

Chương 2

Biến ngẫu nhiên và quy luật phân phối xác suất

Mục tiêu

1. Thông qua các công cụ giải tích, cung cấp cho sinh viên khái niệm về biến ngẫu nhiên, phân loại các biến ngẫu nhiên, các quy luật phân phối xác suất của biến ngẫu nhiên, các tham số đặc trưng của biến ngẫu nhiên cùng một số quy luật phân phối xác suất thông dụng của biến ngẫu nhiên.
2. Với các kiến thức nền tảng đó, sinh viên biết tính các tham số đặc trưng của biến ngẫu nhiên; hiểu và vận dụng được ý nghĩa của các đặc trưng của biến ngẫu nhiên cùng các quy luật phân phối xác suất trong các bài toán xác suất thuộc các lĩnh vực kỹ thuật, kinh tế, xã hội...

Nội dung

Hai nội dung quan trọng nhất của chương là quy luật phân phối xác suất và các tham số đặc trưng của một biến ngẫu nhiên.

1. Định nghĩa và phân loại biến ngẫu nhiên
2. Quy luật phân phối xác suất của biến ngẫu nhiên (bảng phân phối xác suất, hàm phân phối xác suất, hàm mật độ xác suất)
3. Các tham số đặc trưng của biến ngẫu nhiên (kỳ vọng, phương sai, độ lệch chuẩn, môđ, trung vị...)
4. Một số phân phối xác suất thông dụng

Thời lượng: 8 tiết

BÀI 5 (2 tiết)

2.1 Định nghĩa và phân loại biến ngẫu nhiên

2.1.1 Định nghĩa biến ngẫu nhiên

Khái niệm biến ngẫu nhiên (random variable) rất thông dụng trong giải tích. Vì vậy ta tìm cách đưa vào khái niệm biến ngẫu nhiên như một đại lượng phụ thuộc vào kết cục của một phép thử ngẫu nhiên nào đó.

Ví dụ 2.1. Gieo một con xúc sắc. Nếu ta gọi biến ngẫu nhiên là "số chấm xuất hiện" thì nó phụ thuộc vào kết cục của phép thử và nhận các giá trị nguyên từ 1 đến 6.

Về mặt hình thức, có thể định nghĩa biến ngẫu nhiên như một hàm số có giá trị thực xác định trên không gian các sự kiện sơ cấp.

Ký hiệu biến ngẫu nhiên là X, Y, Z, X_1, X_2, \dots . Các giá trị có thể có của chúng ký hiệu là x, y, z, x_1, x_2, \dots .

Tập hợp tất cả các giá trị của X gọi là miền giá trị của X , ký hiệu là S_X .

Nhận xét 2.1. (a) X được gọi là biến ngẫu nhiên vì trước khi tiến hành phép thử ta chưa có thể nói một cách chắc chắn nó sẽ nhận một giá trị bằng bao nhiêu mà chỉ dự đoán điều đó với một xác suất nhất định. Nói cách khác, việc biến ngẫu nhiên X nhận một giá trị nào đó $(X = x_1), (X = x_2), \dots, (X = x_n)$ về thực chất là các sự kiện ngẫu nhiên.

(b) Nếu biến ngẫu nhiên X chỉ nhận các giá trị x_1, x_2, \dots, x_n thì các sự kiện $(X = x_1), (X = x_2), \dots, (X = x_n)$ tạo nên một hệ đầy đủ.

2.1.2 Phân loại biến ngẫu nhiên

Biến ngẫu nhiên được phân làm hai loại: biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục.

Định nghĩa 2.1 (Biến ngẫu nhiên rời rạc). X là biến ngẫu nhiên rời rạc (discrete random variable) nếu tập giá trị S_X của nó là tập hợp hữu hạn hoặc vô hạn đếm được phần tử. Nói cách khác, ta có thể liệt kê tất cả các giá trị của biến ngẫu nhiên đó.

Định nghĩa 2.2 (Biến ngẫu nhiên liên tục). X là biến ngẫu nhiên liên tục (continuous random variable) nếu tập giá trị S_X có thể có của nó lấp đầy một khoảng trên trục số.

Ví dụ 2.2. (a) Gọi X là số chấm xuất hiện khi gieo một con xúc sắc cân đối đồng chất thì X là biến ngẫu nhiên rời rạc có thể nhận một trong các giá trị 1, 2, 3, 4, 5 và 6.

(b) Một người phải tiến hành thí nghiệm cho tới khi thành công thì dừng. Gọi Y là số lần tiến hành thí nghiệm. Khi đó Y là biến ngẫu nhiên rời rạc có thể nhận các giá trị 1, 2, \dots, n, \dots .

- (c) Bắn một viên đạn vào bia có bán kính là 20cm và giả sử viên đạn trúng vào bia. Gọi Z là khoảng cách từ tâm bia tới điểm bia trúng đạn thì Z là biến ngẫu nhiên liên tục có thể nhận các giá trị thuộc $(0; 20)$.

2.2 Quy luật phân phối xác suất của biến ngẫu nhiên

Định nghĩa 2.3 (Quy luật phân phối xác suất). Bất kỳ một hình thức nào cho phép biểu diễn mối quan hệ giữa các giá trị có thể có của biến ngẫu nhiên và xác suất tương ứng để biến ngẫu nhiên nhận các giá trị đó đều được gọi là quy luật phân phối xác suất (probability distribution) của biến ngẫu nhiên.

Một số phương pháp mô tả quy luật phân phối xác suất của biến ngẫu nhiên:

1. Bảng phân phối xác suất (áp dụng cho biến ngẫu nhiên rời rạc).
2. Hàm phân phối xác suất (áp dụng cho cả biến ngẫu nhiên rời rạc và liên tục).
3. Hàm mật độ xác suất (áp dụng cho biến ngẫu nhiên liên tục).

2.2.1 Bảng phân phối xác suất của biến ngẫu nhiên rời rạc

Định nghĩa 2.4 (Hàm khối lượng xác suất). Cho biến ngẫu nhiên rời rạc X . Đặt

$$p_X(x) = P(X = x), \quad x \in \mathbb{R} \quad (2.1)$$

Hàm $p_X(x)$ được gọi là hàm khối lượng xác suất (probability mass function) của biến ngẫu nhiên rời rạc X .

Hàm khối lượng xác suất có tính chất sau.

Định lý 2.1. (a) $p_X(x) > 0$ với mọi $x \in \mathbf{S}_X$;

(b) $\sum_{x \in \mathbf{S}_X} p_X(x) = 1$;

(c) $p_X(x) = 0$ với mọi $x \notin \mathbf{S}_X$.

Định nghĩa 2.5 (Bảng phân phối xác suất). Bảng phân phối xác suất (probability distribution) của biến ngẫu nhiên rời rạc X là bảng ghi sự tương ứng giữa các giá trị mà biến ngẫu nhiên nhận được với giá trị của hàm khối lượng xác suất tương ứng.

1. Nếu X là biến ngẫu nhiên rời rạc có hữu hạn (n) phần tử thì bảng phân phối xác suất của biến ngẫu nhiên X là:

X	x_1	x_2	\dots	x_n
$P(X = x_i)$	p_1	p_2	\dots	p_n

(2.2)

trong đó $\{x_1, x_2, \dots, x_n\}$ là tập các giá trị của X đã sắp xếp theo thứ tự tăng dần, $p_i = P(X = x_i), i = 1, 2, \dots, n$.

2. Nếu X là biến ngẫu nhiên rời rạc có vô hạn đếm được phần tử thì bảng phân phối xác suất của biến ngẫu nhiên X là:

X	x_1	x_2	\dots	x_n	\dots
$P(X = x_i)$	p_1	p_2	\dots	p_n	\dots

(2.3)

trong đó $\{x_1, x_2, \dots, x_n, \dots\}$ là tập các giá trị của X đã sắp xếp theo thứ tự tăng dần, $p_n = P(X = x_n), n = 1, 2, \dots$.

Nhận xét 2.2. Trong (2.2), $\sum_{i=1}^n p_i = 1$ và trong (2.3), $\sum_{n=1}^{\infty} p_n = 1$.

Ví dụ 2.3. Một xạ thủ có 3 viên đạn được yêu cầu bắn lần lượt từng viên cho đến khi trúng mục tiêu hoặc hết cả 3 viên thì thôi. Tìm bảng phân phối xác suất của số đạn đã bắn, biết rằng xác suất bắn trúng đích của mỗi lần bắn là 0,8.

Lời giải Ví dụ 2.3 Gọi X là số đạn đã bắn, X là biến ngẫu nhiên rời rạc nhận các giá trị 1, 2, 3.

Gọi A_i là sự kiện "bắn trúng mục tiêu ở lần bắn thứ i ", $i = 1, 2, 3$. Khi đó,

$$P(X = 1) = P(A_1) = 0,8.$$

$$P(X = 2) = P(\bar{A}_1 A_2) = P(\bar{A}_1)P(A_2) = 0,2 \times 0,8 = 0,16.$$

$$P(X = 3) = P(\bar{A}_1 \bar{A}_2 (A_3 + \bar{A}_3)) = P(\bar{A}_1)P(\bar{A}_2)P(A_3 + \bar{A}_3) = 0,2 \times 0,2 \times (0,8 + 0,2) = 0,04.$$

Vậy bảng phân phối xác suất của X là

X	1	2	3
$P(X = x_i)$	0,8	0,16	0,04

Ví dụ 2.4. Một người đem 10 nghìn VNĐ đi đánh một số đề. Nếu trúng thì thu được 700 nghìn VNĐ, nếu trượt thì không được gì. Gọi X (nghìn VNĐ) là số tiền thu được. Ta có bảng phân phối xác suất của X

X	0	700
$P(X = x_i)$	$99/100$	$1/100$

Ví dụ 2.5. Một chùm chìa khóa gồm 4 chiếc giống nhau, trong đó chỉ có một chiếc mở được cửa. Người ta thử ngẫu nhiên từng chiếc cho đến khi mở được cửa. Gọi X là số lần thử. Tìm phân phối xác suất của X .

Lời giải Ví dụ 2.5 X có thể nhận các giá trị 1, 2, 3, 4.

Gọi A_i là sự kiện "mở được cửa ở lần thử thứ i ", $i = 1, 2, 3, 4$. Khi đó,

$$P(X = 1) = P(A_1) = \frac{1}{4}$$

$$P(X = 2) = P(\bar{A}_1 A_2) = P(\bar{A}_1)P(A_2|\bar{A}_1) = \frac{3 \times 1}{4 \times 3} = \frac{1}{4}$$

$$P(X = 3) = P(\bar{A}_1 \bar{A}_2 A_3) = P(\bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(A_3|\bar{A}_1 \bar{A}_2) = \frac{3 \times 2 \times 1}{4 \times 3 \times 2} = \frac{1}{4}$$

$$P(X = 4) = P(\bar{A}_1 \bar{A}_2 \bar{A}_3 A_4) = P(\bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(\bar{A}_3|\bar{A}_1 \bar{A}_2)P(A_4|\bar{A}_1 \bar{A}_2 \bar{A}_3) = \frac{1}{4}.$$

Vậy bảng phân phối xác suất của X là

X	1	2	3	4
$P(X = x_i)$	1/4	1/4	1/4	1/4

2.2.2 Hàm phân phối xác suất

Định nghĩa 2.6 (Hàm phân phối xác suất). Hàm phân phối xác suất (cumulative distribution function) của biến ngẫu nhiên X , ký hiệu là $F_X(x)$, được định nghĩa như sau:

$$F_X(x) = P(X \leq x), \quad x \in \mathbb{R} \quad (2.4)$$

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì hàm phân phối (tích lũy) là:

$$F_X(x) = \begin{cases} 0, & x \leq x_1, \\ p_1, & x_1 < x \leq x_2, \\ p_1 + p_2, & x_2 < x \leq x_3, \\ \dots & \\ 1, & x > x_n. \end{cases} \quad (2.5)$$

2. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.3) thì hàm phân phối (tích lũy) là:

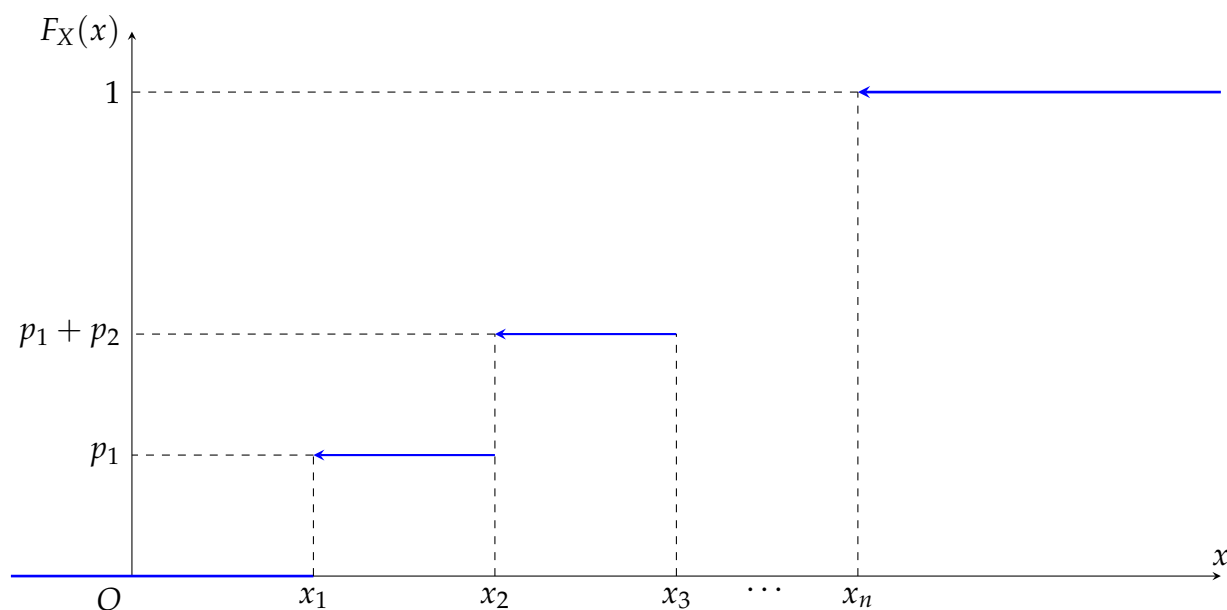
$$F_X(x) = \begin{cases} 0, & x \leq x_1, \\ p_1, & x_1 < x \leq x_2, \\ p_1 + p_2, & x_2 < x \leq x_3, \\ \dots & \\ \sum_{i=1}^n p_i, & x_n < x \leq x_{n+1}, \\ \dots & \end{cases} \quad (2.6)$$

Nhận xét 2.3. 1. Hàm phân phối xác suất $F_X(x)$ phản ánh mức độ tập trung xác suất ở bên trái của một số thực x nào đó.

2. Đồ thị của hàm phân phối xác suất của biến ngẫu nhiên rời rạc có dạng bậc thang (Hình 2.1)

Ví dụ 2.6. (a) Tìm hàm phân phối xác suất của biến ngẫu nhiên ở Ví dụ 2.3. (b) Vẽ đồ thị hàm phân phối.

Lời giải Ví dụ 2.6

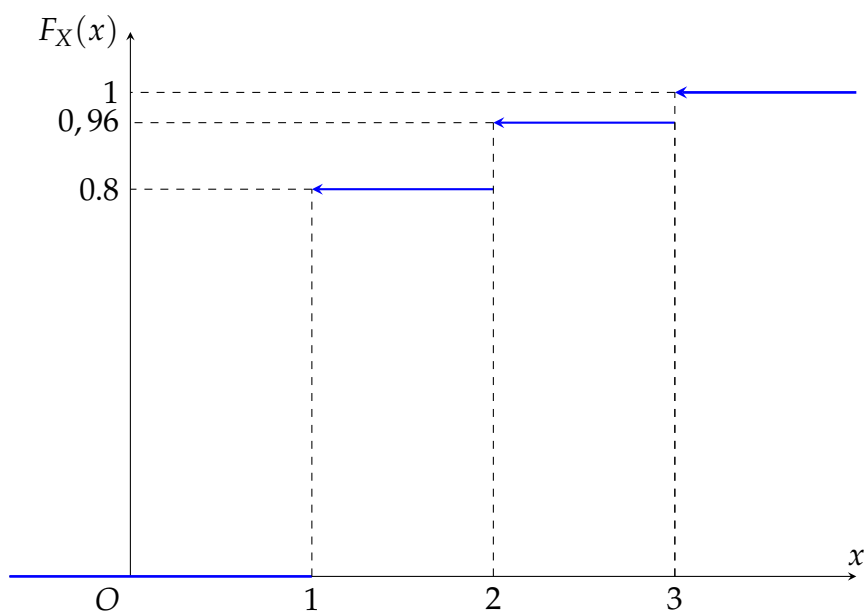


Hình 2.1: Đồ thị của hàm phân phối xác suất (2.5)

(a) Từ bảng phân phối xác suất ở Ví dụ 2.3, sử dụng (2.5) suy ra

$$F_X(x) = \begin{cases} 0, & x \leq 1, \\ 0,8, & 1 < x \leq 2, \\ 0,96, & 2 < x \leq 3, \\ 1, & x > 3. \end{cases}$$

(b) Đồ thị của hàm $F_X(x)$ có dạng bậc thang:



Hình 2.2: Đồ thị hàm phân phối xác suất trong Ví dụ 2.6

Hàm phân phối có các tính chất sau.

Định lý 2.2. 1. $0 \leq F_X(x) \leq 1$ với mọi $x \in \mathbb{R}$.

2. $F_X(x)$ là hàm không giảm, liên tục bên trái, nghĩa là với mọi $x_1, x_2 \in \mathbb{R}$, $x_1 < x_2$ thì $F_X(x_1) \leq F_X(x_2)$ và với mọi $a \in \mathbb{R}$, $F_X(a^-) = F_X(a)$, với $F_X(a^-) = \lim_{x \rightarrow a^-} F_X(x)$.

Nếu X là biến ngẫu nhiên liên tục thì $F_X(x)$ là hàm liên tục.

3. $P(a \leq X < b) = F_X(b) - F_X(a)$;

Nếu X là biến ngẫu nhiên liên tục thì $P(X = a) = 0$ và

$$P(a \leq X < b) = P(a \leq X \leq b) = P(a < X \leq b) = P(a < X < b) = F_X(b) - F_X(a).$$

4. $F_X(-\infty) = 0, F_X(+\infty) = 1$.

