Chương 2

Biến ngẫu nhiên và quy luật phân phối xác suất

TUẦN 5

2.1 Định nghĩa và phân loại biến ngẫu nhiên

2.1.1 Định nghĩa biến ngẫu nhiên

Khái niệm biến ngẫu nhiên 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 $X, Y, Z, X_1, X_2, \ldots$. Các giá trị có thể có của chúng ký kiệu là $x, y, z, x_1, x_2, \ldots$.

Tập hợp tất cả các giá trị của X gọi là miền giá trị của X, ký hiệu là $\mathbf{S}_X$.

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 chắn chắc 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 $(X = x_1), (X = x_2), \ldots, (X = x_n)$ về thực chất là các sự kiện ngẫu nhiên.

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

(c) Hai biến ngẫu nhiên X và Y là độc lập nếu X nhận các giá trị nào đó không phụ thuộc Y và ngược lại.

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.

(a) **Biến ngẫu nhiên rời rạc:** X là biến ngẫu nhiên rời rạc nếu tập giá trị $\mathbf{S}_X$ 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.

(b) **Biến ngẫu nhiên liên tục:** X là biến ngẫu nhiên liên tục nếu tập giá trị $\mathbf{S}_X$ 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 $1, 2, \ldots, n, \ldots$.

(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 2.1 (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 của biến ngẫu nhiên.

Một số phương pháp mô tả quy luật phân phối xác suất

(a) Bảng phân phối xác suất (áp dụng cho biến ngẫu nhiên rời rạc).

(b) Hàm phân phối xác suất (áp dụng cho cả biến ngẫu nhiên rời rạc và liên tục).

(c) 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 2.2 (Hàm khối lượng xác suất)**.** Cho biến ngẫu nhiên rời rạc X. Đặt

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

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.

Tính chất 2.1. (a) $p_X(x_k) > 0$ với mọi $x_k \in \mathbf{S}_X$;

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

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

Định nghĩa 2.3 (Bảng phân phối xác suất)**.** Bảng phân phối xác suất 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.

(i) 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	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, $p_i = p_X(x_i) = P(X = x_i)$, $i = 1, 2 \dots, n$.

(ii) 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	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, $p_n = p_X(x_n) = P(X = x_n)$, $n = 1, 2 \dots$.

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: 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,

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

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

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

Vậy bảng phân phối xác suất của X là

X	1	2	3
p	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	$^{99}/_{100}$	$^{1}/_{100}$

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: 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,

$$\begin{aligned}
&P(X=1) = P(A_1) = \frac{1}{4}\\
&P(X=2) = P(\overline{A}_1A_2) = P(\overline{A}_1)P(A_2|\overline{A}_1) = \frac{3\times 1}{4\times 3} = \frac{1}{4}\\
&P(X=3) = P(\overline{A}_1\overline{A}_2A_3) = P(\overline{A}_1)P(\overline{A}_2|)P(A_3|\overline{A}_1\overline{A}_2) = \frac{3\times 2\times 1}{4\times 3\times 2} = \frac{1}{4}\\
&P(X=4) = P(\overline{A}_1\overline{A}_2\overline{A}_3A_4) = P(\overline{A}_1)P(\overline{A}_2|\overline{A}_1)P(\overline{A}_3|\overline{A}_1\overline{A}_2)P(A_4|\overline{A}_1\overline{A}_2\overline{A}_3) = \frac{1}{4}.
\end{aligned}$$

Suy ra bảng phân phối xác suất của X

X	1	2	3	4
p	$^{1}/_{4}$	$^{1}/_{4}$	$^{1}/_{4}$	$^{1}/_{4}$

2.2.2 Hàm phân phối xác suất

Định nghĩa 2.4 (Hàm phân phối xác suất)**.** Hàm phân phối xác suất của biến ngẫu nhiên X, ký hiệu là $F_X(x)$, được định nghĩa như sau:

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

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

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

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

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

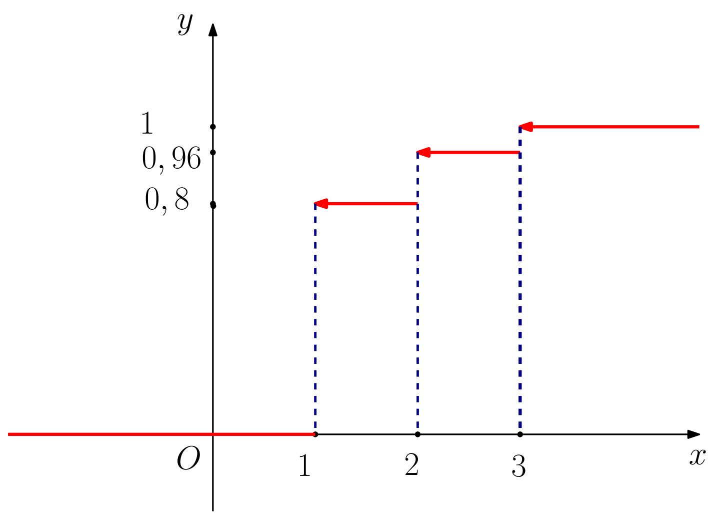
Nhận xét 2.2. Hàm phân phối xác suất $F_X(x)$ phản ánh mức độ tập trung xác suất ở bên trái của một số thực x nào đó.

Ví dụ 2.6. Tìm hàm phân phối xác suất của biến ngẫu nhiên ở Ví dụ 2.3.

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

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

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



Hình 2.1: Đồ thị hàm phân phối xác suất trong Ví dụ 2.6

Hàm phân phối có các tính chất sau.

Tính chất 2.2. (a) $0 \le F_X(x) \le 1$.

(b) $F_X(x)$ 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 $F_X(a^-) = \lim_{x \to a^-} F_X(x)$.

