



BÀI GIẢNG XÁC SUẤT THỐNG KÊ

Giảng viên: TS. PHÙNG MINH ĐỨC

(Bộ môn Toán Lý)

Chương 2: Biến ngẫu nhiên

- 2.1 Biến ngẫu nhiên
- 2.2 Biến ngẫu nhiên rời rạc
- 2.3 Biến ngẫu nhiên liên tục
- 2.4 Các tham số của biến ngẫu nhiên
- 2.5 Các phân phối thường gặp
- 2.6 Hàm đặc trưng

2.1 Biến ngẫu nhiên

Định nghĩa 2.1.1

Cho một phép thử có không gian mẫu là Ω . Biến ngẫu nhiên X (random variable - đại lượng ngẫu nhiên) một phép gán một kết quả (outcome) của Ω với một số.

Định nghĩa 2.1.1

Cho một phép thử có không gian mẫu là Ω . Biến ngẫu nhiên X (random variable - đại lượng ngẫu nhiên) một phép gán một kết quả (outcome) của Ω với một số.

Ví dụ 2.1.2

Xét một phép thử gieo 3 đồng xu và tính số mặt sấp. Đặt X là số mặt sấp. Ta thấy rằng X có thể nhận các giá trị 0,1,2,3.

Định nghĩa 2.1.1

Cho một phép thử có không gian mẫu là Ω . Biến ngẫu nhiên X (random variable - đại lượng ngẫu nhiên) một phép gán một kết quả (outcome) của Ω với một số.

Ví dụ 2.1.2

Xét một phép thử gieo 3 đồng xu và tính số mặt sấp. Đặt X là số mặt sấp. Ta thấy rằng X có thể nhận các giá trị 0,1,2,3.

Ví dụ 2.1.3

Trong một cái túi có 1 bi đỏ và 99 bi đen. Người chơi cần bốc ra 10000 đồng để lấy ngẫu nhiên 1 bi. Nếu lấy trúng bi đỏ thì được thưởng 100000 đồng. Nếu lấy trúng bi đen thì không được thưởng. Gọi X là số tiền mà người chơi có sau 1 lần lấy bi. Khi đó X nhận các giá trị 90000 và -10000.

Định nghĩa 2.1.4

- ① Biến ngẫu nhiên được gọi là **rời rạc** nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử.

Định nghĩa 2.1.4

- ① Biến ngẫu nhiên được gọi là **rời rạc** nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử.
- ② Biến ngẫu nhiên được gọi là **liên tục** nếu tập giá trị của nó lấp kín một khoảng trên trực số.

Định nghĩa 2.1.4

- ① Biến ngẫu nhiên được gọi là **rời rạc** nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử.
- ② Biến ngẫu nhiên được gọi là **liên tục** nếu tập giá trị của nó lấp kín một khoảng trên trực số.

Ví dụ 2.1.5

- ① Đặt X là số chấm xuất hiện khi gieo một con xúc xắc. Khi đó X nhận các giá trị là 1,2,3,4,5,6. Do đó, X là một biến ngẫu nhiên rời rạc.

Định nghĩa 2.1.4

- ① Biến ngẫu nhiên được gọi là **rời rạc** nếu tập giá trị của nó là một tập hữu hạn hoặc vô hạn đếm được các phần tử.
- ② Biến ngẫu nhiên được gọi là **liên tục** nếu tập giá trị của nó lấp kín một khoảng trên trực số.

Ví dụ 2.1.5

- ① Đặt X là số chấm xuất hiện khi gieo một con xúc xắc. Khi đó X nhận các giá trị là 1,2,3,4,5,6. Do đó, X là một biến ngẫu nhiên rời rạc.
- ② Đặt Y là nhiệt độ trong một ngày. Khi đó Y là một số ngẫu nhiên từ $-90^{\circ}C$ đến $71^{\circ}C$.^a Do đó, Y là một biến ngẫu nhiên liên tục.

^aNăm 1983, nhiệt độ thấp kỷ lục từng được ghi nhận tại trạm Vostok (Nga) là $-89,2^{\circ}C$. Sa mạc Lut tại Iran là vùng đất nóng nhất trên Trái đất với nhiệt độ lên tới $70,7^{\circ}C$ (vào năm 2005).

2.2 Biến ngẫu nhiên rời rạc

Dịnh nghĩa 2.2.1

Cho X là một biến ngẫu nhiên rời rạc có miền giá trị là $\{x_1, x_2, \dots\}$.
Bảng phân phối xác suất của biến ngẫu nhiên X là

X			
$P(X = x_i)$			

trong đó $\sum_{i=1}^{+\infty} p_i = 1$.

Dịnh nghĩa 2.2.1

Cho X là một biến ngẫu nhiên rời rạc có miền giá trị là $\{x_1, x_2, \dots\}$.
Bảng phân phối xác suất của biến ngẫu nhiên X là

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

trong đó $\sum_{i=1}^{+\infty} p_i = 1$.

Ví dụ 2.2.2

Xét phép thử là gieo một con xúc xắc. Đặt X là “số chấm xuất hiện” và ta thấy nó nhận các giá trị nguyên từ 1 đến 6.

Dịnh nghĩa 2.2.1

Cho X là một biến ngẫu nhiên rời rạc có miền giá trị là $\{x_1, x_2, \dots\}$.
Bảng phân phối xác suất của biến ngẫu nhiên X là

X		x_1	x_2	...
$P(X = x_i)$		p_1	p_2	...

trong đó $\sum_{i=1}^{+\infty} p_i = 1$.

Ví dụ 2.2.2

Xét phép thử là gieo một con xúc xắc. Đặt X là “số chấm xuất hiện” và ta thấy nó nhận các giá trị nguyên từ 1 đến 6.

Bảng phân phối của biến ngẫu nhiên X là

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

Ví dụ 2.2.3

Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Hãy lập bảng phân phối xác suất của số đạn cần bắn.

Ví dụ 2.2.3

Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Hãy lập bảng phân phối xác suất của số đạn cần bắn.

Giải. Đặt X là số đạn cần bắn. Khi đó X nhận các giá trị 1, 2, 3 và X là một biến ngẫu nhiên rời rạc.

Ví dụ 2.2.3

Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Hãy lập bảng phân phối xác suất của số đạn cần bắn.

Giải. Đặt X là số đạn cần bắn. Khi đó X nhận các giá trị 1, 2, 3 và X là một biến ngẫu nhiên rời rạc.

- Nếu $X = 1$ thì người đó bắn viên thứ nhất trúng và do đó $P(X = 1) = 0,6$.

Ví dụ 2.2.3

Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Hãy lập bảng phân phối xác suất của số đạn cần bắn.

Giải. Đặt X là số đạn cần bắn. Khi đó X nhận các giá trị 1, 2, 3 và X là một biến ngẫu nhiên rời rạc.

- Nếu $X = 1$ thì người đó bắn viên thứ nhất trúng và do đó $P(X = 1) = 0,6$.
- Nếu $X = 2$ thì viên thứ nhất bắn không trúng và viên thứ hai bắn trúng. Do đó $P(X = 2) = 0,4 \cdot 0,6 = 0,24$.

Ví dụ 2.2.3

Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Hãy lập bảng phân phối xác suất của số đạn cần bắn.

Giải. Đặt X là số đạn cần bắn. Khi đó X nhận các giá trị 1, 2, 3 và X là một biến ngẫu nhiên rời rạc.

- Nếu $X = 1$ thì người đó bắn viên thứ nhất trúng và do đó $P(X = 1) = 0,6$.
- Nếu $X = 2$ thì viên thứ nhất bắn không trúng và viên thứ hai bắn trúng. Do đó $P(X = 2) = 0,4 \cdot 0,6 = 0,24$.
- Nếu $X = 3$ thì 2 viên đạn đầu không trúng và viên thứ 3 có thể trúng hay không trúng. Do đó $P(X = 3) = 0,4 \cdot 0,4 = 0,16$.

Ví dụ 2.2.3

Một xạ thủ chỉ có 3 viên đạn. Anh ta được yêu cầu bắn từng phát cho đến khi trúng mục tiêu thì dừng bắn, biết rằng xác suất trúng của mỗi lần bắn là 0,6. Hãy lập bảng phân phối xác suất của số đạn cần bắn.

Giải. Đặt X là số đạn cần bắn. Khi đó X nhận các giá trị 1, 2, 3 và X là một biến ngẫu nhiên rời rạc.