Chứng minh. 1. Suy trực tiếp từ định nghĩa (2.4) và tính chất của xác suất.

2. Giả sử $x_1 < x_2$, xét sự kiện $(X < x_2) = (X < x_1) + (x_1 \leq X < x_2)$. Do tính xung khắc của các sự kiện suy ra

$$P(X < x_2) = P(X < x_1) + P(x_1 \leq X < x_2).$$

Từ đây kết hợp với (2.4) suy ra

$$F_X(x_2) - F_X(x_1) = P(x_1 \leq X < x_2) \geq 0.$$

3. Suy trực tiếp từ chứng minh tính chất trên.

4. $F_X(-\infty) = P(X < -\infty) = P(\emptyset) = 0, F_X(+\infty) = P(X < +\infty) = P(S) = 1$.

□

Ví dụ 2.7. Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \leq -1, \\ A + B \arcsin x, & -1 < x < 1, \\ 1, & x \geq 1. \end{cases}$$

Hãy xác định A và B ?

Lời giải Ví dụ 2.7 Sử dụng Định lý 2.2(1), $0 \leq A + B \arcsin x \leq 1$ và theo Định lý 2.2(2) vì $F_X(x)$ liên tục nên $A - \frac{\pi}{2} \times B = 0, A + \frac{\pi}{2} \times B = 1$.

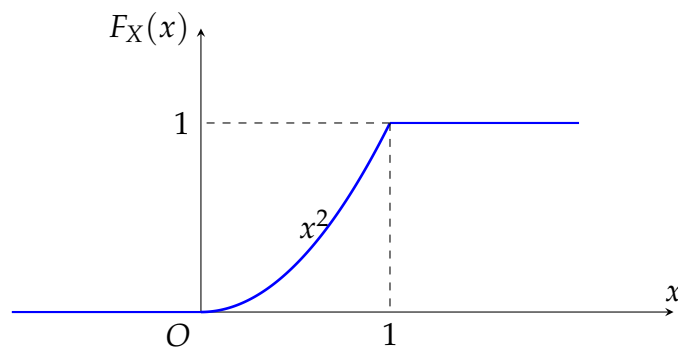
$$\text{Suy ra } A = \frac{1}{2}, B = \frac{1}{\pi}.$$

Ví dụ 2.8. Xét phép thử ném phi tiêu vào một đĩa tròn có bán kính bằng 1(m). Ký hiệu X là biến ngẫu nhiên đo khoảng cách từ điểm mũi phi tiêu cắm vào đĩa đến tâm của đĩa. Giả sử mũi phi tiêu luôn cắm vào đĩa và đồng khả năng tại mọi điểm của đĩa. (a) Tìm miền giá trị của X . (b) Tìm hàm phân phối $F_X(x)$ và vẽ đồ thị của $F_X(x)$.

Lời giải Ví dụ 2.8

(a) $S_X = \{x \in \mathbb{R} : 0 \leq x < 1\}$.

(b) Sử dụng định nghĩa $F_X(x) = P(X < x) = \begin{cases} 0, & x \leq 0, \\ x^2, & 0 < x \leq 1, \\ 1, & x > 1. \end{cases}$



Hình 2.3: Đồ thị hàm phân phối xác suất của Ví dụ 2.8

BÀI 6 (2 tiết)

2.2.3 Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Định nghĩa 2.7 (Hàm mật độ xác suất). Giả sử X là một biến ngẫu nhiên liên tục có hàm phân phối xác suất $F_X(x)$, $x \in \mathbb{R}$. Nếu tồn tại hàm $f_X(x)$ sao cho

$$F_X(x) = \int_{-\infty}^x f_X(t) dt, \quad x \in \mathbb{R} \quad (2.7)$$

thì $f_X(x)$ được gọi là hàm mật độ xác suất (probability density function) của biến ngẫu nhiên X .

Như vậy, hàm mật độ xác suất của biến ngẫu nhiên liên tục X là đạo hàm bậc nhất của hàm phân phối xác suất của biến ngẫu nhiên đó,

$$f_X(x) = F'_X(x), \quad x \in \mathbb{R} \quad (2.8)$$

Nhận xét 2.4. Hàm mật độ xác suất của biến ngẫu nhiên X tại mỗi điểm x cho biết mức độ tập trung xác suất tại điểm đó.

Hàm mật độ xác suất có các tính chất sau.

Định lý 2.3. 1. $f_X(x) \geq 0$ với mọi $x \in \mathbb{R}$.

$$2. P(a < X < b) = \int_a^b f_X(x) dx.$$

$$3. \int_{-\infty}^{+\infty} f_X(x) dx = 1.$$

Chứng minh. 1. Vì $f_X(x)$ là đạo hàm của hàm không giảm.

2. Suy từ Định lý 2.2(3).

$$3. \int_{-\infty}^{+\infty} f_X(x) dx = F_X(+\infty) = 1.$$

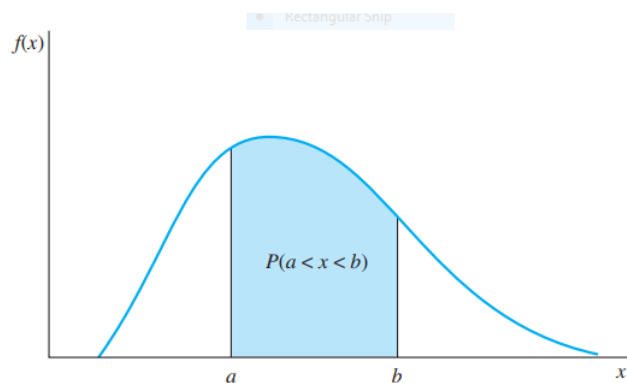
□

Ví dụ 2.9. Hàm phân phối xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x) = a + b \arctan x, \quad (-\infty < x < +\infty).$$

(a) Tìm a và b . (b) Tìm hàm mật độ xác suất $f_X(x)$. (c) Tìm xác suất để khi tiến hành 3 phép thử độc lập có 2 lần X nhận giá trị trong khoảng $(-1; 1)$.

Lời giải Ví dụ 2.9



Hình 2.4: $P[a < X < b]$ là diện tích miền tô màu dưới đường cong $y = f_X(x)$

(a) Sử dụng Định lý 2.2(4) ta tìm được $a = \frac{1}{2}, b = \frac{1}{\pi}$.

(b) Sử dụng (2.8) ta được $f_X(x) = \frac{1}{\pi(1+x^2)}$.

(c) Theo Định lý 2.3(2)

$$p = P(-1 < X < 1) = \int_{-1}^1 \frac{1}{\pi} \times \frac{dx}{1+x^2} = \frac{1}{2}.$$

Bài toán thỏa mãn lược đồ Béc-nu-li. Áp dụng công thức (1.19) ta tính được

$$P_3(2) = C_3^2 \times p^2 \times (1-p)^1 = \frac{3}{8}.$$

Ví dụ 2.10. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất là

$$f_X(x) = \begin{cases} a \cos x, & x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \\ 0, & x \notin \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. \end{cases}$$

(a) Tìm a . (b) Tìm hàm phân phối xác suất tương ứng. (c) Tìm xác suất để X nhận giá trị trong khoảng $\left(0, \frac{\pi}{4}\right)$.

Lời giải Ví dụ 2.10

(a) Sử dụng Định lý 2.3(1),(3) tính được $a = \frac{1}{2}$.

(b) Áp dụng (2.7).

$$\text{Nếu } x \leq -\frac{\pi}{2} \text{ thì } F_X(x) = \int_{-\infty}^x 0 du = 0.$$

$$\text{Nếu } -\frac{\pi}{2} < x \leq \frac{\pi}{2} \text{ thì } F_X(x) = \int_{-\infty}^x f_X(u) du = \int_{-\frac{\pi}{2}}^x \frac{1}{2} \cos u du = \frac{1}{2}(\sin x + 1).$$

Nếu $x > \frac{\pi}{2}$ thì $F_X(x) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2} \cos x = 1$. Vậy

$$F_X(x) = \begin{cases} 0, & x \leq -\frac{\pi}{2}, \\ \frac{1}{2}(\sin x + 1), & -\frac{\pi}{2} < x \leq \frac{\pi}{2}, \\ 1, & x > \frac{\pi}{2}. \end{cases}$$

$$(c) P(0 < X < \frac{\pi}{4}) = \int_0^{\frac{\pi}{4}} \frac{1}{2} \cos x dx = \frac{\sqrt{2}}{4}.$$

2.3 Các tham số đặc trưng của biến ngẫu nhiên

Đặc trưng quan trọng nhất của biến ngẫu nhiên là phân phối xác suất của nó. Nhưng trong thực tế nhiều khi không xác định được hàm phân phối và không phải cứ nhất thiết phải biết hàm phân phối. Vì vậy nảy sinh vấn đề phải đặc trưng cho biến ngẫu nhiên bằng một hoặc nhiều số, mỗi số hạng đặc trưng phản ánh được các tính chất cơ bản nhất của biến ngẫu nhiên X . Trong mục này ta chỉ xét một vài tham số quan trọng nhất.

2.3.1 Kỳ vọng

Định nghĩa 2.8 (Kỳ vọng). Kỳ vọng (expected value) của biến ngẫu nhiên X , ký hiệu là $E(X)$ (hoặc μ_X hoặc đơn giản là μ) được xác định như sau:

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$\mu_X = E(X) = \sum_{i=1}^n x_i p_i \quad (2.9)$$

2. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.3) thì

$$\mu_X = E(X) = \sum_{n=1}^{\infty} x_n p_n \quad (2.10)$$

nếu chuỗi về phải hội tụ.

3. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$, $x \in \mathbb{R}$ thì

$$\mu_X = E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx \quad (2.11)$$

nếu tích phân về phải hội tụ.

Nhận xét 2.5. 1. Kỳ vọng mang ý nghĩa là giá trị trung bình của biến ngẫu nhiên. Kỳ vọng là số xác định.

Thật vậy, giả sử đối với biến ngẫu nhiên X , tiến hành n phép thử, trong đó n_1 lần X nhận giá trị x_1 , n_2 lần X nhận giá trị x_2, \dots, n_k lần X nhận giá trị x_k , $n_1 + n_2 + \dots + n_k = n$. Giá trị trung bình của biến ngẫu nhiên X trong n phép thử này là

$$\begin{aligned}\bar{X} &= \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n} = x_1\frac{n_1}{n} + x_2\frac{n_2}{n} + \dots + x_k\frac{n_k}{n} \\ &\simeq x_1p_1 + x_2p_2 + \dots + x_kp_k = E(X).\end{aligned}$$

2. Khái niệm kỳ vọng được áp dụng rộng rãi trong nhiều lĩnh vực. Trong kinh doanh và quản lý, kỳ vọng được ứng dụng dưới dạng lợi nhuận kỳ vọng hay doanh số kỳ vọng.

Ví dụ 2.11. Theo thống kê việc một người Mỹ 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, còn xác suất để người đó chết trong vòng một năm tới là 0,008. Một chương trình bảo hiểm đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả 1000\$, còn tiền đóng là 10\$. Hỏi lợi nhuận trung bình của công ty bảo hiểm nhận được là bao nhiêu?

Lời giải Ví dụ 2.11 Gọi X là lợi nhuận của công ty bảo hiểm nhận được. Khi đó X là biến ngẫu nhiên rời rạc có thể nhận giá trị -990, 10. Bảng phân phối xác suất của X là

X	-990	10
$P(X = x_i)$	0,008	0,992

Suy ra $E(X) = -990 \times 0,008 + 10 \times 0,992 = 2\$$. Ta thấy lợi nhuận trung bình bằng 2\$ (một số dương) vì vậy công ty bảo hiểm có thể làm ăn có lãi.

Ví dụ 2.12. Kỳ vọng của biến ngẫu nhiên X trong Ví dụ 2.4 là $\mu_X = E(X) = 0 \times \frac{99}{100} + 700 \times \frac{1}{100} = 7$ nghìn VNĐ. Như vậy bỏ ra 10 nghìn VNĐ, trung bình thu được 7 nghìn VNĐ, người chơi về lâu dài sẽ lỗ 30% tổng số tiền chơi.

Ví dụ 2.13. Xét trò chơi trả lời hai câu hỏi A và B; người chơi có quyền chọn câu hỏi nào để trả lời đầu tiên. Câu hỏi A được trả lời đúng với xác suất 0,8 và khi đó người chơi sẽ được thưởng 100 USD, câu hỏi B được trả lời đúng với xác suất 0,6 và người chơi được thưởng 200 USD. Nếu không trả lời đúng lần thứ nhất sẽ không được trả lời tiếp. Vậy người chơi nên chọn câu hỏi nào trả lời đầu tiên để tiền thưởng trung bình nhận được cao hơn.

Lời giải Ví dụ 2.13 Gọi X là số tiền thưởng nhận được khi người chơi chọn câu hỏi A trả lời đầu tiên,

X	0	100	300
$P(X = x_i)$	0,2	0,32	0,48

và $E(X) = 0 \times 0,2 + 100 \times 0,32 + 300 \times 0,48 = 176$ USD.

Gọi Y là số tiền thưởng nhận được khi người chơi chọn câu hỏi B trả lời đầu tiên,

Y	0	200	300
$P(Y = y_i)$	0,4	0,12	0,48

và $E(Y) = 0 \times 0,4 + 200 \times 0,12 + 300 \times 0,48 = 168$ USD.

Vậy nên chọn câu hỏi A để trả lời đầu tiên để có khả năng nhận thưởng cao hơn.

Ví dụ 2.14. Theo thống kê ở một cửa hàng đậu tương, người ta thấy số lượng đậu tương bán ra X là một biến ngẫu nhiên rời rạc có bảng phân phối là:

X (kg)	10	13	16	19	22
$P(X = x_i)$	0,15	0,2	0,35	0,2	0,1

Nếu giá nhập là 10000 VNĐ/kg thì cửa hàng sẽ lãi 5000 VNĐ/kg, nếu đến cuối ngày không bán được sẽ lỗ 8000 VNĐ/kg. (a) Tìm hàm phân phối xác suất của X . (b) Mỗi ngày cửa hàng nên nhập bao nhiêu kg để thu được lãi nhiều nhất.

Lời giải Ví dụ 2.14

(a) Từ bảng phân phối xác suất ta có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \leq 10, \\ 0,15, & 10 < x \leq 13, \\ 0,35, & 13 < x \leq 16, \\ 0,7, & 16 < x \leq 19, \\ 0,9, & 19 < x \leq 22, \\ 1, & x > 22. \end{cases}$$

(b) Số lượng đậu tương nhập trong ngày theo các phương án 10, 13, 16, 19, 22. Gọi T_i là "số tiền lời thu được ứng với phương án i ", $i = 1, 2, \dots, 5$, trong đó phương án 1, 2, 3, 4, 5 tương ứng là nhập 10, 13, 16, 19, 22 (kg).

(b1) Phương án nhập 10kg: chắc chắn cửa hàng sẽ bán hết vì $P(X < 10) = 0$. Do đó $E(T_1) = 1 \times 50000 = 50000$ VNĐ.

(b2) Phương án nhập 13kg: do không có thống kê số lượng bán 11, 12kg, nên xem như cửa hàng đó chỉ có 2 phương án hoặc bán 10kg, hoặc bán 13kg. Do chỉ nhập 13kg nên xem như số lượng bán trên 13kg là số lượng bán được 13kg. Suy ra

$$E(T_2) = 26000 \times 0,15 + 65000 \times 0,85 = 59150 \text{ VNĐ.}$$

(b3) Phương án nhập 16kg: số lượng bán ra có thể là 10, 13, 16 với xác suất tương ứng là 0,15; 0,2 và 0,65. Suy ra

$$E(T_3) = 2000 \times 0,15 + 41000 \times 0,2 + 80000 \times 0,65 = 60500 \text{ VNĐ.}$$

(b4) Phương án nhập 19kg: số lượng bán ra có thể là 10, 13, 16, 19 với xác suất tương ứng là 0,15; 0,2; 0,35 và 0,3. Suy ra

$$E(T_4) = (-22000) \times 0,15 + 17000 \times 0,2 + 56000 \times 0,35 + 95000 \times 0,3 = 48200 \text{ VNĐ.}$$

(b5) Phương án nhập 22kg: số lượng bán ra có thể là 10, 13, 16, 19, 22 với xác suất tương ứng là 0,15; 0,2; 0,35; 0,2 và 0,1. Suy ra

$$E(T_5) = (-46000) \times 0,15 + (-7000) \times 0,2 + 32000 \times 0,35 + 71000 \times 0,2 + 110000 \times 0,1 = 28100 \text{ VNĐ.}$$

Từ các kết quả trên, ta thấy $E(T_3)$ là cao nhất nên phương án nhập hiệu quả nhất là 16kg.

Chú ý 2.1. Nếu trong bảng phân phối xác suất mà giá trị nào của biến ngẫu nhiên X không được đề cập đến thì xem như xác suất tại đó bằng 0.

Ví dụ 2.15. Một dây chuyền tự động khi hoạt động bình thường có thể sản xuất ra phế phẩm với xác suất $p = 0,001$ và được điều chỉnh ngay lập tức khi phát hiện có phế phẩm. Tính số trung bình các sản phẩm được sản xuất giữa 2 lần điều chỉnh.

Lời giải Ví dụ 2.15 Gọi X là số sản phẩm được sản xuất giữa hai lần điều chỉnh. Khi đó X là biến ngẫu nhiên rời rạc có thể nhận các giá trị $1, 2, \dots$ với xác suất tương ứng

$$P(X = 1) = 0,001, \quad P(X = 2) = 0,999 \times 0,001, \quad P(X = 3) = (0,999)^2 \times 0,001 \dots$$

Vậy

$$\begin{aligned} E(X) &= 1 \times 0,001 + 2 \times 0,999 \times 0,001 + 3 \times (0,999)^2 \times 0,001 + \dots \\ &= 0,001 \times \sum_{n=1}^{\infty} n \times (0,999)^{n-1} = 1000, \end{aligned}$$

ở đây ta sử dụng tính chất của chuỗi lũy thừa và công thức tính tổng của cấp số nhân lùi vô hạn với $x = 0,999$:

$$\sum_{n=1}^{\infty} nx^n = \sum_{n=1}^{\infty} (x^n)' = \left(\sum_{n=1}^{\infty} x^n \right)' = \left(\frac{x}{1-x} \right)' = \frac{1}{(1-x)^2}.$$

Ví dụ 2.16. Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên (đơn vị là tháng) với hàm mật độ xác suất như sau:

$$f_X(x) = \begin{cases} kx^2(4-x), & x \in [0, 4], \\ 0, & x \notin [0, 4]. \end{cases}$$

Tìm tuổi thọ trung bình của loại côn trùng trên.