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

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

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

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

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

Chứng minh. (a) Suy trực tiếp từ định nghĩa hàm phân phối và tính chất của xác suất.

(b) Giả sử $x_1 < x_2$ và xét sự kiện $(X < x_2) = (X < x_1) + (x_1 \leq X < x_2)$. Khi đó 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 định nghĩa hàm phân phối xác suất (2.4) suy ra

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

(c) Suy trực tiếp từ chứng minh tính chất (b).

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

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: Sử dụng Tính chất 2.2(a) của hàm phân phối xác suất, $0 \leq A + B \arcsin x \leq 1$ và theo Tính chất 2.2(b) vì $F_X(x)$ liên tục nên $A - \dfrac{\pi}{2} \times B = 0$, $A + \dfrac{\pi}{2} \times B = 1$.

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

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.

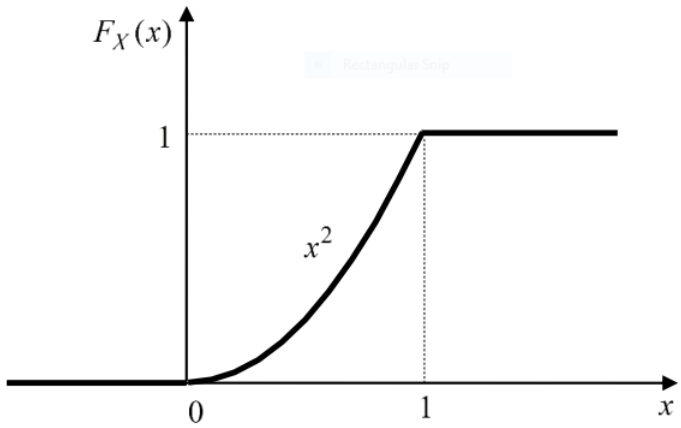
(b) Tìm hàm phân phối $F_X(x)$ và vẽ đồ thị của $F_X(x)$.

Lời giải:

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

(b) Sử dụng định nghĩa $F_X(x) = P(X < x)$,

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



Hình 2.2: Hàm phân phối xác suất của Ví dụ 2.8

TUẦN 6

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

Định nghĩa 2.5 (Hàm mật độ)**.** 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

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

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

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,

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

Nhận xét 2.3. 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.

Tính chất 2.3. (a) $f_X(x) \geq 0$ với mọi $x \in \mathbb{R}$.

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

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

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

(b) Được suy từ Tính chất 2.2(c).

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

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

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

(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:

(a) Tương tự như Ví dụ 2.7 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 Tính chất 2.3(b)

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

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 $\left(0, \dfrac{\pi}{4}\right)$.

Lời giải:

(a) Sử dụng Tính chất 2.3(a),(c) tính được $a = \dfrac{1}{2}$.

(b) Áp dụng (2.7).
Nếu $x \le -\dfrac{\pi}{2}$ thì $F_X(x) = \int_{-\infty}^{x} 0du = 0$.
Nếu $-\dfrac{\pi}{2} < x \le \dfrac{\pi}{2}$ thì $F_X(x) = \int_{-\infty}^{x} f_X(u)du = \int_{-\frac{\pi}{2}}^{x} \frac{1}{2}\cos u du = \dfrac{1}{2}(\sin x + 1)$.
Nếu $x > \dfrac{\pi}{2}$ thì $F_X(x) = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}\cos x = 1$. Vậy

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

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

Ví dụ 2.11 (Đề thi MI2020 giữa kỳ 20191)**.** Biến ngẫu nhiên liên tục X có hàm mật độ xác suất

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

(a) Tìm hằng số k.

(b) 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 $\left(0; \dfrac{1}{2}\right)$.

Lời giải:

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

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

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

2.3 Các tham số đặc trưng của biến ngẫu nhiên

Đặc trưng quan trọng nhất của biến ngẫu nhiên là hàm phân phối 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 2.6 (Kỳ vọng). Kỳ vọng (expected value) của biến ngẫu nhiên X, ký hiệu là $E(X)$, được xác định như sau:

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

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

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

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

nếu chuỗi vế phải hội tụ.

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

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

nếu tích phân vế phải hội tụ.

Nhận xét 2.4. (a) 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 $x_2, \ldots, n_k$ lần X nhận giá trị x_k, $n_1 + n_2 + \cdots + n_k = n$. Giá trị trung bình của biến ngẫu nhiên X trong n phép thử này là

$$\begin{aligned}\overline{X} &= \frac{n_1 x_2 + n_2 x_2 + \cdots + n_k x_k}{n} = x_1 \frac{n_1}{n} + x_2 \frac{n_2}{n} + \cdots + x_k \frac{n_k}{n} \\ &\simeq x_1 p_1 + x_2 p_2 + \cdots + x_k p_k = E(X).\end{aligned}$$

(b) 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.

Tính chất 2.4. (a) $E(C) = C$ với C là hằng số.

(b) $E(CX) = CE(X)$ với C là hằng số.

(c) $E(X+Y) = E(X) + E(Y)$; mở rộng $E(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} E(X_i)$.

(d) $E(XY) = E(X)E(Y)$ nếu X và Y là các biến ngẫu nhiên độc lập.

Chứng minh. Ta chứng minh Tính chất 2.4(b). Giả sử X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất (2.2) Khi đó CX sẽ là một biến ngẫu nhiên rời rạc có bảng phân phối xác suất là

CX	Cx_1	Cx_2	$\dots$	Cx_n
p	p_1	p_2	$\dots$	p_n

vì $P(CX = Cx_i) = P(X = x_i) = p_i$, $i = 1, 2, \dots, n$. Áp dụng (2.9) ta được

$$E(CX) = \sum_{i=1}^{n} Cx_i p_i = C \sum_{i=1}^{n} x_i p_i = CE(X).$$

Các tính chất khác được chứng minh tương tự.