- Nếu $X = 1$ thì người đó bắn viên thứ nhất trúng và do đó $P(X = 1) = 0,6$.
- Nếu $X = 2$ thì viên thứ nhất bắn không trúng và viên thứ hai bắn trúng. Do đó $P(X = 2) = 0,4 \cdot 0,6 = 0,24$.
- Nếu $X = 3$ thì 2 viên đạn đầu không trúng và viên thứ 3 có thể trúng hay không trúng. Do đó $P(X = 3) = 0,4 \cdot 0,4 = 0,16$.

Bảng phân phối của biến ngẫu nhiên X là

X	1	2	3
$P(X = x_i)$	0,6	0,24	0,16

Ví dụ 2.2.4

Gieo một con xúc xắc và gọi X là số lần gieo đến khi gặp mặt 6 chấm.
Lập bảng phân phối xác suất của X .

Ví dụ 2.2.4

Gieo một con xúc xắc và gọi X là số lần gieo đến khi gặp mặt 6 chấm.
Lập bảng phân phối xác suất của X .

Giải. Biết ngẫu nhiên X nhận các giá trị $1, 2, 3, \dots$ (vô hạn đếm được).

Ví dụ 2.2.4

Gieo một con xúc xắc và gọi X là số lần gieo đền khi gặp mặt 6 chấm.
Lập bảng phân phối xác suất của X .

Giải. Biết ngẫu nhiên X nhận các giá trị $1, 2, 3, \dots$ (vô hạn đếm được).
Xác suất để lần thứ n xuất hiện mặt 6 chấm là

$$P(X = n) = \left(\frac{5}{6}\right)^{n-1} \cdot \frac{1}{6}.$$

Ví dụ 2.2.4

Gieo một con xúc xắc và gọi X là số lần gieo đền khi gặp mặt 6 chấm.
Lập bảng phân phối xác suất của X .

Giải. Biết ngẫu nhiên X nhận các giá trị $1, 2, 3, \dots$ (vô hạn đếm được).
Xác suất để lần thứ n xuất hiện mặt 6 chấm là

$$P(X = n) = \left(\frac{5}{6}\right)^{n-1} \cdot \frac{1}{6}.$$

Bảng phân phối xác suất của X là

X	1	2	\dots	n	\dots
$P(X = x_i)$	$\frac{1}{6}$	$\frac{5}{6} \cdot \frac{1}{6}$	\dots	$\left(\frac{5}{6}\right)^{n-1} \cdot \frac{1}{6}$	\dots

Định nghĩa 2.2.5 (Hàm phân phối xác suất)

Hàm phân phối xác suất của biến ngẫu nhiên X , ký hiệu là $F(x)$, được xác định như sau:

$$F(x) = P(X \leq x).$$

Định nghĩa 2.2.5 (Hàm phân phối xác suất)

Hàm phân phối xác suất của biến ngẫu nhiên X , ký hiệu là $F(x)$, được xác định như sau:

$$F(x) = P(X \leq x).$$

Cho X là một biến ngẫu nhiên rời rạc nhận các giá trị $\{x_1, x_2, \dots\}$. Nếu $x_1 < x_2 < \dots$ thì

$$F(x_n) = \sum_{i=1}^n P(X = x_i).$$

Ví dụ 2.2.6 Gieo một con xúc xắc. Đặt biến ngẫu nhiên X là “số chấm xuất hiện.” Bảng phân phối của biến ngẫu nhiên X là

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

Hàm phân phối xác suất của X được xác định bởi

$$F(1) = \frac{1}{6}; F(2) = \frac{2}{6}; F(3) = \frac{3}{6}; F(4) = \frac{4}{6}; F(5) = \frac{5}{6}; F(6) = 1$$

Ví dụ 2.1.8. Có hai lô hàng: lô 1 có 10 sản phẩm trong đó có 3 sản phẩm loại A; lô 2 có 12 sản phẩm trong đó có 4 sản phẩm loại A.

a. Lấy ngẫu nhiên mỗi lô một sản phẩm. Gọi X là số sản phẩm loại A lấy được. Lập bảng phân phối xác suất (luật phân phối xác suất) của X .

b. Lấy ngẫu nhiên một lô hàng, từ đó lấy ngẫu nhiên 3 sản phẩm. Đặt Y là số sản phẩm loại A lấy được. Lập bảng phân phối xác suất của Y .

Hướng dẫn: a. Tập giá trị của X là $\{0, 1, 2\}$.

- $P(X = 0)$ là xác suất lấy 2 sản phẩm không có sản phẩm loại A

$$P(X = 0) = \dots$$

- $P(X = 2)$ là xác suất lấy 2 sản phẩm có 2 sản phẩm loại A

$$P(X = 2) = \dots$$

- $P(X = 1)$ là xác suất lấy 2 sản phẩm có 1 sản phẩm loại A

$$P(X = 1) = \dots$$

Bảng phân phối xác suất của X là

X	0	1	2
$P(X = x_i)$	7/15.	9/10	1.....

b. Dùng công thức xác suất toàn phần.

Gọi A_i là biến cố "Lấy lô i " với $i = 1, 2$. Ta có

$$P(A_1) = P(A_2) = \dots$$

$$P(Y = 0) = \dots$$

(tiếp tục...)

2.3 Biến ngẫu nhiên liên tục

Định nghĩa 2.3.1 (Hàm phân phối xác suất)

Hàm phân phối xác suất của một biến ngẫu nhiên liên tục X được xác định như sau

$$F(x) = P(X \leq x).$$

Định nghĩa 2.3.1 (Hàm phân phối xác suất)

Hàm phân phối xác suất của một biến ngẫu nhiên liên tục X được xác định như sau

$$F(x) = P(X \leq x).$$

Định nghĩa 2.3.2 (Hàm mật độ)

Hàm mật độ xác suất của một biến ngẫu nhiên liên tục là một hàm số $f(x)$ thỏa mãn

Định nghĩa 2.3.1 (Hàm phân phối xác suất)

Hàm phân phối xác suất của một biến ngẫu nhiên liên tục X được xác định như sau

$$F(x) = P(X \leq x).$$

Định nghĩa 2.3.2 (Hàm mật độ)

Hàm mật độ xác suất của một biến ngẫu nhiên liên tục là một hàm số $f(x)$ thỏa mãn

- ① $f(x) \geq 0,$

Định nghĩa 2.3.1 (Hàm phân phối xác suất)

Hàm phân phối xác suất của một biến ngẫu nhiên liên tục X được xác định như sau

$$F(x) = P(X \leq x).$$

Định nghĩa 2.3.2 (Hàm mật độ)

Hàm mật độ xác suất của một biến ngẫu nhiên liên tục là một hàm số $f(x)$ thỏa mãn

(1) $f(x) \geq 0,$

(2) $\int_{-\infty}^{+\infty} f(x)dx = 1,$

Định nghĩa 2.3.1 (Hàm phân phối xác suất)

Hàm phân phối xác suất của một biến ngẫu nhiên liên tục X được xác định như sau

$$F(x) = P(X \leq x).$$

Định nghĩa 2.3.2 (Hàm mật độ)

Hàm mật độ xác suất của một biến ngẫu nhiên liên tục là một hàm số $f(x)$ thỏa mãn

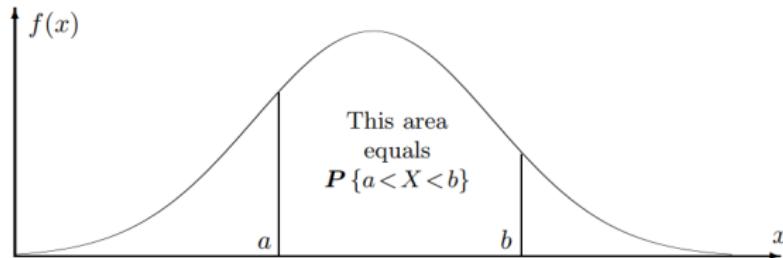
① $f(x) \geq 0,$

② $\int_{-\infty}^{+\infty} f(x)dx = 1,$

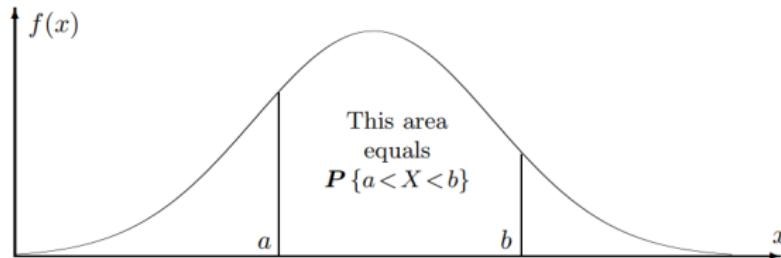
③ $P(a \leq X \leq b) = \int_a^b f(x)dx$

Hai điều kiện 1 và 2 trong định nghĩa trên có nghĩa là phần diện tích giới hạn bởi đồ thị của $f(x)$ (không nằm dưới trục hoành) và trục hoành bằng 1.