Lời giải Ví dụ 2.16 Vì $f_X(x)$ là hàm mật độ xác suất của biến ngẫu nhiên liên tục X nên theo Định lý 2.3(1),(3), $k = \frac{3}{64}$.

Sử dụng công thức (2.11), tuổi thọ trung bình của loại côn trùng trên là

$$E(X) = \frac{3}{64} \int_0^4 x^3(4-x)dx = \frac{12}{5} \quad (\text{tháng}).$$

Hàm của một biến ngẫu nhiên

Bây giờ ta xét một biến ngẫu nhiên mới $g(X)$, phụ thuộc vào X ; nghĩa là, mỗi giá trị của $g(X)$ được xác định bởi giá trị của X . Chẳng hạn, $g(X)$ có thể là X^2 hoặc $3X - 1$ và giả sử X nhận giá trị 2, thì $g(X)$ sẽ nhận giá trị $g(2)$.

Ví dụ 2.17. Cho X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất là

X	-1	0	1	2
$P(X = x_i)$	p_{-1}	p_0	p_1	p_2

và $g(X) = X^2$, thì $g(X)$ nhận các giá trị 0, 1, 2 với

$$P(g(X) = 0) = P(X = 0) = p_0,$$

$$P(g(X) = 1) = P(X = -1) + P(X = 1) = p_{-1} + p_1,$$

$$P(g(X) = 4) = P(X = 2) = p_2.$$

Bảng phân phối xác suất của $g(X)$ là

$g(X)$	0	1	4
$P\{g(X) = g(x_i)\}$	p_0	$p_{-1} + p_1$	p_2

Từ đây, theo Định nghĩa 2.8(1) suy ra

$$\begin{aligned} E[g(X)] &= (0)(p_0) + (1)(p_{-1} + p_1) + (4)(p_2) \\ &= (-1)^2(p_{-1}) + (0)^2(p_0) + (1)^2(p_1) + (2)^2(p_2) = \sum_i g(x_i)p_i, \quad i = -1, 0, 1, 2. \end{aligned}$$

Kết quả này được mở rộng trong định lý dưới đây cho cả biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục.

Định lý 2.4. Cho X là một biến ngẫu nhiên và $Y = g(X)$ là một hàm của X .

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$E(Y) = E[g(X)] = \sum_{i=1}^n g(x_i)p_i \quad (2.12)$$

2. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$ thì

$$E(Y) = E[g(X)] = \int_{-\infty}^{+\infty} g(x)f_X(x)dx. \quad (2.13)$$

Ví dụ 2.18. Gọi X là số trang trong một bản fax. Một công ty điện thoại tính cước như sau: 10 xu cho trang thứ nhất, 9 xu cho trang thứ hai, ..., 6 xu cho trang thứ năm. Những bản fax từ 6 đến 10 trang có phí là 50 xu (công ty không nhận những bản fax quá 10 trang). Gọi Y là chi phí phải trả cho một bản fax. (a) Xác định Y . (b) Tính $E(Y)$ nếu X có phân phối là

X	1	2	3	4
$P(X = x_i)$	0,25	0,25	0,25	0,25

Lời giải Ví dụ 2.18

(a) Y là một hàm của X xác định bởi

$$Y = g(X) = \begin{cases} 10,5X - 0,5X^2, & 1 \leq X \leq 5, \\ 50, & 6 \leq X \leq 10. \end{cases}$$

(b) Theo Định lý 2.4(1),

$$\begin{aligned} E[Y] &= \sum_{i=1}^4 g(x_i)p_i \\ &= (0,25)[(10,5)(1) - (0,5)(1)^2] + (0,25)[(10,5)(2) - (0,5)(2)^2] \\ &\quad + (0,25)[(10,5)(3) - (0,5)(3)^2] + (0,25)[(10,5)(4) - (0,5)(4)^2] \\ &= (0,25)[10 + 19 + 27 + 34] = 22,5 \text{ (xu)}. \end{aligned}$$

Ví dụ 2.19. Với biến ngẫu nhiên X trong Ví dụ 2.16, kỳ vọng của biến ngẫu nhiên $Y = X^2$ là

$$E(Y) = E(X^2) = \frac{3}{64} \int_0^4 x^4(4-x)dx = \frac{32}{5}.$$

Sau đây là một số tính chất hữu ích giúp đơn giản hóa trong tính toán kỳ vọng của các biến ngẫu nhiên. Các tính chất này đúng cho cả biến ngẫu nhiên rời rạc và liên tục. Các chứng minh được đưa ra cho biến ngẫu nhiên liên tục.

Định lý 2.5. Nếu a và b là các hằng số thì

$$E(aX + b) = aE(X) + b.$$

Chứng minh. Theo định nghĩa,

$$E(aX + b) = \int_{-\infty}^{+\infty} (ax + b)f_X(x)dx = a \int_{-\infty}^{+\infty} xf_X(x)dx + b \int_{-\infty}^{+\infty} f_X(x)dx = aE(X) + b$$

vì $\int_{-\infty}^{+\infty} f_X(x)dx = 1.$

□

Hệ quả 2.1. 1. Nếu $a = 0, E(b) = b.$

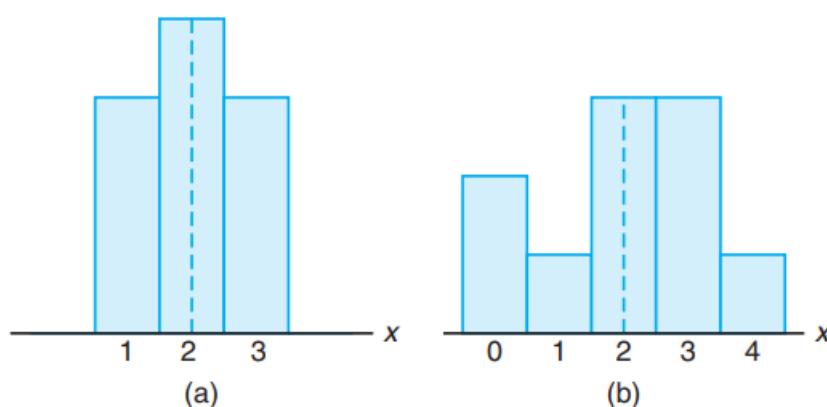
2. Nếu $b = 0, E(aX) = aE(X).$

Định lý 2.6. Cho X là một biến ngẫu nhiên, $h(X), g(X)$ là các hàm của X . Khi đó,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

2.3.2 Phương sai

Giá trị kỳ vọng hay giá trị trung bình của một biến ngẫu nhiên X có tầm quan trọng đặc biệt trong thống kê vì nó phản ánh giá trị trung tâm của phân phối xác suất của biến ngẫu nhiên. Tuy nhiên, giá trị trung bình không đưa ra một mô tả đầy đủ về hình dạng của phân phối. Trong Hình 2.5, ta có biểu đồ của hai phân phối xác suất rời rạc có cùng giá trị trung bình, $\mu = 2$, nhưng khác nhau đáng kể về độ biến thiên hoặc độ phân tán của các quan sát của chúng so với giá trị trung bình. Do đó cần xác định mức độ phân tán của các giá trị của biến ngẫu nhiên xung quanh giá trị trung bình của nó.



Hình 2.5: Phân phối rời rạc với kỳ vọng bằng nhau nhưng độ phân tán khác nhau

Công thức quan trọng nhất về tính biến thiên của biến ngẫu nhiên X có được bằng cách áp dụng Định lý 2.4 với $g(X) = [X - E(X)]^2$. Đại lượng này được gọi là phương sai của biến ngẫu nhiên X hoặc phương sai của phân phối xác suất của X , ký hiệu là $V(X)$ hoặc σ_X^2 , hoặc đơn giản là σ^2 .

Định nghĩa 2.9 (Phương sai). Phương sai (variance) của biến ngẫu nhiên X được định nghĩa như sau:

$$\sigma_X^2 = V(X) = E[X - E(X)]^2 \quad (2.14)$$

Vì $X - E(X)$ là một hàm của biến ngẫu nhiên X , nên từ Định nghĩa 2.9 và Định lý 2.4 ta nhận được các công thức sau đây:

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$V(X) = \sum_{i=1}^n [x_i - E(X)]^2 p_i \quad (2.15)$$

2. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.3) thì

$$V(X) = \sum_{n=1}^{\infty} [x_n - E(X)]^2 p_n \quad (2.16)$$

nếu chuỗi về phải hội tụ.

3. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$, $x \in \mathbb{R}$ thì

$$V(X) = \int_{-\infty}^{+\infty} [x - E(X)]^2 f_X(x) dx \quad (2.17)$$

nếu tích phân về phải hội tụ.

Ví dụ 2.20. Cho X là biến ngẫu nhiên chỉ số lượng ô tô được sử dụng cho mục đích kinh doanh chính thức trong mỗi ngày làm việc. Phân phối xác suất của công ty A , xem Hình 2.5(a), là

X_A	1	2	3
$P(X_A = x_i)$	0,3	0,4	0,3

và của công ty B , xem Hình 2.5(b), là

X_B	0	1	2	3	4
$P(X_B = y_j)$	0,2	0,1	0,3	0,3	0,1

Chỉ ra rằng phương sai của phân phối xác suất của công ty B lớn hơn so với công ty A .

Lời giải Ví dụ 2.20

Từ số liệu của công ty A ta tính

$$E(X_A) = (1)(0,3) + (2)(0,4) + (3)(0,3) = 2,0$$

và

$$V(X_A) = (1 - 2)^2(0,3) + (2 - 2)^2(0,4) + (3 - 2)^2(0,3) = 0,6.$$

Với công ty B ta có

$$E(X_B) = (0)(0,2) + (1)(0,1) + (2)(0,3) + (3)(0,3) + (4)(0,1) = 2,0,$$

và

$$\begin{aligned} V(X_B) &= (0-2)^2(0,2) + (1-2)^2(0,1) + (2-2)^2(0,3) + (3-2)^2(0,3) + (4-2)^2(0,1) \\ &= 1,6. \end{aligned}$$

Công thức tương đương của (2.14) được cho trong định lý dưới đây.

Định lý 2.7.

$$V(X) = E(X^2) - [E(X)]^2 \quad (2.18)$$

Chứng minh. Giả sử X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2),

$$\begin{aligned} V(X) &= \sum_{i=1}^n [x_i - E(X)]^2 p_i = \sum_{i=1}^n \{x_i^2 - 2E(X)x_i + [E(X)]^2\} p_i \\ &= \sum_{i=1}^n x_i^2 p_i - 2E(X) \sum_{i=1}^n x_i p_i + [E(X)]^2 \sum_{i=1}^n p_i = E(X^2) - [E(X)]^2, \end{aligned}$$

vì theo định nghĩa $E(X) = \sum_{i=1}^n x_i p_i$, $E(X^2) = \sum_{i=1}^n x_i^2 p_i$ và $\sum_{i=1}^n p_i = 1$.

Đối với biến ngẫu nhiên liên tục ta chứng minh tương tự. □

Hệ quả 2.2. 1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$V(X) = \sum_{i=1}^n x_i^2 p_i - \left(\sum_{i=1}^n x_i p_i \right)^2 \quad (2.19)$$

2. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.3) thì

$$V(X) = \sum_{n=1}^{\infty} x_n^2 p_n - \left(\sum_{n=1}^{\infty} x_n p_n \right)^2 \quad (2.20)$$

nếu các chuỗi về phải hội tụ.

3. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$, $x \in \mathbb{R}$ thì

$$V(X) = \int_{-\infty}^{+\infty} x^2 f_X(x) dx - \left(\int_{-\infty}^{+\infty} x f_X(x) dx \right)^2 \quad (2.21)$$

nếu các tích phân về phải hội tụ.

Ví dụ 2.21. Dùng Định lý 2.7 tính phương sai của biến ngẫu nhiên xét trong Ví dụ (2.11).

Lời giải Ví dụ 2.21 $E(X^2) = (-990)^2 \times 0,008 + (10)^2 \times 0,992 = 7940$. Suy ra

$$V(X) = E(X^2) - [E(X)]^2 = 7940 - 4 = 7936.$$

Điều này nói lên rằng mặc dù kinh doanh bảo hiểm có lãi nhưng rủi ro khá lớn.

Ví dụ 2.22. Dùng Định lý 2.7 tính phương sai của biến ngẫu nhiên xét trong Ví dụ (2.16).

Lời giải Ví dụ 2.22 Từ kết quả của Ví dụ 2.16 và 2.19 suy ra

$$V(X) = E(X^2) - [E(X)]^2 = \frac{32}{5} - \left(\frac{12}{5}\right)^2 = \frac{16}{25}.$$

Chú ý 2.2. Phương sai của biến ngẫu nhiên là một giá trị xác định không âm.

Nhận xét 2.6. 1. Phương sai chính là trung bình số học của bình phương các sai lệch giữa các giá trị có thể có của biến ngẫu nhiên so với giá trị trung bình của các giá trị đó. Nó phản ánh mức độ phân tán của các giá trị của biến ngẫu nhiên xung quanh giá trị trung tâm của nó là kỳ vọng.

2. Trong kỹ thuật phương sai đặc trưng cho mức độ phân tán của các chi tiết gia công hay sai số của thiết bị. Trong quản lý và kinh doanh thì phương sai đặc trưng cho mức độ rủi ro của các quyết định.

Bây giờ ta sẽ mở rộng khái niệm về phương sai của biến ngẫu nhiên X cho biến ngẫu nhiên liên quan đến X , biến ngẫu nhiên $g(X)$.

Định lý 2.8. Cho X là một biến ngẫu nhiên và $Y = g(X)$ là một hàm của X .

1. Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) thì

$$V(Y) = V[g(X)] = \sum_{i=1}^n \{g(x_i) - E[g(X)]\}^2 p_i. \quad (2.22)$$

2. Nếu X là biến ngẫu nhiên liên tục có hàm mật độ xác suất $f_X(x)$ thì

$$V(Y) = V[g(X)] = \int_{-\infty}^{+\infty} \{g(x) - E[g(X)]\}^2 f_X(x) dx. \quad (2.23)$$

Chứng minh. Sử dụng Định lý 2.4 và Định nghĩa 2.9 ta được điều cần chứng minh. \square

Phương sai của biến ngẫu nhiên X có tính chất sau.

Định lý 2.9. Nếu a và b là các hằng số thì

$$1. V(aX) = a^2 V(X).$$

$$2. V(b) = 0.$$

Hoặc $V(aX + b) = a^2 V(X)$.

Chứng minh. 1. $V(aX) = E(a^2 X^2) - [E(aX)]^2 = a^2 E(X^2) - a^2 [E(X)]^2 = a^2 V(X)$.

$$2. V(b) = E(b^2) - [E(b)]^2 = b^2 - b^2 = 0.$$

\square

2.3.3 Độ lệch chuẩn

Định nghĩa 2.10 (Độ lệch chuẩn). Độ lệch chuẩn (standard deviation) của biến ngẫu nhiên X , ký hiệu là $\sigma(X)$, được định nghĩa như sau:

$$\sigma(X) = \sqrt{V(X)} \quad (2.24)$$

Nhận xét 2.7. Khi cần đánh giá mức độ phân tán của biến ngẫu nhiên theo đơn vị đo của nó người ta dùng độ lệch chuẩn vì độ lệch chuẩn có cùng đơn vị đo với đơn vị đo của biến ngẫu nhiên.

Ví dụ 2.23. Tính phương sai và độ lệch chuẩn của biến ngẫu nhiên xét trong Ví dụ (2.11).

Lời giải Ví dụ 2.23 Từ kết quả trong Ví dụ 2.21, $\sigma(X) = \sqrt{V(X)} = \sqrt{7936} \simeq 89,08$.

Ví dụ 2.24. Tính độ lệch chuẩn của biến ngẫu nhiên xét trong Ví dụ (2.16).

Lời giải Ví dụ 2.24 Từ kết quả của Ví dụ 2.22 suy ra $\sigma(X) = \sqrt{V(X)} = \frac{4}{5}$.

2.3.4 Một số đặc trưng khác

2.3.4a Một (mode)

Định nghĩa 2.11 (Một). 1. Nếu X là biến ngẫu nhiên rời rạc thì một là giá trị của X ứng với xác suất lớn nhất.

2. Nếu X là biến ngẫu nhiên liên tục thì một là giá trị làm hàm mật độ đạt max.

Ký hiệu $modX$.

2.3.4b Trung vị (median)

Định nghĩa 2.12 (Trung vị). Trung vị của biến ngẫu nhiên X , ký hiệu là $medX$, là giá trị của biến ngẫu nhiên X chia phân phối thành hai phần có xác suất giống nhau, nghĩa là

$$P(X < medX) = P(X \geq medX) = \frac{1}{2} \quad (2.25)$$

Nhận xét 2.8. 1. Từ định nghĩa hàm phân phối, để tìm trung vị ta cần giải phương trình $F_X(x) = \frac{1}{2}$.

2. Trong nhiều trường hợp ứng dụng, trung vị là đặc trưng vị trí tốt nhất, nhiều khi tốt hơn cả kỳ vọng, nhất là khi trong số liệu có những sai sót. Trung vị còn có tên là phân vị 50% của phân phối.

Ví dụ 2.25. Cho hàm mật độ xác suất của biến ngẫu nhiên liên tục X :

$$f_X(x) = \begin{cases} \frac{3}{4}x(2-x), & 0 \leq x \leq 2, \\ 0, & \text{nếu trái lại.} \end{cases}$$

Tìm $medX$ và $modX$.

Lời giải Ví dụ 2.25 Theo (2.7), hàm phân phối xác suất của X là

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ \frac{3}{4}\left(x^2 - \frac{x^3}{3}\right), & 0 < x \leq 2, \\ 1, & x > 2. \end{cases}$$

Khi đó $medX$ là nghiệm của phương trình $F_X(x) = \frac{1}{2}$. Hay $x^3 - 3x^2 + 2 = 0$ với $0 < x \leq 2$. Suy ra $medX = 1$.

Hàm mật độ xác suất $f_X(x)$ có

$$f'_X(x) = \begin{cases} 0, & x \leq 0, \\ \frac{3}{2}(1-x), & 0 < x < 2, \\ 0, & \text{nếu trái lại,} \end{cases}$$

đổi dấu từ dương sang âm khi đi qua $x = 1$, do đó đạt cực đại tại điểm này, nên $modX = 1$.

