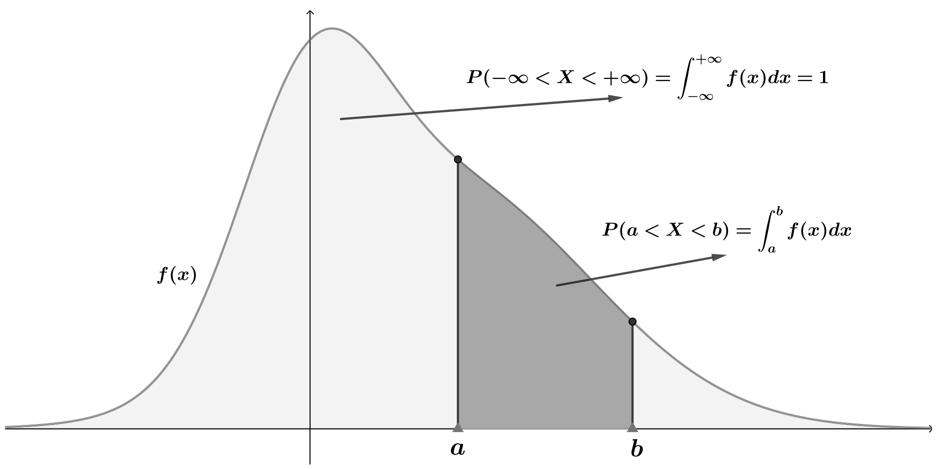
X	0	1
P	0,68	0,32

4. Biến ngẫu nhiên liên tục

Cho biến ngẫu nhiên liên tục X với hàm phân phối $F_X(x)$. Khi đó tồn tại hàm $f(x)$ sao cho ta có biểu diễn:

$$F_X(x) = \int_{-\infty}^{x} f(t)dt,\ x \in \mathbb{R}$$

Hàm $f(x)$ được gọi là hàm mật độ của X.
Tính chất của hàm mật độ:
i) $f(x) \geq 0$ ii) $\int_{-\infty}^{+\infty} f(x)dx = 1$ iii) $f(x) = F'_X(x)$
Ngược lại, một hàm số thực $f(x)$ thỏa 2 điều kiện i) và ii) ở trên thì nó là hàm mật độ của một biến ngẫu nhiên nào đó.
Nhận xét: i) $P(X = c) = 0$ với mọi hằng số c
ii) $P(a \leq X \leq b) = \ldots = P(a < X < b) = F_X(b) - F_X(a) = \int_a^b f(x)dx$
iii) Nếu $f(x)$ là hàm mật độ thì hàm phân phối: $F_X(x) = \int_{-\infty}^{x} f(t)dt$.



Ví dụ 7

Tuổi thọ (năm) của một thiết bị là biến ngẫu nhiên X có hàm mật độ:

$$f(x) = \begin{cases} 0.6e^{-0.6x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

Chọn ngẫu nhiên 1 thiết bị. Tính xác suất tuổi thọ của thiết bị này:
a) Nhỏ hơn 1 năm.   b) Lớn hơn 2 năm.

Giải. a. Xác suất tuổi thọ thiết bị này nhỏ hơn 1 năm:

$$P(0 < X < 1) = \int_0^1 f(x)dx = \int_0^1 0,6e^{-0,6x}dx = 1 - e^{-0,6}$$

b. Xác suất tuổi thọ thiết bị này lớn hơn 2 năm:

$$P(X > 2) = \int_2^{+\infty} f(x)dx = \int_2^{+\infty} 0,6e^{-0,6x}dx = -e^{-0,6x}\Big|_2^{+\infty} = e^{-1,2}$$

Ví dụ 8

Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} a\,x, & x \in [0,1] \\ 0, & x \notin [0,1] \end{cases}$$

a) Tìm hệ số a và hàm phân phối $F_X(x)$.
b) Thực hiện 10 phép thử độc lập. Tính xác suất có 3 lần xảy ra biến cố $(1/2 < X < 1)$.

Giải. a. Theo tính chất: $\int_{-\infty}^{+\infty} f(x)dx = \int_0^1 axdx = \frac{a}{2} = 1 \Leftrightarrow a = 2$.
Hàm phân phối:

$$F_X(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \int_{-\infty}^{x} 0dt, & x < 0 \\ \int_{-\infty}^{0} 0dt + \int_0^x 2tdt, & x \in [0,1] \\ \int_{-\infty}^{0} 0dt + \int_0^1 2tdt + \int_1^x 0dt, & x > 1 \end{cases}$$

$$F_X(x) = \begin{cases} 0, & x < 0 \\ x^2, & x \in [0,1] \\ 1, & x > 1 \end{cases}$$

b. Ta có mô hình Bernoulli với $n = 10$ và $p = P(1/2 < X < 1)$.

$$p = \int_{1/2}^{1} f(x)dx = \int_{1/2}^{1} 2xdx = 3/4$$

Xác suất cần tìm:

$$p_{10}(3) = C_{10}^3 \left(\frac{3}{4}\right)^3 \left(\frac{1}{4}\right)^7$$

Ví dụ 9

Cho biến ngẫu nhiên liên tục X có hàm phân phối:

$$F(x) = \begin{cases} 1 - a/x^3, & x \geq 2, \\ 0, & x < 2 \end{cases}$$

a. Tìm a, hàm mật độ $f(x)$. b. Tìm x thỏa $P(X > x) = 1/4$.

Giải. a. Vì $F(x)$ liên tục trái tại $x = 2$ nên:

$$\lim_{x \to 2^-} F(x) = F(2) \Leftrightarrow 1 - a/8 = 0 \Leftrightarrow a = 8$$

Hàm mật độ:

$$f(x) = F'(x) = \begin{cases} 24/x^4, & x \geq 2, \\ 0, & x < 2 \end{cases}$$

b. Ta có:

$$P(X > x) = 1 - P(X \leq x) = 1 - F(x) = 1/4 \Leftrightarrow F(x) = 3/4$$

$$\Leftrightarrow \begin{cases} 1 - 8/x^3 = 3/4, \\ x \geq 2 \end{cases} \Leftrightarrow x = 2\sqrt[3]{4}$$

5. Sự độc lập của các biến ngẫu nhiên

- Hai biến ngẫu nhiên X và Y được gọi là độc lập nếu các biến cố $(X < a)$ và $(Y < b)$ độc lập với mọi cặp giá trị (a, b), tức là:

$$P(X < a, Y < b) = P(X < a)P(Y < b)$$

- Nhóm n biến ngẫu nhiên $\{X_1, X_2, ..., X_n\}$ được gọi là độc lập nếu các biến cố $(X_1 < a_1), ..., (X_n < a_n)$ độc lập với mọi bộ giá trị $(a_1, a_2, ..., a_n)$.

6. Các số đặc trưng

a. Kỳ vọng toán

Định nghĩa: Kỳ vọng toán của biến ngẫu nhiên X, kí hiệu $E(X)$, là một số được xác định như sau:
- Nếu X có phân phối rời rạc với phân phối xác suất $P(X = x_k) = p_k$, $k = 1, 2, ...$ thì

$$E(X) = \sum_{i=1}^{+\infty} x_i p_i$$

- Nếu X là biến ngẫu nhiên liên tục với hàm mật độ $f(x)$ thì

$$E(X) = \int_{-\infty}^{+\infty} x\, f(x) dx$$

Trong trường hợp $\sum\limits_{i=1}^{+\infty}|x_i|p_i$ hoặc $\int\limits_{-\infty}^{+\infty}|x|\,f_X(x)dx$ phân kì thì ta nói biến ngẫu nhiên X không có kỳ vọng.

Tính chất:

i) $E(c) = c$ với c là hằng số.
ii) $E(cX) = cE(X)$ với c là hằng số.
iii) $E(X \pm Y) = E(X) \pm E(Y)$ với mọi biến ngẫu nhiên X, Y.
iv) $E(XY) = E(X).E(Y)$ nếu X, Y độc lập.

