

Chương 5

PHÂN PHỐI XÁC SUẤT RỜI RẠC

5.1 Phân phối xác suất

Trong chương 1 ta đã biết về khái niệm biến là một đặc điểm hoặc một thuộc tính có thể cho các giá trị khác nhau. Biến được kí hiệu bằng các chữ cái in hoa như $A, B, \dots, X, Y, \dots$. Trong chương này ta khảo sát các biến gắn với các giá trị nào đó, khi các giá trị này thay đổi ta được các biến ngẫu nhiên. Ta xét ví dụ sau:

Gieo một con súc sắc một lần, gọi X là số chấm xuất hiện của con súc sắc. Khi đó X có thể nhận một trong những giá trị $\{1, 2, 3, 4, 5, 6\}$.

Vì khi ta chưa gieo con súc sắc thì chúng ta không thể biết trước được con súc sắc xuất hiện mặt bao nhiêu chấm, có nghĩa là ta chưa thể biết X nhận giá trị bao nhiêu, nên nó được gọi là biến ngẫu nhiên.

Biến ngẫu nhiên là biến mà giá trị của nó được xác định một cách ngẫu nhiên.

Cũng trong chương 1 chúng ta cũng đã phân loại biến ngẫu nhiên thành biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục. Ta nhắc lại:

Biến ngẫu nhiên rời rạc là biến ngẫu nhiên X mà các giá trị có thể của X là một tập hữu hạn hoặc vô hạn đếm được. Chẳng hạn, số cuộc gọi điện thoại đến một công ty trong một ngày hay là số sinh viên trong một trường đại học, ...

Biến ngẫu nhiên liên tục là biến ngẫu nhiên X mà tại các giá trị có thể của X là một hoặc một số khoảng trên trục số. Chẳng hạn, tốc độ của một xe ô tô hay nhiệt độ của một khu vực trong một ngày, ...

Trong chương này ta sẽ đi khảo sát các biến ngẫu nhiên rời rạc. Biến ngẫu nhiên liên tục sẽ được tìm hiểu trong chương 6.

Phân bố xác suất rời rạc bao gồm các giá trị của biến ngẫu nhiên có thể giả định và xác suất tương ứng của các giá trị đó. Các xác suất được xác định theo lý thuyết hoặc theo quan sát.

Phân phối xác suất có thể được thể hiện bằng cách sử dụng đồ thị hoặc một bảng cũng có thể được sử dụng bằng một công thức.

Ví dụ 5.1.1. Lập một phân phối xác suất về số mặt xuất hiện của một con súc sắc khi được gieo.

Lời giải:

Gọi X là số chấm xuất hiện của con súc sắc.

Vì không gian mẫu của phép thử là $\Omega = \{1, 2, 3, 4, 5, 6\}$ và xác suất của mỗi kết quả trong không gian mẫu đều bằng $\frac{1}{6}$ nên ta có bảng phân phối xác suất như sau:

[illegible]

Ví dụ 5.1.2. Giải bóng chày World Series được chơi giữa đội vô địch American League và đội vô địch National League. Đội đầu tiên giành chiến thắng 4 trận đấu sẽ là đội chiến thắng trong giải World Series. Nói cách khác có 7 trận đấu, đội giành được chiến thắng trong giải World Series là đội chiến thắng từ 4 đến 7 trận thắng. Dữ liệu dưới đây thể hiện số trận thắng trong World Series từ năm 1965 đến năm 2005 (không có giải World Series vào năm 1994). Với X là số lượng các trận thắng. Tìm xác suất $P(X)$ cho mỗi giá trị X , lập bảng phân bố xác suất và vẽ đồ thị cho dữ liệu.

X	Số các trận thi đấu
4	8
5	7
6	9
7	16
	40

Lời giải:

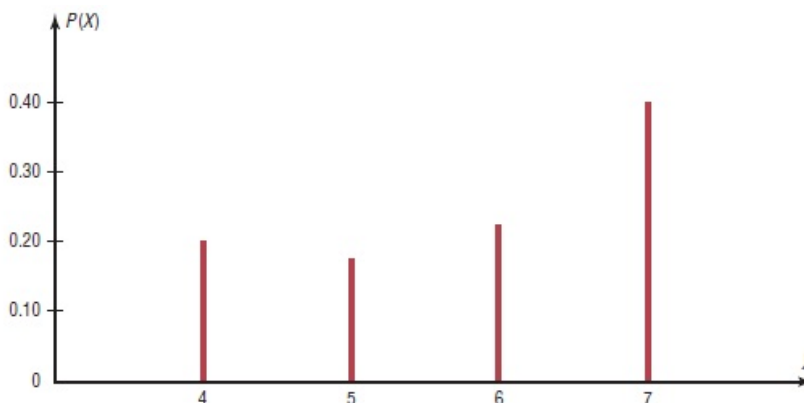
$$P(X = 4) = \frac{8}{40} = 0.200 \quad P(X = 6) = \frac{9}{40} = 0.225$$

$$P(X = 5) = \frac{7}{40} = 0.175 \quad P(X = 7) = \frac{16}{40} = 0.400$$

Bảng phân phối xác suất cho biến ngẫu nhiên X là:

X	4	5	6	7
$P(X)$	0.200	0.225	0.175	0.400

Biểu đồ cho phân bố xác suất:



Tính chất 5.1.1. Trong một phân phối xác suất ta luôn có:

- i) Tổng của các xác suất của tất cả các biến cố trong không gian mẫu phải bằng 1, tức là: $\sum P(X) = 1$.
- ii) Xác suất của mỗi biến cố trong không gian mẫu phải nằm trong đoạn từ 0 đến 1, tức là: $0 \leq P(X) \leq 1$.

Ví dụ 5.1.3. Xác định phân phối xác suất trong các câu sau?

a. X	0	5	10	15	20
$P(X)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$

c. X	1	2	3	4
$P(X)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{9}{16}$

b. X	0	2	4	6
$P(X)$	-1.0	1.5	0.3	0.2

d. X	2	3	7
$P(X)$	0.5	0.3	0.4

5.2 Trung bình, phương sai, độ lệch chuẩn, kỳ vọng của một phân phối xác suất

5.2.1 Trung bình của một phân phối xác suất

Như trong chương 3 ta đã biết được trung bình của một mẫu hay trung bình của một tổng thể được tính bằng cách lấy tổng tất cả các giá trị rồi chia cho số các giá trị, tức là:

$$\bar{X} = \frac{\sum X_i}{n}; \quad \mu = \frac{\sum X_i}{N}.$$

Nhưng làm thế nào để tính trung bình số chấm xuất hiện trên con súc sắc khi nó được gieo? Bạn có thể thử gieo con súc sắc 10 lần và ghi lại số chấm xuất hiện rồi tính trung bình. Tuy nhiên trung bình này chỉ là giá trị xấp xỉ với giá trị chính xác của trung bình. Trên thực tế nếu số lần gieo tăng lên thì sự xấp xỉ này càng chính xác hơn. Và câu hỏi đặt ra là muốn có được chính xác trung bình thì cần phải gieo con súc sắc bao nhiêu lần? Câu trả lời phải là gieo vô hạn lần. Nhưng việc này là không thể thực hiện được, do đó các công thức trước đây sẽ không được sử dụng bởi vì mẫu thức ở các công thức sẽ dẫn đến vô cùng. Do đó, một phương pháp mới để tính trung bình là cần thiết. Phương pháp mà nó có thể đưa ra giá trị lý thuyết chính xác của trung bình như thể con súc sắc có thể gieo vô hạn lần.