Hai điều kiện 1 và 2 trong định nghĩa trên có nghĩa là phần diện tích giới hạn bởi đồ thị của $f(x)$ (không nằm dưới trục hoành) và trục hoành bằng 1. Xác suất khi biến ngẫu nhiên X nhận các giá trị thuộc đoạn $[a, b]$ bằng diện tích hình giới hạn bởi của đồ thị $f(x)$, trục hoành và hai đường thẳng $x = a; x = b$.



Hai điều kiện 1 và 2 trong định nghĩa trên có nghĩa là phần diện tích giới hạn bởi đồ thị của $f(x)$ (không nằm dưới trục hoành) và trục hoành bằng 1. Xác suất khi biến ngẫu nhiên X nhận các giá trị thuộc đoạn $[a, b]$ bằng diện tích hình giới hạn bởi của đồ thị $f(x)$, trục hoành và hai đường thẳng $x = a; x = b$.



Định lý 2.3.3

Cho X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và hàm phân phối xác suất $F(x)$. Khi đó

Định lý 2.3.3

Cho X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và hàm phân phối xác suất $F(x)$. Khi đó

$$\textcircled{1} \quad P(X = a) = \int_a^a f(x)dx = 0$$

Định lý 2.3.3

Cho X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và hàm phân phối xác suất $F(x)$. Khi đó

$$\textcircled{1} \quad P(X = a) = \int_a^a f(x)dx = 0$$

$$\textcircled{2} \quad P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$

Định lý 2.3.3

Cho X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và hàm phân phối xác suất $F(x)$. Khi đó

$$\textcircled{1} \quad P(X = a) = \int_a^a f(x)dx = 0$$

$$\textcircled{2} \quad P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$

$$\textcircled{3} \quad F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Định lý 2.3.3

Cho X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và hàm phân phối xác suất $F(x)$. Khi đó

$$\textcircled{1} \quad P(X = a) = \int_a^a f(x)dx = 0$$

$$\textcircled{2} \quad P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b).$$

$$\textcircled{3} \quad F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

$$\textcircled{4} \quad P(X > a) = 1 - P(X \leq a) = \int_a^{+\infty} f(t)dt.$$

Ví dụ 2.3.4

Thời gian sử dụng của một thiết bị điện tử (tính bằng năm) là một biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} \frac{k}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases}$$

- a. Tìm k .
- b. Tìm hàm phân phối xác suất.
- c. Tìm xác suất thiết bị đó sử dụng được hơn 5 năm.

a. Tìm k . Ta có

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_1^{+\infty} f(x)dx = \dots$$

Do đó $k = \dots$

b. Hàm phân phối xác suất

$$F(x) = \int_{-\infty}^x f(t)dt = \int_1^x \frac{k}{t} dt = \dots$$

c. Xác suất để thiết bị đó sử dụng được nhiều hơn 5 năm là

$$P(X > 5) = \int_5^{+\infty} f(x)dx = \dots$$

(Ta có thể sử dụng $P(X > 5) = 1 - F(5) = \dots$)

Bài tập 2.3.5

Thời gian phục vụ mỗi khách hàng tại một cửa hàng là một biến ngẫu nhiên liên tục X (tính bằng phút) có hàm mật độ xác suất như sau

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

- Tìm k và hàm phân phối xác suất của X .
- Tính xác suất thời gian phục vụ một khách hàng trong khoảng thời gian từ 1 đến 2 phút.

Ký hiệu: $\exp(-4x) = e^{-4x}$.

Bài tập 2.3.6

Cho biến ngẫu nhiên X là thời gian sử dụng của một loại pin (tính bằng năm). Giả sử hàm mật độ xác suất của X được cho bởi

$$f(x) = \begin{cases} \frac{1}{2} \exp(-\frac{x}{2}), & x > 0 \\ 0, & x \leq 0. \end{cases}$$

- Tính xác suất để loại pin này sử dụng được ít hơn 2 năm hoặc nhiều hơn 4 năm.
- Giả sử có một viên pin loại này đã được sử dụng 2 năm. Tính xác suất để pin này được sử dụng ít nhất 3 năm.

Từ hàm mật độ xác suất, ta tìm được hàm phân phối xác suất

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Từ hàm mật độ xác suất, ta tìm được hàm phân phối xác suất

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Nếu biết hàm phân phối xác suất $F(x)$ thì ta tìm hàm được mật độ xác suất

$$f(x) = F'(x).$$

Từ hàm mật độ xác suất, ta tìm được hàm phân phối xác suất

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt.$$

Nếu biết hàm phân phối xác suất $F(x)$ thì ta tìm hàm được mật độ xác suất

$$f(x) = F'(x).$$

Ví dụ 2.3.7 Đặt X là thời gian (tháng) sử dụng của một loại thiết bị điện tử. Giả sử biến ngẫu nhiên X có hàm phân phối xác suất

$$F(x) = \begin{cases} 1 - e^{-0,03x^{1,2}}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$

- Chọn ngẫu nhiên một thiết bị. Tính xác suất thiết bị đó sử dụng được ít nhất 12 tháng.
- Tìm hàm mật độ xác suất của X .

2.4 Các tham số của biến ngẫu nhiên

Kỳ vọng

Định nghĩa 2.4.1 (Kỳ vọng)

Kỳ vọng (expectation) của biến ngẫu nhiên X , ký hiệu là $E(X)$, được xác định như sau:

- Nếu X là biến ngẫu nhiên rời rạc thì

$$E(X) = \sum_{i \geq 1} x_i P(X = x_i)$$

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

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

Ta thường ký hiệu $E(X) = \mu$.

Kỳ vọng còn được gọi là giá trị trung bình (Mean), giá trị trung bình có trọng số (Weighted Average), giá mong đợi (Expected Value).

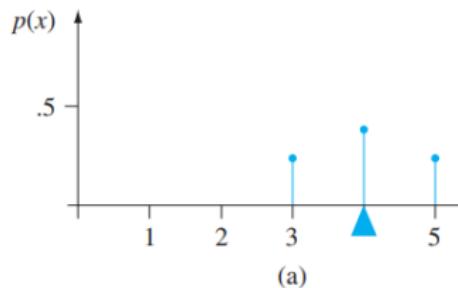
Kỳ vọng còn được gọi là giá trị trung bình (Mean), giá trị trung bình có trọng số (Weighted Average), giá mong đợi (Expected Value).

Ý nghĩa. Kỳ vọng của biến ngẫu nhiên là **giá trị trung bình có trọng số** của biến ngẫu nhiên (phân biệt với trung bình cộng của các giá trị).

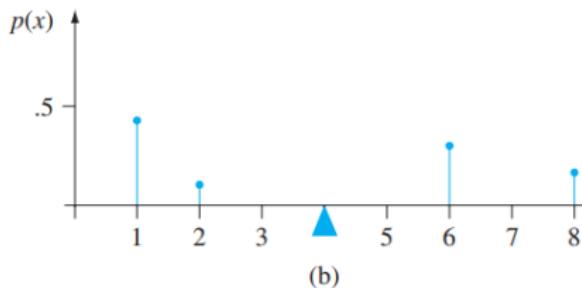
Quan sát 2 hình dưới đây biểu diễn phân phối xác suất của 2 biến ngẫu nhiên. Giả sử tại vị trí tam giác là kỳ vọng của biến ngẫu nhiên.

X	3	4	5
$P(X = x_i)$	0,3	0,4	0,3

Y	1	2	6	8
$P(Y = y_i)$	0,4	0,1	0,3	0,2



(a)



(b)

Mặc dù cả hai phân phối xác suất trong trên có cùng kỳ vọng $E(X)$ nhưng phân phối của Hình (b) có độ phân tán (tức là độ biến thiên) lớn hơn so với của Hình (a). Chúng ta sẽ sử dụng **phương sai** (variance) của X để đánh giá lượng phân tán trong (phân phối của) X .