BÀI 7 (2 tiết)

2.4 Một số phân phối xác suất thông dụng

2.4.1 Phân phối đều

2.4.1a Phân phối đều rời rạc

Định nghĩa 2.13 (Phân phối đều rời rạc). Biến ngẫu nhiên X được gọi là tuân theo luật phân phối đều rời rạc (discrete uniform distribution) với tham số n , ký hiệu là $X \sim \mathcal{U}(n)$, nếu X có bảng phân phối xác suất

X	1	2	...	n
$P(X = x_i)$	$\frac{1}{n}$	$\frac{1}{n}$...	$\frac{1}{n}$

(2.26)

Sử dụng Định nghĩa 2.8(1) và Định nghĩa 2.9(1) ta nhận được kỳ vọng và phương sai của biến ngẫu nhiên phân phối đều.

Định lý 2.10. Kỳ vọng, phương sai của biến ngẫu nhiên có phân phối đều rời rạc:

$$E(X) = \frac{n+1}{2}, \quad V(X) = \frac{n^2-1}{12} \quad (2.27)$$

2.4.1b Phân phối đều liên tục

Định nghĩa 2.14 (Phân phối đều liên tục). Biến ngẫu nhiên X được gọi là tuân theo luật phân phối đều liên tục (continuous uniform distribution) trên $[a, b]$ ($a < b$), ký hiệu là $X \sim \mathcal{U}([a, b])$, nếu X có hàm mật độ

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b]. \end{cases} \quad (2.28)$$

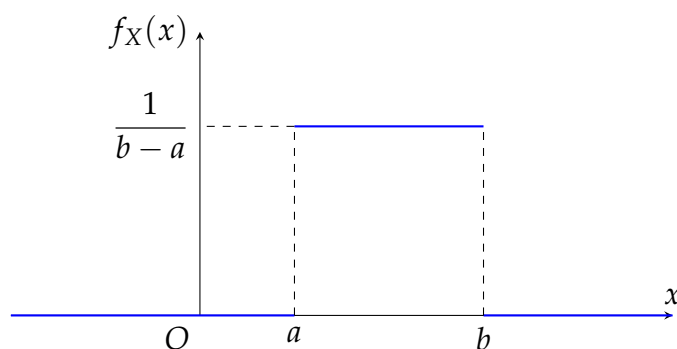
Định lý 2.11. Nếu X là biến ngẫu nhiên có phân phối đều liên tục trên $[a, b]$ thì

1. Hàm phân phối xác suất là

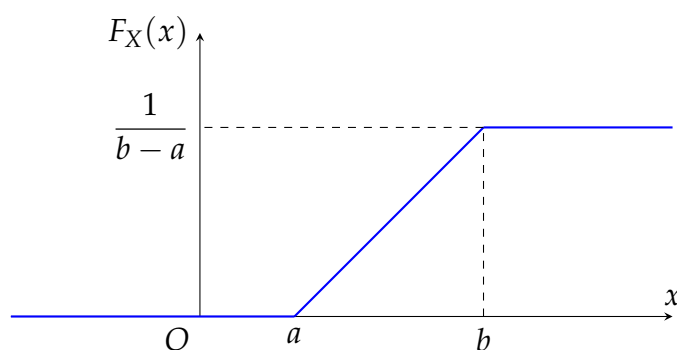
$$F_X(x) = \int_{-\infty}^x f_X(u) du = \begin{cases} 0, & x \leq a, \\ \frac{x-a}{b-a}, & a < x \leq b, \\ 1, & x > b. \end{cases} \quad (2.29)$$

2. Kỳ vọng $E(X) = \frac{a+b}{2}$.

3. Phương sai $V(X) = \frac{(b-a)^2}{12}$ và độ lệch chuẩn $\sigma(X) = \sqrt{\frac{(b-a)^2}{12}}$.



Hình 2.6: Đồ thị hàm mật độ (2.28) của biến ngẫu nhiên có phân phối đều



Hình 2.7: Đồ thị hàm phân phối (2.29) của biến ngẫu nhiên có phân phối đều

Nhận xét 2.9. 1. X có khả năng nhận giá trị trong khoảng (a, b) là "đều nhau".

2. Phân phối đều có nhiều ứng dụng trong thống kê toán như mô phỏng thống kê, đặc biệt trong phương pháp phi tham số.
3. Trong một số lý thuyết kết luận thống kê người ta thường xuất phát từ quy tắc sau đây: Nếu ta không biết gì về giá trị của tham số cần ước lượng, mỗi giá trị có thể có của tham số đó là đồng khả năng, điều đó dẫn đến việc quan niệm tham số cần ước lượng như một biến ngẫu nhiên có phân phối đều.

Ví dụ 2.26. Lịch chạy của xe bus tại một trạm xe bus như sau: chiếc xe bus đầu tiên trong ngày sẽ khởi hành từ trạm này lúc 7 giờ, cứ sau 15 phút sẽ có một xe khác đến trạm. Giả sử một hành khách đến trạm ngẫu nhiên trong khoảng thời gian từ 7 giờ đến 7 giờ 30. Tìm xác suất để hành khách này chờ (a) Ít hơn 5 phút; (b) Ít nhất 12 phút.

Lời giải Ví dụ 2.26 Gọi X là số phút từ 7 giờ đến 7 giờ 30 hành khách đến trạm, ta có $X \sim \mathcal{U}([0, 30])$.

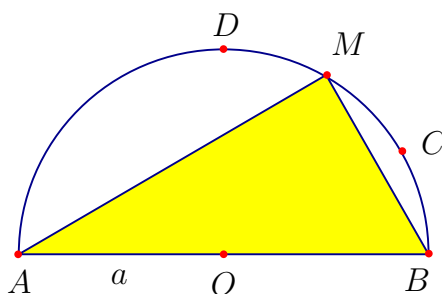
- (a) Hành khách chờ ít hơn 5 phút nếu đến trạm giữa 7 giờ 10 và 7 giờ 15 hoặc giữa 7 giờ 25 và 7 giờ 30. Do đó xác suất cần tìm là:

$$P(10 < X \leq 15) + P(25 < X \leq 30) = \frac{5}{30} + \frac{5}{30} = \frac{1}{3}.$$

(b) Hành khách chờ ít nhất 12 phút nếu đến trạm giữa 7 giờ và 7 giờ 03 hoặc giữa 7 giờ 15 và 7 giờ 18. Xác suất cần tìm là:

$$P(0 < X \leq 3) + P(15 < X \leq 18) = \frac{3}{30} + \frac{3}{30} = 0,2.$$

Ví dụ 2.27. Lấy ngẫu nhiên một điểm M trên nửa đường tròn tâm O , đường kính $AB = 2a$. Biết rằng xác suất điểm M rơi vào cung CD bất kì của nửa đường tròn AMB chỉ phụ thuộc vào độ dài cung CD . (a) Tìm hàm phân phối xác suất của biến ngẫu nhiên Y chỉ diện tích tam giác AMB . (b) Tìm kỳ vọng của Y .



Hình 2.8: Minh họa cho Ví dụ 2.27

Lời giải Ví dụ 2.27 (a) Theo định lý hàm số sin, ta có $S_{AMB} = a^2 \sin \varphi$, ở đây φ là góc giữa trục Ox và OM . Từ giả thiết ta có φ là biến ngẫu nhiên liên tục có phân phối đều $\mathcal{U}[0, \pi]$ có hàm mật độ xác suất

$$f_{\varphi}(x) = \begin{cases} \frac{1}{\pi}, & x \in [0, \pi], \\ 0, & x \notin [0, \pi]. \end{cases}$$

Do đó, hàm phân phối xác suất của φ là

$$F_{\varphi}(x) = \begin{cases} 0, & x \leq 0, \\ \frac{x}{\pi}, & 0 < x \leq \pi, \\ 1, & x > \pi. \end{cases}$$

Biến ngẫu nhiên $Y = a^2 \sin \varphi$, nên Y là biến ngẫu nhiên liên tục nhận giá trị trong đoạn $[0, a^2]$. Hàm phân phối xác suất của Y là

$$F_Y(x) = P(Y < x) = \begin{cases} 0, & x \leq 0, \\ \frac{2}{\pi} \arcsin \frac{x}{a^2}, & 0 < x \leq a^2, \\ 1, & x > a^2, \end{cases}$$

vì với $x \in (0, a^2]$,

$$\begin{aligned} F_Y(x) &= P(Y < x) = P(a^2 \sin \varphi < x) = P(\sin \varphi < \frac{x}{a^2}) \\ &= P\left(0 < \varphi < \arcsin \frac{x}{a^2}\right) + P\left(\pi - \arcsin \frac{x}{a^2} < \varphi < \pi\right) = \frac{2}{\pi} \arcsin \frac{x}{a^2}. \end{aligned}$$

(b) Hàm mật độ xác suất của X

$$f_X(x) = \begin{cases} \frac{2}{\pi \sqrt{a^4 - x^2}}, & x \in [0, a^2], \\ 0, & x \notin [0, a^2]. \end{cases}$$

Suy ra kỳ vọng của X là

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x) dx = \frac{2}{\pi} \int_0^{a^2} \frac{x}{\sqrt{a^4 - x^2}} dx = \frac{2}{\pi} a^2.$$

2.4.2 Phân phối nhị thức

2.4.2a Phân phối Béc-nu-li

Định nghĩa 2.15 (Phân phối Béc-nu-li). Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối Béc-nu-li (Bernoulli distribution) với tham số p , ký hiệu là $X \sim \mathcal{B}(1, p)$, nếu X nhận hai giá trị 0, 1 với xác suất tương ứng

$$P(X = k) = p^k q^{1-k}, \quad k = 0, 1, \quad (2.30)$$

ở đây $0 < p < 1, q = 1 - p$.

Định lý 2.12. Nếu X là biến ngẫu nhiên có phân phối Béc-nu-li $\mathcal{B}(1; p)$ thì

$$\boxed{E(X) = p, \quad V(X) = p(1 - p)} \quad (2.31)$$

Chứng minh. Theo định nghĩa kỳ vọng và phương sai của biến ngẫu nhiên rời rạc,

$$E(X) = (0)(1 - p) + (1)(p) = p \quad \text{và} \quad V(X) = (0)^2(1 - p) + (1)^2(p) - p^2 = p(1 - p).$$

□

Nhận xét 2.10. Xét phép thử Béc-nu-li với sự thành công của phép thử là sự xuất hiện của sự kiện A và giả sử xác suất xuất hiện A trong mỗi lần thử là p . Gọi X là số lần thành công trong một lần thử thì X là biến ngẫu nhiên rời rạc có phân bố Béc-nu-li tham số p . Biến ngẫu nhiên X còn được gọi là tuân theo phân phối không - một $\mathcal{A}(p)$.

2.4.2b Phân phối nhị thức

Định nghĩa 2.16 (Phân phối nhị thức). Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối nhị thức (binomial distribution) với tham số n và p , ký hiệu là $X \sim \mathcal{B}(n, p)$, nếu X có bảng phân phối xác suất

X	0	1	...	k	...	n
$P(X = x_i)$	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$...	$C_n^k p^k q^{n-k}$...	$C_n^n p^n q^0$

(2.32)

ở đây $q = 1 - p$ và $P(X = k) = C_n^k p^k q^{n-k}$ được tính bằng công thức Béc-nu-li (1.19).

Phân phối nhị thức xuất phát từ tên thực tế của khai triển nhị thức $(p + q)^n$ có $n + 1$ số hạng:

$$(p + q)^n = C_n^0(p)^0(q)^n + C_n^1(p)^1(q)^{n-1} + \dots + C_n^n(p)^n(q)^0.$$

Nếu $p + q = 1$ thì $\sum_{k=0}^n C_n^k(p)^k(q)^{n-k} = 1$, đây là điều kiện cần thiết của phân phối (2.32).

Nhận xét 2.11. 1. Thực hiện n phép thử Béc-nu-li với xác suất thành công của sự kiện A trong mỗi lần thử là p . Với mỗi $i = 1, 2, \dots, n$, nếu ở lần thử thứ i sự kiện A xuất hiện ta cho X_i nhận giá trị 1, nếu sự kiện A không xuất hiện ta cho X_i nhận giá trị 0. Như vậy $X_i \sim \mathcal{B}(1, p)$. Gọi X là số lần thành công trong n phép thử Béc-nu-li này thì

$$X = X_1 + X_2 + \dots + X_n \sim \mathcal{B}(n, p). \quad (2.33)$$

2. Nếu $X \sim \mathcal{B}(n_1, p)$ và $Y \sim \mathcal{B}(n_2, p)$ và nếu X, Y độc lập thì $X + Y \sim \mathcal{B}(n_1 + n_2, p)$ (xem Chương 3).

Định lý 2.13. Kỳ vọng và phương sai của biến ngẫu nhiên có phân phối nhị thức là

$$\boxed{E(X) = np, \quad V(X) = npq} \quad (2.34)$$

Chứng minh. Từ Nhận xét 2.11(1), biến ngẫu nhiên X xác định bởi (2.33) có phân phối nhị thức $\mathcal{B}(n; p)$. Theo Hệ quả 3.1, 3.3 (Chương 3) và Định lý 2.12,

$$E(X) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = np$$

và

$$V(X) = \sum_{i=1}^n V(X_i) = \sum_{i=1}^n pq = npq.$$

□

Ví dụ 2.28. Tỷ lệ phế phẩm của lô hàng là 4%. Chọn ngẫu nhiên 20 sản phẩm để kiểm tra. Gọi X là số phế phẩm phát hiện được. (a) X có phân phối gì? (b) Tính xác suất có đúng 5 phế phẩm phát hiện được. (c) Lô hàng được xem là đạt tiêu chuẩn nếu số phế phẩm phát hiện được không nhiều hơn 2. Tính xác suất để lô hàng đạt tiêu chuẩn.

Lời giải Ví dụ 2.28 Có thể xem việc kiểm tra chất lượng mỗi sản phẩm là thực hiện một phép thử Béc-nu-li với sự thành công của phép thử là phát hiện ra phế phẩm. Theo giả thiết xác suất thành công của mỗi lần thử là 0,04. Kiểm tra 20 sản phẩm là thực hiện 20 phép thử.

(a) Số phế phẩm phát hiện được là số lần thành công trong 20 phép thử này. Vậy X có phân phối nhị thức $X \sim \mathcal{B}(n, p)$, với $n = 20$, $p = 0,04$.

(b) $P(X = 5) = C_{20}^5 \times (0,04)^5 \times (0,96)^{15} = 0,0008$.

(c) Xác suất để lô hàng đạt tiêu chuẩn là

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= C_{20}^0 (0,04)^0 (0,96)^{20} + C_{20}^1 (0,04)^1 (0,96)^{19} + C_{20}^2 (0,04)^2 (0,96)^{18} \simeq 0,956. \end{aligned}$$

2.4.3 Phân phối Poa-xông

Các phép thử mang lại các giá trị số cho biến ngẫu nhiên X , chỉ số các kết quả xảy ra trong một khoảng thời gian nhất định nào đó, được gọi là phép thử Poa-xông. Khoảng thời gian nhất định có thể là một phút, một ngày, thậm trí một năm. Chẳng hạn phép thử đếm số cuộc điện thoại gọi đến một tổng đài trong vòng 5 phút là một phép thử Poa-xông.

Một phép thử Poa-xông có nguồn gốc từ quá trình Poa-xông và có các tính chất sau.

1. Số lượng kết quả xảy ra trong một khoảng thời gian không phụ thuộc vào số lượng kết quả xảy ra trong bất kỳ khoảng thời gian nào khác. Do đó, quá trình Poa-xông có tính chất không nhớ.
2. Xác suất xảy ra một kết quả trong một khoảng thời gian ngắn tỷ lệ thuận với độ dài của khoảng thời gian và không phụ thuộc vào số lượng kết quả xảy ra bên ngoài khoảng thời gian này.
3. Xác suất có nhiều hơn một kết quả sẽ xảy ra trong một khoảng thời gian ngắn là không đáng kể.

Định nghĩa 2.17 (Phân phối Poa-xông). Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối Poa-xông (Poisson distribution) với tham số λ , ký hiệu là $X \sim \mathcal{P}(\lambda)$, nếu X có bảng phân phối xác suất

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

(2.35)

trong đó $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$ được tính bằng công thức Poa-xông, λ là số kết quả trung bình trên mỗi đơn vị thời gian và $e = 2,71828 \dots$

Định lý 2.14. Giả sử X là biến ngẫu nhiên có phân phối Poa-xông tham số λ . Khi đó,

1. Hàm phân phối xác suất của X là $F_X(x) = e^{-\lambda} \sum_{k=0}^n \frac{\lambda^k}{k!}$, $n < x \leq n+1$.
2. Kỳ vọng $E(X) = \lambda$.
3. Phương sai $V(X) = \lambda$ và độ lệch chuẩn $\sigma(X) = \sqrt{\lambda}$.

Chứng minh. Ta sẽ chứng minh hai tính chất đầu của định lý.

1. Suy trực tiếp từ Định nghĩa (2.4) và Định nghĩa 2.17.
2. Sử dụng Định nghĩa 2.8

$$E(X) = \sum_{k=0}^{\infty} kP(X=k) = \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda,$$

$$\text{vì } \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \sum_{l=0}^{\infty} \frac{\lambda^l}{l!} = e^{\lambda}.$$

□

Nhận xét 2.12. 1. Phân phối Poa-xông được ứng dụng rộng rãi trong nhiều lĩnh vực thực tế như kiểm tra chất lượng sản phẩm, lý thuyết sắp hàng, các hệ phục vụ đám đông, các bài toán chuyển mạch trong tổng đài ...

2. Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân phối Poa-xông tham số lần lượt λ_1, λ_2 , thì $X_1 + X_2$ cũng có phân phối Poa-xông tham số $\lambda_1 + \lambda_2$ (xem Chương 3).
3. Trong thực tế với một số giả thiết thích hợp thì các biến ngẫu nhiên là các quá trình đếm sau: số cuộc gọi đến một tổng đài; số khách hàng đến một điểm phục vụ; số xe cộ qua một ngã tư; số tai nạn (xe cộ); số các sự cố xảy ra ở một địa điểm ... trong một khoảng thời gian xác định nào đó sẽ có phân phối Poa-xông với tham số λ , trong đó λ là tốc độ trung bình diễn ra trong khoảng thời gian này.