Định nghĩa 5.2.1. Trung bình của biến ngẫu nhiên X với phân phối xác suất rời rạc được xác định bởi:

$$\begin{aligned} \mu &= X_1.P(X_1) + X_2.P(X_2) + \dots + X_n.P(X_n) \\ &= \sum X.P(X) \end{aligned}$$

Ví dụ 5.2.1. Tìm số chấm trung bình xuất hiện trên mặt con súc sắc khi gieo một con súc sắc.

Lời giải: Gọi X là số chấm xuất hiện trên mặt con súc sắc. Khi đó ta có phân phối xác suất của X là:

X	1	2	3	4	5	6
$P(X)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Do đó số chấm trung bình xuất hiện trên mặt con súc sắc là:

$$\begin{aligned}\mu &= \sum X.P(X) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} \\ &= \frac{21}{6} = 3.5\end{aligned}$$

Ví dụ 5.2.2. Trong một gia đình có hai con, hãy tìm trung bình của số trẻ em sẽ là con gái.

Lời giải: Gọi X là số con gái trong một gia đình. Khi đó X có thể nhận các giá trị là $\{0, 1, 2\}$. Ta có bảng phân phối xác suất cho biến ngẫu nhiên X như sau:

X	0	1	2
$P(X)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Vậy trung bình số con gái trong một gia đình là:

$$\mu = \sum X.P(X) = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

5.2.2 Phương sai và độ lệch chuẩn của một phân phối xác suất

Trong chương 3 chúng ta đã biết, để đo độ phân tán của dữ liệu ta sử dụng các số đo phương sai và độ lệch chuẩn. Công thức của nó là:

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}; \quad \sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

Các công thức này không thể được sử dụng cho một biến ngẫu nhiên với phân bố xác suất vì N là vô hạn, vì vậy phương sai và độ lệch chuẩn phải được tính theo một công thức khác.

Định nghĩa 5.2.2. Phương sai của biến ngẫu nhiên X với phân phối xác suất được xác định:

$$\sigma^2 = \sum [X^2.P(X)] - \mu^2.$$

Độ lệch chuẩn của biến ngẫu nhiên X với phân phối xác suất được xác định:

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum [X^2.P(X)] - \mu^2}.$$

Ví dụ 5.2.3. Tìm phương sai và độ lệch chuẩn cho phân phối xác suất trong ví dụ 5.2.1.

Lời giải: Theo ví dụ 5.2.1 ta đã có $\mu = 3.5$. Khi đó, phương sai của phân phối xác suất là:

$$\sigma^2 = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6} - (3.5)^2 = 2.9$$

độ lệch chuẩn là:

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.9} = 1.7$$

Ví dụ 5.2.4. Một hộp chứa 5 quả bóng. Hai quả được đánh số 3, một quả được đánh số 4 và hai quả được đánh số 5. Các quả bóng được trộn lẫn vào nhau và chọn ngẫu nhiên ra một quả. Sau khi một quả được chọn, số của nó được ghi lại. Sau đó nó được hoàn lại trong hộp. Nếu phép thử được lặp lại nhiều lần, hãy tìm phương sai và độ lệch chuẩn của các con số trên quả bóng.

Lời giải: Gọi X là số của các quả bóng. Khi đó phân phối xác suất của X là:

X	3	4	5
$P(X)$	$\frac{2}{5}$	$\frac{1}{5}$	$\frac{2}{5}$

Trung bình là:

$$\mu = \sum X.P(X) = 3 \cdot \frac{2}{5} + 4 \cdot \frac{1}{5} + 5 \cdot \frac{2}{5} = 4.$$

Phương sai là:

$$\begin{aligned}\sigma^2 &= \sum [X^2.P(X)] - \mu^2 \\ &= 3^2 \cdot \frac{2}{5} + 4^2 \cdot \frac{1}{5} + 5^2 \cdot \frac{2}{5} - 4^2 = \frac{4}{5}.\end{aligned}$$

Độ lệch chuẩn là: $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{4}{5}} = 0.894$.

5.2.3 Kỳ vọng của một phân phối xác suất

Một khái niệm liên quan đến trung bình cho phân phối xác suất là giá trị kỳ vọng hoặc kỳ vọng. Giá trị kỳ vọng được sử dụng trong nhiều loại trò chơi may rủi, bảo hiểm, và trong các lĩnh vực khác.

Định nghĩa 5.2.3. Giá trị mong đợi của một biến ngẫu nhiên rời rạc của một phân bố xác suất là trung bình theo lý thuyết của biến ngẫu nhiên rời rạc đó. Ký hiệu là $E(X)$, được xác định bởi:

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

Ví dụ 5.2.5. Một nghìn vé được bán với giá 1 (đô la) mỗi vé, để nhận một chiếc tivi có giá 350 (đô la). Tìm giá trị kỳ vọng của lợi ích nếu bạn mua một vé?

Lời giải: Gọi X là lợi ích của việc mua một vé. Ta có:

	Chiến thắng	Thua
X	349	-1
$P(X)$	$\frac{1}{1000}$	$\frac{999}{1000}$

Khi đó giá trị kỳ vọng là:

$$E(X) = 349 \cdot \frac{1}{1000} + (-1) \cdot \frac{999}{1000} = -0.65.$$

Dấu trừ ở đây chứng tỏ là khi ta mua một vé ta sẽ bị lỗ 0.65 (đô la).

5.3 Phân phối nhị thức

Trong thực tế có nhiều loại vấn đề xác suất mà chỉ có hai kết quả hoặc có thể được giảm xuống còn hai kết quả. Chẳng hạn, khi tung một đồng xu thì nó có hai kết quả là sấp hoặc ngửa hay khi khi nghiên cứu về giới tính của một em bé chào đời thì nó sẽ nhận hai giá trị là nam hoặc nữ, ... Các tình huống như vậy được gọi là các phép thử nhị thức (*binomial experiments*).

Định nghĩa 5.3.1. Phép thử nhị thức là một dãy n phép thử xác suất thỏa mãn bốn yêu cầu sau:

- + Số các phép thử phải cố định, tức n không đổi.
- + Trong mỗi phép thử ta chỉ quan tâm đến sự xuất hiện của một biến cố A nào đó với xác suất $P(A) = p$ (hoặc A không xuất hiện với xác suất $q = 1 - p$).
- + Các kết quả của mỗi phép thử phải độc lập nhau.
- + Xác suất của biến cố A xảy ra trong mỗi phép thử là không đổi.

Định nghĩa 5.3.2. Các kết quả của phép thử nhị thức và xác suất tương ứng của các kết quả này được gọi là **phân phối nhị thức** (*binomial distribution*). Kí hiệu: $B(n, p)$.

Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện biến cố A , X sẽ nhận các giá trị $0, 1, 2, \dots, n$ và có công thức xác suất là:

$$P(X = x) = C_n^x \cdot p^x \cdot q^{n-x}.$$

Mệnh đề 5.3.1. Trung bình, phương sai và độ lệch chuẩn của một biến ngẫu nhiên X tuân theo phân bố nhị thức được xác định bằng các công thức sau:

$$\mu = n \cdot p \quad \sigma^2 = n \cdot p \cdot q \quad \sigma = \sqrt{n \cdot p \cdot q}.$$

Ví dụ 5.3.1. Trong các câu dưới đây, câu nào là phép thử nhị thức?