Ví dụ 10

Tính kỳ vọng của biến ngẫu nhiên X trong hai trường hợp sau:
a. X có phân phối rời rạc với bảng phân phối xác suất:

X	1	2	3
P	0,2	0,7	0,1

b. Khi X có phân phối liên tục với hàm mật độ:

$$f(x) = \begin{cases} \sin x, & x \in [0, \pi/2] \\ 0, & x \notin [0, \pi/2] \end{cases}$$

Giải.

a. $E(X) = \sum_i p_i x_i = 0,2 * 1 + 0,7 * 2 + 0,1 * 3 = 1,9$
b. Ta có:

$$E(X) = \int\limits_{-\infty}^{+\infty} x\, f(x)dx = \int_0^{\pi/2} x \sin x dx = \int_0^{\pi/2} xd(-\cos x)$$

$$= -x\cos x\Big|_0^{\pi/2} + \int_0^{\pi/2} \cos x dx = 1$$

Nhận xét:

i) Kì vọng của biến ngẫu nhiên thể hiện giá trị trung bình của biến ngẫu nhiên đó, tức là khi thực hiện một số lớn lần các phép thử thì giá trị trung bình thu được của các kết quả sẽ xấp xỉ với kì vọng.
ii) Xét trò chơi may rủi với số tiền đặt cược và xác suất thắng trong mỗi ván không đổi. Trò chơi được gọi là công bằng (có lợi hay có hại) đối với người chơi nếu kỳ vọng số tiền nhận được trong mỗi lần chơi bằng (lớn hơn hay bé hơn) số tiền đặt cược trong mỗi ván chơi.
iii) Nếu $g(x)$ là hàm liên tục thì $g(X)$ cũng là biến ngẫu nhiên và kì vọng của nó được tính là:

- $Eg(X) = \sum\limits_{i=1}^{+\infty} g(x_i).p_i$ nếu X là biến ngẫu nhiên rời rạc.
- $Eg(X) = \int\limits_{-\infty}^{+\infty} g(x)\, f_X(x)dx$ nếu X là biến ngẫu nhiên liên tục với hàm mật độ $f(x)$.

Ví dụ 11

Trong hộp có 7 bút xanh và 3 bút đỏ. Một sinh viên rút ngẫu nhiên 2 bút để mua. Giá bút xanh và đỏ lần lượt 2000 đồng và 3000 đồng. Tìm số tiền trung bình sinh viên này phải trả.

Giải. Gọi X (ngàn đồng) là số tiền sinh viên này phải trả. Ta có X nhận các giá trị: 4, 5 và 6.
$P(X=4) = C_7^2/C_{10}^2 = 7/15$
$P(X=5) = C_7^1 C_3^1/C_{10}^2 = 7/15$
$P(X=6) = C_3^2/C_{10}^2 = 1/15$
Bảng phân phối xác suất của X:

X	4	5	6
P	7/15	7/15	1/15

Số tiền trung bình phải trả:

$$E(X) = \sum p_i x_i = 7/15 * 4 + 7/15 * 5 + 1/15 * 6 = 4,6 \text{(ngàn đồng)}$$

Ví dụ 12

Thời gian xếp hàng chờ mua hàng của khách hàng là 1 biến ngẫu nhiên T (phút) có hàm mật độ:

$$f(t) = \begin{cases} a.t^3, & t \in [0,3] \\ 0, & t \notin [0,3] \end{cases}$$

Tìm a và tính thời gian chờ trung bình của khách hàng.

Giải. Theo tính chất của hàm mật độ:

$$\int_{-\infty}^{+\infty} f(t)dt = \int_0^3 at^3 dt = 81a/4 = 1 \Leftrightarrow a = 4/81$$

Thời gian chờ trung bình của khách hàng:

$$E(T) = \int_{-\infty}^{+\infty} tf(t)dt = \int_0^3 t * \frac{4}{81} t^3 dt = 2,4 \text{(phút)}$$

Ví dụ 13

Một người tham gia trò chơi may rủi với tiền cược mỗi ván là 10000 đồng. Người này tung ngẫu nhiên 2 đồng xu, nếu được i mặt sấp người này thu về $(i+1) * 5000$ đồng, $i = 1, 2$. Ngược lại, người này sẽ mất tiền. Hỏi người này có nên chơi trò này thường xuyên hay không?

Giải. Gọi X (ngàn đồng) là số tiền người này nhận được trong 1 lần chơi. Ta có X nhận các giá trị: 0, 10 và 15. Bảng phân phối của X:

X	0	10	15
P	1/4	1/2	1/4

Do đó: $E(X) = \sum p_i x_i = 8,75$ (ngàn đồng). Vì số tiền này nhỏ hơn số tiền đặt cược nên nếu người này chơi càng nhiều thì thua càng lớn.
Vậy, người này không nên chơi trò này thường xuyên.

b. Phương sai

Định nghĩa: Giả sử biến ngẫu nhiên X có kỳ vọng EX. Nếu tồn tại kỳ vọng $E(X-EX)^2$ thì ta gọi giá trị này là *phương sai* của X, kí hiệu là DX (hay $Var(X), V(X)$), tức là:

$$DX = E(X - EX)^2$$

Giá trị $\sigma(X) = \sqrt{DX}$ được gọi là *độ lệch chuẩn*.

Tính chất:

i) $D(c) = 0$ với c là hằng số

ii) $D(cX) = c^2 D(X)$ với c là hằng số

iii) $DX = E(X^2) - (EX)^2$

iv) $D(X \pm Y) = D(X) + D(Y)$ với X, Y độc lập

Nhận xét: Trong thực hành, ta hay sử dụng công thức sau: $D(X) = E(X^2) - (EX)^2$, trong đó:

$$E(X^2) = \begin{cases} \sum p_i x_i^2, & \text{nếu X là biến ngẫu nhiên rời rạc} \\ \int\limits_{-\infty}^{+\infty} x^2 f(x)dx, & \text{nếu X là biến ngẫu nhiên liên tục} \end{cases}$$

Ý nghĩa:

- Phương sai cũng như độ lệch chuẩn là đại lượng đặc trưng cho mức độ tập trung của các giá trị của biến ngẫu nhiên X quanh kỳ vọng EX. Phương sai càng nhỏ thì các giá trị khuếch tán càng gần với kỳ vọng hơn.
- Trong kỹ thuật phương sai đặc trưng cho sai số của các thiết bị hoặc của các phép đo. Trong quản lý và kinh doanh, nó đặc trưng cho mức độ rủi ro của các quyết định.

Ví dụ 14

Cho biến ngẫu nhiên X có phân phối

X	1	2	a
P	b	0.3	0.5

a. Tìm a và b, biết $E(X) = 2,3$. *b. Tính $D(2X-3)$, $D(X^2)$*

Giải. a. Ta có hệ:

$$\begin{cases} \sum p_i = 1 \\ E(X) = \sum p_i x_i = 2,3 \end{cases} \Leftrightarrow \begin{cases} b + 0,3 + 0,5 = 1 \\ b + 0,6 + 0,5a = 2,3 \end{cases} \Leftrightarrow \begin{cases} a = 3 \\ b = 0,2 \end{cases}$$

b. Ta có:

$D(2X-3) = D(2X) + D(3) = 4D(X) = 4[E(X^2) - (E(X))^2] = 4[E(X^2) - 5,29]$

$E(X^2) = \sum p_i x_i^2 = 0,2 * 1^2 + 0,3 * 2^2 + 0,5 * 3^2 = 5,9$

$D(2X-3) = 4(5,9 - 5,29) = 2,44$

Tương tự: $D(X^2) = E(X^4) - (E(X^2))^2 = \sum p_i x_i^4 - 5,9^2 = 10,69$

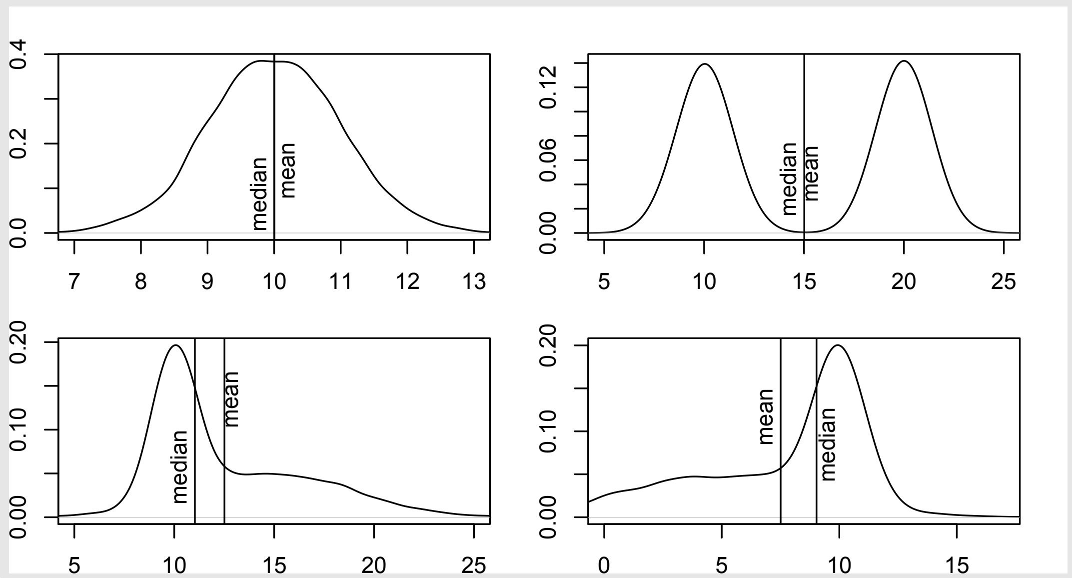
c. Trung vị (median)

Định nghĩa: Trung vị (hay Median) của biến ngẫu nhiên X, được kí hiệu $med(X)$ xác định theo hệ thức:

$$P(X < med(X)) \leq \frac{1}{2} \text{ và } P(X > med(X)) \leq \frac{1}{2}$$

Nhận xét: Theo định nghĩa trên thì X có thể có nhiều trung vị và trong trường hợp X là biến ngẫu nhiên liên tục thì $med(X)$ chính là nghiệm của phương trình

$$F_X(x) = \frac{1}{2}$$



Ví dụ 15

Tìm trung vị của biến ngẫu nhiên X biết:

a. X có bảng phân phối:

X	1	2	3
P	0.4	0.3	0.3

b. X có bảng phân phối:

X	1	2	3	4
P	0.2	0.3	0.4	0.1

c. X có hàm mật độ:

$$f(x) = \begin{cases} 2e^{-2x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

Giải. a. Ta có $med(X) = 2$ vì:

$$P(X < 2) = P(X = 1) = 0,4 \leq 1/2; P(X > 2) = P(X = 3) = 0,3 \leq 1/2$$

b. $med(X) = x_0$ với x_0 bất kì thuộc đoạn $[2,3]$.
c. Hàm phân phối:

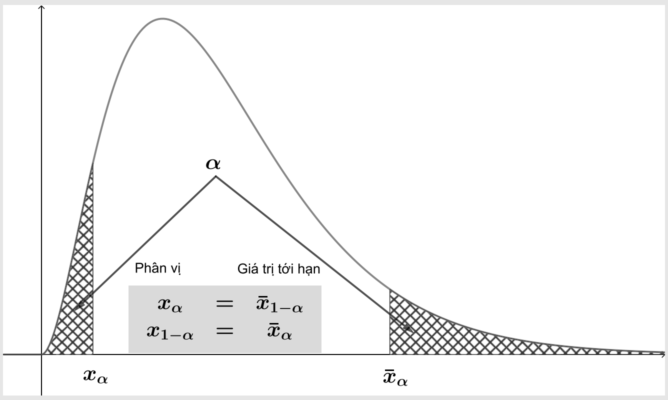
$$F(x) = \int_{-\infty}^{x} f(t)dt = \begin{cases} \int_0^x 2e^{-2t}dt, & x > 0 \\ 0, & x \leq 0 \end{cases} = \begin{cases} 1 - e^{-2x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Xét phương trình: $F(x) = 1/2$

$$\Leftrightarrow \begin{cases} 1 - e^{-2x} = 1/2 \\ x > 0 \end{cases} \quad \Leftrightarrow x = \frac{ln2}{2}$$

Vậy, $med(X) = \frac{ln2}{2}$.

d. Phân vị và giá trị tới hạn



- Phân vị mức $\alpha,\ 0 < \alpha < 1$ của phân phối tương ứng với biến ngẫu nhiên X là giá trị x_α thỏa $P(X < x_\alpha) = \alpha$.
- Giá trị tới hạn mức $\alpha,\ 0 < \alpha < 1$ của phân phối tương ứng với biến ngẫu nhiên X là giá trị $\bar{x}_\alpha$ thỏa $P(X > \bar{x}_\alpha) = \alpha$.

7. Một số phân phối quan trọng

7.1 Phân phối Bernoulli

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối Bernoulli với tham số $p \in (0,1)$, kí hiệu $X \sim Ber(p)$, nếu X có bảng phân phối xác suất:

X	0	1
P	1-p	p

Tính chất: Nếu $X \sim Ber(p)$ thì $E(X) = p$ và $DX = p(1-p)$.

7.2 Phân phối nhị thức

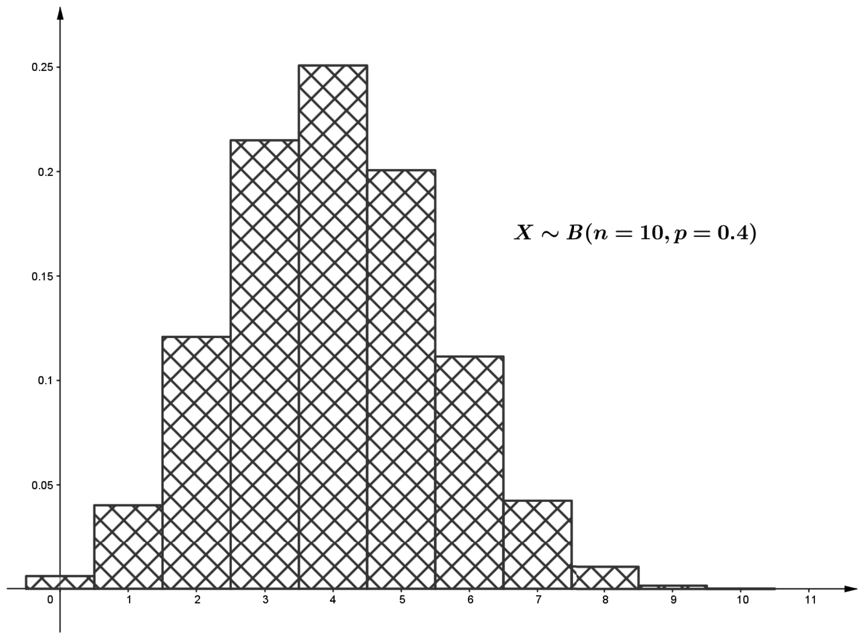
Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối nhị thức với tham số n và p, kí hiệu $X \sim B(n,p)$ nếu X nhận các giá trị $\{0,1,2,...,n\}$ với xác suất:

$$P(X=k) = p_n(k) = C_n^k p^k (1-p)^{n-k}, k = \overline{0,n}$$

Tính chất:
i) Nếu $X \sim B(n,p)$ thì $E(X) = np, D(X) = np(1-p)$.
ii) Nếu các biến ngẫu nhiên $X_i, i = \overline{1,n}$ độc lập và $X_i \sim Ber(p)$ thì

$$X = X_1 + ... + X_n \sim B(n,p)$$

Nhận xét:
i) $B(1,p)$ chính là phân phối $Ber(p)$.
ii) Xét dãy n phép thử Bernoulli với xác suất thành công là p. Lúc đó, nếu gọi X là biến ngẫu nhiên chỉ số lần thành công trong dãy n phép thử này thì $X \sim B(n,p)$.



Ví dụ 16

Một thành phố A có 70% gia đình có tivi. Chọn ngẫu nhiên 20 gia đình và gọi X là số gia đình có tivi.
a) Tính xác suất có đúng 10 gia đình có tivi.
b) Tính xác suất để có ít nhất 2 gia đình có tivi.