Phương sai

Định nghĩa 2.4.2 (Phương sai)

Cho X là biến ngẫu nhiên và kỳ vọng μ . Phương sai (variance) của X , kí hiệu là $V(X)$ hoặc σ^2 , được xác định bởi

Phương sai

Định nghĩa 2.4.2 (Phương sai)

Cho X là biến ngẫu nhiên và kỳ vọng μ . Phương sai (variance) của X , kí hiệu là $V(X)$ hoặc σ^2 , được xác định bởi

- ① Nếu X là biến ngẫu nhiên rời rạc có miền giá trị $\{x_1, x_2, \dots\}$ thì

$$V(X) = \sigma^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 P(X = x_i) = E((X - \mu)^2)$$

Phương sai

Định nghĩa 2.4.2 (Phương sai)

Cho X là biến ngẫu nhiên và kỳ vọng μ . Phương sai (variance) của X , kí hiệu là $V(X)$ hoặc σ^2 , được xác định bởi

- ① Nếu X là biến ngẫu nhiên rời rạc có miền giá trị $\{x_1, x_2, \dots\}$ thì

$$V(X) = \sigma^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 P(X = x_i) = E((X - \mu)^2)$$

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

$$V(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = E((X - \mu)^2)$$

Phương sai

Định nghĩa 2.4.2 (Phương sai)

Cho X là biến ngẫu nhiên và kỳ vọng μ . Phương sai (variance) của X , ký hiệu là $V(X)$ hoặc σ^2 , được xác định bởi

- ① Nếu X là biến ngẫu nhiên rời rạc có miền giá trị $\{x_1, x_2, \dots\}$ thì

$$V(X) = \sigma^2 = \sum_{i=1}^{\infty} (x_i - \mu)^2 P(X = x_i) = E((X - \mu)^2)$$

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

$$V(X) = \sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx = E((X - \mu)^2)$$

Độ lệch chuẩn (standard deviation) của X được ký hiệu và xác định bởi $Std(X) = \sqrt{V(X)} = \sigma$.

Mode

Định nghĩa 2.4.3 (Mode)

Mode của biến ngẫu nhiên X là một giá trị của biến ngẫu nhiên X , ký hiệu $\text{Mod}(X)$, mà tại đó biến ngẫu nhiên X nhận xác suất lớn nhất (nếu X là rời rạc) hoặc tại đó hàm mật độ đạt giá trị lớn nhất (nếu X là biến liên tục).

Mode

Định nghĩa 2.4.3 (Mode)

Mode của biến ngẫu nhiên X là một giá trị của biến ngẫu nhiên X , ký hiệu $\text{Mod}(X)$, mà tại đó biến ngẫu nhiên X nhận xác suất lớn nhất (nếu X là rời rạc) hoặc tại đó hàm mật độ đạt giá trị lớn nhất (nếu X là biến liên tục).

Ví dụ 2.4.4

Trong một hộp có 99 tấm thẻ đen và 1 tấm thẻ đỏ. Mỗi lần rút thẻ người chơi sẽ phải trả 10000 đồng. Nếu rút trúng thẻ màu đỏ thì người chơi được thưởng 100000 đồng.

- Nếu người chơi chỉ rút 1 lần thì kỳ vọng bằng bao nhiêu?
- Nếu người chơi rút 2 lần (rút lần thứ nhất rồi bỏ lại vào hộp) thì kỳ vọng bằng bao nhiêu?

a. Gọi X là "Số tiền thu được sau khi rút". Nếu rút trúng thẻ đỏ thì $X = 90000$. Nếu rút được thẻ đen thì $X = -10000$.

a. Gọi X là "Số tiền thu được sau khi rút". Nếu rút trúng thẻ đỏ thì $X = 90000$. Nếu rút được thẻ đen thì $X = -10000$. Bảng phân phối xác suất

X	90000	-10000
$P(X = x_i)$	$\frac{1}{100}$	$\frac{99}{100}$

a. Gọi X là "Số tiền thu được sau khi rút". Nếu rút trúng thẻ đỏ thì $X = 90000$. Nếu rút được thẻ đen thì $X = -10000$. Bảng phân phối xác suất

X	90000	-10000
$P(X = x_i)$	$\frac{1}{100}$	$\frac{99}{100}$

Kỳ vọng $E(X) = 90000 \cdot \frac{1}{100} + (-10000) \cdot \frac{99}{100} = -9000$
(Trung bình mỗi lần chơi, người chơi mất 9000 đồng).

a. Gọi X là "Số tiền thu được sau khi rút". Nếu rút trúng thẻ đỏ thì $X = 90000$. Nếu rút được thẻ đen thì $X = -10000$. Bảng phân phối xác suất

X	90000	-10000
$P(X = x_i)$	$\frac{1}{100}$	$\frac{99}{100}$

$$\text{Kỳ vọng } E(X) = 90000 \cdot \frac{1}{100} + (-10000) \cdot \frac{99}{100} = -9000$$

(Trung bình mỗi lần chơi, người chơi mất 9000 đồng).

b. Gọi Y là "Số tiền thu được sau hai lần rút".

a. Gọi X là "Số tiền thu được sau khi rút". Nếu rút trúng thẻ đỏ thì $X = 90000$. Nếu rút được thẻ đen thì $X = -10000$. Bảng phân phối xác suất

X	90000	-10000
$P(X = x_i)$	$\frac{1}{100}$	$\frac{99}{100}$

$$\text{Kỳ vọng } E(X) = 90000 \cdot \frac{1}{100} + (-10000) \cdot \frac{99}{100} = -9000$$

(Trung bình mỗi lần chơi, người chơi mất 9000 đồng).

b. Gọi Y là "Số tiền thu được sau hai lần rút". Khi đó Y nhận các giá trị: 180000, 80000, -20000. Bảng phân phối xác suất của Y là

Y	180000	80000	-20000
$P(Y = y_i)$	$\frac{1}{10000}$	$\frac{198}{10000}$	$\frac{9801}{10000}$

a. Gọi X là "Số tiền thu được sau khi rút". Nếu rút trúng thẻ đỏ thì $X = 90000$. Nếu rút được thẻ đen thì $X = -10000$. Bảng phân phối xác suất

X	90000	-10000
$P(X = x_i)$	$\frac{1}{100}$	$\frac{99}{100}$

$$\text{Kỳ vọng } E(X) = 90000 \cdot \frac{1}{100} + (-10000) \cdot \frac{99}{100} = -9000$$

(Trung bình mỗi lần chơi, người chơi mất 9000 đồng).

b. Gọi Y là "Số tiền thu được sau hai lần rút". Khi đó Y nhận các giá trị: 180000, 80000, -20000. Bảng phân phối xác suất của Y là

Y	180000	80000	-20000
$P(Y = y_i)$	$\frac{1}{10000}$	$\frac{198}{10000}$	$\frac{9801}{10000}$

Kỳ vọng của Y là

$$E(Y) = 180000 \cdot \frac{1}{10000} + 80000 \cdot \frac{198}{10000} + (-20000) \cdot \frac{9801}{10000} = -18000.$$

(Trung bình sau 2 lần chơi, người chơi sẽ mất 18000 đồng)

Ví dụ 2.4.5

Đặt X là lượng cát (tính bằng tấn) do công ty CAT bán ra hàng tuần. Giả sử biến ngẫu nhiên X có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2), & 0 \leq x \leq 1 \\ 0, & x \notin [0; 1] \end{cases}$$

Kỳ vọng của X là

Ví dụ 2.4.5

Đặt X là lượng cát (tính bằng tấn) do công ty CAT bán ra hàng tuần. Giả sử biến ngẫu nhiên X có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2), & 0 \leq x \leq 1 \\ 0, & x \notin [0; 1] \end{cases}$$

Kỳ vọng của X là

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 \frac{3}{2}x(1 - x^2)dx = \frac{3}{8}$$

và phương sai của X là

Ví dụ 2.4.5

Đặt X là lượng cát (tính bằng tấn) do công ty CAT bán ra hàng tuần. Giả sử biến ngẫu nhiên X có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2), & 0 \leq x \leq 1 \\ 0, & x \notin [0; 1] \end{cases}$$

Kỳ vọng của X là

$$E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^1 \frac{3}{2}x(1 - x^2)dx = \frac{3}{8}$$

và phương sai của X là

$$V(X) = \int_{-\infty}^{+\infty} (x - \frac{3}{8})^2 f(x)dx = \int_0^1 \frac{3}{2}(x - \frac{3}{8})^2(1 - x^2)dx = \frac{19}{320}.$$

Dịnh lý 2.4.6

① Nếu X là một biến ngẫu nhiên rời rạc và $g(x)$ là một hàm số thì

$$E(g(X)) = \sum_x g(x)P(X=x)$$

② Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và $g(x)$ là một hàm số thì

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

Dịnh lý 2.4.6

① Nếu X là một biến ngẫu nhiên rời rạc và $g(x)$ là một hàm số thì

$$E(g(X)) = \sum_x g(x)P(X=x)$$

② Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và $g(x)$ là một hàm số thì