Ví dụ 2.12. 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: 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	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 2.13. 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 là

X	0	700
p	$^{99}/_{100}$	$^{1}/_{100}$

Kỳ vọng của X là $E(X) = 0 \times {}^{99}/_{100} + 700 \times {}^{1}/_{100} = 7$ 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 2.14. 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: Gọi X là số tiền thưởng nhận được khi người chơi chọn câu hỏi A trả lời đầu tiên,

X	0	100	300
p	$0,2$	$0,32$	$0,48$

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	$0,4$	$0,12$	$0,48$

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 2.15. 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	$0,15$	$0,2$	$0,35$	$0,2$	$0,1$

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:

(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

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

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 2.16 (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.

(a) 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.

(b) 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:

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

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

(b) 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 2.17. 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: 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 $1, 2, \ldots$ với xác suất tương ứng

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

Vậy

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

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

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

Ví dụ 2.18. 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: Vì $f_X(x)$ là hàm mật độ xác suất nên theo Tính chất 2.3(a),(c), $k = \dfrac{3}{64}$.

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

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

Ví dụ 2.19 (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 $1 \leq x < 2\ldots$).

(a) Tính $P(Y = 0)$.

(b) Tính $E(Y)$.

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

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

2.3.2 Phương sai

Trong thực tế nhiều khi chỉ xác định kỳ vọng toán của biến ngẫu nhiên thì chưa đủ để xác định biến ngẫu nhiên đó. Ta còn phải xác định mức độ phân tán của các giá trị của biến ngẫu nhiên xung quanh giá trị trung bình của nó nữa.

Định nghĩa 2.7 (Phương sai)**.** Phương sai (variance) của biến ngẫu nhiên X, ký hiệu là $V(X)$, được định nghĩa như sau:

$$\boxed{V(X) = E[X - E(X)]^2} \tag{2.12}$$

Công thức tương đương của (2.12)

$$\boxed{V(X) = E(X^2) - [E(X)]^2} \tag{2.13}$$

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

$$\boxed{V(X) = \sum_{i=1}^{n} \big(x_i - E(X)\big)^2 p_i} \tag{2.14}$$

hay

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

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

$$\boxed{V(X) = \sum_{n=1}^{\infty} \big(x_n - E(X)\big)^2 p_n} \tag{2.16}$$

hay

$$\boxed{V(X) = E(X^2) - [E(X)]^2 = \sum_{n=1}^{\infty} x_n^2 p_n - \Big(\sum_{n=1}^{\infty} x_n p_n\Big)^2} \tag{2.17}$$

nếu chuỗi/các chuỗi vế phải hội tụ.

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

$$\boxed{V(X) = \int_{-\infty}^{+\infty} \big(x - E(X)\big)^2 f_X(x)dx} \tag{2.18}$$

hay

$$\boxed{V(X) = E(X^2) - [E(X)]^2 = \int_{-\infty}^{+\infty} x^2 f_X(x)dx - \left(\int_{-\infty}^{+\infty} x f_X(x)dx\right)^2} \tag{2.19}$$

nếu tích phân/các tích phân vế phải hội tụ.

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

Nhận xét 2.5. (a) 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.

(b) 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.

Tính chất 2.5. (a) $V(C) = 0$ với C là hằng số.

(b) $V(CX) = C^2V(X)$ với C là hằng số.

(c) $V(X+Y) = V(X) + V(Y)$ nếu X và Y là các biến ngẫu nhiên độc lập.

Mở rộng: nếu $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên độc lập thì $V(\sum_{i=1}^n X_i) = \sum_{i=1}^n V(X_i)$.

$V(X+C) = V(X)$, C là hằng số.

$V(X-Y) = V(X) + V(Y)$ nếu X và Y là các biến ngẫu nhiên độc lập.

Chứng minh. Sử dụng định nghĩa của phương sai và tính chất của kỳ vọng ta có:
(a) $V(C) = E[C - E(C)]^2 = E(C - C) = E(0) = 0$.
(b) $V(CX) = E[(CX)^2] - [E(CX)]^2 = C^2[E(X^2) - (E(X))^2] = C^2V(X)$.
Các tính chất khác được chứng minh tương tự.

2.3.3 Độ lệch chuẩn

Định nghĩa 2.8 (Độ lệch chuẩn)**.** Độ lệch chuẩn của biến ngẫu nhiên X, ký hiệu là $\sigma(X)$, được định nghĩa như sau:

$$\boxed{\sigma(X) = \sqrt{V(X)}} \tag{2.20}$$

Nhận xét 2.6. 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 2.20. Tính phương sai và độ lệch chuẩn của biến ngẫu nhiên xét trong Ví dụ (2.12).

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

$$\begin{aligned} V(X) &= E(X^2) - (E(X))^2 = 7940 - 4 = 7936, \\ \sigma(X) &= \sqrt{V(X)} = \sqrt{7936} \simeq 89,08. \end{aligned}$$

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

Ví dụ 2.21. Tính phương sai và độ lệch chuẩn của biến ngẫu nhiên xét trong Ví dụ (2.18).

Lời giải: $E(X^2) = \dfrac{3}{64}\displaystyle\int_0^4 x^4(4-x)dx = \dfrac{32}{5}$. Suy ra $V(X) = \dfrac{32}{5} - \left(\dfrac{12}{5}\right)^2 = \dfrac{16}{5}$ và $\sigma(X) = \dfrac{4}{5}$.

TUẦN 7

2.3.4 Một số đặc trưng khác

2.3.4a Mốt (mode)

Định nghĩa 2.9 (Mốt)**.** (a) Nếu X là biến ngẫu nhiên rời rạc thì mốt là giá trị của X ứng với xác suất lớn nhất.