Giải. Ta có mô hình Bernoulli với $n = 20$ và $p = 0,7$. Do đó: $X \sim B(n = 20; p = 0,7)$.
a. Xác suất có đúng 10 gia đình có tivi:

$$P(X = 10) = p_{20}(10) = C_{20}^{10} 0,7^{10} 0,3^{10}$$

b. Xác suất có ít nhất 2 gia đình có tivi:

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - p_{20}(0) - p_{20}(1)$$

$$= 1 - C_{20}^{0} 0,7^0 0,3^{20} - C_{20}^{1} 0,7^1 0,3^{19}$$

Ví dụ 17

Một sinh viên thi vấn đáp trả lời 5 câu hỏi một cách độc lập. Khả năng trả lời đúng mỗi câu hỏi đều bằng 60%. Nếu trả lời đúng thì sinh viên được 4 điểm, ngược lại bị trừ 2 điểm.
a) Tìm xác suất để sinh viên đó trả lời đúng 3 câu.
b) Tìm số điểm trung bình mà sinh viên đó đạt được.
c) Một sinh viên khác vào thi với khả năng trả lời đúng mỗi câu đều như nhau và cho rằng số điểm trung bình đạt được không ít hơn 14. Hỏi sinh viên này phán đoán khả năng trả lời đúng mỗi câu tối thiểu là bao nhiêu ?

Giải. Mô hình Bernoulli với $n = 5$ và $p = 0,6$. Gọi X là số câu trả lời đúng của sinh viên này. Ta có: $X \sim B(n = 5; p = 0,6)$.
a. $P(X = 3) = p_5(3) = C_5^3 0,6^3 0,4^2$.
b. Gọi Y là số điểm sinh viên này đạt được.
Ta có: $Y = 4X - 2(5 - X) = 6X - 10$.
Số điểm trung bình sinh viên này đạt được:
$E(Y) = 6E(X) - 10 = 6np - 10 = 6 * 5 * 0,6 - 10 = 8$ (điểm)

c. Gọi p là xác suất trả lời đúng mỗi câu của sinh viên mới này. Gọi Z và T là số câu trả lời đúng và số điểm đạt được. Tương tự, ta có:

$$Z \sim B(n = 5; p) \text{ và } T = 6Z - 10$$

Theo giả thiết:

$$E(T) \geq 14$$

$$\Leftrightarrow 6np - 10 = 30p - 10 \geq 14 \Leftrightarrow p \geq 0,8$$

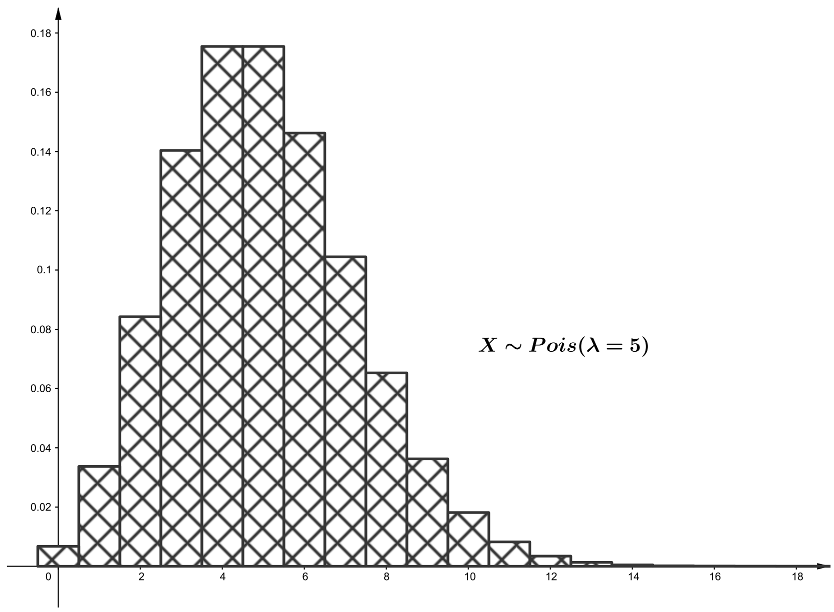
Vậy, sinh viên này dự đoán khả năng trả lời đúng tối thiểu mỗi câu là 80%.

7.3 Phân phối Poisson

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối Poisson với tham số $\lambda > 0$, kí hiệu $X \sim Pois(\lambda)$ nếu X nhận giá trị trong tập $\mathbb{N} = \{0, 1, 2, 3, ...\}$ với xác suất:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!},\ k = 0, 1, 2, ...$$

Tính chất:
i) Nếu $X \sim Pois(\lambda)$ thì $E(X) = \lambda, D(X) = \lambda$.
ii) Cho X_1, X_2 độc lập và có phân phối Poisson với tham số λ_1, λ_2. Khi đó $X = X_1 + X_2 \sim Pois(\lambda_1 + \lambda_2)$.
Nhận xét: Trong thực tế phân phối Poisson phản ánh phân phối số lượng các biến cố xuất hiện trong một khoảng thời gian (số cuộc điện thoại gọi đến tổng đài, số khách hàng đến rút tiền từ một ngân hàng,....) và có tham số tỉ lệ với độ dài khoảng thời gian đó.



Ví dụ 18

Một gara cho thuê ôtô thấy rằng số đơn đặt hàng thuê ôtô vào ngày thứ 7 là một biến ngẫu nhiên có phân phối Poisson với tham số $\lambda = 2$. Giả sử gara có 4 chiếc ôtô. Hãy tìm xác suất để:
a) Có đúng 2 ôtô được thuê.
b) Tất cả 4 ôtô đều được thuê.
c) Gara không đáp ứng được yêu cầu.

Giải. Gọi X là số đơn đặt hàng ôtô vào ngày thứ 7. Ta có $X \sim Pois(\lambda = 2)$.

a. $P(X = 2) = e^{-2}\dfrac{2^2}{2!} = 2e^{-2}$

b. $P(X \geq 4) = 1 - P(X < 4) = 1 - [P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)]$
$= 1 - e^{-2}\left(\dfrac{2^0}{0!} + \dfrac{2^1}{1!} + \dfrac{2^2}{2!} + \dfrac{2^3}{3!}\right) = 1 - 19/3e^{-2}$

c. $P(X > 4) = 1 - P(X \leq 4) = 1 - e^{-2}\left(\dfrac{2^0}{0!} + \dfrac{2^1}{1!} + \dfrac{2^2}{2!} + \dfrac{2^3}{3!} + \dfrac{2^4}{4!}\right) = 1 - 7e^{-2}$

Ví dụ 19

Quan sát tại siêu thị A thấy trung bình 4 phút có 20 khách đến mua hàng.
1) Tính xác suất để trong 7 phút có 30 khách đến siêu thị A ?
2) Tính xác suất để trong 2 phút có từ 3 đến 5 khách đến siêu thị A ?

Giải.
a. Gọi X là số khách đến siêu thị trong 7 phút. Lúc đó, X có phân phối Poisson với tham số $\lambda = 7/4 * 20 = 35$. Xác suất cần tìm:

$$P(X = 30) = e^{-35}\frac{35^{30}}{30!}$$

b. Gọi Y là số khách đến siêu thị trong 2 phút. Lúc đó, Y có phân phối Poisson với tham số $\lambda = 2/4 * 20 = 10$. Xác suất cần tìm:

$$P(3 \leq Y \leq 5) = P(Y = 3) + P(Y = 4) + P(Y = 5) = e^{-10}\left(\frac{10^3}{3!} + \frac{10^4}{4!} + \frac{10^5}{5!}\right)$$

Định lý: (Luật biến cố hiếm)

Cho $X \sim B(n;p)$. Khi p khá nhỏ và n khá lớn, ta có

$$P(X=k) \approx \frac{\lambda^k e^{-\lambda}}{k!} \text{ trong đó} \lambda = np.$$

Điều này có nghĩa là X có phân phối xấp xỉ phân phối Poisson với tham số $\lambda = np$.

Ví dụ 20

Xác suất mỗi trang giấy bị lỗi do in ấn là 0,009. Tính xấp xỉ xác suất trong một quyển sách 300 trang có nhiều nhất 3 trang bị lỗi.