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

Mệnh đề 2.4.7

Cho X là một biến ngẫu nhiên và $a, b \in \mathbb{R}$. Khi đó

Dịnh lý 2.4.6

① Nếu X là một biến ngẫu nhiên rời rạc và $g(x)$ là một hàm số thì

$$E(g(X)) = \sum_x g(x)P(X=x)$$

② Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và $g(x)$ là một hàm số thì

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

Mệnh đề 2.4.7

Cho X là một biến ngẫu nhiên và $a, b \in \mathbb{R}$. Khi đó

① $E(aX + b) = aE(X) + b$

Dịnh lý 2.4.6

① Nếu X là một biến ngẫu nhiên rời rạc và $g(x)$ là một hàm số thì

$$E(g(X)) = \sum_x g(x)P(X=x)$$

② Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và $g(x)$ là một hàm số thì

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

Mệnh đề 2.4.7

Cho X là một biến ngẫu nhiên và $a, b \in \mathbb{R}$. Khi đó

① $E(aX + b) = aE(X) + b$

② $V(aX + b) = a^2V(X)$

Dịnh lý 2.4.6

① Nếu X là một biến ngẫu nhiên rời rạc và $g(x)$ là một hàm số thì

$$E(g(X)) = \sum_x g(x)P(X=x)$$

② Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ và $g(x)$ là một hàm số thì

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

Mệnh đề 2.4.7

Cho X là một biến ngẫu nhiên và $a, b \in \mathbb{R}$. Khi đó

① $E(aX + b) = aE(X) + b$

② $V(aX + b) = a^2V(X)$

③ $V(X) = E(X^2) - E(X)^2$

Ví dụ 2.4.8

Một trung tâm bảo hành laptop thấy rằng số laptop hiệu Potpal cần sửa trong năm tối đa là 3 máy. Cho X là số laptop Potpal cần sửa trong một năm. Dựa trên các số liệu đã có, bảng phân phối xác suất của X như sau

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

- Trong một năm, trung bình có bao nhiêu laptop Potpal cần sửa?
- Tìm phương sai, độ lệch chuẩn và mode.
- Một cửa hàng Z có hợp đồng với trung tâm bảo hành đối với việc sửa laptop Potpal như sau: Phí dịch vụ hàng năm là 200\$ và 50\$ cho mỗi laptop được sửa. Tìm trung bình, phương sai và độ lệch chuẩn của số tiền mà cửa hàng Z trả cho trung tâm bảo hành mỗi năm.

a. Kỳ vọng của X là

a. Kỳ vọng của X là

$$\begin{aligned}E(x) &= \sum_{i=1}^4 x_i P(X = x_i) \\&= 0.0, 2 + 1.0, 3 + 2.0, 4 + 3.0, 1 = 1, 4\end{aligned}$$

Như vậy, số laptop cần sửa trung bình là 1,4 máy trong một năm.

b. Phương sai của X là

a. Kỳ vọng của X là

$$\begin{aligned}E(x) &= \sum_{i=1}^4 x_i P(X = x_i) \\&= 0.0, 2 + 1.0, 3 + 2.0, 4 + 3.0, 1 = 1, 4\end{aligned}$$

Như vậy, số laptop cần sửa trung bình là 1,4 máy trong một năm.

b. Phương sai của X là

$$\begin{aligned}V(X) &= \sum_{i=1}^4 (x_i - \mu)^2 P(X = x_i) \\&= (0 - 1,4)^2 \cdot 0,2 + (1 - 1,4)^2 \cdot 0,3 + (2 - 1,4)^2 \cdot 0,4 + (3 - 1,4)^2 \cdot 0,1 \\&= 0,84\end{aligned}$$

và độ lệch chuẩn là

$$\sigma = \sqrt{V(X)} = \sqrt{0,84} = 0,92.$$

a. Kỳ vọng của X là

$$\begin{aligned} E(x) &= \sum_{i=1}^4 x_i P(X = x_i) \\ &= 0.0, 2 + 1.0, 3 + 2.0, 4 + 3.0, 1 = 1, 4 \end{aligned}$$

Như vậy, số laptop cần sửa trung bình là 1,4 máy trong một năm.

b. Phương sai của X là

$$\begin{aligned} V(X) &= \sum_{i=1}^4 (x_i - \mu)^2 P(X = x_i) \\ &= (0 - 1,4)^2 \cdot 0,2 + (1 - 1,4)^2 \cdot 0,3 + (2 - 1,4)^2 \cdot 0,4 + (3 - 1,4)^2 \cdot 0,1 \\ &= 0,84 \end{aligned}$$

và độ lệch chuẩn là

$$\sigma = \sqrt{V(X)} = \sqrt{0,84} = 0,92.$$

Vì xác suất $P(X = 2)$ là lớn nhất nên $\text{Mod}(X) = 2$.

c. Chi phí hàng năm mà cửa hàng Z phải trả cho trung tâm bảo hành là một hàm số theo biến X như sau

$$200 + 50X.$$

c. Chi phí hàng năm mà cửa hàng Z phải trả cho trung tâm bảo hành là một hàm số theo biến X như sau

$$200 + 50X.$$

Do đó

$$\begin{aligned}E(200 + 50X) &= 50E(X) + 200 \\&= 50 \cdot 1,4 + 200 \\&= 270\end{aligned}$$

c. Chi phí hàng năm mà cửa hàng Z phải trả cho trung tâm bảo hành là một hàm số theo biến X như sau

$$200 + 50X.$$

Do đó

$$\begin{aligned}E(200 + 50X) &= 50E(X) + 200 \\&= 50 \cdot 1,4 + 200 \\&= 270\end{aligned}$$

Như vậy, cửa hàng Z trả cho trung tâm bảo hành trung bình 270\$ mỗi năm.

c. Chi phí hàng năm mà cửa hàng Z phải trả cho trung tâm bảo hành là một hàm số theo biến X như sau

$$200 + 50X.$$

Do đó

$$\begin{aligned}E(200 + 50X) &= 50E(X) + 200 \\&= 50 \cdot 1,4 + 200 \\&= 270\end{aligned}$$

Như vậy, cửa hàng Z trả cho trung tâm bảo hành trung bình 270\$ mỗi năm.

Phương sai của chi phí hàng năm là

$$V(200 + 50X) = 50^2 V(X) = 50^2 \cdot 0,84 = 2100$$

c. Chi phí hàng năm mà cửa hàng Z phải trả cho trung tâm bảo hành là một hàm số theo biến X như sau

$$200 + 50X.$$

Do đó

$$\begin{aligned}E(200 + 50X) &= 50E(X) + 200 \\&= 50 \cdot 1,4 + 200 \\&= 270\end{aligned}$$

Như vậy, cửa hàng Z trả cho trung tâm bảo hành trung bình 270\$ mỗi năm.

Phương sai của chi phí hàng năm là

$$V(200 + 50X) = 50^2 V(X) = 50^2 \cdot 0,84 = 2100$$

và độ lệch chuẩn là

$$\sigma = \sqrt{2100} = 45,83.$$

Ví dụ 2.4.9 Một cửa hàng bán một loại hóa chất thấy rằng lượng hóa chất bán ra hàng tuần (đơn vị tính là trăm lít) là X có hàm mật độ xác suất là

$$f(x) = \begin{cases} \frac{x}{4}, & 0 \leq x \leq 2 \\ \frac{1}{2}, & 2 < x \leq 3 \\ 0, & \text{các trường hợp khác} \end{cases}$$

Giả sử mỗi tuần, chi phí bảo dưỡng cửa hàng là 50\$. Cửa hàng mua mỗi lít hóa chất là 1,75\$ và bán ra 2,7\$.

- Tính lượng hóa chất trung bình bán ra hàng tuần.
 - Tính lợi nhuận trung bình hàng tuần của cửa hàng.
 - Tính mode của X .
-
-
-
-
-

Theorem (Định lý Chebyshev)

Cho X là một biến ngẫu nhiên có kỳ vọng μ và phương sai σ^2 . Khi đó, với mọi $\varepsilon > 0$, ta có

$$P(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

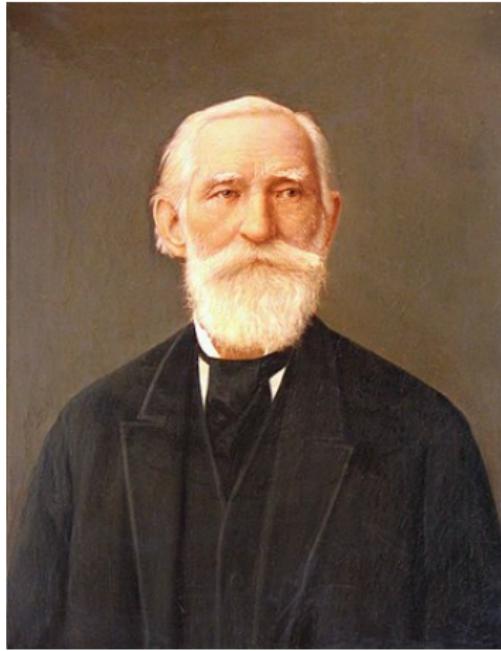
Theorem (Định lý Chebyshev)

Cho X là một biến ngẫu nhiên có kỳ vọng μ và phương sai σ^2 . Khi đó, với mọi $\varepsilon > 0$, ta có

$$P(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}.$$

Định lý Chebyshev cho thấy

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 1 - \frac{\sigma^2}{\varepsilon^2}.$$



Pafnuty Lvovich Chebyshev (1821 - 1894) là nhà toán học người Nga.