Ví dụ 2.29. Ở một tổng đài bưu điện, các cuộc điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau với tốc độ trung bình 2 cuộc gọi trong một phút. Tìm xác suất để (a) Có đúng 5 cuộc điện thoại trong vòng 2 phút; (b) Không có cuộc điện thoại nào trong khoảng thời gian 30 giây; (c) Có ít nhất 1 cuộc điện thoại trong khoảng thời gian 10 giây.

Lời giải Ví dụ 2.29

- (a) Gọi X là số cuộc điện thoại xuất hiện trong vòng 2 phút. $X \sim \mathcal{P}(\lambda)$, λ chính là số cuộc điện thoại trung bình đến trong vòng 2 phút, $\lambda = 4$.

$$P(X=5) = e^{-\lambda} \frac{\lambda^5}{5!} = e^{-4} \frac{4^5}{5!} = 0,156.$$

- (b) Gọi X là số cuộc điện thoại xuất hiện trong vòng 30 giây. $X \sim \mathcal{P}(\lambda)$ với $\lambda = 1$.

$$P(X=0) = e^{-\lambda} \frac{\lambda^0}{0!} = e^{-1} = 0,3679.$$

(c) Gọi X là số cuộc điện thoại xuất hiện trong vòng 10 giây. $X \sim \mathcal{P}(\lambda)$ với $\lambda = 1/3$.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - e^{-1/3} = 0,2835.$$

Ví dụ 2.30. Một gara cho thuê ô tô thấy rằng số người đến thuê ô tô vào thứ bảy cuối tuần là một biến ngẫu nhiên tuân theo luật phân phối Poa-xông với tham số $\lambda = 2$. Giả sử gara có 4 chiếc ô tô. (a) Tìm xác suất để tất cả 4 ô tô đều được thuê vào thứ 7. (b) Tìm xác suất gara không đáp ứng được yêu cầu (thiếu xe cho thuê) vào thứ 7. (c) Trung bình có bao nhiêu ô tô được thuê vào ngày thứ 7?

Lời giải Ví dụ 2.30 Gọi X là biến ngẫu nhiên chỉ "số người đến thuê ô tô vào thứ bảy". Theo giả thiết X là biến ngẫu nhiên phân phối tuân theo quy luật Poa-xông $\mathcal{P}(\lambda)$ với $\lambda = 2$. Gọi Y là biến ngẫu nhiên chỉ "số xe ô tô được thuê vào thứ bảy".

(a) Áp dụng phân phối Poa-xông, $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$,

$$\begin{aligned} P(Y = 4) &= P(X \geq 4) = 1 - P(X < 4) \\ &= 1 - P(X = 0) - P(X = 1) - P(X = 2) - P(X = 3) \\ &= 1 - e^{-2} \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} \right) = 0,1429. \end{aligned}$$

(b) $P(X > 4) = P(X \geq 4) - P(X = 4) = 0,1429 - e^{-2} \frac{2^4}{4!} = 0,0527$.

(c) Y có thể nhận các giá trị 0, 1, 2, 3, 4, với

$$\begin{aligned} P(Y = 0) &= P(X = 0) = 0,1353, & P(Y = 1) &= P(X = 1) = 0,2707, \\ P(Y = 2) &= P(X = 2) = 0,2707, & P(Y = 3) &= P(X = 3) = 0,1804, \\ P(Y = 4) &= P(X \geq 4) = 0,1429. \end{aligned}$$

Bảng phân phối xác suất của Y là:

x	0	1	2	3	4
$P(Y = x)$	0,1353	0,2707	0,2707	0,1804	0,1429

Vậy trung bình số ô tô được thuê trong ngày thứ bảy là $E(Y) = 1,9249$, tức là khoảng 2 chiếc.

Chú ý 2.3. Giá trị xác suất của phân phối Poa-xông được tính sẵn trong bảng Phụ lục 5.

2.4.4 Xấp xỉ phân phối nhị thức bằng phân phối Poa-xông

Trong Mục 1.5.5 ta đã đề cập đến việc tính xấp xỉ công thức Béc-nu-y (1.19) khi số phép thử n khá lớn bởi công thức (1.21). Ở đây ta xét mối liên hệ của hai phân phối tương ứng.

Định lý 2.15. Cho X là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n, p)$. Nếu $n \rightarrow \infty, p \rightarrow 0$ và $np \rightarrow \lambda$ (λ là một hằng số) thì

$$\mathcal{B}(n, p) \rightarrow \mathcal{P}(\lambda) \quad \text{khi } n \rightarrow \infty.$$

Trong thực tế nếu n đủ lớn và $\lambda = np$ đủ nhỏ (thỏa mãn $np < 7$) thì ta có thể xấp xỉ phân phối nhị thức $\mathcal{B}(n, p)$ bằng phân phối Poa-xông $\mathcal{P}(\lambda)$ và

$$P_n(k) = C_n^k(p)^k(1-p)^{n-k} \simeq \frac{(np)^k}{k!} e^{-np} \quad (2.36)$$

Ví dụ 2.31. Giả sử một công ty bảo hiểm nhân thọ bảo hiểm cho cuộc sống của 5000 người đàn ông ở độ tuổi 42. Nghiên cứu của các chuyên gia tính toán cho thấy xác suất để một người đàn ông 42 tuổi sẽ chết trong một năm (xác định) là 0,001. Hãy tìm xác suất mà công ty sẽ phải trả bảo hiểm cho 4 người trong một năm (xác định).

Lời giải Ví dụ 2.31 Gọi X là số người chết trong một năm (xác định). X là biến ngẫu nhiên có phân phối nhị thức tham số $n = 5000$ và $p = 0,001$. Khi đó,

$$P(X = 4) = P_{5000}(4) = C_{5000}^4(0,001)^4(1 - 0,001)^{5000-4} = \frac{5000!}{4!4996!}(0,001)^4(0,999)^{4996}.$$

Vì $n = 5000$ đủ lớn và $\lambda = np = (5000)(0,001) = 5$ nên xác suất trên có thể được xấp xỉ bằng công thức Poa-xông:

$$P(X = 4) \simeq \frac{5^4}{4!} e^{-5} = \frac{(625)(0,006738)}{24} = 0,175.$$

BÀI 8 (2 tiết)

2.4.5 Phân phối mũ

Định nghĩa 2.18 (Phân phối mũ). Biến ngẫu nhiên X được gọi là tuân theo luật phân phối mũ (exponential distribution) với tham số $\lambda > 0$ nếu nó có hàm mật độ xác suất có dạng:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (2.37)$$

Ký hiệu: $X \sim \mathcal{E}(\lambda)$.

Ví dụ 2.32. Biến ngẫu nhiên T có phân phối mũ với hàm phân phối xác suất

$$F_T(t) = \begin{cases} 1 - e^{-t/3}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

(a) Tìm hàm mật độ xác suất của T . (b) Tính $P[2 \leq T \leq 4]$. (c) Tính kỳ vọng, phương sai, độ lệch chuẩn của T .

Lời giải Ví dụ 2.32

(a)

$$f_T(t) = \frac{dF_T(t)}{dt} = \begin{cases} \frac{1}{3}e^{-t/3}, & t \geq 0, \\ 0, & \text{trái lại.} \end{cases}$$

Theo Định nghĩa 2.18, biến ngẫu nhiên T có phân phối mũ với tham số $\lambda = 1/3$.

(b) $P[2 \leq T \leq 4] = F_T(4) - F_T(2) = e^{-2/3} - e^{-4/3} = 0,250$.

(c) Sử dụng phương pháp tích phân từng phần,

$$E(T) = \int_{-\infty}^{+\infty} t f_T(t) dt = \int_0^{+\infty} t \frac{1}{3} e^{-t/3} dt = -te^{-t/3} \Big|_0^{+\infty} + \int_0^{+\infty} e^{-t/3} dt = 3.$$

Để tính phương sai của T , ta tính $E(T^2) = \int_{-\infty}^{+\infty} t^2 f_T(t) dt = \int_0^{+\infty} t^2 \frac{1}{3} e^{-t/3} dt$. Tích phân từng phần ta được

$$E(T^2) = -t^2 e^{-t/3} \Big|_0^{+\infty} + \int_0^{+\infty} 2te^{-t/3} dt = 2 \int_0^{+\infty} te^{-t/3} dt.$$

Vì $E(T) = 3$, nên $\int_0^{+\infty} te^{-t/3} dt = 3E(T) = 9$. Do đó $E(T^2) = 6E(T) = 18$ và

$$V(T) = E(T^2) - [E(T)]^2 = 18 - 3^2 = 9.$$

Độ lệch chuẩn là $\sigma(T) = \sqrt{V(T)} = 3$.

Định lý 2.16. Nếu X là biến ngẫu nhiên có phân phối mũ $\mathcal{E}(\lambda)$ thì

$$1. F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0, \\ 0, & \text{trái lại.} \end{cases}$$

$$2. \mu = E(X) = 1/\lambda.$$

$$3. \sigma^2 = V(X) = 1/\lambda^2.$$

Nhận xét 2.13. 1. $P(X > x) = 1 - P(X \leq x) = 1 - P(X < x) = 1 - F_X(x) = e^{-\lambda x}.$

2. Phân phối mũ có tính chất không nhớ: $P(X > t + s | X > t) = P(X > s).$ Vì

$$P(X > t + s | X > t) = \frac{P[(X > t + s)(X > t)]}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)}.$$

Chú ý 2.4. 1. Phân phối mũ có nhiều ứng dụng trong thực tiễn.

2. Nói chung với một giả thiết nào đó, khoảng thời gian giữa hai lần xuất hiện của một sự kiện E nào đó sẽ có phân phối mũ. Vì lý do này phân phối mũ còn có tên gọi là phân phối của thời gian chờ đợi “Waiting time distribution” (khoảng thời gian giữa 2 ca cấp cứu ở một bệnh viện, khoảng thời gian giữa 2 lần hỏng hóc của một chiếc máy, khoảng thời gian giữa 2 trận lụt hay động đất...).

Ví dụ 2.33. 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 kỳ vọng là 6,25. Thời gian bảo hành của mạch điện tử này là 5 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Lời giải Ví dụ 2.33 Gọi X là tuổi thọ của mạch điện tử trong máy tính. X tuân theo phân phối mũ với tham số $\lambda = \frac{1}{E(X)} = \frac{1}{6,25} = 0,16.$

$P(X \leq 5) = 1 - P(X > 5) = 1 - e^{-5\lambda} = 1 - e^{-0,8} = 0,5506.$ Vậy có khoảng 55,06% mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Ví dụ 2.34. Công ty điện thoại A thu phí 0,15\$ mỗi phút cho các cuộc gọi điện thoại. Với bất kỳ cuộc gọi nào trong vòng một phút, họ sẽ tính phí trong một phút. Công ty điện thoại B cũng tính phí 0,15\$ mỗi phút. Tuy nhiên, công ty điện thoại B tính toán phí dựa trên thời lượng chính xác của một cuộc gọi. Cho T , thời lượng của một cuộc gọi tính bằng phút, là biến ngẫu nhiên có phân phối mũ với tham số $\lambda = 1/3.$ (a) Hàm mật độ xác suất của T là gì? (b) Kỳ vọng của T là bao nhiêu? (c) Doanh thu trung bình cho mỗi cuộc gọi $E(R_A)$ và $E(R_B)$ của công ty A và B là bao nhiêu?

Lời giải Ví dụ 2.34

(a) Vì T có phân phối mũ với $\lambda = 1/3$, $f_T(t) = \begin{cases} \frac{1}{3}e^{-\frac{1}{3}t}, & t \geq 0, \\ 0, & \text{trái lại.} \end{cases}$

(b) Theo Định lý 2.16, $E[T] = \int_{-\infty}^{+\infty} t f_T(t) dt = \frac{1}{\lambda} = 3$ phút/cuộc gọi.

(c) Với công ty B , $E(R_B) = 0,15 \times E(T) = 0,45$ \$/cuộc gọi.

Với công ty A , đặt $K = [T]$ theo nghĩa nếu $0 < T \leq 1$ thì $K = 1$, nếu $1 < T \leq 2$ thì $K = 2, \dots$. Khi đó $E(R_A) = 0,15 \times E(K)$. Để tính $E(K)$, trước hết ta tính

$$P(K = k) = P(k-1 < T \leq k) = \int_{k-1}^k f_T(t) dt = (e^{-\lambda})^{k-1} (1 - e^{-\lambda}).$$

Suy ra $E(K) = \sum_{k=1}^{\infty} k P(K = k) = \sum_{k=1}^{\infty} k (1-p)^{k-1} p = \frac{1}{p}$ với $p = 1 - e^{-\lambda}$. Vậy

$$E(R_A) = \frac{0,15}{p} = \frac{0,15}{0,2834} = (0,15) \times (3,5285) = 0,5292 \text{ $/cuộc gọi.}$$

2.4.6 Phân phối chuẩn

2.4.4a Phân phối chuẩn

Định nghĩa 2.19 (Phân phối chuẩn). Biến ngẫu nhiên liên tục X được gọi là tuân theo luật phân phối chuẩn (normal distribution) với tham số μ, σ^2 , ký hiệu là $X \sim \mathcal{N}(\mu, \sigma^2)$, nếu hàm mật độ xác suất của X có dạng

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R} \quad (2.38)$$

ở đây e và π được lấy xấp xỉ lần lượt là 2,71828 và 3,14159.

Định lý 2.17. Kỳ vọng và phương sai của biến ngẫu nhiên X có phân phối chuẩn

$$E(X) = \mu, \quad V(X) = \sigma^2 \quad (2.39)$$

và độ lệch tiêu chuẩn là $\sigma(X) = \sigma$.

Chứng minh. Để xác định kỳ vọng, trước hết ta tính

$$E(X - \mu) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu) e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Đặt $z = (x - \mu)/\sigma$ và $dx = \sigma dz$, ta nhận được

$$E(X - \mu) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-\frac{z^2}{2}} dz = 0,$$

vì hàm số dưới dấu tích phân là hàm lẻ của z . Do đó,

$$E(X) = \mu.$$

Phương sai của biến ngẫu nhiên tuân theo luật phân phối chuẩn được cho bởi

$$E[(X - \mu)^2] = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Đặt $z = (x - \mu)/\sigma$ và $dx = \sigma dz$, ta nhận được

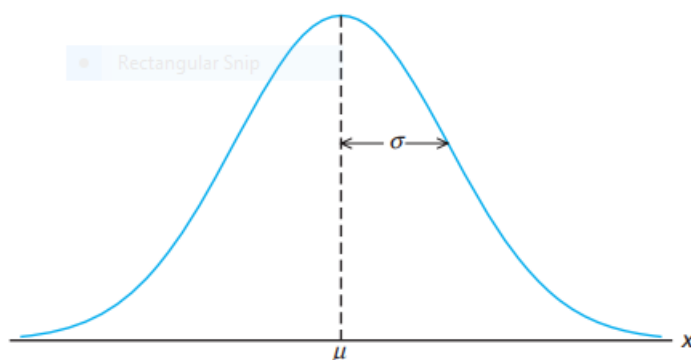
$$E[(X - \mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 e^{-\frac{z^2}{2}} dz.$$

Tích phân từng phần với $u = z$ và $dv = ze^{-z^2/2} dz$ suy ra $du = dz$ và $v = -e^{-z^2/2}$, ta tìm được

$$E[(X - \mu)^2] = \frac{\sigma^2}{\sqrt{2\pi}} \left(-ze^{-z^2/2} \Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} dz \right) = \sigma^2(0 + 1) = \sigma^2.$$

□

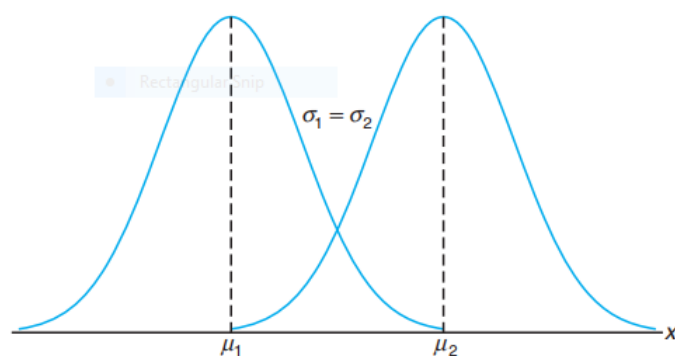
Nhận xét 2.14. Phân phối liên tục quan trọng nhất trong lĩnh vực thống kê là phân phối chuẩn. Đồ thị của hàm mật độ xác suất $f_X(x)$ của biến ngẫu nhiên tuân theo luật phân phối chuẩn, được gọi là đường cong chuẩn, có dạng hình chuông (xem Hình 2.9), mô tả gần đúng nhiều hiện tượng trong tự nhiên, công nghiệp và nghiên cứu.



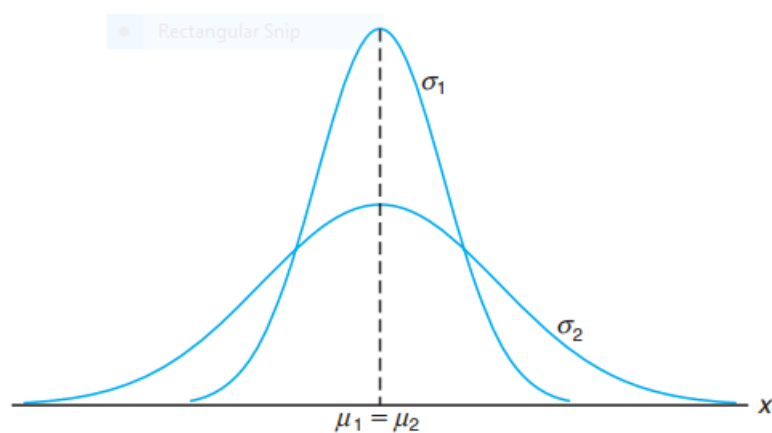
Hình 2.9: Đường cong chuẩn

Hình 2.10 mô tả hai đường cong chuẩn có cùng độ lệch chuẩn nhưng kỳ vọng khác nhau. Hai đường cong giống hệt nhau về hình thức nhưng được tập trung tại các vị trí khác nhau dọc theo trục hoành.