- a. Khảo sát 100 người để xác định liệu họ có thích xà phòng Sudsy.
- b. Tung một đồng xu 100 lần để xem có bao nhiêu lần xuất hiện mặt ngửa.
- c. Rút một con bài từ bộ bài 52 con và xem có nhận được hay không con bài cơ.
- d. Hỏi 1000 người về nhãn hiệu thuốc là mà họ hút.
- e. Thử nghiệm bốn nhãn hiệu khác nhau của loại thuốc "aspirin" để xem thương hiệu nào có hiệu quả.
- f. Thử nghiệm một nhãn hiệu thuốc "aspirin" bằng cách sử dụng 10 người để xác định liệu có hiệu quả hay không.

Lời giải:

Phép thử nhị thức gồm các câu: a, b, c, f.

Ví dụ 5.3.2. Một tổ chức báo cáo rằng có 5% người Mỹ sợ ở một mình trong nhà vào ban đêm. Chọn ngẫu nhiên một mẫu 20 người Mỹ, hãy tìm xác suất để:

- Có chính xác 5 người trong mẫu chọn sợ ở một mình vào ban đêm.
- Có tối đa 3 người trong mẫu chọn sợ ở một mình vào ban đêm.
- Có ít nhất 3 người trong mẫu chọn sợ ở một mình vào ban đêm.

Lời giải: Gọi X là biến ngẫu nhiên số người trong mẫu chọn sợ ở một mình vào ban đêm. Khi đó X có phân phối nhị thức.

- Theo phép thử nhị thức, xác suất để có chính xác 5 người trong mẫu chọn sợ ở một mình vào ban đêm là:

$$P(X = 5) = C_{20}^5 \cdot (0.05)^5 \cdot (1 - 0.05)^{15} \approx 0.002.$$

- Theo phép thử nhị thức, xác suất để có tối đa 3 người trong mẫu chọn sợ ở một mình vào ban đêm là:

$$\begin{aligned} P(X \leq 3) &= P(0) + P(1) + P(2) + P(3) \\ &= C_{20}^0 \cdot (0.05)^0 \cdot (0.95)^{20} + C_{20}^1 \cdot (0.05)^1 \cdot (0.95)^{19} + C_{20}^2 \cdot (0.05)^2 \cdot (0.95)^{18} \\ &\quad + C_{20}^3 \cdot (0.05)^3 \cdot (0.95)^{17} \\ &= 0.358 + 0.377 + 0.189 + 0.060 \\ &= 0.984. \end{aligned}$$

- Theo phép thử nhị thức, xác suất để có ít nhất 3 người trong mẫu chọn sợ ở một mình vào ban đêm là:

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - P(0) - P(1) - P(2) \\ &= 1 - C_{20}^0 \cdot (0.05)^0 \cdot (0.95)^{20} - C_{20}^1 \cdot (0.05)^1 \cdot (0.95)^{19} - C_{20}^2 \cdot (0.05)^2 \cdot (0.95)^{18} \\ &= 0.076. \end{aligned}$$

Ví dụ 5.3.3. Tung một đồng tiền 4 lần. Tìm trung bình, phương sai và độ lệch chuẩn của số lần xuất hiện mặt ngửa.

Lời giải: Gọi X là biến ngẫu nhiên chỉ số lần xuất hiện mặt ngửa của đồng tiền trong 4 lần tung. Khi đó, X có phân phối nhị thức.

Theo công thức của phân phối nhị thức với $n = 4, p = \frac{1}{2}, q = \frac{1}{2}$, ta có:

$$\begin{aligned} \mu &= n \cdot p = 4 \cdot \frac{1}{2} = 2; \\ \sigma^2 &= n \cdot p \cdot q = 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = 1; \\ \sigma &= \sqrt{1} = 1. \end{aligned}$$

5.4 Các loại phân phối khác

5.4.1 Phân phối đa thức

Như ta đã biết phân phối nhị thức sử dụng cho phép thử nhị thức, tức là mỗi lần thử chỉ quan tâm đến hai kết quả là sự xuất hiện hay không xuất hiện của một biến

cổ. Trong trường hợp có nhiều hơn 2 kết quả trong mỗi phép thử thì ta không thể dùng phân phối nhị thức được nữa mà ta sẽ tổng quát nó lên và được gọi là **phân phối đa thức** (*multinomial distribution*).

Định nghĩa 5.4.1. Giả sử X bao gồm các biến cố $E_1, E_2, \dots, E_k$ có xác suất xảy ra tương ứng là $p_1, p_2, \dots, p_k$ và X_1 là số lần xuất hiện E_1 , X_2 là số lần xuất hiện E_2 , ..., X_k là số lần xuất hiện E_k . Khi đó xác suất để X xảy ra là:

$$P(X) = \frac{n!}{X_1! \cdot X_2! \cdot \dots \cdot X_k!} \cdot p_1^{X_1} \cdot p_2^{X_2} \cdot \dots \cdot p_k^{X_k}.$$

Ví dụ 5.4.1. Trong một thành phố lớn, có 50% người chọn xem phim, 30% chọn chơi thể thao và 20% chọn mua sắm như một hoạt động giải trí. Chọn ngẫu nhiên một mẫu 5 người từ thành phố đó. Hãy tìm xác suất mà 3 người lên kế hoạch đi xem phim, 1 người chơi thể thao và 1 người đến trung tâm mua sắm.

Lời giải: Gọi các biến cố E_1, E_2, E_3 lần lượt là biến cố người chọn xem phim, người chọn chơi thể thao và người chọn mua sắm. Ta có $n = 5, X_1 = 3, X_2 = 1, X_3 = 1, p_1 = 0.5, p_2 = 0.3, p_3 = 0.2$.

Áp dụng công thức của phân phối đa thức ta được:

$$P(X) = \frac{5!}{3! \cdot 1! \cdot 1!} \cdot 0.5^3 \cdot 0.3^1 \cdot 0.2^1 = 0.15.$$

5.4.2 Phân phối Poisson

Với phân phối nhị thức khi mà số lần thử n đủ lớn, p đủ nhỏ và các biến cố xuất hiện độc lập trong một khoảng thời gian nhất định thì một phân phối xác suất hữu ích được sử dụng đó là phân phối Poisson. Sự khác biệt của phân phối Poisson với phân phối nhị thức là ở chỗ, phân phối nhị thức thông tin cho biết số lần mà một sự kiện nào đó xảy ra trong n lần thử, còn với phân phối Poisson thông tin được cho là trung bình số lần xảy ra thành công của một sự kiện trong một khoảng thời gian nhất định.

Định nghĩa 5.4.2. Gọi X là biến ngẫu nhiên chỉ số lần một biến cố A xuất hiện trên một miền hay một vùng nào đó trong một khoảng thời gian định trước, biết số lần A xuất hiện trung bình trong khoảng thời gian đó là λ . Khi đó ta có công thức tính xác suất:

$$P(X; \lambda) = \frac{e^{-\lambda} \lambda^X}{X!}.$$

Ví dụ 5.4.2. Giả sử có 200 lỗi đánh máy được phân phối ngẫu nhiên trong một bản thảo 500 trang. Tìm xác suất mà một trang nhất định chứa chính xác 3 lỗi đánh máy.