Giải. Gọi X là số trang bị lỗi trong quyển sách 300 trang. Ta có: $X \sim B(n = 300; p = 0,009)$. Vì n khá lớn và p khá nhỏ nên X có phân phối xấp xỉ phân phối Poisson với tham số $\lambda = np = 2,7$. Xác suất cần tìm:

$$\begin{aligned} P(X \le 3) &= P(X=0) + P(X=1) + P(X=2) + P(X=3) \\ &\approx e^{-2,7}\left(\frac{2,7^0}{0!} + \frac{2,7^1}{1!} + \frac{2,7^2}{2!} + \frac{2,7^3}{3!}\right) \end{aligned}$$

7.4 Phân phối hình học

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối hình học với tham số $p \in (0,1)$ nếu X nhận các giá trị 1,2,3,... với xác suất:

$$P(X=k) = (1-p)^{k-1}p, \;\; k = 1,2,3,...$$

Tính chất: $E(X) = \dfrac{1}{p}, D(X) = \dfrac{1-p}{p^2}$

Nhận xét: Xét phép thử và A là một biến cố ở trong phép thử đó với xác suất xảy ra $p = P(A)$. Thực hiện độc lập và liên tiếp các phép thử cho đến khi biến cố A xuất hiện thì dừng. Gọi X là số phép thử đã thực hiện. Khi đó, X có phân phối hình học với tham số p.

7.5 Phân phối siêu bội

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối siêu bội với tham số $(\mathrm{N}, \mathrm{M}, \mathrm{n})$ với $n \le M \le N$, kí hiệu $X \sim H(N,M,n)$ nếu tập giá trị của X là $\{0,1,2,...,n\}$ và

$$P(X=k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = 0,1,...,n$$

Tính chất: $E(X) = n\frac{M}{N}$, $D(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$

Nhận xét:

i) Cho một tập có N phần tử, trong đó có M phần tử có tính chất A, $M \le N$. Chọn ngẫu nhiên lần lượt không hoàn lại n lần, mỗi lần một phần tử trong tập hợp này và gọi X là số phần tử được chọn có tính chất A. Khi đó $X \sim H(N,M,n)$.

ii) Khi giá trị N lớn và số lần lấy n nhỏ thì phương pháp lấy không hoàn lại và lấy có hoàn lại gần như không khác nhau. Vì vậy trong trường hợp N lớn và số lần lấy n nhỏ thì ta có thể xem phân phối $H(N,M,n)$ xấp xỉ phân phối $B(n,p)$ với $p = M/N$.

7.6 Phân phối đều

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối đều trên đoạn $[a, b]$, kí hiệu $X \sim U(a, b)$ nếu hàm mật độ của nó có dạng:

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b] \\ 0, & x \notin [a, b] \end{cases}$$

Tính chất: $EX = \frac{a+b}{2}$, $DX = \frac{(b-a)^2}{12}$

Nhận xét: Hàm random ở các phần mềm lập trình hay ở máy tính bỏ túi là hàm mô phỏng giá trị của biến ngẫu nhiên có phân phối đều trên đoạn $[0, 1]$.

7.7 Phân phối mũ

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối mũ với tham số $\lambda > 0$, kí hiệu $X \sim Exp(\lambda)$ nếu hàm mật độ của nó có dạng:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Tính chất: $E(X) = 1/\lambda$; $D(X) = 1/\lambda^2$

Nhận xét: Trong thực tế, phân phối mũ thường thể hiện phân phối khoảng thời gian chờ giữa các lần xảy ra biến cố hay thời gian sống của các đối tượng.

Ví dụ 21

Giả sử tuổi thọ (tính bằng năm) của một mạch điện tử trong máy tính là một biến ngẫu nhiên có phân phối mũ với tuổi thọ trung bình là 6,25 năm. Thời gian bảo hành của mạch điện tử này là 1 năm.
a) Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải bảo hành ?
b) Một công ty mua 40 mạch điện tử loại này. Tìm số mạch trung bình công ty phải bảo hành.

Giải.a. Gọi X (năm) là tuổi thọ của mạch điện tử. Theo giả thiết, $X \sim Exp(\lambda)$ với $E(X) = 1/\lambda = 6,25 \Rightarrow \lambda = 4/25$. Hàm mật độ:

$$f(x) = \begin{cases} 4/25e^{-4/25x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Tỉ lệ mạch điện tử phải bảo hành:

$$P(0 < X < 1) = \int_0^1 f(x)dx = \int_0^1 4/25e^{-4/25x}dx = 1 - e^{-4/25} \approx 0,15$$

b. Ta có mô hình Bernoulli với $n = 40$ và $p = 0,15$.
Gọi Y là số mạch công ty cần bảo hành. Ta có:

$$Y \sim B(n = 40; p = 0,15)$$

Số mạch trung bình công ty cần bảo hành:

$$E(Y) = np = 40 * 0,15 = 6$$

7.8 Phân phối chuẩn

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối chuẩn với tham số μ và σ^2, kí hiệu $X \sim N(\mu, \sigma^2)$ nếu hàm mật độ của nó có dạng:

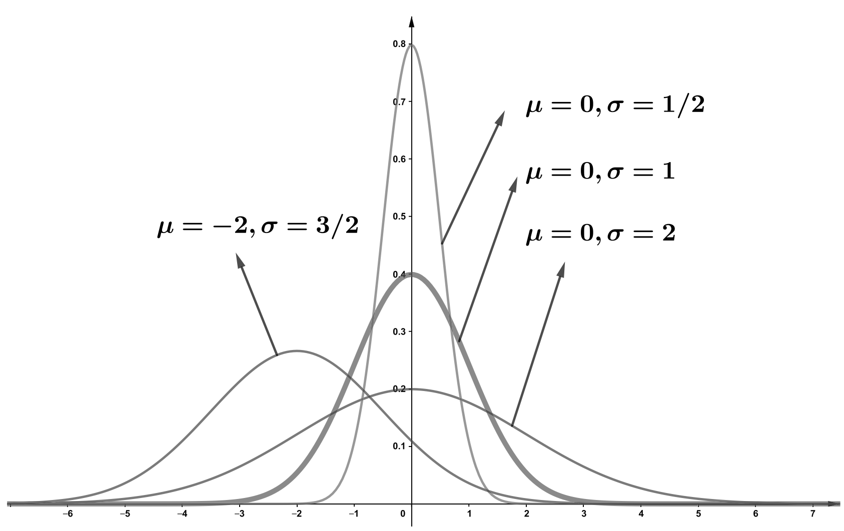
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

Khi $\mu = 0$, $\sigma = 1$ ta bảo X có **phân phối chuẩn tắc** $N(0, 1)$. Hàm mật độ $\varphi(x)$ và hàm phân phối $\Phi(x)$ tương ứng là:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, x \in \mathbb{R}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt, x \in \mathbb{R}$$

Đường cong chuẩn (đường cong hình chuông)



Nhận xét:
i) Biến ngẫu nhiên có phân phối chuẩn tắc thường được kí hiệu là Z.
ii) $\boxed{\Phi(-x) = 1 - \Phi(x), \forall x \in \mathbb{R}.}$
iii) $\Phi(x) = 1/2 + \Phi_0(x)$, trong đó $\Phi_0(x) = \dfrac{1}{\sqrt{2\pi}} \int_0^x e^{-t^2/2} dt$ - hàm Laplace.
Một vài giá trị hay sử dụng:

$$\Phi(0) = 0,5; \ \Phi(1/2) = 0,691; \ \Phi(1) = 0,841; \ \Phi(2) = 0,977$$

Giá trị tới hạn: Giá trị tới hạn mức α của phân phối chuẩn tắc được kí hiệu z_α, tức là:

$$P(Z > z_\alpha) = \alpha \text{ hay } z_\alpha = \Phi^{-1}(1 - \alpha)$$

Giá trị z_α được tra ở bảng.

Tính $\Phi(x)$ bằng máy tính Casio

1) CASIO FX570MS:
- Vào Mode tìm SD: Mode $\to$ Mode $\to$ 1 (SD);
- Shift $\to$ 3 (Distr) $\to$ 1;
- Nhập x.