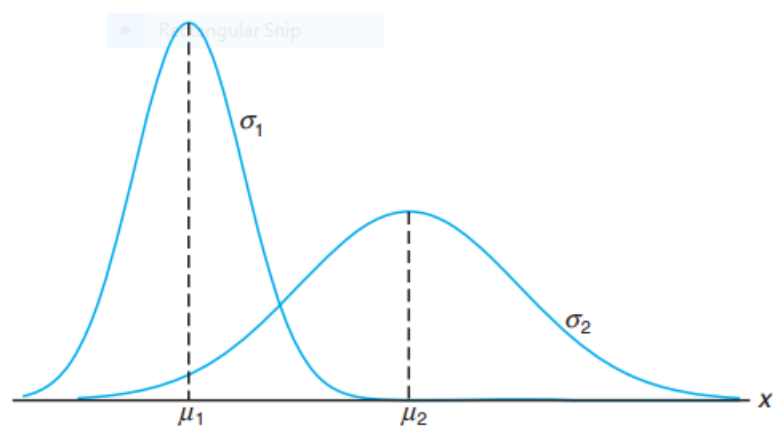
Hình 2.11 mô tả hai đường cong chuẩn có cùng kỳ vọng nhưng độ lệch chuẩn khác nhau. Hình 2.12 mô tả cho trường hợp kỳ vọng và độ lệch chuẩn khác nhau.



Hình 2.10: Đường cong chuẩn với $\mu_1 < \mu_2$ và $\sigma_1 = \sigma_2$



Hình 2.11: Đường cong chuẩn với $\mu_1 = \mu_2$ và $\sigma_1 < \sigma_2$



Hình 2.12: Đường cong chuẩn với $\mu_1 < \mu_2$ và $\sigma_1 < \sigma_2$

Định lý 2.18. Nếu X là biến ngẫu nhiên tuân theo luật phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$, thì biến ngẫu nhiên $Y = aX + b$ tuân theo luật phân phối chuẩn $\mathcal{N}(a\mu + b, a^2\sigma^2)$.

Chú ý 2.5. 1. Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân phối chuẩn $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ thì $X_1 + X_2$ cũng có phân phối chuẩn $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ (xem Chương 3).

2. Nếu n biến ngẫu nhiên độc lập X_i cùng có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$, $i = 1, \dots, n$, thì (xem Chương 3).

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right).$$

2.4.4b Phân phối chuẩn tắc

Định nghĩa 2.20 (Phân phối chuẩn tắc). Phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với $\mu = 0$ và $\sigma = 1$ gọi là phân phối chuẩn tắc $\mathcal{N}(0, 1)$.

Nếu X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ thì

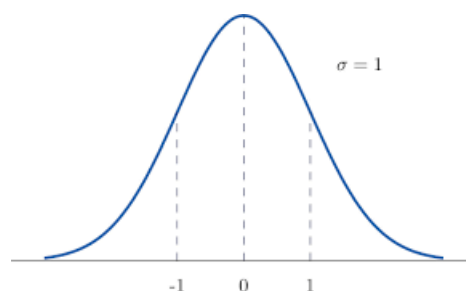
$$U = \frac{X - \mu}{\sigma} \quad (2.40)$$

là biến ngẫu nhiên có phân phối chuẩn tắc $\mathcal{N}(0, 1)$. Do đó các tính toán về X sẽ được quy về U .

Định nghĩa 2.21 (Hàm mật độ xác suất). Hàm mật độ xác suất của biến ngẫu nhiên có phân phối chuẩn tắc là

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R} \quad (2.41)$$

Đây là hàm Gau-xơ với các giá trị được tính sẵn trong Phụ lục 1.



Hình 2.13: Hàm mật độ của biến ngẫu nhiên có phân phối chuẩn tắc $\mathcal{N}(0, 1)$

Định nghĩa 2.22 (Hàm phân phối xác suất). Hàm phân phối xác suất của biến ngẫu nhiên U phân phối chuẩn tắc là

$$\Phi_U(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt, \quad x \in \mathbb{R} \quad (2.42)$$

Giá trị của hàm $\Phi(x)$ được tính sẵn trong Phụ lục 3.

Hàm $\Phi(x)$ có tính chất sau.

Định lý 2.19.

$$\boxed{\Phi(-x) = 1 - \Phi(x)} \quad (2.43)$$

Chứng minh. Từ Định nghĩa 2.22,

$$\Phi(-x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = 1 - \Phi(x).$$

□

2.4.4c Xác suất để biến ngẫu nhiên X có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ nhận giá trị trong khoảng (α, β)

Định lý 2.20. Nếu X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ thì

$$\boxed{P(\alpha < X < \beta) = \phi\left(\frac{\beta - \mu}{\sigma}\right) - \phi\left(\frac{\alpha - \mu}{\sigma}\right), \quad \alpha, \beta \in \mathbb{R}} \quad (2.44)$$

trong đó $\phi(x)$ là hàm số Láp-la-xơ xác định bởi (1.24).

Chứng minh. Sử dụng phép đổi biến $t = \frac{x - \mu}{\sigma}$ ta nhận được

$$\begin{aligned} P(\alpha < X < \beta) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\frac{(x - \mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha - \mu}{\sigma}}^{\frac{\beta - \mu}{\sigma}} e^{-\frac{t^2}{2}} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha - \mu}{\sigma}}^0 e^{-\frac{t^2}{2}} dt + \frac{1}{\sqrt{2\pi}} \int_0^{\frac{\beta - \mu}{\sigma}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_0^{\frac{\beta - \mu}{\sigma}} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int_0^{\frac{\alpha - \mu}{\sigma}} e^{-\frac{t^2}{2}} dt. \end{aligned}$$

Từ đây và (1.24) ta nhận được (2.44). □

Hệ quả 2.3. 1. $P(X < \beta) = P(-\infty < X < \beta) = 0,5 + \phi\left(\frac{\beta - \mu}{\sigma}\right).$

2. $P(X > \alpha) = P(\alpha < X < \infty) = 0,5 - \phi\left(\frac{\alpha - \mu}{\sigma}\right).$

3. $P(|X - \mu| < t\sigma) = 2\phi(t).$

Định lý 2.21. Nếu X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ thì hàm phân phối xác suất của X là

$$F_X(x) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Khi đó, xác suất để biến ngẫu nhiên X nhận giá trị trong khoảng (α, β) là

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - \mu}{\sigma}\right) - \Phi\left(\frac{\alpha - \mu}{\sigma}\right) \quad (2.45)$$

Chú ý 2.6. 1. Các giá trị của hàm Láp-la-xơ (1.24) được tính trong bảng Phụ lục 2 (xem Mục 1.5.5) đối với các giá trị x dương. Hàm $\phi(x)$ là hàm lẻ, tức là $\phi(-x) = -\phi(x)$. Khi $x > 5$ ta có thể lấy $\phi(x) \simeq 0,5$.

2. Hàm phân phối xác suất của biến ngẫu nhiên phân phối chuẩn tắc $\Phi(x)$ xác định bởi (2.42) và hàm Láp-la-xơ $\phi(x)$ xác định bởi (1.24) có mối liên hệ:

$$\Phi(x) = 0,5 + \phi(x), \quad x \geq 0 \quad (2.46)$$

Các giá trị của hàm phân phối chuẩn tắc $\Phi(x)$ được tính sẵn trong bảng Phụ lục 3 đối với các giá trị x dương.

3. Nếu $X \sim \mathcal{N}(\mu, \sigma^2)$ thì các công thức (2.44) và (2.45) là tương đương.

2.4.4d Quy tắc 3σ

Từ Hệ quả 2.3(3) suy ra xác suất để độ lệch tuyệt đối của biến ngẫu nhiên $X \sim \mathcal{N}(\mu, \sigma^2)$ khỏi trị trung bình của nó bé hơn $\varepsilon = t\sigma$ là

$$P(|X - \mu| < \varepsilon) = 2\phi\left(\frac{\varepsilon}{\sigma}\right) \quad \text{hay} \quad P(|X - \mu| < t\sigma = \varepsilon) = 2\phi(t) \quad (2.47)$$

Thay $t = 1, 2, 3$, tra bảng giá trị hàm số Láp-la-xơ (Phụ lục 2) ta nhận được

$$\begin{aligned} P(|X - \mu| < \sigma) &= 2\phi(1) = 0,6827, \\ P(|X - \mu| < 2\sigma) &= 2\phi(2) = 0,9545, \\ P(|X - \mu| < 3\sigma) &= 2\phi(3) = 0,9973. \end{aligned} \quad (2.48)$$

Quy tắc 3σ được phát biểu như sau: *Hầu chắc chắn rằng (với độ tin cậy 0,9973) X có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ lấy giá trị trong khoảng $(\mu - 3\sigma, \mu + 3\sigma)$.*

Trong thực tế, quy tắc 3σ được áp dụng như sau: Nếu quy luật phân phối xác suất của biến ngẫu nhiên được nghiên cứu chưa biết, song nó thỏa mãn điều kiện của Quy tắc 3σ thì có thể xem như nó là biến ngẫu nhiên có phân phối chuẩn.

Chú ý 2.7. 1. Phân phối chuẩn được Gao-xơ tìm ra năm 1809 nên nó còn được gọi là phân phối Gao-xơ.

2. Phân phối chuẩn thường được sử dụng trong các bài toán đo đạc các đại lượng vật lý, thiên văn ...

3. Trong thực tế, nhiều biến ngẫu nhiên tuân theo phân phối chuẩn hoặc tiệm cận chuẩn. Chẳng hạn, trọng lượng, chiều cao của một nhóm người nào đó; điểm thi của thí sinh; năng suất cây trồng; mức lãi suất của một công ty; nhu cầu tiêu thụ của một mặt hàng nào đó; nhiễu trắng trên các kênh thông tin ... là các biến ngẫu nhiên có phân phối chuẩn.

Ví dụ 2.35. Lãi suất (%) đầu tư vào một dự án trong năm 2018 được coi như một biến ngẫu nhiên tuân theo quy luật chuẩn. Theo đánh giá của ủy ban đầu tư thì với xác suất 0,1587 cho lãi suất lớn hơn 20% và với xác suất 0,0228 cho lãi suất lớn hơn 25%. Vậy khả năng đầu tư mà không bị lỗ là bao nhiêu?

Lời giải Ví dụ 2.35 Gọi X là lãi suất (%) của dự án trong năm 2018. Khi đó X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$. Theo đầu bài ta có

$$P(X > 20) = P(20 < X < +\infty) = 0,5 - \Phi\left(\frac{20 - \mu}{\sigma}\right) = 0,1587$$

và

$$P(X > 25) = P(25 < X < +\infty) = 0,5 - \Phi\left(\frac{25 - \mu}{\sigma}\right) = 0,0228.$$

Từ bảng giá trị hàm số Láp-la-xơ (Phụ lục 2) suy ra $\frac{20 - \mu}{\sigma} = 1$ và $\frac{25 - \mu}{\sigma} = 2$. Hay $\mu = 15$, $\sigma = 5$. Vậy khả năng đầu tư không bị lỗ là

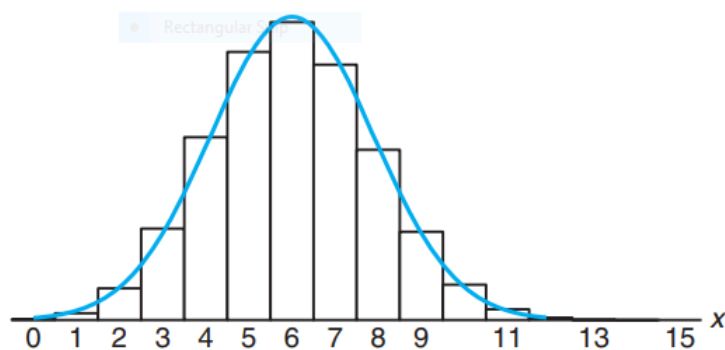
$$P(X \geq 0) = 0,5 + \Phi(3) = 0,5 + 0,49865 = 0,99865.$$

2.4.4e Xấp xỉ phân phối nhị thức bởi phân phối chuẩn

Trong Mục 1.5.5 ta đã đề cập đến việc xấp xỉ công thức Béc-nu-y (1.20) bởi công thức (3.40) khi số phép thử n khá lớn. Ở đây ta xét chi tiết về mối liên hệ giữa phân phối nhị thức và phân phối chuẩn. Phân phối chuẩn có thể dùng xấp xỉ khá tốt cho một số phân phối rời rạc. Ta có định lý sau đây mang tên là Định lý Moa-vơ-Lap-la-xơ.

Định lý 2.22. Giả sử X là biến ngẫu nhiên có phân phối nhị thức $\mathcal{B}(n; p)$. Nếu $np > 5$ và $n(1 - p) > 5$ thì X có phân phối xấp xỉ phân phối chuẩn với tham số $\mu = np$, $\sigma^2 = np(1 - p)$.

Phân phối chuẩn với kỳ vọng $\mu = np$ và phương sai $\sigma^2 = np(1 - p)$ không chỉ xấp xỉ khá tốt cho phân phối nhị thức khi n khá lớn và xác suất p không quá gần 0 hoặc 1 mà còn cung cấp một xấp xỉ khá tốt cho phân phối nhị thức ngay cả khi n nhỏ và p gần 1/2. Để minh họa việc xấp xỉ phân phối chuẩn cho phân phối nhị thức, ta vẽ biểu đồ của $\mathcal{B}(15; 0,4)$ và vẽ đường cong chuẩn có cùng kỳ vọng $\mu = np = 15 \times 0,4 = 6$ và phương sai $\sigma^2 = np(1 - p) = 15 \times 0,4 \times 0,6 = 3,6$ với biến ngẫu nhiên tuân theo luật phân phối nhị thức X (xem Hình 2.14).



Hình 2.14: Xấp xỉ phân phối chuẩn cho phân phối nhị thức $\mathcal{B}(15; 0,4)$

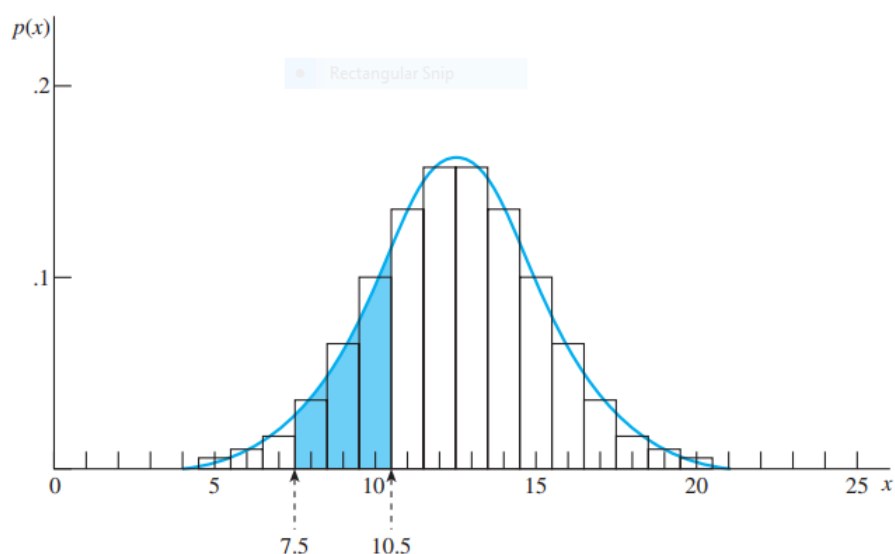
Trong hình minh họa về xấp xỉ phân phối nhị thức bởi phân phối chuẩn, vì ta xấp xỉ một phân phối rời rạc bằng một phân phối liên tục, nên cần một sự hiệu chỉnh để giảm sai số.

Định lý 2.23. Cho X là biến ngẫu nhiên tuân theo luật phân phối nhị thức $\mathcal{B}(n; p)$. Phân phối xác suất của X được xấp xỉ bởi phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ với $\mu = np$ và $\sigma^2 = np(1 - p)$ và

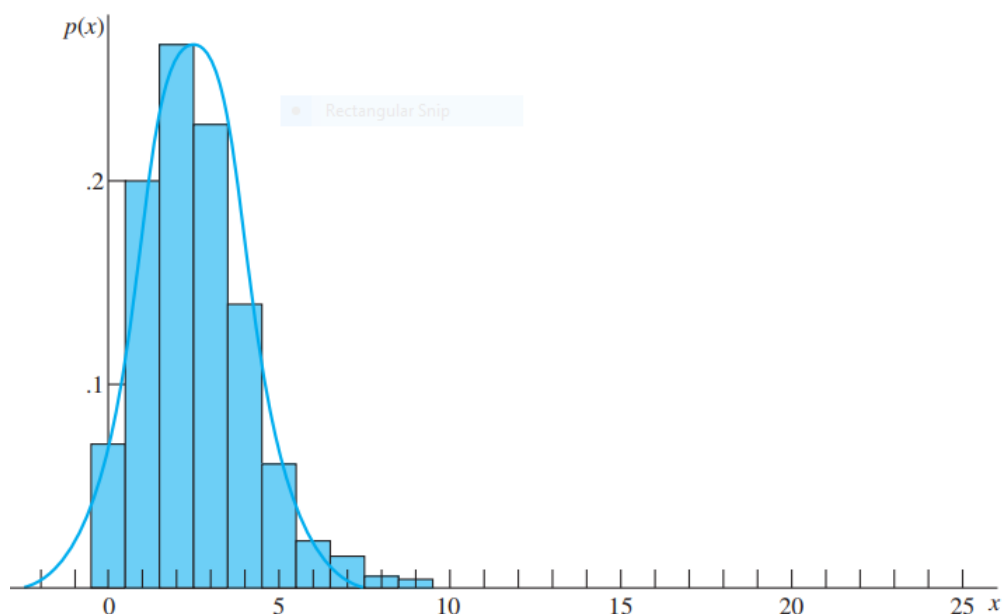
$$P(X = k) = C_n^k p^k (1 - p)^{n-k} \simeq \phi\left(\frac{k + 0,5 - \mu}{\sigma}\right) - \phi\left(\frac{k - 0,5 - \mu}{\sigma}\right) \quad (2.49)$$

$$P(k_1 \leq X \leq k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k (1 - p)^{n-k} \simeq \phi\left(\frac{k_2 + 0,5 - \mu}{\sigma}\right) - \phi\left(\frac{k_1 - 0,5 - \mu}{\sigma}\right) \quad (2.50)$$

Nhận xét 2.15. Hình 2.15 và 2.16 biểu thị biểu đồ xác suất nhị thức với $n = 25$ và $p = 0,5$, $p = 0,1$ tương ứng. Phân phối trong Hình 2.15 là hoàn toàn đối xứng.