Lời giải: Gọi X là biến ngẫu nhiên chỉ số lỗi đánh máy trong một trang giấy. Ta có số lỗi đánh máy trung bình trên một trang giấy là $\lambda = \frac{200}{500} = 0.4$.

Áp dụng công thức của phân phối Poisson với $X = 3$ ta được:

$$P(X; \lambda) = \frac{e^{-\lambda} \lambda^X}{X!} = \frac{e^{-0.4} (0.4)^3}{3!} \approx 0.0072.$$

Ví dụ 5.4.3. Một công ty bán hàng nhận được trung bình 3 cuộc gọi mỗi giờ trên số điện thoại miễn phí.

- a. Tìm xác suất để trong một giờ có tối đa 3 cuộc gọi.
- b. Tìm xác suất để trong một giờ có ít nhất 3 cuộc gọi.
- b. Tìm xác suất để trong một giờ có từ 5 cuộc gọi trở lên.

Ví dụ 5.4.4. Trong một căn phòng có 200 người. Giả sử có trung bình 2% số người thuận tay trái. Tìm xác suất để có chính xác 5 người thuận tay trái.

5.4.3 Phân phối siêu bội (hypergeometric distribution)

Trong thực tế nhiều trường hợp khi lấy mẫu nó được thực hiện mà không được hoàn lại khi đó phân phối nhị thức sẽ không được sử dụng vì các phép thử không độc lập nhau. Lúc đó ta khái quát hóa bằng cách sử dụng một phân bố xác suất đặc biệt gọi là **phân phối siêu bội** (*hypergeometric distribution*). Phân phối siêu bội là sự phân bố của một biến có hai kết quả khi lấy mẫu được thực hiện mà không có sự hoàn lại.

Định nghĩa 5.4.3. Từ một tập hợp gồm N phần tử (trong đó có a phần tử có tính chất A) lấy ngẫu nhiên không hoàn lại một mẫu n phần tử. Gọi X là biến ngẫu nhiên chỉ số phần tử có tính chất A có trong n phần tử của mẫu lấy ra. Khi đó xác suất của X được tính theo công thức:

$$P(X) = \frac{C_a^X \cdot C_{N-a}^{n-X}}{C_N^n}.$$

Ví dụ 5.4.5. Mười người xin việc làm trợ lý quản lý của một nhà hàng. Năm người đã tốt nghiệp đại học và năm người thì không. Giả sử người quản lý chọn ngẫu nhiên 3 người. Hãy tìm xác suất để cả 3 người đều đã tốt nghiệp đại học.

Lời giải: Gọi X là biến ngẫu nhiên chỉ số người đã tốt nghiệp đại học. Khi đó X tuân theo luật phân phối siêu bội. Áp dụng công thức xác suất của phân phối siêu bội, với $N = 10, a = 5, n = 3, X = 3$, ta có:

$$P(X = 3) = \frac{C_5^3 \cdot C_5^0}{C_{10}^3} = \frac{1}{12}.$$

Ví dụ 5.4.6. Một nghiên cứu gần đây cho thấy 2 trong số 10 căn nhà trong khu phố không có bảo hiểm. Nếu 5 ngôi nhà được chọn ngẫu nhiên từ 10 căn nhà trong khu phố đó thì hãy tìm xác suất có chính xác 1 căn nhà sẽ không có bảo hiểm.

Ví dụ 5.4.7. Một lô hàng có 12 sản phẩm, trong đó có 3 phế phẩm. Lấy ngẫu nhiên 3 sản phẩm để kiểm tra. Biết rằng lô hàng sẽ bị từ chối nếu có phế phẩm. Tính xác suất để lô hàng bị từ chối.

Phần bài tập chương 5

1. Hãy xác định phân phối nào biểu diễn một phân phối xác suất? Nếu không thì tại sao lại không phải?

a.

X	1	2	3	4	5
$P(X)$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{1}{10}$	$\frac{2}{10}$	$\frac{3}{10}$

b.

X	10	20	30
$P(X)$	0.1	0.4	0.3

c.

X	8	12	16	20
$P(X)$	$\frac{5}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$

2. Số cuộc gọi khẩn cấp vào một sở cảnh sát đại phương trong vòng 24 giờ được phân phối như sau. Xây dựng một đồ thị cho dữ liệu.

Số cuộc gọi	10	11	12	13	14
Xác suất $P(X)$	0.02	0.12	0.40	0.31	0.15

3. Dưới đây là phân phối số lượng đơn xin thẻ tín dụng nhận được cho mỗi nhân viên trong 8 giờ đồng hồ.

X	0	1	2	3	4	5
$P(X)$	0.27	0.28	0.20	0.15	0.08	0.02

- a. Tính xác suất mà nhân viên sẽ nhận được từ 2 đến 3 đơn xin thẻ tín dụng?
b. Tìm trung bình, phương sai và độ lệch chuẩn cho phân phối xác suất này.

4. Một ngân hàng có dịch vụ "drive-through" (là một hình thức dịch vụ mua hàng mà khách hàng có thể mua hàng trực tiếp ngay trên xe của mình). Số lượng khách hàng đến mua trong khoảng thời gian 15 phút được cho trong phân phối sau. Tìm trung bình, phương sai và độ lệch chuẩn cho phân phối.

Số khách hàng X	0	1	2	3	4
Xác suất $P(X)$	0.12	0.20	0.31	0.25	0.12

5. Tại một bảo tàng cộng đồng nhỏ, số lượng thăm quan trong ngày có phân phối được cho như sau. Tìm trung bình, phương sai và độ lệch chuẩn của phân phối.

Số khách thăm quan X	13	14	15	16	17
Xác suất $P(X)$	0.12	0.15	0.29	0.25	0.19

6. Trong đợt bán sơn gần đây nhất tại một cửa hàng sơn, số lượng lon sơn được bán ra được phân phối như sau. Tìm trung bình, phương sai và độ lệch chuẩn của phân phối.