2) CASIO FX570ES, FX570ES - PLUS, FX570VN - PLUS:
- Vào Mode tìm 1-Var: Mode $\to$ 3 (Stat) $\to$ 1 (1-Var)$\to$ AC
- Shift $\to$ 1(Stat) $\to$ 5 (Distr) $\to$ 1;
- Nhập x.

Ví dụ 22

Tính $\Phi(1.96), \Phi(2), \Phi(3)$

Đáp số. $\Phi(1.96) = 0,975, \Phi(2) = 0,977, \Phi(3) = 0,998$

Tính chất:
i)Nếu $X \sim N(\mu, \sigma^2)$ thì $E(X) = \mu, D(X) = \sigma^2$.
ii) Nếu $X \sim N(\mu, \sigma^2)$ thì $Y = aX + b \sim N(a\mu + b, a^2\sigma^2), a \neq 0$.
iii) Nếu $X_1, ..., X_n$ là các biến ngẫu nhiên **độc lập** có **phân phối chuẩn** $X_i \sim N(\mu_i, \sigma_i^2)$, $i = \overline{1,n}$ thì biến ngẫu nhiên $X = \lambda_1 X_1 + ... + \lambda_n X_n + C$, $\sum\limits_{i=1}^n \lambda_i^2 \neq 0$ cũng có phân phối chuẩn với

$$\begin{cases} \mu_X = E(X) = E\left(\sum\limits_{i=1}^n \lambda_i X_i + C\right) = \sum\limits_{i=1}^n \lambda_i \mu_i + C, \\ \sigma_X^2 = D(X) = D\left(\sum\limits_{i=1}^n \lambda_i X_i + C\right) = \sum\limits_{i=1}^n \lambda_i^2 \sigma_i^2 \end{cases}$$

Hệ quả: («chuẩn hóa») Nếu $X \sim N(\mu, \sigma^2)$ thì

$$\frac{X-\mu}{\sigma} \sim N(0;\ 1)$$

Ứng dụng: Cho $X \sim N(\mu, \sigma^2)$. Lúc đó:

$$P(X < a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

Ví dụ 23

Cho biến ngẫu nhiên $X \sim N(4, 3^2)$.
Tính xác suất: $P(X < 5), P(X > 1), P(|X-3| < 1)$.

Giải. Ta có:

$$P(X < 5) = \Phi\Big(\frac{5-4}{3}\Big) = 0,63$$

$$P(X > 1) = 1 - P(X \le 1) = 1 - \Phi\Big(\frac{1-4}{3}\Big) = 0,841$$

$$P(|X-3| < 1) = P(2 < X < 4) = \Phi\Big(\frac{4-4}{3}\Big) - \Phi\Big(\frac{2-4}{3}\Big) = 0,2475$$

Ví dụ 24

Giả sử số đo chiều dài của một sợi dây kim loại do một máy tự động cắt ra là một biến ngẫu nhiên chuẩn với $\mu = 10mm$, $\sigma^2 = 4mm^2$.
a) Tính xác suất lấy ra được một sợi dây có chiều dài lớn hơn 13mm.
b) Chọn ngẫu nhiên 10 sợi dây loại này. Tính xác suất có đúng 3 sợi có chiều dài từ $8,5mm$ đến $12,5mm$.

Giải. Gọi X là số đo chiều dài sợi dây kim loại. Ta có: $X \sim N(10; 2^2)$.

a. $P(X > 13) = 1 - P(X \le 13) = 1 - \Phi\Big(\dfrac{13-10}{2}\Big) = 0,067$

b. Mô hình Bernoulli với $n = 10$ và $p = P(8,5 < X < 12,5)$. Ta có:

$$p = \Phi\Big(\frac{12,5-10}{2}\Big) - \Phi\Big(\frac{8,5-10}{2}\Big) = 0,668$$

Xác suất cần tìm:

$$p_{10}(3) = C_{10}^3 0,668^3 (1-0,668)^7$$

Ví dụ 25

Cho X là biến ngẫu nhiên có phân phối chuẩn. Biết rằng $P(X > 4) = 0,159$ và $P(X < 3) = 0,309$. Tính xác suất $P(X > 1)$. Cho biết $\Phi(1/2) = 0,691$ và $\Phi(1) = 0,841$.

Giải. Giả sử $X \sim N(\mu; \sigma^2)$. Khi đó:

$$P(X > 4) = 1 - \Phi\Big(\frac{4-\mu}{\sigma}\Big) = 0,159 \Leftrightarrow \Phi\Big(\frac{4-\mu}{\sigma}\Big) = 0,841 = \Phi(1) \Leftrightarrow 4 - \mu = \sigma\,(1)$$

$$P(X < 3) = \Phi\Big(\frac{3-\mu}{\sigma}\Big) = 0,309 \Leftrightarrow \Phi\Big(-\frac{3-\mu}{\sigma}\Big) = 1 - \Phi\Big(\frac{3-\mu}{\sigma}\Big) = 0,691 = \Phi(1/2)$$

$$\Leftrightarrow 3 - \mu = -\sigma/2\,(2)$$

Giải hệ (1),(2) ta được: $\mu = 10/3; \sigma = 2/3$.

Vậy, $P(X > 1) = 1 - P(X \leq 1) = 1 - \Phi\Big(\dfrac{1 - 10/3}{2/3}\Big) = 0,99976$

Ví dụ 26

Cho $X \sim N(15,\ 4)$, $Y \sim N(10,\ 1)$, X và Y độc lập.
a) Tính $P(2X > 3Y)$.
b) Tìm a và b biết $T = X + aY + b$ và $E(T) = 30,\ \ D(T) = 5$.

Giải. a. Đặt $Z = 2X - 3Y$. Theo tính chất của phân phối chuẩn, Z cũng có phân phối chuẩn với tham số:

$$\begin{cases} \mu_Z = E(Z) = 2E(X) - 3E(Y) = 2 * 15 - 3 * 10 = 0 \\ \sigma_Z^2 = D(Z) = 4D(X) + 9D(Y) = 4 * 4 + 9 * 1 = 25 \end{cases}$$

Do đó, $Z \sim N(0; 25)$. Vậy, $P(2X > 3Y) = P(Z > 0) = 1 - \Phi\Big(\dfrac{0-0}{5}\Big) = 1/2$.

b. Sử dụng tính chất của kỳ vọng và phương sai:

$$\begin{cases} E(T) = 30 \\ D(T) = 5 \end{cases} \Leftrightarrow \begin{cases} 15 + 10a + b = 30 \\ 4 + a^2 = 5 \end{cases} \Leftrightarrow \left[\begin{array}{l} a = 1; b = 5 \\ a = -1; b = 25 \end{array}\right.$$

Ví dụ 27

Chiều cao X (mét) của nam thanh niên trưởng thành ở quốc gia A tuân theo quy luật phân bố chuẩn $N(\mu; 0,1^2)$. Chọn ngẫu nhiên 100 nam thanh niên của quốc gia A. Tính xác suất sai số tuyệt đối giữa chiều cao trung bình của 100 nam thanh niên được chọn với μ không vượt quá $0,03$.