(b) Nếu X là biến ngẫu nhiên liên tục thì mốt là giá trị làm hàm mật độ đạt max.

Ký hiệu $modX$.

2.3.4b Trung vị (median)

Định nghĩa 2.10 (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

$$\boxed{P(X < medX) = P(X \geq medX) = \frac{1}{2}} \tag{2.21}$$

Nhận xét 2.7. (a) Từ định nghĩa hàm phân phối, để tìm trung vị ta cần giải phương trình $F(x) = \dfrac{1}{2}$.

(b) 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 2.22. Cho hàm mật độ xác suất của biến ngẫu nhiên liên tục X:

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

Tìm $medX$ và $modX$.

Lời giải: Theo (2.7), hàm phân phối xác suất của X là

$$F_X(x) = \begin{cases} 0, & x \le 0, \\ \dfrac{3}{4}\Big(x^2 - \dfrac{x^3}{3}\Big), & 0 < x \le 2, \\ 1, & x > 2. \end{cases}$$

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

Hàm mật độ xác suất $f_X(x)$ có

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

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

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 2.11 (Phân phối đều rời rạc)**.** Biến ngẫu nhiên X được gọi là tuân theo luật phân phối đều rời rạc 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	$\frac{1}{n}$	$\frac{1}{n}$	...	$\frac{1}{n}$

(2.22)

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

$$\boxed{E(X) = \frac{n+1}{2}, \quad V(X) = \frac{n^2-1}{12}} \tag{2.23}$$

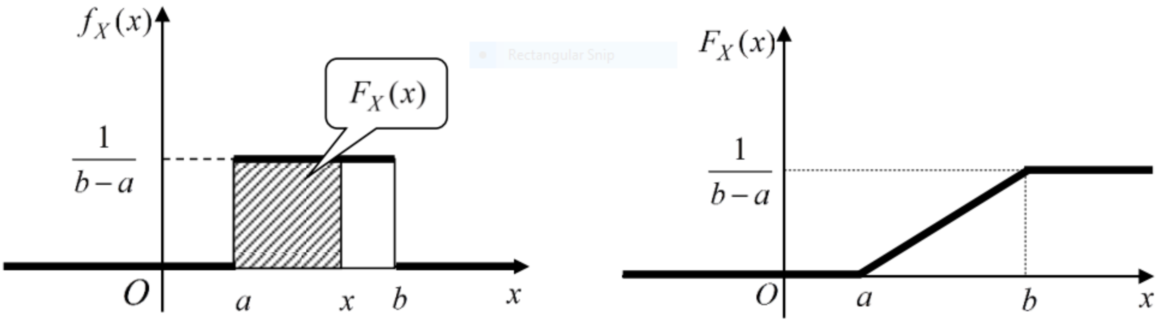
2.4.1b Phân phối đều liên tục

Định nghĩa 2.12 (Phân phối đều liên tục). Biến ngẫu nhiên X được gọi là tuân theo luật phân phối đều liên tục trên $[a,b]$ ($a<b$), ký hiệu là $X \sim \mathcal{U}([a,b])$, nếu X có hàm mật độ

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a,b], \\ 0, & x \notin [a,b]. \end{cases} \tag{2.24}$$

Hàm phân phối xác suất tương ứng

$$F_X(x) = \int_{-\infty}^{x} f_X(u)du = \begin{cases} 0, & x \le a, \\ \dfrac{x-a}{b-a}, & a < x \le b, \\ 1, & x > b. \end{cases} \tag{2.25}$$



Hình 2.3: Hàm mật độ và hàm phân phối xác suất của biến ngẫu nhiên có phân phối đều

Kỳ vọng, phương sai của biến ngẫu nhiên có phân phối đều liên tục:

$$\boxed{E(X) = \frac{a+b}{2}, \quad V(X) = \frac{(b-a)^2}{12}} \tag{2.26}$$

Nhận xét 2.8. (a) X có khả năng nhận giá trị trong khoảng (a,b) là "đều nhau".

(b) Phân phối đều có nhiều ứng dụng trong thống kê toán như mô phỏng thống kê, đặc biệt trong phương pháp phi tham số.

(c) 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 2.23. 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: Gọi X là số phút từ 7 giờ đến 7 giờ 30 hành khách đến trạm, ta có $X \sim \mathcal{U}([0,30])$.

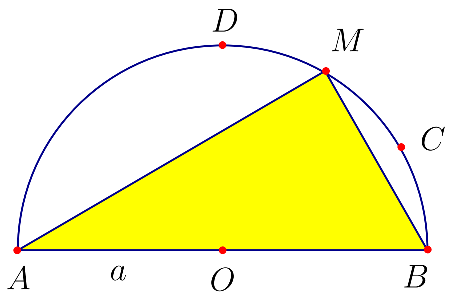
(a) Hành khách chờ ít hơn 5 phút nếu đến trạm giữa 7 giờ 10 và 7 giờ 15 hoặc giữa 7 giờ 25 và 7 giờ 30. Do đó xác suất cần tìm là:

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

(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:

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

Ví dụ 2.24. 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. Tìm hàm phân phối xác suất của biến ngẫu nhiên Y chỉ diện tích tam giác AMB.



Hình 2.4: Minh họa cho Ví dụ 2.24

Lời giải: (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

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

Do đó, hàm phân phối xác suất của φ là

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

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

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

vì với $x \in (0, a^2]$,

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

2.4.2 Phân phối nhị thức

2.4.2a Phân phối Béc–nu–li

Định nghĩa 2.13 (Phân phối Béc–nu–li)**.** Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối Béc–nu–li 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

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

ở đây $0 < p < 1, q = 1 - p$.