Số lon sơn bán ra X	1	2	3	4	5
Xác suất	0.42	0.27	0.15	0.10	0.06

7. Một nhà sản xuất dự định tổ chức một cuộc đua thuyền ngoài trời vào ngày 3 tháng 5. Chi phí cho cuộc đua thuyền là 8,000 đô (bao gồm chi phí cho quảng cáo, an ninh, vé,...). Nhà sản xuất có kế hoạch kiếm được 15,000 đô lợi nhuận nếu mọi việc suôn sẻ. Tuy nhiên, nếu trời mưa thì cuộc đua thuyền sẽ phải hủy. Theo dự báo thời tiết xác suất trời mưa là 0.3. Tìm lợi nhuận mong muốn của nhà sản xuất.
8. Một trò chơi được thiết kế như sau: tất cả các con bài cơ được lấy ra khỏi một bộ bài và 13 con bài này được đặt vào một cái túi. Các thẻ này được trộn lẫn và sau đó một thẻ được chọn ngẫu nhiên (và sau đó được hoàn lại vào túi). Người chơi thắng theo các quy tắc sau:
Nếu con bài A được rút ra thì người chơi mất 20 đô.
Nếu một con bài hình được rút ra thì người chơi sẽ thắng 10 đô.
Nếu bạn rút được một thẻ khác (từ con 2 đến con 10) thì người chơi sẽ thắng 2 đô.
Bạn phải trả bao nhiêu tiền để chơi trò chơi này để hòa tiền?
9. Cho X là biến ngẫu nhiên có phân phối nhị thức với $n = 12$ và $p = 0.3$. Tìm:
- $P(X = 8)$.
 - $P(X < 5)$.
 - $P(X \geq 10)$.
 - $P(4 < X \leq 9)$.
10. Gieo một con súc sắc 360 lần. Tìm trung bình, phương sai và độ lệch chuẩn xuất hiện mặt chấm 4 trên con súc sắc.
11. Một báo cáo từ Bộ y tế và dịch vụ nhân sinh cho biết 70% số người tử vong do tai nạn giao thông vào ban đêm của những ngày cuối tuần có liên quan đến việc say rượu. Chọn ngẫu nhiên một mẫu 15 người tử vong do tai nạn giao thông xảy ra vào ban đêm của những ngày cuối tuần. Hãy tìm xác suất để:
- Có chính xác 12 người tử vong liên quan đến việc say rượu.
 - Có ít tối đa 14 người tử vong liên quan đến việc say rượu.
 - Có từ 2 đến 4 người tử vong liên quan đến việc say rượu.
12. Một thống kê cho biết rằng, có 14% người sử dụng điện thoại để truy cập internet. Trong một mẫu ngẫu nhiên gồm 10 người sử dụng điện thoại, xác suất để có 2 người sử dụng điện thoại để truy cập internet là bao nhiêu? Hơn 2 người sử dụng điện thoại để truy cập internet là bao nhiêu?
13. Giả sử có 75% sinh viên điều dưỡng có thể vượt qua bài kiểm tra về tính lượng thuốc. Tìm trung bình, phương sai và độ lệch chuẩn của số sinh viên vượt qua bài kiểm tra trong một mẫu gồm 180 sinh viên điều dưỡng.
14. Đã có một nghiên cứu đưa ra rằng, có 63% người từ 65 tuổi trở lên bị cảm cúm vào năm ngoái. Trong một mẫu ngẫu nhiên 300 người từ 65 tuổi trở lên, hãy tìm trung bình, phương sai và độ lệch chuẩn cho số người bị cảm cúm.

15. Cơ hội mà một cảnh sát trưởng Hoa Kỳ tin rằng án tử hình "làm giảm đáng kể số vụ giết người" là $1/4$. Nếu chọn một mẫu ngẫu nhiên 8 cảnh sát trưởng Hoa Kỳ thì hãy tìm xác suất mà có tối đa 3 người tin rằng án tử hình làm giảm đáng kể án số vụ giết người.
16. Tạp chí Năng Lượng Hoa Kỳ báo cáo rằng, có 27% số hộ gia đình Mỹ đốt gỗ. Nếu một mẫu ngẫu nhiên 500 hộ gia đình Mỹ được lựa chọn thì hãy tìm trung bình, phương sai và độ lệch chuẩn của số gia đình đốt gỗ.
17. Một ngàn vé được bán với giá 1 (đô la) cho bốn giải thưởng là 100 (đô la), 50 (đô la), 25 (đô la) và 10 (đô la). Sau mỗi lần mở thưởng, vé thắng sau đó được bỏ lại vào thùng vé. Tìm giá trị kỳ vọng nếu bạn mua hai vé?
18. Một cố vấn tài chính cho thấy khách hàng của mình chọn một trong hai loại trái phiếu để đầu tư 5000 (đô la). Trái phiếu X có tỉ lệ hoàn vốn là 4% và có tỷ lệ vỡ nợ là 2%. Trái phiếu Y có tỷ lệ hoàn vốn là 2,5% và tỷ lệ vỡ nợ là 1%. Tìm kỳ vọng của tỷ suất lợi nhuận và quyết định xem trái phiếu nào sẽ là một khoản đầu tư tốt hơn. Biết rằng khi trái phiếu bị phá sản thì chủ đầu tư mất toàn bộ khoản đầu tư.
19. Có 3 trong 4 người Mỹ dưới 35 tuổi đã ăn pizza cho bữa ăn sáng. Nếu một mẫu ngẫu nhiên 20 người dưới 35 tuổi được lựa chọn thì hãy tìm xác suất để có 16 người đã ăn bánh pizza cho bữa ăn sáng.
20. Theo các tài liệu điều tra, có 75.4% phụ nữ trong độ tuổi 20 – 24 chưa bao giờ lập gia đình. Một mẫu ngẫu nhiên 250 phụ nữ trong độ tuổi 20 – 24 thì hãy tìm trung bình, phương sai và độ lệch chuẩn cho số người trong độ tuổi 20 – 24 đã kết hôn.
21. Sau một cuộc bầu cử quốc gia gần đây, các cử tri được hỏi về mức độ tin tưởng như thế nào về việc đếm chính xác các phiếu bầu trong các tiểu bang. Kết quả thu được như sau:
46% rất tin tưởng
41% khá tin tưởng
9% không tin tưởng
3% hoàn toàn không tin tưởng.
Nếu 10 cử tri được lựa chọn một cách ngẫu nhiên thì hãy tìm xác suất để có 5 người rất tin tưởng, 3 người khá tin tưởng, 1 người không tin tưởng và 1 người hoàn toàn không tin tưởng.
22. Trong một cửa hàng âm nhạc, người quản lý thấy rằng xác suất mà một người không mua đĩa CD là 0.3, mua 1 đĩa CD là 0.6, mua từ 2 đĩa CD trở lên là 0.1. Giả sử có 6 người vào cửa hàng, tìm xác suất mà có 1 người không mua đĩa CD, 3 người mua 1 đĩa CD và 2 người mua từ 2 đĩa CD trở lên.
23. Một hộp chứa 4 quả bóng trắng, 3 quả bóng đỏ và 3 quả bóng màu xanh. Một quả bóng được chọn ngẫu nhiên và màu của nó được ghi lại. Tìm xác suất nếu 5 quả bóng được chọn (chọn có hoàn lại) trong đó có 2 quả bóng màu trắng, 2 màu đỏ và 1 màu xanh.

24. Trước khi máy ghi âm rời khỏi nhà máy sản xuất nó sẽ được kiểm tra chất lượng. Xác suất mà một máy ghi âm có 0, 1 hoặc 2 lỗi kỹ thuật tương ứng là 0.90, 0.06 và 0.04. Trong một mẫu ngẫu nhiên gồm 12 máy ghi âm, tìm xác suất để có 8 máy 0 có lỗi, 3 máy có 1 lỗi và 1 máy có 2 lỗi.
25. Cơ quan vận tải hàng không thống kê rằng có 8.25 trong số 1000 hành khách hàng không bị mất hành lý trong tất cả các chuyến bay năm ngoái. Nếu chọn ngẫu nhiên một mẫu 400 hành khách hàng không thì xác suất có 5 hành khách bị mất hành lý là bao nhiêu?
26. Cho biến ngẫu nhiên X xác định bởi phân phối xác suất như sau. Hãy tìm trung bình cho biến ngẫu nhiên X ?