Giải. Gọi X_i là chiều cao của thanh niên thứ $i, i = \overline{1,100}$ và $\bar{X} = (X_1 + ... + X_{100})/100$. Ta có: X_i độc lập và $X_i \sim N(\mu; 0,1^2)$. Theo tính chất của phân phối chuẩn, $\bar{X}$ cũng có phân phối chuẩn với tham số:

$$\begin{cases} \mu_{\bar{X}} = E(\bar{X}) = \mu \\ \sigma_{\bar{X}}^2 = D(\bar{X}) = 0,1^2/100 = 10^{-4} \end{cases}$$

Do đó, $\bar{X} \sim N(\mu; 10^{-4})$. Xác suất cần tìm:

$$P(|\bar{X} - \mu| \leq 0,03) = \Phi\left(\frac{0,03}{10^{-2}}\right) - \Phi\left(-\frac{0,03}{10^{-2}}\right) = 0,9973$$

Qui tắc 2σ**:** Cho $X \sim N(\mu, \sigma^2)$. Khi đó

$$P(|X-\mu| < 2\sigma) = P\left(\frac{|X-\mu|}{\sigma} < 2\right) = 2\Phi(2) - 1 \approx 0,9545$$

hay $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0,9545$

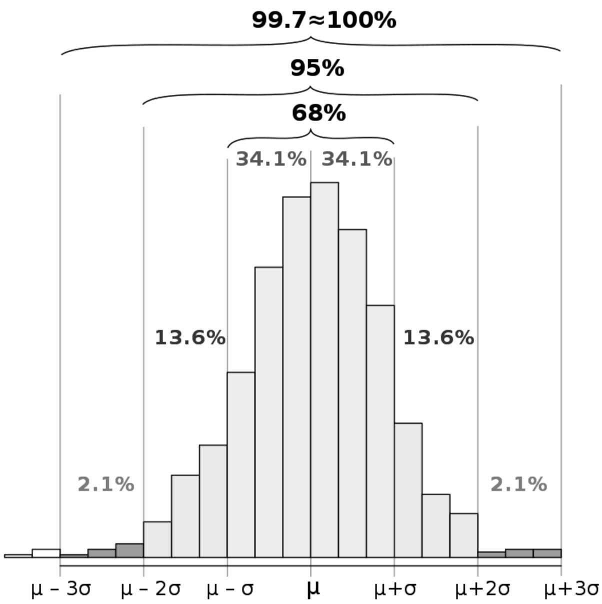
Điều này có nghĩa xác suất X nhận giá trị trong khoảng $(\mu - 2\sigma, \mu + 2\sigma)$ bằng 95,45%.

Qui tắc 3σ**:** Cho $X \sim N(\mu, \sigma^2)$. Khi đó

$$P(|X-\mu| < 3\sigma) = P\left(\frac{|X-\mu|}{\sigma} < 3\right) = 2\Phi(3) - 1 \approx 0,9973$$

hay $P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 0,9973$

Điều này có nghĩa xác suất X nhận giá trị trong khoảng $(\mu - 3\sigma, \mu + 3\sigma)$ bằng 99,73%.



7.9 Phân phối χ^2

Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối χ^2 với bậc tự do n, kí hiệu $X \sim \chi^2_n$ nếu như hàm mật độ của nó có dạng:

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \, x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

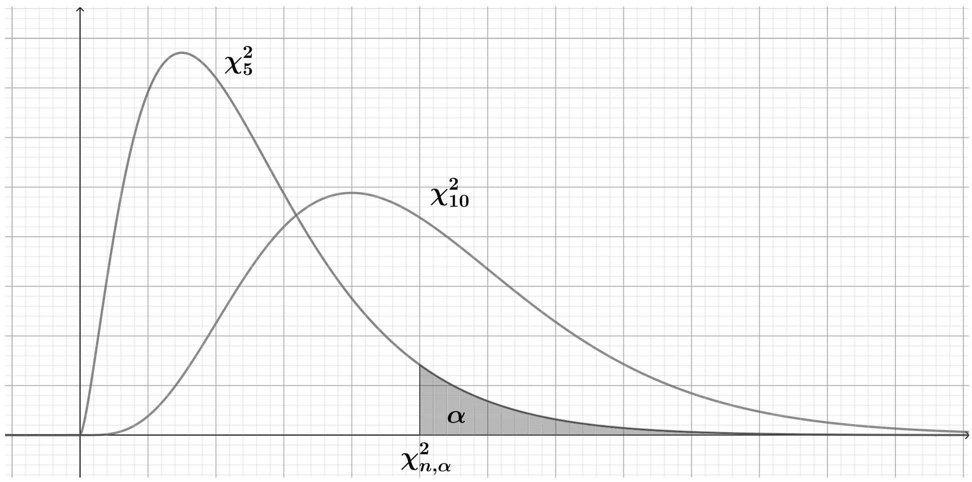
trong đó hàm gamma: $\Gamma(a) = \int\limits_0^{+\infty} x^{a-1} e^{-x} dx$, $a > 0$.

Giá trị tới hạn mức α của phân phối χ^2_n được kí hiệu $\chi^2_{n;\alpha}$ và được tra ở bảng.

Tính chất:

i) $E(X) = n$, $D(X) = 2n$

ii) Cho $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập và có cùng phân phối $N(0,1)$. Khi đó biến ngẫu nhiên $X = X_1^2 + X_2^2 + ... + X_n^2$ có phân phối $\chi^2{}_n$.



7.10 Phân phối student

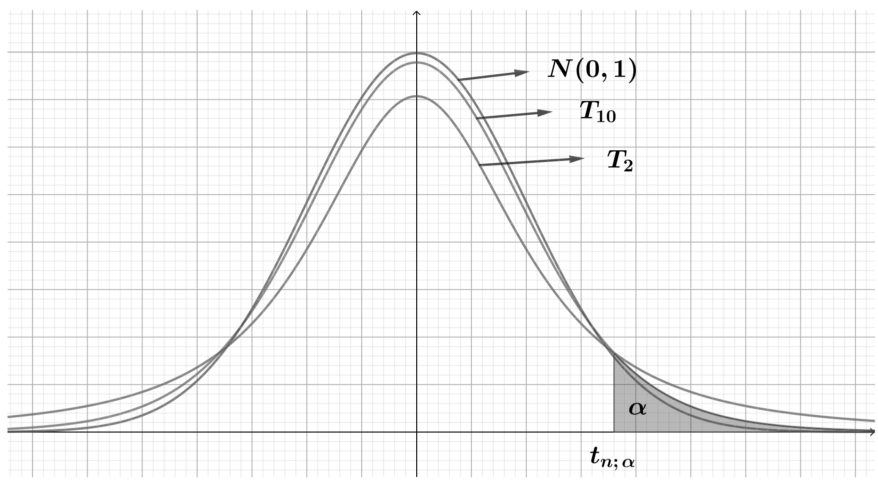
Định nghĩa: Biến ngẫu nhiên X được gọi là có phân phối Student với bậc tự do n, kí hiệu $X \sim T_n$ nếu như hàm mật độ của nó có dạng:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, x \in \mathbb{R}.$$

Giá trị tới hạn mức α của phân phối T_n được kí hiệu $t_{n;\alpha}$ và được tra ở bảng. Khi $n > 30$ người ta thường sử dụng công thức xấp xỉ:

$$t_{n;\alpha} \approx z_\alpha$$

Tính chất:
i) $E(X) = 0, D(X) = \frac{n}{n-2}$, $n > 2$
ii) Cho $X \sim N(0,1)$, $Y \sim \chi^2{}_n$ và X, Y độc lập. Khi đó biến ngẫu nhiên $Z = \frac{X}{\sqrt{Y/n}}$ sẽ có phân phối T_n.



8. Các định lý giới hạn

8.1 Luật số lớn

Bất đẳng thức Chebyshev: Cho X là biến ngẫu nhiên có kỳ vọng hữu hạn. Khi đó với mọi $\varepsilon > 0$ ta có:

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

Từ đó, dễ thấy $P(|X - E(X)| > 3\sigma(X)) \leq 1/9$.
Định lý:(Luật yếu số lớn) Nếu $\{X_n, n \geq 1\}$ là dãy biến ngẫu nhiên độc lập, cùng phân phối xác suất với biến X có kỳ vọng $E(X) = \mu$ và phương sai $D(X) = \sigma^2$ hữu hạn, thì

$$\lim_{n\to+\infty} P(|\bar{X} - \mu| > \varepsilon) = 0,$$

trong đó $\bar{X} = (X_1 + ... + X_n)/n$.
Hệ quả: (Định lý Bernoulli) Nếu gọi X là số lần xuất hiện biến cố A trong dãy n phép thử Bernoulli với $p = P(A)$ thì $\lim\limits_{n\to\infty} P\left(|f_n - p| > \varepsilon\right) = 0, \forall \varepsilon > 0$ trong đó $f_n = \frac{X}{n}$.