Kỳ vọng, phương sai của biến ngẫu nhiên có phân phối Béc–nu–li:

$$\boxed{E(X) = p, \quad V(X) = pq} \tag{2.27}$$

Nhận xét 2.9. 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 2.14 (Phân phối nhị thức)**.** Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối nhị thức 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	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	...	$C_n^k p^k q^{n-k}$	...	$C_n^n p^n q^0$

(2.28)

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

Kỳ vọng, phương sai của biến ngẫu nhiên có phân phối nhị thức:

$$\boxed{E(X) = np, \quad V(X) = npq} \tag{2.29}$$

Nhận xét 2.10. (a) 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 $i = 1, 2, \ldots, n$, 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 $X = X_1 + X_2 + \cdots + X_n \sim \mathcal{B}(n, p)$.

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

Ví dụ 2.25. 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: 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

Định nghĩa 2.15 (Phân phối Poa–xông). Biến ngẫu nhiên rời rạc X được gọi là tuân theo luật phân phối Poa-xông với tham số $\lambda = np$, ký hiệu là $X \sim \mathcal{P}(\lambda)$, nếu X có bảng phân phối xác suất

X	0	1	...	k	...	n	...
p	$\frac{\lambda^0}{0!}e^{-\lambda}$	$\frac{\lambda^1}{1!}e^{-\lambda}$	...	$\frac{\lambda^k}{k!}e^{-\lambda}$	...	$\frac{\lambda^n}{n!}e^{-\lambda}$	...

(2.30)

Kỳ vọng, phương sai của biến ngẫu nhiên có phân phối Poa–xông:

$$\boxed{E(X) = \lambda, \quad V(X) = \lambda} \tag{2.31}$$

Nhận xét 2.11. (a) Phân phối Poa–xông xuất hiện trong dãy phép thử Béc–nu–li khi số phép thử n khá lớn và xác xuất p khá bé (thỏa mãn $\lambda = np < 7$).

(b) 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 ...

(c) 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 $\lambda_1 + \lambda_2$.

(d) 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 2.26. Xác suất để trong khi vận chuyển mỗi chai rượu bị vỡ là 0,001. Người ta tiến hành vận chuyển 2000 chai rượu đến cửa hàng. Tìm số chai vỡ trung bình khi vận chuyển.

Lời giải: Bài toán thỏa mãn lược đồ Béc–nu–li với $n = 2000$ khá lớn và $p = 0,001$ khá bé, $np = 2 < 7$. Gọi X là số chai rượu bị vỡ khi vận chuyển thì X là biến ngẫu nhiên phân phối theo quy luật Poa–xông.

Số chai vỡ trung bình khi vận chuyển là $E(X) = \lambda = np = 2000 \times 0,001 = 2$.

Ví dụ 2.27. 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:

(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 2.28. Một ga ra cho thuê ôtô thấy rằng số người đến thuê ôtô vào thứ bảy cuối tuần là một biến ngẫu nhiên tuân theo luật phân phối Poát-xông với tham số $\lambda = 2$. Giả sử gara có 4 chiếc ôtô.

(a) Tìm xác suất để tất cả 4 ôtô 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 ôtô được thuê vào ngày thứ 7?

Lời giải: Gọi X là biến ngẫu nhiên chỉ "số người đến thuê ôtô vào thứ bảy". Theo giả thiết X là biến ngẫu nhiên phân phối tuân theo quy luật Poát-xông $\mathcal{P}(\lambda)$. Gọi Y là biến ngẫu nhiên chỉ "số xe được thuê vào thứ bảy".

(a) Áp dụng công thức (2.31),

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

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

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

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

Bảng phân phối xác suất của Y là:

X	0	1	2	3	4
p	0,1353	0,2707	0,2707	0,1804	0,1429

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

Ví dụ 2.29 (Đề thi MI2020 kỳ 20191)**.** Số khách hàng đến một cửa hàng bán lẻ là một biến ngẫu nhiên có phân phối Poisson 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: Gọi X là "số khách hàng đến cửa hàng bán lẻ trong vòng 30 phút". $X \sim \mathcal{P}(\lambda)$. Xác suất cần tìm $p = P(X \geq 3)$ với $\lambda = 3$. Vậy,

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

Chú ý 2.3. Giá trị xác suất của phân phối Poa–xông được tính sẵn trong bảng giá trị khối lượng xác suất Poa–xông (Phụ lục 5).

TUẦN 8

2.4.4 Phân phối chuẩn

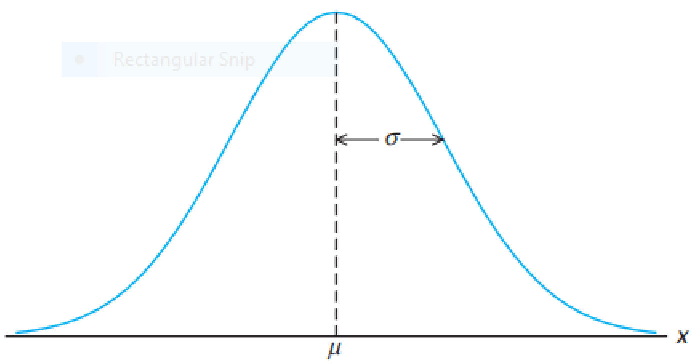
2.4.4a Phân phối chuẩn

Định nghĩa 2.16 (Phân phối chuẩn)**.** Biến ngẫu nhiên liên tục X được gọi là tuân theo luật phân phối chuẩn 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

$$\boxed{f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}} \tag{2.32}$$

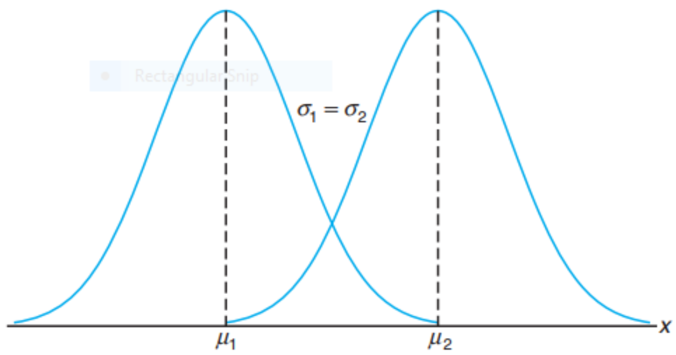
ở đây e và π được lấy xấp xỉ lần lượt là 2.71828 và 3.14159.