X	0	1	2	3	4
$P(X)$	0.06	0.70	0.20	0.03	0.01

27. Một xạ thủ có 3 viên đạn. Anh ta bắn từng phát cho tới khi trúng mục tiêu hoặc hết cả 3 viên đạn thì thôi. Hãy lập phân phối xác suất cho số viên đạn đã bắn và tính trung bình số viên đạn đã bắn. Biết rằng xác suất trúng đích của mỗi phát là 0.8.
28. Đường dây nóng của một công ty tư vấn về vấn đề máy tính nhận trung bình 6 cuộc gọi yêu cầu mỗi giờ. Dữ liệu có phân phối Poisson. Chọn ngẫu nhiên một giờ bất kỳ, hãy tìm xác suất mà công ty sẽ nhận được:
- ít nhất 6 cuộc gọi.
 - từ 4 cuộc gọi trở lên.
 - tối đa 5 cuộc gọi.
29. Số vụ tai nạn chèo thuyền trên hồ Emilie có phân phối Poisson. Xác suất xảy ra một vụ tai nạn là 0.003. Nếu có 1000 chiếc thuyền trên hồ trong một tháng hè, thì hãy tìm xác suất xảy ra 6 vụ tai nạn.
30. Nếu 5 thẻ được rút ra từ một bộ bài gồm 52 cây, thì hãy tìm xác suất mà có 2 thẻ chất cơ được rút ra.
31. Trong số 50 chiếc xe ô tô có 10 chiếc màu trắng. Chọn ngẫu nhiên 5 chiếc để bán đấu giá tại một cuộc bán đấu giá. Hãy tìm xác suất chọn ra được chính xác 2 chiếc xe màu trắng.
32. Tại một cửa hàng thực phẩm, các mặt hàng được tặng bao gồm 10 lon canh, 10 lon rau và 8 lon trái cây. Nếu 3 lon được chọn một cách ngẫu nhiên, thì hãy tìm xác suất để nhận được một lon rau và 2 lon trái cây.

Chương 6

PHÂN PHỐI CHUẨN

Các nhà nghiên cứu y khoa đã xác định được cái gọi là khoảng bình thường (*normal interval*) cho huyết áp, cholesterol, triglyceride và những thứ tương tự. Ví dụ huyết áp tâm thu bình thường là từ 110 đến 140. Khoảng bình thường đối với triglycerides của một người là từ 30 đến 200 miligam mỗi dichilitry (mg/dl). Bằng cách đo các biến số này, bác sĩ có thể xác định liệu số liệu thống kê quan trọng của bệnh nhân nằm trong khoảng bình thường hay liệu cần điều trị để điều chỉnh tình trạng và tránh bệnh tật trong tương lai. Vì vậy một câu hỏi đặt ra là làm thế nào để xác định cái gọi là khoảng bình thường?

Trong chương này, sẽ trình bày cách mà các nhà nghiên cứu xác định khoảng bình thường cho các xét nghiệm y khoa cụ thể bằng cách sử dụng một phân bố chuẩn (*normal distribution*). Đồng thời cũng trình bày các phương pháp tương tự để xác định tuổi thọ của pin, sức mạnh của dây thừng và nhiều tính năng khác được nhắc đến.

6.1 Phân phối chuẩn

Định nghĩa 6.1.1. Biến ngẫu nhiên X được gọi là tuân theo luật phân phối chuẩn, kí hiệu $N(\mu, \sigma)$, nếu hàm mật độ xác suất của X có dạng:

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

trong đó:

x là các giá trị của X

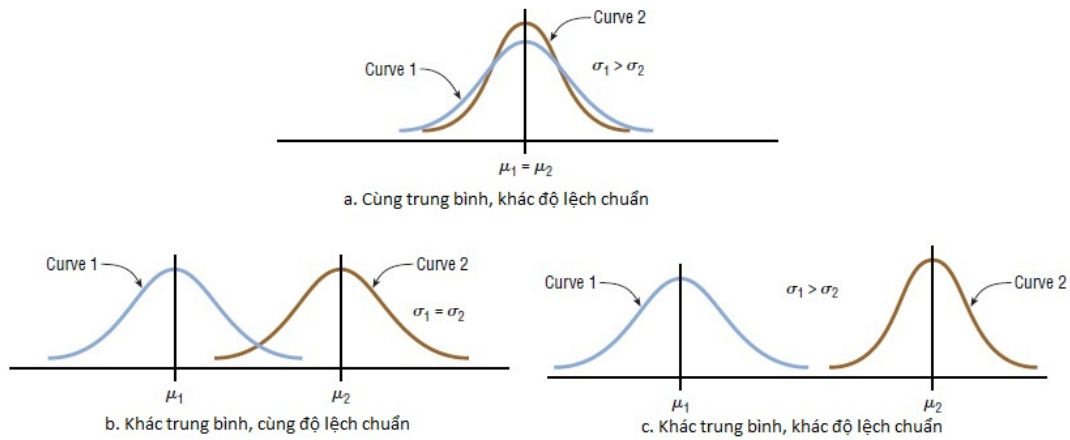
$e \approx 2.718$

$\pi \approx 3.14$

μ trung bình của tổng thể

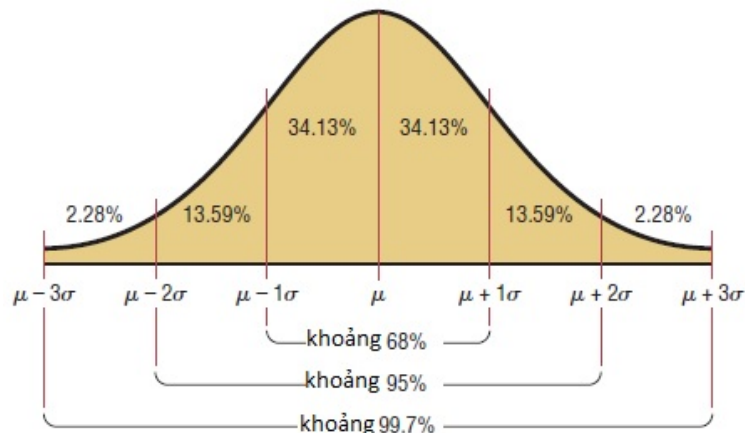
σ độ lệch chuẩn của tổng thể.

Các dạng đồ thị của phương trình đường phân phối chuẩn:



Tính chất 6.1.1. Một phân phối chuẩn có các tính chất sau:

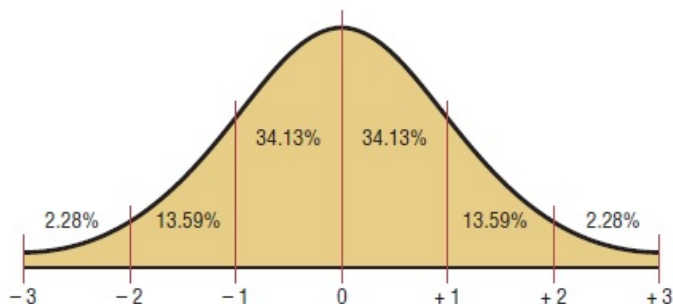
1. Đồ thị của phân phối chuẩn có dạng hình chuông.
2. Trung bình, trung vị và mode đều bằng nhau và nằm ở trung tâm của phân bố.
3. Phân phối chuẩn là một đường cong chỉ có một mode duy nhất.
4. Đồ thị là một đường cong đối xứng qua đường thẳng đứng đi qua trung bình.
5. Là một đường cong liên tục.
6. Là đường cong nhận trục Ox làm tiệm cận ngang.
7. Tổng diện tích phần phía dưới đường cong phân phối chuẩn luôn bằng 1.
8. Diện tích nằm dưới đường cong phân phối chuẩn nằm trong khoảng 1 độ lệch chuẩn là xấp xỉ 0.68; trong khoảng 2 độ lệch chuẩn khoảng 0.95 và trong khoảng 3 độ lệch chuẩn khoảng 0.997. Xem hình 6.1.