Ví dụ 2.4.10

Khảo sát tình hình sản xuất trong năm 2021, giám đốc một nhà máy thấy rằng trung bình một ngày làm được 120 sản phẩm với độ lệch chuẩn là 10.

- Tỉ lệ những ngày nhà máy sản xuất được từ 100 đến 140 sản phẩm có nhiều hơn 70% không?
- Tìm khoảng số lượng sản phẩm ngắn nhất để chứa ít nhất 90% mức sản xuất hàng ngày.

a. Đặt X là số sản phẩm sản xuất trong một ngày. Ta có $E(X) = \mu = 120$ và $\sigma = 10$. Vì $X \in [100; 140] = [\mu - \varepsilon, \mu + \varepsilon]$ nên $\varepsilon = 20$. Khi đó, theo Định lý Chebyshev,

$$P(100 \leq X \leq 140) \geq 1 - \frac{10^2}{20^2} = \frac{3}{4} = 75\%$$

Như vậy, tỉ lệ những ngày làm được từ 100 đến 140 sản phẩm không nhỏ hơn 75%. Do đó, tỉ lệ này lớn hơn 70%.

a. Đặt X là số sản phẩm sản xuất trong một ngày. Ta có $E(X) = \mu = 120$ và $\sigma = 10$. Vì $X \in [100; 140] = [\mu - \varepsilon, \mu + \varepsilon]$ nên $\varepsilon = 20$. Khi đó, theo Định lý Chebyshev,

$$P(100 \leq X \leq 140) \geq 1 - \frac{10^2}{20^2} = \frac{3}{4} = 75\%$$

Như vậy, tỉ lệ những ngày làm được từ 100 đến 140 sản phẩm không nhỏ hơn 75%. Do đó, tỉ lệ này lớn hơn 70%.

b. Ta cần tìm ε sao cho

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 90\%$$

a. Đặt X là số sản phẩm sản xuất trong một ngày. Ta có $E(X) = \mu = 120$ và $\sigma = 10$. Vì $X \in [100; 140] = [\mu - \varepsilon, \mu + \varepsilon]$ nên $\varepsilon = 20$. Khi đó, theo Định lý Chebyshev,

$$P(100 \leq X \leq 140) \geq 1 - \frac{10^2}{20^2} = \frac{3}{4} = 75\%$$

Như vậy, tỉ lệ những ngày làm được từ 100 đến 140 sản phẩm không nhỏ hơn 75%. Do đó, tỉ lệ này lớn hơn 70%.

b. Ta cần tìm ε sao cho

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 90\%$$

hay

$$1 - \frac{\sigma^2}{\varepsilon^2} = 0,9$$

a. Đặt X là số sản phẩm sản xuất trong một ngày. Ta có $E(X) = \mu = 120$ và $\sigma = 10$. Vì $X \in [100; 140] = [\mu - \varepsilon, \mu + \varepsilon]$ nên $\varepsilon = 20$. Khi đó, theo Định lý Chebyshev,

$$P(100 \leq X \leq 140) \geq 1 - \frac{10^2}{20^2} = \frac{3}{4} = 75\%$$

Như vậy, tỉ lệ những ngày làm được từ 100 đến 140 sản phẩm không nhỏ hơn 75%. Do đó, tỉ lệ này lớn hơn 70%.

b. Ta cần tìm ε sao cho

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 90\%$$

hay

$$1 - \frac{\sigma^2}{\varepsilon^2} = 0,9$$

hay $\varepsilon = \sqrt{1000} \approx 31,6$.

a. Đặt X là số sản phẩm sản xuất trong một ngày. Ta có $E(X) = \mu = 120$ và $\sigma = 10$. Vì $X \in [100; 140] = [\mu - \varepsilon, \mu + \varepsilon]$ nên $\varepsilon = 20$. Khi đó, theo Định lý Chebyshev,

$$P(100 \leq X \leq 140) \geq 1 - \frac{10^2}{20^2} = \frac{3}{4} = 75\%$$

Như vậy, tỉ lệ những ngày làm được từ 100 đến 140 sản phẩm không nhỏ hơn 75%. Do đó, tỉ lệ này lớn hơn 70%.

b. Ta cần tìm ε sao cho

$$P(\mu - \varepsilon \leq X \leq \mu + \varepsilon) \geq 90\%$$

hay

$$1 - \frac{\sigma^2}{\varepsilon^2} = 0,9$$

hay $\varepsilon = \sqrt{1000} \approx 31,6$. Suy ra

$$X \in [120 - 31,6; 120 + 31,6] = [88,4; 151,6].$$

Định nghĩa 2.4.11 (Trung vị)

Trung vị (median) là giá trị của biến ngẫu nhiên X mà nó chia phân phối của biến ngẫu nhiên thành hai phần bằng nhau, kí hiệu $\text{Med}(X)$.

Định nghĩa 2.4.11 (Trung vị)

Trung vị (median) là giá trị của biến ngẫu nhiên X mà nó chia phân phối của biến ngẫu nhiên thành hai phần bằng nhau, kí hiệu $\text{Med}(X)$.

Nếu X là biến ngẫu nhiên rời rạc thì giá trị x_k sẽ là trung vị nếu điều kiện sau được thỏa mãn:

$$F(x_{k-1}) < 0,5 \leq F(x_{k+1}).$$

Định nghĩa 2.4.11 (Trung vị)

Trung vị (median) là giá trị của biến ngẫu nhiên X mà nó chia phân phối của biến ngẫu nhiên thành hai phần bằng nhau, kí hiệu $\text{Med}(X)$.

Nếu X là biến ngẫu nhiên rời rạc thì giá trị x_k sẽ là trung vị nếu điều kiện sau được thỏa mãn:

$$F(x_{k-1}) < 0,5 \leq F(x_{k+1}).$$

Nếu X là biến ngẫu nhiên liên tục thì trung vị của X là giá trị m thỏa mãn điều kiện:

$$\int_{-\infty}^m f(x)dx = 0,5.$$

Định nghĩa 2.4.11 (Trung vị)

Trung vị (median) là giá trị của biến ngẫu nhiên X mà nó chia phân phối của biến ngẫu nhiên thành hai phần bằng nhau, kí hiệu $\text{Med}(X)$.

Nếu X là biến ngẫu nhiên rời rạc thì giá trị x_k sẽ là trung vị nếu điều kiện sau được thỏa mãn:

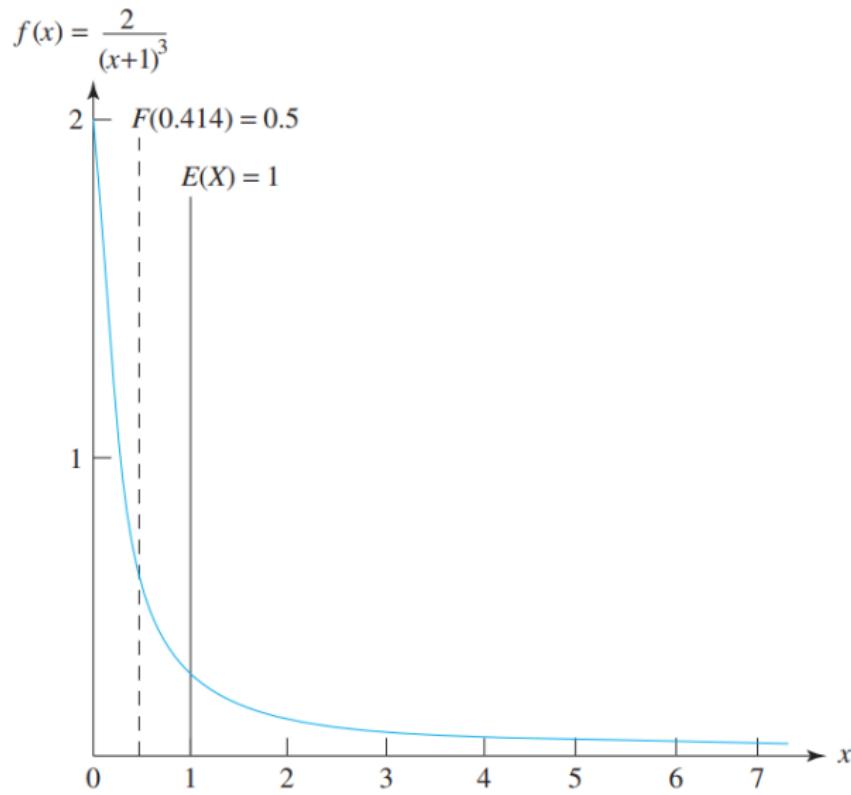
$$F(x_{k-1}) < 0,5 \leq F(x_{k+1}).$$

Nếu X là biến ngẫu nhiên liên tục thì trung vị của X là giá trị m thỏa mãn điều kiện:

$$\int_{-\infty}^m f(x)dx = 0,5.$$

Nếu X là một biến ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$ là đối xứng qua điểm μ thì cả trung vị và kỳ vọng của biến ngẫu nhiên đều bằng μ .

Hình minh họa trung vị và kỳ vọng



2.5 Các phân phối xác suất thường gặp

2.5.1 Phân phối Bernoulli

Định nghĩa 2.5.1 (Phép thử Bernoulli)

Một **phép thử Bernoulli** là phép thử mà ta chỉ quan tâm đến hai biến cố: A (thành công) và \bar{A} (thất bại) trong đó $P(A) = p$.

2.5.1 Phân phối Bernoulli

Định nghĩa 2.5.1 (Phép thử Bernoulli)

Một **phép thử Bernoulli** là phép thử mà ta chỉ quan tâm đến hai biến cố: A (thành công) và \bar{A} (thất bại) trong đó $P(A) = p$.

Định nghĩa 2.5.2 (Phân phối Bernoulli)

Một phân phối xác suất của biến ngẫu nhiên rời rạc X được gọi là phân phối Bernoulli nếu X chỉ nhận hai giá trị 1 (nếu thành công) hoặc 0 (nếu thất bại), trong đó $P(X = 1) = p$ (gọi là xác suất thành công) và $P(X = 0) = 1 - p$ (gọi là xác suất thất bại). Ký hiệu $X \sim B(p)$.

Bảng phân phối xác suất của biến ngẫu nhiên X có phân phối Bernoulli là

X	0	1
$P(X = x_i)$	$1 - p$	p

Bảng phân phối xác suất của biến ngẫu nhiên X có phân phối Bernoulli là

X	0	1
$P(X = x_i)$	$1 - p$	p

Cho biến ngẫu nhiên $X \sim B(p)$. Khi đó

$$E(X) = p \text{ và } V(X) = (1 - p)p.$$

Ví dụ 2.5.3

Một câu hỏi trắc nghiệm có 4 phương án trả lời trong đó có một phương án đúng. Một sinh viên chọn ngẫu nhiên một phương án để trả lời câu hỏi đó. Đặt A là biến cố "sinh viên chọn phương án đúng". Khi đó việc trả lời của sinh viên là một phép thử Bernoulli và $p = P(A) = \frac{1}{4}$ và $P(\bar{A}) = \frac{3}{4}$.

Ví dụ 2.5.3

Một câu hỏi trắc nghiệm có 4 phương án trả lời trong đó có một phương án đúng. Một sinh viên chọn ngẫu nhiên một phương án để trả lời câu hỏi đó. Đặt A là biến cố "sinh viên chọn phương án đúng". Khi đó việc trả lời của sinh viên là một phép thử Bernoulli và $p = P(A) = \frac{1}{4}$ và $P(\bar{A}) = \frac{3}{4}$.

Gọi X là biến ngẫu nhiên xác định bởi

$$X = \begin{cases} 1 & \text{nếu sinh viên trả lời đúng} \\ 0 & \text{nếu sinh viên trả lời sai.} \end{cases}$$

Ví dụ 2.5.3

Một câu hỏi trắc nghiệm có 4 phương án trả lời trong đó có một phương án đúng. Một sinh viên chọn ngẫu nhiên một phương án để trả lời câu hỏi đó. Đặt A là biến cố "sinh viên chọn phương án đúng". Khi đó việc trả lời của sinh viên là một phép thử Bernoulli và $p = P(A) = \frac{1}{4}$ và $P(\bar{A}) = \frac{3}{4}$.

Gọi X là biến ngẫu nhiên xác định bởi

$$X = \begin{cases} 1 & \text{nếu sinh viên trả lời đúng} \\ 0 & \text{nếu sinh viên trả lời sai.} \end{cases}$$

Khi đó $X \sim B\left(\frac{1}{4}\right)$ và $E(X) = \frac{1}{4}$, $V(X) = \frac{3}{4} \cdot \frac{1}{4} = \frac{3}{16}$.

2.5.2 Phân phối nhị thức (Binomial distribution)

Định nghĩa 2.5.4 (Dãy phép thử Bernoulli)

Dãy các phép thử lặp lại, độc lập, trong mỗi phép thử chỉ có một trong hai biến cỗ A và \bar{A} , xác suất $P(A) = p$ không đổi được gọi là **dãy phép thử Bernoulli**.

2.5.2 Phân phối nhị thức (Binomial distribution)

Định nghĩa 2.5.4 (Dãy phép thử Bernoulli)

Dãy các phép thử lặp lại, độc lập, trong mỗi phép thử chỉ có một trong hai biến cő A và \bar{A} , xác suất $P(A) = p$ không đổi được gọi là **đãy phép thử Bernoulli**.

Xác suất biến cő A xảy ra k lần trong n phép thử là:

$$P_n(k) = C_n^k p^k (1-p)^{n-k}$$

Công thức nêu trên được gọi là **công thức Bernoulli**.

Ví dụ 2.5.5 Trong phân xưởng có 5 máy hoạt động độc lập. Xác suất để trong mỗi ca mỗi máy bị hỏng đều bằng 0,1.

- Tìm xác suất để trong ca đó có đúng 2 máy hỏng.
- Biết rằng trong một ca có đúng 2 máy hỏng. Tính xác suất để máy thứ nhất không hỏng.

Ví dụ 2.5.5 Trong phân xưởng có 5 máy hoạt động độc lập. Xác suất để trong mỗi ca mỗi máy bị hỏng đều bằng 0,1.

- Tìm xác suất để trong ca đó có đúng 2 máy hỏng.
- Biết rằng trong một ca có đúng 2 máy hỏng. Tính xác suất để máy thứ nhất không hỏng.

Giải. a. Ta xem sự hoạt động của mỗi máy là một phép thử. Ta có 5 phép thử độc lập; trong mỗi phép thử chỉ có 2 trường hợp: hoặc máy hỏng, hoặc máy không hỏng. Xác suất hỏng của mỗi máy đều bằng 0,1.

Ví dụ 2.5.5 Trong phân xưởng có 5 máy hoạt động độc lập. Xác suất để trong mỗi ca mỗi máy bị hỏng đều bằng 0,1.

- Tìm xác suất để trong ca đó có đúng 2 máy hỏng.
- Biết rằng trong một ca có đúng 2 máy hỏng. Tính xác suất để máy thứ nhất không hỏng.

Giải. a. Ta xem sự hoạt động của mỗi máy là một phép thử. Ta có 5 phép thử độc lập; trong mỗi phép thử chỉ có 2 trường hợp: hoặc máy hỏng, hoặc máy không hỏng. Xác suất hỏng của mỗi máy đều bằng 0,1. Như vậy, bài toán thỏa mãn dãy phép thử Bernoulli với $n = 5$; $p = 0,1$ và $k = 2$.

Ví dụ 2.5.5 Trong phân xưởng có 5 máy hoạt động độc lập. Xác suất để trong mỗi ca mỗi máy bị hỏng đều bằng 0,1.

- Tìm xác suất để trong ca đó có đúng 2 máy hỏng.
- Biết rằng trong một ca có đúng 2 máy hỏng. Tính xác suất để máy thứ nhất không hỏng.