Nhận xét 2.12. 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 2.5), mô tả gần đúng nhiều hiện tượng trong tự nhiên, công nghiệp và nghiên cứu.



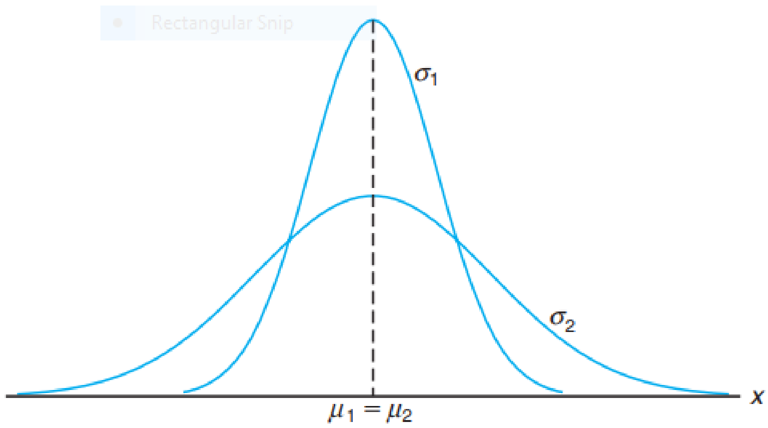
Hình 2.5: Đường cong chuẩn

Hình 2.6 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 2.6: Đường cong chuẩn với $\mu_1 < \mu_2$ và $\sigma_1 = \sigma_2$

Hình 2.7 mô tả hai đường cong chuẩn có cùng kỳ vọng nhưng độ lệch chuẩn khác nhau. Hình 2.8 mô tả cho trường hợp kỳ vọng và độ lệch chuẩn khác nhau.



Hình 2.7: Đường cong chuẩn với $\mu_1 = \mu_2$ và $\sigma_1 < \sigma_2$

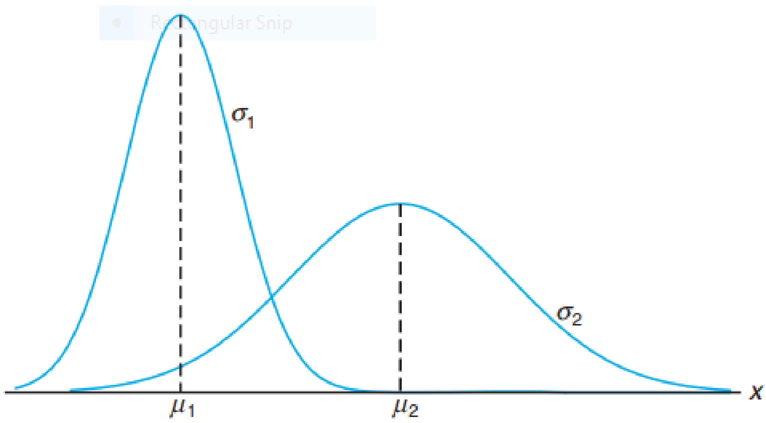
Định lý 2.1. Kỳ vọng và phương sai của biến ngẫu nhiên X có phân phối chuẩn

$$\boxed{E(X) = \mu, \quad V(X) = \sigma^2} \tag{2.33}$$

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

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

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



Hình 2.8: Đường cong chuẩn với $\mu_1 < \mu_2$ và $\sigma_1 < \sigma_2$

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

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

□

Định lý 2.2. 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 2.4. (a) 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 $X_1 + X_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

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

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \sim \mathcal{N}\Big(\mu, \frac{\sigma^2}{n}\Big).$$

2.4.4b 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

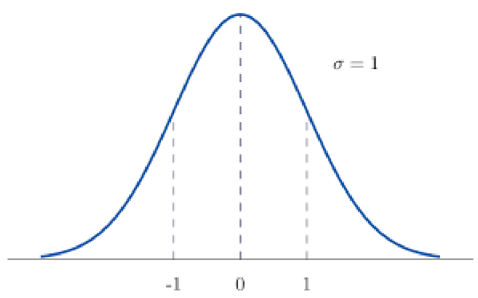
$$U = \frac{X-\mu}{\sigma} \tag{2.34}$$

là biến ngẫu nhiên có phân phối chuẩn tắc $\mathcal{N}(0,1)$.

Hàm mật độ xác suất của biến ngẫu nhiên có phân phối chuẩn tắc là

$$\boxed{\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}} \tag{2.35}$$

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



Hình 2.9: Hàm mật độ của biến ngẫu nhiên có phân phối chuẩn tắc $\mathcal{N}(0,1)$

Hàm phân phối xác suất của biến ngẫu nhiên U phân phối chuẩn tắc là

$$\boxed{\Phi_U(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt, \quad x \in \mathbb{R}} \tag{2.36}$$

với các giá trị được tính sẵn trong Phụ lục 3.