Hình 6.1: Diện tích dưới đường phân phối chuẩn.

Phân phối chuẩn tắc (*standard normal distribution*)

Định nghĩa 6.1.2. Phân phối chuẩn tắc là phân phối chuẩn với trung bình bằng 0 và độ lệch chuẩn bằng 1.



Khi đó hàm số cho phân phối chuẩn tắc là:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

Tất cả các biến phân phối chuẩn có thể được biến đổi thành biến phân phối chuẩn tắc bằng cách sử dụng công thức cho điểm chuẩn (*z-core*):

$$z = \frac{X - \mu}{\sigma}.$$

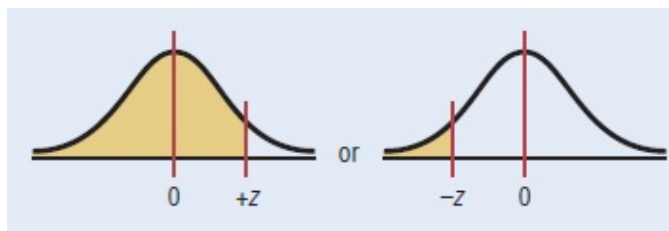
Tìm diện tích dưới đường phân phối chuẩn tắc

Để giải quyết một số vấn đề khi sử dụng phân phối chuẩn tắc ta cần biết cách tra bảng phân phối chuẩn tắc theo quy trình sau:

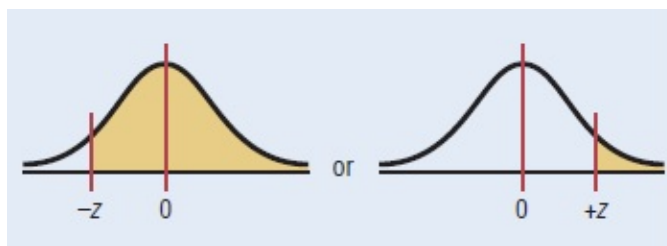
- Bước 1: Vẽ đường cong phân phối chuẩn tắc và bôi đậm phần diện tích cần tìm.
- Bước 2: Tìm số thích hợp trong bảng và làm theo hướng dẫn sau:

Có ba vấn đề cơ bản khi tra bảng:

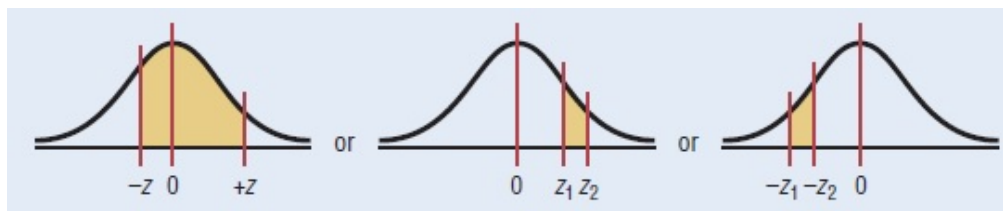
- (a) Diện tích ở bên trái của một giá trị z bất kì: Khi đó tra giá trị z trong bảng và sử dụng vùng đã cho trong bảng.



- (b) Diện tích ở bên phải của một giá trị z bất kì: Khi đó tra giá trị z và lấy 1 trừ cho diện tích vừa tìm được.



(c) Diện tích nằm giữa hai giá trị z bất kì: Khi đó tra cả hai giá trị z và lấy hiệu tương ứng hai phần đó.



Phụ lục cho diện tích phần nằm dưới đường phân phối chuẩn ở bên trái của bất kỳ giá trị z với hai chữ số thập phân.

Ví dụ 6.1.1. Tìm diện tích phần bên trái với $z = 1.99$.

Ví dụ 6.1.2. Tìm diện tích phần giữa giá trị $z = 1.68$ và $z = -1.37$.

Đường phân phối chuẩn như một đường cong phân phối xác suất

Một đường phân phối chuẩn có thể được sử dụng như một đường phân phối xác suất cho các biến phân phối chuẩn. Diện tích phần dưới đường phân phối chuẩn tắc cũng có thể được xem là xác suất. Ta dùng kí hiệu $P(a < z < b)$ để chỉ xác suất của giá trị z bất kì từ a đến b .

Chú ý 1. Với biến liên tục thì xác suất tại một giá trị z luôn bằng 0. Do đó $P(a \leq z \leq b) = P(a < z < b)$.

Ví dụ 6.1.3. Tìm xác suất trong các trường hợp sau:

- $P(0 < z < 2.32)$.
- $P(z < 1.65)$.
- $P(z > 1.91)$.

Ví dụ 6.1.4. Tìm giá trị z sao cho diện tích phần nằm dưới đường phân phối chuẩn tắc giữa giá trị 0 và z bằng 0.2123.

6.2 Các ứng dụng của phân phối chuẩn

Đường cong phân phối chuẩn tắc có thể được sử dụng để giải quyết một loạt các vấn đề thực tế. Yêu cầu duy nhất là biến phải tuân theo phân phối chuẩn hoặc tiệm

cận phân phối chuẩn. Một số bài toán kiểm định toán học đặt ra là xác định liệu một biến có được phân phối chuẩn hay không. Đối với tất cả các vấn đề được trình bày trong chương này, các biến được giả định là phân phối chuẩn hoặc tiệm cận phân phối chuẩn.

Để giải quyết các vấn đề bằng cách sử dụng phân phối chuẩn tắc thì ta biến đổi biến ban đầu thành một biến phân bố chuẩn hóa bằng cách sử dụng công thức:

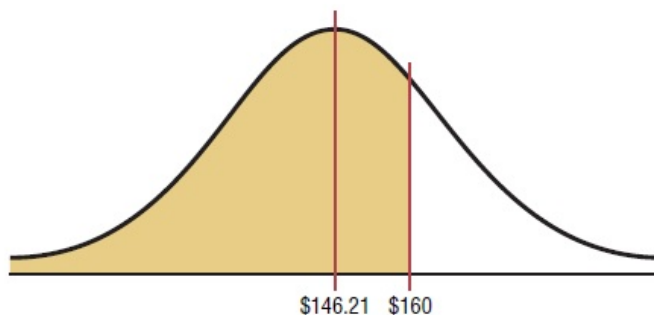
$$z = \frac{X - \mu}{\sigma}.$$

Đây là công thức để chuyển đổi một biến bất kì thành biến có phân phối chuẩn. Chẳng hạn, giả sử điểm số của một bài kiểm tra chuẩn.

Ví dụ 6.2.1. Một cuộc điều tra của Liên đoàn bán lẻ Quốc gia cho thấy phụ nữ chi trung bình 146.21 đô la cho những ngày lễ Giáng sinh. Giả sử độ lệch chuẩn là 29.44 đô la. Tìm số phần trăm phụ nữ chi tiêu dưới 160 đô la. Giả sử biến có phân phối chuẩn.

Lời giải:

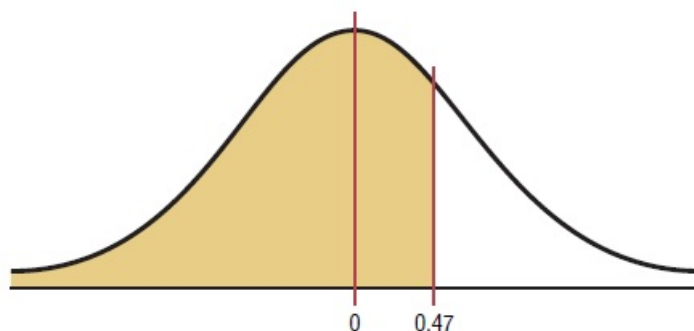
- Bước 1: Vẽ hình và mô tả diện tích của bài toán như hình vẽ.