8.2 Định lý giới hạn trung tâm

Định lý: Nếu $\{X_n, n \geq 1\}$ là dãy biến ngẫu nhiên **độc lập**, **cùng phân phối xác suất** với biến X có **kỳ vọng** $E(X) = \mu$ **và phương sai** $D(X) = \sigma^2$ **hữu hạn**, thì:

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} < x\right) = \Phi(x),$$

trong đó $S_n = X_1 + X_2 + ... + X_n$.
Ứng dụng: Nếu $\{X_n, n \geq 1\}$ là dãy biến ngẫu nhiên độc lập, cùng phân phối xác suất với biến X có kỳ vọng $E(X) = \mu$ và phương sai $D(X) = \sigma^2$ hữu hạn, thì với n lớn $(n \geq 30)$ $S_n = X_1 + ... + X_n$ có phân phối xấp xỉ phân phối chuẩn $N(n\mu, n\sigma^2)$.

Ví dụ 28

Tuổi thọ của một loại bóng đèn là một biến ngẫu nhiên X có $E(X) = 250h$ và độ lệch chuẩn $\sigma(X) = 50h$.
a) Một cửa hàng mua 30 bóng đèn để khi hỏng có thể thay thế. Tính xấp xỉ xác suất cửa hàng có thể duy trì ánh sáng liên tục ít nhất 8750h.
b) Chủ cửa hàng phải mua dự trữ ít nhất bao nhiêu bóng đèn để duy trì ánh sáng liên tục ít nhất 8750h với xác suất lớn hơn 0,9772. Cho biết $\Phi(2) = 0,9772$.

Giải. a. Gọi X_i là tuổi thọ của bóng đèn thứ $i, i \geq 1$. Ta có các biến ngẫu nhiên X_i độc lập và có cùng phân phối với X. Do đó, áp dụng định lý giới hạn trung tâm:

$$T = X_1 + ... + X_{30}$$

có phân phối xấp xỉ phân phối chuẩn $N(30 * 250; 30 * 50^2) = N(7500; 75000)$
Xác suất cần tìm:

$$P(T \geq 8750) = 1 - P(T < 8750) \approx 1 - \Phi\left(\frac{8750 - 7500}{\sqrt{75000}}\right) \approx 0$$

b. Gọi n là số bóng đèn cần mua dự trữ. Tương tự, áp dụng định lý giới hạn trung tâm:

$$T_n = X_1 + ... + X_n$$

có phân phối xấp xỉ phân phối chuẩn $N(250n; 50^2n)$.
Từ giả thiết, ta có:

$$P(T_n \geq 8750) = 1 - P(T_n < 8750) \approx 1 - \Phi\left(\frac{8750 - 250n}{50\sqrt{n}}\right) \geq 0,9772 = \Phi(2)$$

$$\Leftrightarrow \Phi\left(-\frac{8750 - 250n}{50\sqrt{n}}\right) \geq 0,9772 = \Phi(2)$$

Giải bất phương trình:

$$-\frac{8750 - 250n}{50\sqrt{n}} \geq 2$$

ta được: $n \geq 37,44$.
Vậy, $n = 38$.

Ví dụ 29

Tung ngẫu nhiên một con xúc xắc cân đối 200 lần độc lập. Tính xấp xỉ xác suất tổng số chấm xuất hiện (ở mặt trên cùng) lớn hơn 720.

Giải. Gọi X_i là số chấm xuất hiện ở lần tung thứ $i, i = \overline{1,200}$. Đặt $S = X_1 + ... + X_{200}$.
Ta có: $X_1, ..., X_{200}$ là các biến ngẫu nhiên độc lập cùng phân phối.

X_i	1	2	3	4	5	6
P	1/6	1/6	1/6	1/6	1/6	1/6

Khi đó: $E(X_i) = 7/2; D(X_i) = E(X_i^2) - (EX_i)^2 = 35/12$.
Theo định lý giới hạn trung tâm, S sẽ có phân phối xấp xỉ phân phối chuẩn $N(200 * 7/2; 200 * 35/12) = N(700; 1750/3)$. Do đó, xác suất cần tìm:

$$P(S > 720) = 1 - P(S \leq 720) \approx 1 - \Phi\left(\frac{720 - 700}{\sqrt{1750/3}}\right) \approx 0,2$$

Hệ quả: (Định lý giới hạn tích phân Moivre-Laplace) Cho $X \sim B(n,p)$. Khi n khá lớn ta có:

$$P(k_1 < X < k_2) \approx \Phi\left(\frac{k_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{np(1-p)}}\right).$$

Hay nói cách khác, khi n lớn X có phân phối xấp xỉ phân phối chuẩn $N(\mu, \sigma^2)$ với $\mu = E(X) = np$ và $\sigma^2 = D(X) = np(1-p)$.

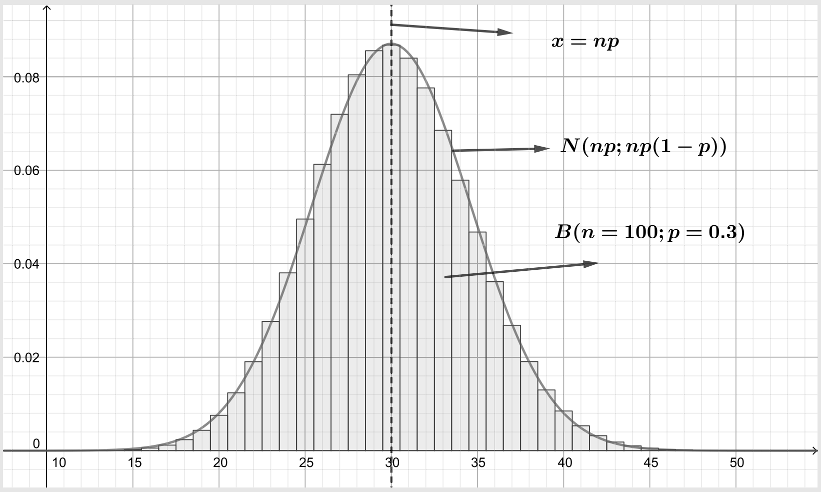
Định lý giới hạn địa phương Moivre-Laplace (đọc thêm)

Cho $X \sim B(n,p)$. Khi n lớn, ta có:

$$P(X = k) \approx \frac{1}{\sqrt{np(1-p)}}\varphi\left(\frac{k - np}{\sqrt{np(1-p)}}\right),$$

trong đó $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ - hàm mật độ phân phối chuẩn tắc N(0,1).

Các định lý giới hạn Moivre-Laplace



Ví dụ 30

Một xạ thủ bắn độc lập 200 viên đạn vào một mục tiêu. Xác suất bắn trúng mỗi viên là 0,6. Tính xấp xỉ xác suất có ít nhất 100 viên trúng đích.

Giải. Gọi X là số viên đạn bắn trúng. Ta có: $X \sim B(n = 200; p = 0,6)$. Áp dụng định lý Moivre-Laplace, X có phân phối xấp xỉ chuẩn với $\mu = np = 120; \sigma^2 = np(1-p) = 48$. Do đó:

$$P(100 \leq X \leq 200) \approx \Phi\left(\frac{200-120}{\sqrt{48}}\right) - \Phi\left(\frac{100-120}{\sqrt{48}}\right) = 0,998$$

Ví dụ 31

Một công ty bảo hiểm xe máy có 10 000 khách hàng. Mỗi chủ xe phải nộp 120.000 đồng/1 năm và trung bình nhận lại là 1.000.000 đồng nếu xe của họ bị tai nạn giao thông. Qua thống kê biết tỉ lệ để 1 xe máy bị tai nạn giao thông trong một năm là 0,006. Tìm xác suất để sau một năm hoạt động công ty bị thất bại.

Giải. Gọi X là số xe bị tai nạn và Y (ngàn đồng) là số tiền lãi của công ty trong 1 năm. Ta có: $X \sim B(n = 10000; p = 0,006)$ và $Y = 12 * 10^5 - 1000X$.
Áp dụng định lý Moivre-Laplace, X có phân phối xấp xỉ chuẩn với tham số

$$\mu = np = 60; \sigma^2 = np(1-p) = 59,64$$

Xác suất để sau một năm hoạt động công ty bị thất bại:

$$P(Y < 0) = P(12 * 10^5 - 1000X < 0) = P(X > 1200) \approx 1 - \Phi\left(\frac{1200-60}{\sqrt{59,64}}\right) \approx 0$$