Chú ý 2.5. (a) Mối liên hệ giữa hàm phân phối xác suất của biến ngẫu nhiên phân phối chuẩn tắc (2.36) và hàm Láp–la–xơ (1.24) là

$$\boxed{\Phi(x) = 0,5 + \phi(x), \quad x \geq 0} \tag{2.37}$$

(b) Nếu n biến ngẫu nhiên độc lập X_i cùng có phân phối chuẩn $\mathcal{N}(\mu,\sigma^2)$, $i = 1,\dots,n$, và $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ thì biến ngẫu nhiên $U = \dfrac{\overline{X}-\mu}{\sigma}\sqrt{n}$ có phân phối chuẩn tắc $\mathcal{N}(0,1)$.

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)

$$\boxed{P(\alpha < X < \beta) = \phi\Big(\frac{\beta - \mu}{\sigma}\Big) - \phi\Big(\frac{\alpha - \mu}{\sigma}\Big)} \tag{2.38}$$

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

Thật vậy, sử dụng phép đổi biến $t = \dfrac{x - \mu}{\sigma}$ ta nhận được

$$P(\alpha < X < \beta) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\alpha}^{\beta} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}} \int_{\frac{\alpha-\mu}{\sigma}}^{\frac{\beta-\mu}{\sigma}} e^{-\frac{t^2}{2}} dt.$$

Từ đây và (1.24) ta nhận được (2.38).

2.4.4d Quy tắc 3σ

Từ (2.38) suy ra $P(|X - \mu| < t\sigma) = 2\phi(t)$, thay $t = 1, 2, 3$, tra bảng hàm số Láp–la–xơ (Phụ lục 2) ta nhận được

$$\begin{aligned} P(|X - \mu| < \sigma) &= 2\phi(1) = 0,6827, \\ P(|X - \mu| < 2\sigma) &= 2\phi(2) = 0,9545, \\ P(|X - \mu| < 3\sigma) &= 2\phi(3) = 0,9973. \end{aligned} \tag{2.39}$$

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 2.6. (a) Phân phối chuẩn được Gao–xơ tìm ra năm 1809 nên nó còn được gọi là phân phối Gao–xơ.

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

(c) 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 2.30. 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: 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

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

và

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

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

Phân phối chuẩn có thể dùng xấp xỉ khá tốt cho một số phân phối rời rạc.

Định lý 2.3. Nếu X là biến ngẫu nhiên tuân theo luật phân phối chuẩn với kỳ vọng $\mu = np$ và phương sai $\sigma^2 = npq$, thì giới hạn của phân phối xác suất của biến ngẫu nhiên

$$Z = \frac{X - np}{\sqrt{npq}},$$

khi $n \to \infty$ tuân theo luật phân phối chuẩn tắc $\mathcal{N}(0, 1)$.

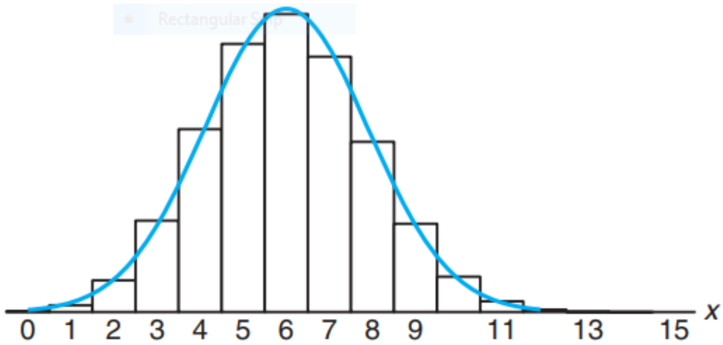
Phân phối chuẩn với kỳ vọng $\mu = np$ và phương sai $\sigma^2 = npq$ không chỉ xấp xỉ khá tốt cho phân phối nhị thức khi n khá lớn và xác suất p không quá gần 0 hoặc 1 mà còn cung cấp một xấp xỉ khá tốt cho phân phối nhị thức ngay cả khi n nhỏ và p gần $1/2$.

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 $\sigma^2 = npq = 15 \times 0,4 \times 0,6 = 3,6$ với biến ngẫu nhiên tuân theo luật phân phối nhị thức X (xem Hình 2.10).

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 nghĩa 2.17. 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

$$\boxed{P(X = k) = C_n^k p^k (1-p)^{n-k} \simeq \phi\Big(\frac{k + 0,5 - \mu}{\sigma}\Big) - \phi\Big(\frac{k - 0,5 - \mu}{\sigma}\Big)} \qquad (2.40)$$

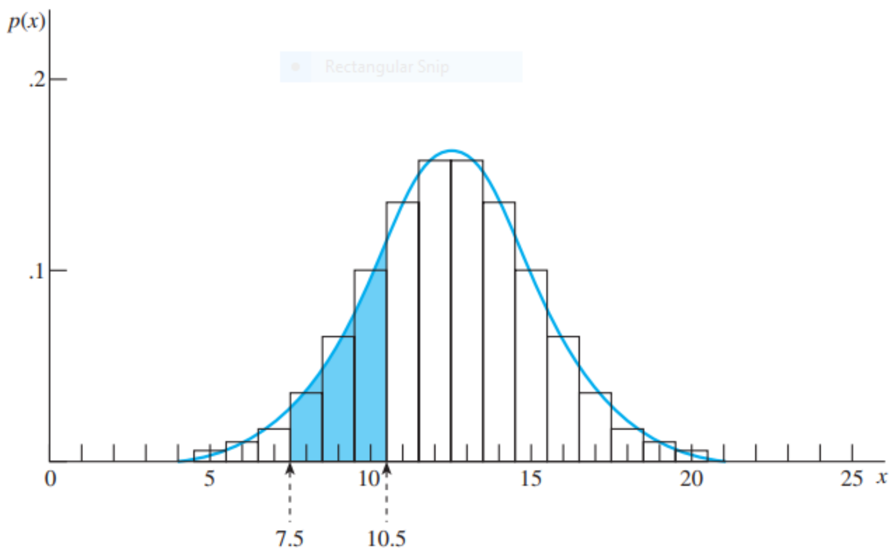


Hình 2.10: Xấp xỉ phân phối chuẩn cho phân phối nhị thức $\mathcal{B}(15; 0,4)$

$$\boxed{P(k_1 \le X \le k_2) = \sum_{k=k_1}^{k_2} C_n^k p^k (1-p)^{n-k} \simeq \phi\Big(\frac{k_2 + 0,5 - \mu}{\sigma}\Big) - \phi\Big(\frac{k_1 - 0,5 - \mu}{\sigma}\Big)} \tag{2.41}$$