- Bước 2: Tìm giá trị z tương ứng với 160 đô la.

$$z = \frac{X - \mu}{\sigma} = \frac{160 - 146.21}{29.44} = 0.47$$

Do đó, ta có phân phối tương ứng với z là:



- Bước 3: Tìm diện tích bằng cách sử dụng bảng phụ lục ta có diện tích phần bên trái dưới đường phân phối z ứng với $z = 0.47$ là 0.6808.

Vậy, có 0.6808 hay 68.08% số phụ nữ chi tiêu ít hơn 160 đô la vào mùa Giáng sinh.

Ví dụ 6.2.2. Mỗi tháng một gia đình người Mỹ thải ra trung bình 28 (pound) giấy loại. Giả sử độ lệch chuẩn là 2 (pound). Nếu một hộ gia đình được lựa chọn một cách ngẫu nhiên, hãy tìm xác suất để:

- gia đình đó thải ra từ 27 đến 31 (pound) giấy loại mỗi tháng,
- gia đình đó thải ra hơn 30.2 (pound) mỗi tháng.

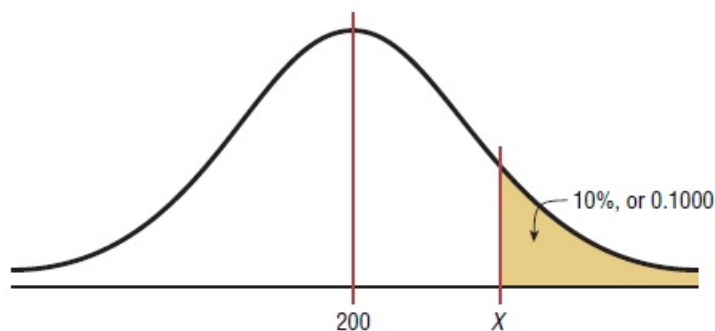
Ví dụ 6.2.3. Hiệp hội Ô tô Hoa Kỳ báo cáo rằng thời gian trung bình để thực hiện cuộc gọi khẩn cấp là 25 phút. Giả sử rằng biến là tiệm cận phân phối chuẩn và độ lệch chuẩn là 4.5 phút. Nếu 80 cuộc gọi được chọn ngẫu nhiên thì có khoảng bao nhiêu sẽ trả lời điện thoại ít hơn 15 phút?

6.2.1 Tìm giá trị dữ liệu được cho bởi xác suất cụ thể

Một phân phối chuẩn cũng có thể được sử dụng để tìm các giá trị dữ liệu cụ thể cho các tỷ lệ phần trăm nhất định. Ứng dụng này được hiển thị trong ví dụ 6.2.4.

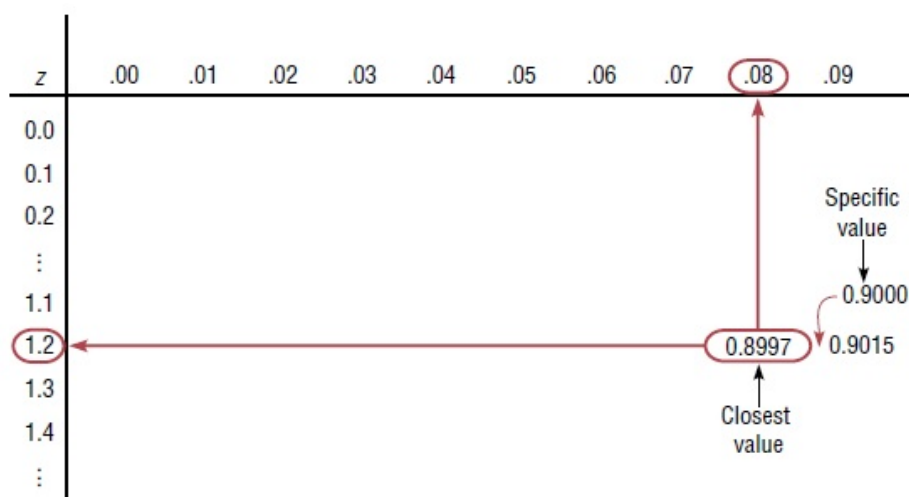
Ví dụ 6.2.4. Để đủ tiêu chuẩn cho một học viện cảnh sát, thí sinh phải ghi được 10% trên tổng số kiểm tra về khả năng. Kiểm tra này có trung bình là 200 và độ lệch chuẩn là 20. Tìm điểm số thấp nhất có thể để đủ điều kiện. Giả sử các điểm kiểm tra có phân phối chuẩn.

Lời giải: Theo giả thuyết thì phần diện tích cần tìm được thể hiện ở hình sau:



Để giải quyết bài toán này ta thực hiện các bước sau:

- Bước 1: Tìm diện tích phần bên trái theo phân phối chuẩn ứng với X là: $1.0000 - 0.1000 = 0.9000$.
- Bước 2: Tra bảng phụ lục để tìm giá trị của z tương ứng với diện tích 0.9000. Nếu không tìm được giá trị cụ thể thì hãy tìm giá trị gần nhất, trong trường hợp này là 0.8997 như trong hình. Như vậy giá trị z tương ứng là 1.28.



- Bước 3: Sử dụng công thức:

$$z = \frac{X - \mu}{\sigma}$$

$$\Leftrightarrow 1.28 = \frac{X - 200}{20}$$

$$\Leftrightarrow X = 226.$$

Như vậy, bất kì ai có điểm từ 226 trở lên đều đủ điều kiện.

Ví dụ 6.2.5. Đối với một nghiên cứu y tế, một nhà nghiên cứu muốn chọn 60% dân số có mức huyết áp quanh giá trị trung bình của nó. Biết rằng huyết áp trung bình là 120 và độ lệch chuẩn là 8. Hãy tìm số người trên và dưới đủ điều kiện tham gia nghiên cứu.

6.2.2 Xác định tính chuẩn

Biểu đồ phân phối chuẩn hay hình chuông chỉ là một trong nhiều hình dạng mà một phân phối có thể giả định; tuy nhiên, nó là rất quan trọng vì nhiều phương pháp thống kê đòi hỏi sự phân bố các giá trị (thể hiện trong các chương tiếp theo) là biểu đồ chuẩn hoặc tiệm cận chuẩn.

Có một số cách để thống kê kiểm tra tính chuẩn. Một trong những cách đơn giản nhất là vẽ một biểu đồ cho dữ liệu và kiểm tra hình dạng của nó.

Độ lệch có thể được kiểm tra bằng cách sử dụng chỉ số Pearson PI . Công thức được xác định:

$$PI = \frac{3(\bar{X} - MD)}{s}.$$

Nếu chỉ số này nằm ngoài khoảng $(-1, 1)$ thì có thể kết luận rằng dữ liệu bị lệch đáng kể. Ngoài ra, dữ liệu cần được kiểm tra cho các giá trị ngoại biên bằng cách sử dụng phương pháp được trình bày trong chương 3. Vì các giá trị ngoại biên nó ảnh hưởng lớn đến tính chuẩn của dữ liệu.