Hình 2.15: Phân phối nhị thức với $n = 25$ và $p = 0,5$ xấp xỉ bởi phân phối chuẩn với $\mu = 12,5$ và $\sigma = 2,5$



Hình 2.16: Phân phối nhị thức và xấp xỉ phân phối chuẩn với $n = 25$ và $p = 0,1$

Việc thêm $+0,5$ và $-0,5$ chính là yếu tố hiệu chỉnh và gọi là hiệu chỉnh liên tục.

Ví dụ 2.36. Sử dụng phân phối chuẩn xấp xỉ xác suất $X = 8, 9$, hoặc 10 cho biến ngẫu nhiên X tuân theo luật phân phối nhị thức với $n = 25$ và $p = 0,5$. So sánh với công thức tính chính xác.

Lời giải Ví dụ 2.36 Vì X là biến ngẫu nhiên tuân theo luật phân phối nhị thức với $n = 25$ và $p = 0,5$,

$$P(X = 8) + P(X = 9) + P(X = 10) = \left(C_{25}^8 + C_{25}^9 + C_{25}^{10} \right) \times (0,5)^{25} \simeq 0,190535.$$

Sử dụng công thức xấp xỉ (3.42) với $\mu = np = 12,5$, $\sigma = \sqrt{npq} = 2,5$ ta nhận được

$$P(8 \leq X \leq 10) \simeq \phi(-0,8) - \phi(-2) = 0,18911.$$

Giá trị xấp xỉ $0,18911$ với giá trị thực $0,190535$ là khá gần nhau.

Ví dụ 2.37. Kiểm tra chất lượng 1000 sản phẩm với tỷ lệ chính phẩm $0,95$. Tìm xác suất để số chính phẩm trong lô kiểm tra từ 940 đến 960.

Lời giải Ví dụ 2.37 Gọi X là biến ngẫu nhiên chỉ số chính phẩm trong lô sản phẩm kiểm tra, ta có $X \sim \mathcal{B}(1000; 0,95)$. Với $n = 1000$, $p = 0,95$, ta có $np = 950$ và $np(1 - p) = 47,5$ đủ lớn nên ta xấp xỉ bởi $X \sim \mathcal{N}(950; 47,5)$:

$$\begin{aligned} P(940 \leq X \leq 960) &= \phi\left(\frac{960 + 0,5 - 950}{\sqrt{47,5}}\right) - \phi\left(\frac{940 - 0,5 - 950}{\sqrt{47,5}}\right) \\ &= \phi(1,52) - \phi(-1,52) = 2\phi(1,52) = 0,8716. \end{aligned}$$

2.4.7 Phân phối khi bình phương

Định nghĩa 2.23 (Phân phối khi bình phương). Biến ngẫu nhiên liên tục X được gọi là tuân theo luật phân phối khi bình phương với n bậc tự do, ký hiệu là $X \sim \chi_n^2$, nếu hàm mật độ xác suất của X có dạng

$$f_X(x) = \frac{1}{2} e^{-\frac{x}{2}} \frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)}, \quad x > 0, n > 0 \quad (2.51)$$

ở đây

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt, \quad x > 0,$$

là hàm Gamma (đã đề cập trong Giải tích 2).

Định nghĩa sau cho cách nhận biết một biến ngẫu nhiên có phân phối khi bình phương xuất phát từ n biến ngẫu nhiên độc lập có cùng phân phối chuẩn tắc.

Định nghĩa 2.24. Nếu X_1, X_2, \dots, X_n là các biến ngẫu nhiên độc lập có cùng phân phối chuẩn tắc $\mathcal{N}(0, 1)$ thì

$$U_n = X_1^2 + X_2^2 + \dots + X_n^2 \sim \chi_n^2 \quad (2.52)$$

(U_n có phân phối khi bình phương với n bậc tự do).

Kỳ vọng và phương sai của biến ngẫu nhiên U_n :

$$E(U_n) = n, \quad V(U_n) = 2n \quad (2.53)$$

Tính chất 2.1. 1. Nếu X_1 và X_2 là hai biến ngẫu nhiên độc lập có phân phối khi bình phương với n_1, n_2 bậc tự do thì biến ngẫu nhiên $X_1 + X_2$ có phân phối khi bình phương với $n_1 + n_2$ bậc tự do (xem Chương 3).

2. Biến ngẫu nhiên $\frac{U_n - n}{\sqrt{2n}}$ có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0, 1)$ khi n đủ lớn.

3. Một hệ quả quan trọng được dùng nhiều trong thống kê (xem Chương 3): Nếu X_1, X_2, \dots, X_n là các biến ngẫu nhiên độc lập có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ và

$$\bar{X} = \frac{1}{n} (X_1 + X_2 + \dots + X_n)$$

thì

$$\frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi_{(n-1)}^2. \quad (2.54)$$

Việc tính toán với phân phối χ_n^2 đưa về việc sử dụng bảng Phụ lục 4.

2.4.8 Phân phối Student

Định nghĩa 2.25 (Phân phối Student). Biến ngẫu nhiên liên tục X được gọi là tuân theo luật phân phối Student với n bậc tự do, ký hiệu là $X \sim t(n)$, nếu hàm mật độ xác suất của X có dạng

$$f_X(x) = \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}, \quad -\infty < x < +\infty \quad (2.55)$$

ở đây $\Gamma(x)$ là hàm Gamma.

Để nhận biết một biến ngẫu nhiên có phân phối Student ta sử dụng định nghĩa sau.

Định nghĩa 2.26. Nếu X và Y là hai biến ngẫu nhiên độc lập tuân theo luật $\mathcal{N}(0, 1)$ và χ_n^2 tương ứng thì

$$T_n = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t(n) \quad (2.56)$$

(T_n có phân phối Student với n bậc tự do).

Kỳ vọng và phương sai của biến ngẫu nhiên T_n có phân phối Student:

$$E(T_n) = 0, \quad n > 1, \quad V(T_n) = \frac{n}{n-2}, \quad n > 2 \quad (2.57)$$

Tính chất 2.2. Biến ngẫu nhiên T_n có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0, 1)$ khi n đủ lớn.

- Nhận xét 2.16.**
1. Phân phối Student có cùng dạng và tính đối xứng như phân phối chuẩn nhưng nó phản ánh tính biến đổi của phân phối sâu sắc hơn. Phân phối chuẩn không thể dùng để xấp xỉ phân phối khi mẫu có kích thước nhỏ. Trong trường hợp này ta dùng phân phối Student.
 2. Khi bậc tự do n tăng lên ($n \geq 30$) thì phân phối Student tiến nhanh về phân phối chuẩn. Do đó khi $n \geq 30$ ta có thể dùng phân phối chuẩn thay thế cho phân phối Student.
 3. Một hệ quả quan trọng được dùng nhiều trong thống kê (xem Chương 3): Nếu $X_0, X_1, X_2, \dots, X_n$ là các biến ngẫu nhiên độc lập có phân phối chuẩn tắc $\mathcal{N}(0, 1)$ thì

$$\frac{X_0}{\sqrt{\frac{1}{n} \sum_{i=1}^n X_i^2}} \sim t(n). \quad (2.58)$$

2.4.9 Phân phối Fisher

Định nghĩa 2.27. Cho X_1, X_2, \dots, X_n và Y_1, Y_2, \dots, Y_m là $n + m$ biến ngẫu nhiên độc lập, trong đó $X_i \sim \mathcal{N}(0; 1)$ và $Y_j \sim \mathcal{N}(0; 1)$, $i = 1, \dots, n, j = 1, \dots, m$. Khi đó biến ngẫu nhiên

$$F_{n,m} = \frac{m \sum_{i=1}^n X_i^2}{n \sum_{j=1}^m Y_j^2} \quad (2.59)$$

có phân phối Fisher với (n, m) bậc tự do (xem Chương 3).

2.5 Tổng hợp một số đề thi

Ví dụ 2.38 (Đề thi MI2020 kỳ 20183). Một kiện hàng có 10 sản phẩm, trong đó có 7 sản phẩm loại I và 3 sản phẩm loại II. Tiền lãi khi bán được mỗi sản phẩm loại I là 50 nghìn đồng, mỗi sản phẩm loại II là 20 nghìn đồng.

- Ngày thứ nhất lấy ngẫu nhiên từ kiện hàng ra 3 sản phẩm và đã bán hết cả 3 sản phẩm đó. Tìm kỳ vọng của số tiền lãi thu được.
- Ngày thứ hai lấy ngẫu nhiên từ kiện hàng ra 2 sản phẩm. Tính xác suất để thu được 100 nghìn đồng tiền lãi khi bán 2 sản phẩm này.

Lời giải Ví dụ 3.24

- Gọi X là "số tiền lãi thu được", X nhận các giá trị 60, 90, 120, 150. Khi đó,

$$E(X) = 60 \times \frac{1}{120} + 90 \times \frac{21}{120} + 120 \times \frac{63}{120} + 150 \times \frac{35}{120} = 123.$$

- Gọi A là sự kiện "ngày thứ hai thu được 100 nghìn đồng tiền lãi khi bán 2 sản phẩm"; A_i : "ngày thứ nhất lấy được i sản phẩm loại I", $i = 0, 1, 2, 3$; A_0, A_1, A_2, A_3 lập thành hệ đầy đủ và $P(A) = P(A_0)P(A|A_0) + P(A_1)P(A|A_1) + P(A_2)P(A|A_2) + P(A_3)P(A|A_3)$. Khi đó,

$$P(A) = \frac{1}{120} \times \frac{21}{21} + \frac{21}{120} \times \frac{15}{21} + \frac{63}{120} \times \frac{10}{21} + \frac{35}{120} \times \frac{6}{21} = \frac{7}{15} \simeq 0,4667.$$

Ví dụ 2.39 (Đề thi MI2020 giữa kỳ 20191). Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} kx^2(1-x), & \text{nếu } x \in [0, 1], \\ 0, & \text{nếu } x \notin [0, 1]. \end{cases}$$

- Tìm hằng số k .
- Tính xác suất để sau 3 lần lặp lại phép thử một cách độc lập có đúng 1 lần X nhận giá trị trong khoảng $\left(0; \frac{1}{2}\right)$.

Lời giải Ví dụ 3.25

(a) Sử dụng Tính chất 2.3(a),(c) tính được $k = 12$.

$$(b) P\left(0 < X < \frac{1}{2}\right) = \int_0^{\frac{1}{2}} 12(x^2 - x^3)dx = \frac{5}{16} = 0,3125.$$

$$\text{Vậy, } P_3(1) = C_3^1 \times p^1 \times (1-p)^2 = C_3^1 \times (0,3125)^1 \times (0,6875)^2 = \frac{1815}{4096} \simeq 0,44312.$$

Ví dụ 2.40 (Đề thi MI2020 kỳ 20191). Cho hàm mật độ xác suất $f_X(x) = \begin{cases} 3e^{-3x}, & \text{nếu } x \geq 0 \\ 0, & \text{nếu } x < 0 \end{cases}$ của biến ngẫu nhiên liên tục X và định nghĩa $Y = [X]$ là số nguyên lớn nhất không vượt quá X (nghĩa là $[x] = 0$ nếu $0 \leq x < 1$, $[x] = 1$ nếu $1 \leq x < 2 \dots$).

(a) Tính $P(Y = 0)$.

(b) Tính $E(Y)$.

Lời giải Ví dụ 2.40 (a) $P(Y = 0) = P(0 \leq X < 1) = \int_0^1 3e^{-3x}dx = 1 - e^{-3}$.

$$(b) \text{ Với } k \geq 0, P(Y = k) = e^{-3k}(1 - e^{-3}) \text{ và } E(X) = \sum_{k=0}^{\infty} kP(Y = k) = \frac{1}{e^3 - 1}.$$

Ví dụ 2.41 (Đề thi MI2020 kỳ 20191). Số khách hàng đến một cửa hàng bán lẻ là một biến ngẫu nhiên có phân phối Poa-xông với trung bình 6 khách hàng đến trong vòng một giờ. Nếu có đúng 5 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:00 thì xác suất để có ít nhất 8 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:30 là bao nhiêu?

Lời giải Ví dụ 2.41 Gọi X là "số khách hàng đến cửa hàng bán lẻ trong vòng 30 phút". Khi đó X là biến ngẫu nhiên có phân phối Poa-xông, $X \sim \mathcal{P}(\lambda)$, với $\lambda = 3$. Xác suất cần tìm $P(X \geq 3)$.

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)] \\ &= 1 - e^{-3} \left[\frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!} \right] \\ &= 1 - 0,42319 = 0,57681. \end{aligned}$$

Ví dụ 2.42 (Đề thi MI2021 kỳ 20193). Số máy D bán được trong ngày của một siêu thị là biến ngẫu nhiên X tuân theo phân phối Poisson tham số λ với $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, $k \in \mathbb{N}$. Biết rằng xác suất bán được máy D trong một ngày là 39,35%. (a) Tính số máy D bán được trung bình trong một ngày của siêu thị đó. (b) Nếu khảo sát 30 ngày thì số ngày bán được máy D có khả năng xảy ra cao nhất là bao nhiêu?

Lời giải Ví dụ 2.42 (a) Gọi X là "số máy D bán được trong một ngày", $X \sim P(\lambda)$.

$$P(X \geq 1) = 0,3935 \Rightarrow 0,6065 = P(X = 0) = e^{-\lambda}.$$

Trung bình số máy D bán được trong ngày là $\lambda = -\ln(0,6065) = 0,5$.

(b) Gọi Y là "số ngày bán được máy D (trong 30 ngày)"; $Y \sim B(n; p)$ với $n = 30$; $p = 0,3935$.

Vì $(n+1) \times p - 1 \leq \text{mod}(Y) \leq (n+1) \times p$ nên

$11,1985 \leq \text{mod}(Y) \leq 12,1985$ hay $\text{mod}(Y) = 12$.

Bài tập Chương 2

Bài tập 2.1. Một chùm chìa khóa gồm 4 chiếc giống nhau, trong đó chỉ có một chiếc mở được cửa. Người ta thử ngẫu nhiên từng chiếc cho đến khi mở được cửa. Gọi X là số lần thử.

(a) Tìm phân phối xác suất của X .

(b) Tìm kỳ vọng và phương sai của X .

(b) Viết hàm phân phối xác suất của X .

Bài tập 2.2. Một xạ thủ có 5 viên đạn. Anh ta phải bắn vào bia với quy định khi nào có 2 viên trúng bia hoặc hết đạn thì dừng. Biết xác suất bắn trúng bia ở mỗi lần bắn là 0,4 và gọi X là số đạn cần bắn.

(a) Tìm phân phối xác suất của X .

(b) Tìm kỳ vọng, phương sai và viết hàm phân phối xác suất của X .

Bài tập 2.3. Tỷ lệ cử tri ủng hộ ứng cử viên A trong một cuộc bầu cử tổng thống là 40%. Người ta hỏi ý kiến 20 cử tri được chọn một cách ngẫu nhiên. Gọi X là số người bỏ phiếu cho ông A trong 20 người đó.

(a) Tìm giá trị trung bình, độ lệch chuẩn của X và $\text{mod}X$.

(b) Tìm $P(X = 10)$.

Bài tập 2.4. Biến ngẫu nhiên rời rạc X chỉ có 2 giá trị x_1 và x_2 ($x_1 < x_2$). Xác suất để X nhận giá trị x_1 là 0,2. Tìm luật phân phối xác suất của X , biết kỳ vọng $E(X) = 2,6$ và độ lệch tiêu chuẩn $\sigma(X) = 0,8$.

Bài tập 2.5. Mỗi khách uống cà phê tại quán cà phê mỗi ngày đều được phát ngẫu nhiên một vé bốc thăm, xác suất khách hàng trúng thăm là 0,1. Nếu khách hàng trúng thăm liên tục trong 5 ngày (từ thứ hai đến thứ sáu) sẽ nhận được 100\$, nếu không sẽ không được gì. An uống cà phê liên tục tại quán này 4 tuần liên tiếp. Gọi X là số tiền An được thưởng khi bốc thăm trong 4 tuần đó. Xác định kỳ vọng và phương sai của X .

Bài tập 2.6. Tung đồng xu 10 lần. Biến ngẫu nhiên X được định nghĩa như sau: ($X = 1$) nếu sự kiện đúng 3 lần ra mặt sấp xảy ra và ($X = 0$) trong trường hợp còn lại. Tính kỳ vọng $E(X)$ và phương sai $V(X)$.

Bài tập 2.7. Có 5 sản phẩm trong đó có 4 chính phẩm và 1 phế phẩm. Người ta lấy ra lần lượt hai sản phẩm (lấy không hoàn lại).

- (a) Gọi X là "số chính phẩm gặp phải". Lập bảng phân phối xác suất của X . Tính $E(X)$ và $V(X)$.
- (b) Gọi Y là "số phế phẩm gặp phải". Lập hệ thức cho mối quan hệ giữa X và Y .

Bài tập 2.8. Người ta đặt ngẫu nhiên 10 thẻ (trong đó có 5 thẻ màu đỏ và 5 thẻ màu xanh) vào 10 phong bì (5 phong bì có màu đỏ và 5 phong bì có màu xanh), mỗi phong bì một thẻ. Gọi X là số phong bì có chứa một thẻ cùng màu. Tính giá trị:

- (a) $P(X = 1)$.
- (b) $E(X)$.

Bài tập 2.9. Có 2 kiện hàng. Kiện I có 3 sản phẩm tốt và 2 sản phẩm xấu. Kiện II có 2 sản phẩm tốt và 3 sản phẩm xấu. Lấy ngẫu nhiên từ kiện I ra 2 sản phẩm và từ kiện II ra 1 sản phẩm. Lập bảng phân phối xác suất cho biến ngẫu nhiên chỉ số sản phẩm tốt trong 3 sản phẩm lấy ra.

Bài tập 2.10. Có hai kiện hàng. Kiện thứ nhất có 8 sản phẩm tốt và 2 sản phẩm xấu. Kiện thứ hai có 5 sản phẩm tốt và 3 sản phẩm xấu. Lấy ngẫu nhiên 2 sản phẩm từ kiện I bỏ sang kiện II. Sau đó từ kiện II lấy ngẫu nhiên ra 2 sản phẩm. Lập bảng phân phối xác suất của biến ngẫu nhiên chỉ số sản phẩm tốt có trong 2 sản phẩm lấy ra từ kiện II.

Bài tập 2.11. Gieo hai con xúc sắc đồng chất 5 lần, gọi X là số lần xuất hiện hai mặt 6.