Xấp xỉ là khá tốt nếu $np \ge 5$ và $n(1-p) \ge 5$.

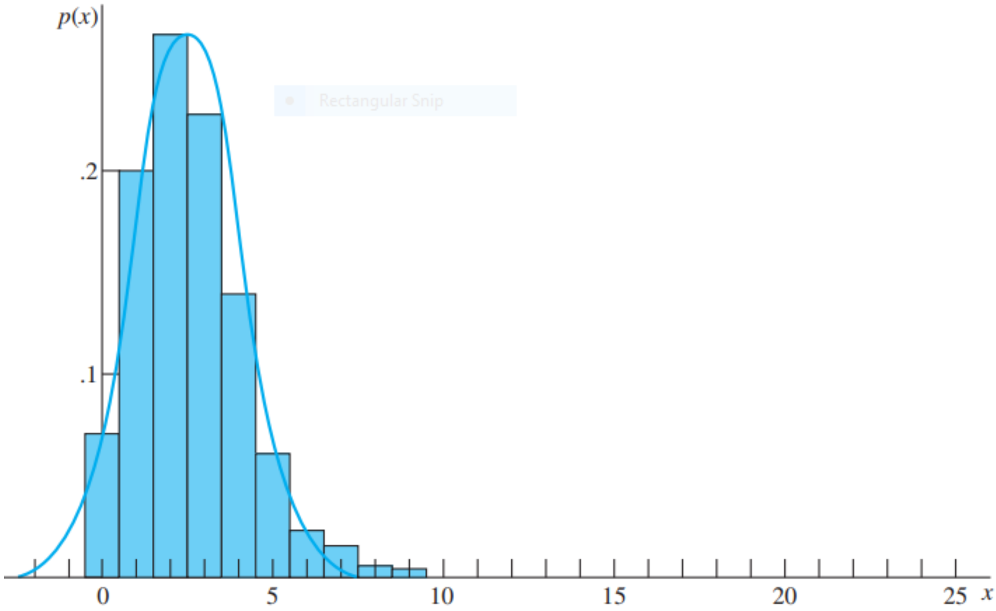
Nhận xét 2.13. Hình 2.11 và 2.12 biểu thị biểu đồ xác suất nhị thức với $n = 25$ và $p = 0.5$, $p = 0.1$ tương ứng. Phân phối trong Hình 2.11 là hoàn toàn đối xứng.



Hình 2.11: Phân phối nhị thức với $n = 25$ và $p = 0,5$ xấp xỉ bởi phân phối chuẩn với $\mu = 12,5$ và $\sigma = 2,5$

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

Ví dụ 2.31. 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.



Hình 2.12: Phân phối nhị thức và xấp xỉ phân phối chuẩn với $n = 25$ và $p = 0,1$

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

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

Sử dụng công thức xấp xỉ (2.41) với $\mu = np = 12,5$, $\sigma = \sqrt{npq} = 2,5$ ta nhận được

$$P(8 \le X \le 10) \simeq \phi(-0.8) - \phi(-2) = 0,18911.$$

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

Ví dụ 2.32. 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: 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 $X \sim \mathcal{N}(950; 47,5)$:

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

2.4.5 Phân phối khi bình phương

Định nghĩa 2.18 (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

$$\boxed{f(x) = \frac{1}{2}e^{-\frac{x}{2}}\frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)}, \quad x > 0} \tag{2.42}$$

ở đây

$$\Gamma(x) = \int\limits_0^{+\infty} t^{x-1}e^{-t}dt, \quad x > 0$$

là hàm Gamma.

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

$$\boxed{U_n = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi_n^2} \tag{2.43}$$

Kỳ vọng và phương sai của biến ngẫu nhiên U_n có phân phối khi bình phương:

$$\boxed{E(U_n) = n, \quad V(U_n) = 2n} \tag{2.44}$$

Tính chất 2.6. (a) 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.

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

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

$$\overline{X} = \frac{1}{n}\Big(X_1 + X_2 + \cdots + X_n\Big)$$

thì

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}\Big(X_i - \overline{X}\Big)^2 \sim \chi^2_{(n-1)}.$$

2.4.6 Phân phối Student

Định nghĩa 2.20 (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 $X \sim \mathcal{T}^n$, nếu hàm mật độ xác suất của X có dạng

$$\boxed{f(x) = \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}, \quad -\infty < x < +\infty} \tag{2.45}$$

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

$$\boxed{T_n = \frac{X}{\sqrt{\frac{Y}{n}}} \sim \mathcal{T}^n} \tag{2.46}$$

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

$$\boxed{E(T_n) = 0,\ n > 1, \quad V(T_n) = \frac{n}{n-2}, \quad n > 2} \tag{2.47}$$

Tính chất 2.7. Biến ngẫu nhiên T_n có phân phối xấp xỉ phân phối chuẩn tắc $\mathcal{N}(0,1)$ khi n đủ lớn.

Nhận xét 2.14. (a) 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.

(b) 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.