- (a) Tính xác suất của sự kiện số lần xuất hiện hai mặt 6 ít nhất là 2.
- (b) Tính $E(X)$, $V(X)$.
- (c) Viết hàm phân phối $F_X(x)$.

Bài tập 2.12. Một thanh niên nam vào cửa hàng thấy 5 máy thu thanh giống nhau. Anh ta đề nghị cửa hàng cho anh ta thử lần lượt các máy đến khi chọn được máy tốt thì mua, nếu cả 5 lần đều xấu thì thôi. Biết rằng xác suất để một máy xấu là 0,6 và các máy xấu tốt độc lập với nhau. Gọi X là số lần thử. Lập bảng phân phối xác suất của X .

Bài tập 2.13. Có hai hộp bi. Hộp I có 2 bi trắng, 3 bi đỏ. Hộp II có 2 bi trắng, 2 bi đỏ. Lấy ngẫu nhiên 2 bi từ hộp I bỏ sang hộp II, sau đó lại lấy ngẫu nhiên 3 bi từ hộp II bỏ vào hộp I. Lập bảng phân phối xác suất của biến ngẫu nhiên chỉ số bi trắng có mặt ở hộp I, hộp II sau khi đã chuyển xong.

Bài tập 2.14. Một người đi làm từ nhà đến cơ quan phải qua 3 ngã tư. Xác suất để người đó gặp đèn đỏ ở các ngã tư tương ứng là 0,2; 0,4 và 0,5. Gọi X là số đèn đỏ mà người đó gặp phải trong một lần đi làm (giả sử 3 đèn giao thông ở ngã tư hoạt động độc lập với nhau).

- (a) Lập bảng phân phối xác suất của X . Tính kỳ vọng, phương sai của X . Tìm hàm phân phối xác suất của X .
- (b) Hỏi thời gian trung bình phải ngừng trên đường là bao nhiêu biết rằng mỗi khi gặp đèn đỏ người ấy phải đợi khoảng 3 phút.

Bài tập 2.15. Một người chơi trò chơi tung con xúc sắc cân đối đồng chất ba lần. Nếu cả ba lần đều xuất hiện mặt 6 thì thu về 36\$, nếu hai lần xuất hiện mặt 6 thì thu về 2,8\$, nếu một lần xuất hiện mặt 6 thì thu về 0,4\$. Biết rằng khi chơi người đó phải nộp x \$.

- (a) Tìm x sao cho trò chơi là vô thưởng vô phạt.
- (b) x bằng bao nhiêu thì trung bình mỗi lần chơi, người chơi mất 1\$?

Bài tập 2.16. Một kiện hàng có 12 sản phẩm, trong đó có 7 sản phẩm loại I và 5 sản phẩm loại II. Khi bán được một sản phẩm loại I thì được lãi 50 ngàn đồng; còn nếu bán được một sản phẩm loại II thì được lãi 20 ngàn đồng. Lấy ngẫu nhiên từ kiện hàng ra 3 sản phẩm.

- (a) Tìm quy luật phân phối xác suất của số tiền lãi thu được do bán 3 sản phẩm đó; tính kỳ vọng, phương sai của số tiền lãi thu được do bán 3 sản phẩm đó.
- (b) Viết hàm phân phối, vẽ đồ thị hàm phân phối của số tiền lãi thu được khi bán 3 sản phẩm đó.

Bài tập 2.17. Một hộp đựng 15 quả bóng bàn trong đó có 10 quả còn mới. Lần đầu ta lấy ra 3 quả để thi đấu, sau đó lại trả 3 quả đó vào hộp. Lần thứ hai lại lấy ra 3 quả. Gọi X là biến ngẫu nhiên chỉ số quả bóng mới trong 3 quả lấy ra. Lập bảng phân phối xác suất, tính kỳ vọng, phương sai của X .

Bài tập 2.18. Một cơ sở thí nghiệm có 3 phòng thí nghiệm như nhau. Xác suất thực hiện thành công một thí nghiệm của các phòng lần lượt là 0,6; 0,7 và 0,8. Một sinh viên chọn một phòng thí nghiệm bất kỳ và tiến hành 3 thí nghiệm độc lập. Gọi X là số thí nghiệm thành công.

- (a) Lập bảng phân phối xác suất của X , tính kỳ vọng $E(X)$ và phương sai $V(X)$.
- (b) Theo anh (chị) thì khả năng chắc chắn sẽ thành công mấy thí nghiệm?

Bài tập 2.19. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} k \sin 3x, & x \in \left(0, \frac{\pi}{3}\right), \\ 0, & x \notin \left(0, \frac{\pi}{3}\right). \end{cases}$$

- (a) Xác định k và hàm phân phối $F_X(x)$.
- (b) Tính $P(\pi/6 \leq X < \pi/3)$.

Bài tập 2.20. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \frac{c}{e^x + e^{-x}}.$$

Xác định hằng số c và sau đó tính kỳ vọng của X .

Bài tập 2.21. Biến ngẫu nhiên liên tục X có hàm mật độ là $f_X(x) = ae^{-|x|}$, $(-\infty < x < +\infty)$.

- (a) Xác định a .
- (b) Tìm hàm phân phối của biến ngẫu nhiên X , biến ngẫu nhiên $Y = X^2$.
- (c) Tìm $E(X)$, $V(X)$.
- (d) Tính xác suất để sau ba lần lặp lại phép thử một cách độc lập có 2 lần X nhận giá trị trong khoảng $(0; \ln 3)$.

Bài tập 2.22. Nhu cầu hàng năm về loại hàng A là biến ngẫu nhiên liên tục X có hàm mật độ xác suất như sau (đơn vị: ngàn sản phẩm):

$$f_X(x) = \begin{cases} k(30 - x), & x \in (0, 30), \\ 0, & x \notin (0, 30). \end{cases}$$

- (a) Tìm k .
- (b) Tìm hàm phân phối $F_X(x)$.
- (c) Tìm nhu cầu trung bình hàng năm về loại hàng đó.

Bài tập 2.23. Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \leq 0, \\ \frac{1}{2} - k \cos x, & 0 < x \leq \pi, \\ 1, & x > \pi. \end{cases}$$

- (a) Tìm k .
- (b) Tìm $P\left(0 < X < \frac{\pi}{2}\right)$.
- (c) Tìm $E(X)$.

Bài tập 2.24. Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất

$$F_X(x) = \begin{cases} 0, & x \leq -a, \\ A + B \arcsin \frac{x}{a}, & x \in (-a, a), \\ 1, & x \geq a. \end{cases}$$

- (a) Tìm A và B .
- (b) Tìm hàm mật độ xác suất $f_X(x)$.

Bài tập 2.25. Hàm phân phối xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x) = a + b \arctan x, \quad (-\infty < x < +\infty).$$

- (a) Tìm hệ số a và b .
- (b) Tìm hàm mật độ xác suất $f_X(x)$.
- (c) Tìm xác suất để khi tiến hành 3 phép thử độc lập có 2 lần X nhận giá trị trong khoảng $(-1; 1)$.

Bài tập 2.26. Biến ngẫu nhiên X liên tục trên toàn trục số và có hàm phân phối xác suất $F_X(x) = 1/2 + 1/\pi \arctan x/2$. Tìm giá trị có thể có của x_1 thỏa mãn điều kiện $P(X > x_1) = 1/4$.

Bài tập 2.27. Thu nhập của dân cư tại một vùng là biến ngẫu nhiên liên tục có hàm phân phối xác suất như sau:

$$F_X(x) = \begin{cases} 1 - \left(\frac{x_0}{x}\right)^\alpha, & x \geq x_0, \alpha > 0, \\ 0, & x < x_0. \end{cases}$$

Hãy xác định mức thu nhập sao cho lấy ngẫu nhiên một người ở vùng đó thì thu nhập của người này vượt quá mức trên với xác suất 0,5.

Bài tập 2.28. Thời gian phục vụ mỗi khách hàng tại một cửa hàng ăn nhanh là biến ngẫu nhiên X tuân theo quy luật lũy thừa với hàm mật độ xác suất

$$f_X(x) = \begin{cases} 5e^{-5x}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$

với x được tính bằng phút/khách hàng.

- (a) Tìm xác suất để thời gian phục vụ một khách hàng nào đó sẽ nằm trong khoảng $(0, 4; 1)$ (phút).
- (b) Tính thời gian trung bình để phục vụ một khách hàng.

Bài tập 2.29. Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

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

- (a) Tính $P(X \geq 5)$.

(b) Xác định hàm phân phối xác suất của biến ngẫu nhiên $Y = -2X + 5$.

Bài tập 2.30. Cho hàm mật độ xác suất

$$f_X(x) = \begin{cases} 3e^{-3x}, & \text{nếu } x \geq 0, \\ 0, & \text{nếu } x < 0 \end{cases}$$

của biến ngẫu nhiên liên tục X và định nghĩa $Y = [X]$ là số nguyên lớn nhất không vượt quá X (nghĩa là $[x] = 0$ nếu $0 \leq x < 1$, $[x] = 1$ nếu $1 \leq x < 2 \dots$).

(a) Tính $P(Y = 0)$.

(b) Tính $E(Y)$.

Bài tập 2.31. Bắn 5 viên đạn vào một mục tiêu. Xác suất trúng đích của mỗi lần bắn như nhau và bằng 0,2. Muốn phá hủy mục tiêu phải có ít nhất 3 viên trúng mục tiêu. Tìm xác suất mục tiêu bị phá hủy.

Bài tập 2.32. Xác suất để một sinh viên chậm giờ thi là 0,02. Tìm số sinh viên chậm giờ thi có khả năng xảy ra nhiều nhất trong 855 sinh viên dự thi.

Bài tập 2.33. Có 10 máy sản xuất sản phẩm (độc lập nhau), mỗi máy sản xuất ra 2% phế phẩm.

(a) Từ mỗi máy sản xuất lấy ngẫu nhiên ra một sản phẩm. Hỏi xác suất lấy được nhiều nhất 2 phế phẩm trong 10 sản phẩm này là bao nhiêu?

(b) Trung bình có bao nhiêu sản phẩm được sản xuất bởi máy đầu tiên trước khi nó tạo ra phế phẩm đầu tiên (giả sử các sản phẩm sản xuất ra là độc lập)?

Bài tập 2.34. Một ga ra cho thuê ô tô thấy rằng số người đến thuê ô tô vào thứ bảy cuối tuần là một biến ngẫu nhiên có phân bố Poa-xông với tham số $\lambda = 2$. Giả sử gara có 4 chiếc ô tô.

(a) Tìm xác suất để tất cả 4 ô tô đều được thuê vào thứ 7.

(b) Tìm xác suất gara không đáp ứng được yêu cầu (thiếu xe cho thuê) vào thứ 7.

(c) Trung bình có bao nhiêu ô tô được thuê vào ngày thứ 7?

Bài tập 2.35. Số khách hàng đến một cửa hàng bán lẻ là một biến ngẫu nhiên có phân phối Poa-xông với trung bình 6 khách hàng đến trong vòng một giờ.

(a) Nếu có đúng 5 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:00 thì xác suất để có ít nhất 8 khách hàng đến trong khoảng thời gian từ 10:00 đến 11:30 là bao nhiêu?

(b) Nếu có ít hơn 6 khách hàng đến trong khoảng thời gian từ 10:00 đến 12:00 thì cửa hàng được xem như là không có lợi nhuận. Tìm xác suất để cửa hàng có đúng 1 ngày có lãi trong một tuần (giả sử cửa hàng mở cửa 6 ngày trong tuần).

Bài tập 2.36. Gọi biến ngẫu nhiên Y là tỷ lệ người trong 1000 người Mỹ xác nhận rằng có uống nhiều hơn 5 cốc bia mỗi ngày. Giả sử rằng tỷ lệ đúng là 10% trên toàn bộ dân số Mỹ. Tính $E(Y)$, $V(Y)$.

Bài tập 2.37. Giả sử X là biến ngẫu nhiên tuân theo luật phân phối chuẩn với trung bình là 3 và phương sai là 0,16.

(a) Hãy tính $P(X > 3)$, $P(X > 3,784)$.

(b) Tìm c sao cho $P(3 - c < X < 3 + c) = 0,9$.

Bài tập 2.38. Cho biên độ dao động của một vật là biến ngẫu nhiên liên tục có hàm phân phối xác suất là

$$F_X(x) = \begin{cases} 1 - e^{-\frac{x^2}{2\sigma^2}}, & \text{nếu } x \geq 0, \\ 0, & \text{nếu } x < 0, \end{cases}$$

trong đó σ là tham số đã biết. Tính xác suất để biên độ dao động đó lớn hơn trị trung bình của nó.

Bài tập 2.39. Lãi suất (%) đầu tư vào một dự án trong năm 2019 được coi như một biến ngẫu nhiên tuân theo quy luật chuẩn. Theo đánh giá của ủy ban đầu tư thì với xác suất 0,1587 cho lãi suất lớn hơn 20% và với xác suất 0,0228 cho lãi suất lớn hơn 25%. Vậy khả năng đầu tư mà không bị lỗ là bao nhiêu?

Bài tập 2.40. Tung một đồng xu vô hạn lần, xác suất thu được mặt ngửa mỗi lần là p .

(a) Gọi X là số lần tung đến khi xuất hiện mặt ngửa lần đầu tiên (tại lần tung thứ X). Tính $E(X)$.

(b) Tính xác suất xuất hiện đúng 6 lần ngửa trong 10 lần tung.

(c) Tính xác suất để lần xuất hiện mặt ngửa thứ 6 rơi vào lần tung thứ 10.

Bài tập 2.41. Xét một phần tư hình tròn tâm $O(0,0)$ bán kính bằng a , ký hiệu là OAB , với tọa độ tương ứng là $A(a,0)$ và $B(0,a)$.

(a) Trên đoạn OA lấy ngẫu nhiên một điểm C . Tìm phân phối xác suất của độ dài đoạn OC .

(b) Vẽ một đường thẳng đi qua C , vuông góc với OA và cắt cung tròn tại điểm D . Tính kỳ vọng và phương sai của độ dài đoạn CD .

Bài tập 2.42. Lấy ngẫu nhiên một điểm M trên nửa đường tròn tâm O , đường kính $AB = 2a$. Biết rằng xác suất điểm M rơi vào cung CD bất kỳ của nửa đường tròn AMB chỉ phụ thuộc vào độ dài cung CD .

(a) Tìm hàm phân phối xác suất của biến ngẫu nhiên Y chỉ diện tích tam giác AMB .

(b) Tìm giá trị trung bình của diện tích tam giác ấy.

Bài tập 2.43. Từ điểm $A(0, -a)$ ($a > 0$) trong nửa mặt phẳng tọa độ xOy phần $x \geq 0$, người ta kẻ ngẫu nhiên một tia At hợp với tia Oy một góc φ . Biết φ là biến ngẫu nhiên có phân phối đều trong khoảng $(0, \pi/4)$. Tia At cắt Ox tại điểm M .

(a) Tìm hàm phân phối xác suất của biến ngẫu nhiên X chỉ diện tích tam giác AOM .

(b) Tìm giá trị trung bình của diện tích trên.

Bài tập 2.44. Một công ty kinh doanh mặt hàng A dự định sẽ áp dụng một trong hai phương án kinh doanh: Phương án 1: Gọi X_1 (triệu đồng/tháng) là lợi nhuận thu được. X_1 có phân phối chuẩn $\mathcal{N}(140; 2500)$. Phương án 2: Gọi X_2 (triệu đồng/tháng) là lợi nhuận thu được. X_2 có phân phối chuẩn $\mathcal{N}(200; 3600)$. Biết rằng công ty tồn tại và phát triển thì lợi nhuận thu được từ mặt hàng A phải đạt ít nhất 80 triệu đồng/tháng. Hỏi nên áp dụng phương án nào để rủi ro thấp hơn.

Bài tập 2.45. Trọng lượng của một loại trái cây tuân theo luật phân phối chuẩn với trọng lượng trung bình là 250g, độ lệch chuẩn là 5g. Trái cây loại I là trái cây có trọng lượng không nhỏ hơn 260g.

(a) Một người lấy 1 trái từ trong sọt trái cây ra. Tính xác suất người này lấy được trái cây loại I.

(b) Nếu lấy được trái loại I thì người này sẽ mua sọt đó. Người ngày kiểm tra 100 sọt. Tính xác suất người này mua được 6 sọt.

Bài tập 2.46. Một dây chuyền tự động khi hoạt động bình thường có thể sản xuất ra phế phẩm với xác suất $p = 0,001$ và được điều chỉnh ngay lập tức khi phát hiện có phế phẩm. Tính số trung bình các sản phẩm được sản xuất giữa 2 lần điều chỉnh.

Bài tập 2.47. Trong một kỳ thi điểm số trung bình của các sinh viên là 80 và độ lệch chuẩn là 10. Giả sử điểm thi của sinh viên tuân theo luật phân phối chuẩn.

(a) Nếu giáo viên muốn 25% số sinh viên đạt điểm A (nhóm điểm cao nhất) thì điểm số thấp nhất để đạt điểm A là bao nhiêu?

(b) Chọn ngẫu nhiên 50 sinh viên, tính xác suất trong đó có nhiều hơn 10 sinh viên đạt điểm A (điểm A lấy ở câu (a)).

Bài tập 2.48. Đường kính của một loại chi tiết do một máy sản xuất tuân theo luật phân phối chuẩn, với kỳ vọng là 20mm và độ lệch chuẩn là 0,2mm. Tính xác suất để lấy ngẫu nhiên một chi tiết có đường kính trong khoảng 19,9mm đến 20,3mm.

Bài tập 2.49. Chiều cao của nam giới khi trưởng thành là biến ngẫu nhiên tuân theo luật phân phối chuẩn với chiều cao trung bình là 160cm và độ lệch chuẩn là 6cm. Tìm xác suất để đo ngẫu nhiên 4 người thì có ít nhất một người có chiều cao nằm trong khoảng (158–162)cm.

Bài tập 2.50. Dùng hai phương pháp để tính sai số của một biến ngẫu nhiên. Phương pháp 1: Cho sai số đó bằng $2X$ với X là biến ngẫu nhiên có phân phối chuẩn $\mathcal{N}(0; 25)$. Phương pháp 2: Cho sai số đó bằng tổng hai biến ngẫu nhiên độc lập $Y = Y_1 + Y_2$ trong đó $E(Y_1) = E(Y_2) = 0$ và $\sigma(Y_1) = \sigma(Y_2) = 5$. Hỏi phương pháp nào được ưa dùng hơn?