Giải. a. Ta xem sự hoạt động của mỗi máy là một phép thử. Ta có 5 phép thử độc lập; trong mỗi phép thử chỉ có 2 trường hợp: hoặc máy hỏng, hoặc máy không hỏng. Xác suất hỏng của mỗi máy đều bằng 0,1. Như vậy, bài toán thỏa mãn dãy phép thử Bernoulli với $n = 5$; $p = 0,1$ và $k = 2$. Theo công thức Bernoulli, ta có xác suất trong ca đó có 2 máy hỏng là

$$C_5^2(0,1)^2 \cdot (1 - 0,1)^3 = 0,0729.$$

b. Gọi A là biến cố "Có đúng 2 máy hỏng" và B là biến cố "Máy thứ nhất không hỏng". Ta cần tính $P(B|A)$.

b. Gọi A là biến cố "Có đúng 2 máy hỏng" và B là biến cố "Máy thứ nhất không hỏng". Ta cần tính $P(B|A)$.

Ta có $P(A) = 0,0729$ và $P(B) = 0,9$. Hơn nữa, $A \cap B$ là biến cố máy thứ nhất không hỏng và có đúng 2 máy hỏng. Do đó

$$P(A \cap B) = 0,9 \cdot C_4^2 \cdot 0,1^2 \cdot 0,9^2 = 0,04374$$

b. Gọi A là biến cố "Có đúng 2 máy hỏng" và B là biến cố "Máy thứ nhất không hỏng". Ta cần tính $P(B|A)$.

Ta có $P(A) = 0,0729$ và $P(B) = 0,9$. Hơn nữa, $A \cap B$ là biến cố máy thứ nhất không hỏng và có đúng 2 máy hỏng. Do đó

$$P(A \cap B) = 0,9 \cdot C_4^2 \cdot 0,1^2 \cdot 0,9^2 = 0,04374$$

Như vậy,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0,04374}{0,0729} = 0,6.$$

Trong dãy phép thử Bernoulli, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là **số có khả năng nhất** (hay số lần xuất hiện chắc chắn nhất).

Trong dãy phép thử Bernoulli, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là **số có khả năng nhất** (hay số lần xuất hiện chắc chắn nhất).

- Nếu $(n + 1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = (n + 1)p - 1$ và $x_0 = (n + 1)p$.

Trong dãy phép thử Bernoulli, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là **số có khả năng nhất** (hay số lần xuất hiện chắc chắn nhất).

- Nếu $(n+1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = (n+1)p - 1$ và $x_0 = (n+1)p$.
- Nếu $(n+1)p \notin \mathbb{Z}$ thì $x_0 = [(n+1)p - 1] + 1$, trong đó $[(n+1)p - 1]$ là phần nguyên của $(n+1)p - 1$.

Ví dụ 2.5.6 Tỷ lệ mắc Covid-19 ở một thành phố nọ là 10%. Trong đợt khám bệnh cho vùng đó, người ta đã khám 100 người. Tìm số người bị Covid-19 có khả năng nhất? Tính xác suất tương ứng.

Trong dãy phép thử Bernoulli, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là **số có khả năng nhất** (hay số lần xuất hiện chắc chắn nhất).

- Nếu $(n + 1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = (n + 1)p - 1$ và $x_0 = (n + 1)p$.
- Nếu $(n + 1)p \notin \mathbb{Z}$ thì $x_0 = [(n + 1)p - 1] + 1$, trong đó $[(n + 1)p - 1]$ là phần nguyên của $(n + 1)p - 1$.

Ví dụ 2.5.6 Tỷ lệ mắc Covid-19 ở một thành phố nọ là 10%. Trong đợt khám bệnh cho vùng đó, người ta đã khám 100 người. Tìm số người bị Covid-19 có khả năng nhất? Tính xác suất tương ứng.

Giải. Đây là một dãy phép thử Bernoulli với $n = 100; p = 0,1$ và cần tìm k sao cho $P_n(k)$ là lớn nhất.

Trong dãy phép thử Bernoulli, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là **số có khả năng nhất** (hay số lần xuất hiện chắc chắn nhất).

- Nếu $(n+1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = (n+1)p - 1$ và $x_0 = (n+1)p$.
- Nếu $(n+1)p \notin \mathbb{Z}$ thì $x_0 = [(n+1)p - 1] + 1$, trong đó $[(n+1)p - 1]$ là phần nguyên của $(n+1)p - 1$.

Ví dụ 2.5.6 Tỷ lệ mắc Covid-19 ở một thành phố nọ là 10%. Trong đợt khám bệnh cho vùng đó, người ta đã khám 100 người. Tìm số người bị Covid-19 có khả năng nhất? Tính xác suất tương ứng.

Giải. Đây là một dãy phép thử Bernoulli với $n = 100; p = 0,1$ và cần tìm k sao cho $P_n(k)$ là lớn nhất. Ta có $(n+1)p = (100+1).0,1 \notin \mathbb{Z}$.

Trong dãy phép thử Bernoulli, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là **số có khả năng nhất** (hay số lần xuất hiện chắc chắn nhất).

- Nếu $(n+1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = (n+1)p - 1$ và $x_0 = (n+1)p$.
- Nếu $(n+1)p \notin \mathbb{Z}$ thì $x_0 = [(n+1)p - 1] + 1$, trong đó $[(n+1)p - 1]$ là phần nguyên của $(n+1)p - 1$.

Ví dụ 2.5.6 Tỷ lệ mắc Covid-19 ở một thành phố nọ là 10%. Trong đợt khám bệnh cho vùng đó, người ta đã khám 100 người. Tìm số người bị Covid-19 có khả năng nhất? Tính xác suất tương ứng.

Giải. Đây là một dãy phép thử Bernoulli với $n = 100; p = 0,1$ và cần tìm k sao cho $P_n(k)$ là lớn nhất. Ta có $(n+1)p = (100+1).0,1 \notin \mathbb{Z}$. Do đó, số có khả năng lớn nhất là

$$x_0 = [(n+1)p - 1] + 1 = [9,1] + 1 = 9 + 1 = 10$$

Trong dãy phép thử Bernoulli, số x_0 mà tại đó xác suất đạt giá trị lớn nhất gọi là **số có khả năng nhất** (hay số lần xuất hiện chắc chắn nhất).

- Nếu $(n+1)p \in \mathbb{Z}$ thì có hai số có khả năng nhất $x_0 = (n+1)p - 1$ và $x_0 = (n+1)p$.
- Nếu $(n+1)p \notin \mathbb{Z}$ thì $x_0 = [(n+1)p - 1] + 1$, trong đó $[(n+1)p - 1]$ là phần nguyên của $(n+1)p - 1$.

Ví dụ 2.5.6 Tỷ lệ mắc Covid-19 ở một thành phố nọ là 10%. Trong đợt khám bệnh cho vùng đó, người ta đã khám 100 người. Tìm số người bị Covid-19 có khả năng nhất? Tính xác suất tương ứng.

Giải. Đây là một dãy phép thử Bernoulli với $n = 100; p = 0,1$ và cần tìm k sao cho $P_n(k)$ là lớn nhất. Ta có $(n+1)p = (100+1).0,1 \notin \mathbb{Z}$. Do đó, số có khả năng lớn nhất là

$$x_0 = [(n+1)p - 1] + 1 = [9,1] + 1 = 9 + 1 = 10$$

và xác suất tương ứng là $P_{100}(10) = C_{100}^{10} \cdot 0,1^{10} \cdot 0,9^{90} \approx 0,1319$.

Định nghĩa 2.5.7 (Phân phối nhị thức)

Xét dãy n phép thử Bernoulli độc lập và xác suất của biến cố A trong mỗi phép thử là p . Đặt X là số lần xuất hiện biến cố A trong n phép thử. Ta nói X là một biến ngẫu nhiên có **phân phối nhị thức**, ký hiệu $X \sim B(n, p)$, với các tham số p, n . Xác suất trong n phép thử có k lần xuất hiện A là

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, 2, \dots, n.$$

Định nghĩa 2.5.7 (Phân phối nhị thức)

Xét dãy n phép thử Bernoulli độc lập và xác suất của biến cố A trong mỗi phép thử là p . Đặt X là số lần xuất hiện biến cố A trong n phép thử. Ta nói X là một biến ngẫu nhiên có **phân phối nhị thức**, ký hiệu $X \sim B(n, p)$, với các tham số p, n . Xác suất trong n phép thử có k lần xuất hiện A là

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, 2, \dots, n.$$

Mệnh đề 2.5.8

Cho biến ngẫu nhiên X có phân phối nhị thức $X \sim B(n, p)$. Khi đó

$$E(X) = np \text{ và } V(X) = n(1 - p)p.$$

Ví dụ 2.5.9

Một đề thi môn xác suất thống kê có 20 câu trắc nghiệm (4 phương án, có 1 phương án đúng). Sinh viên B chọn ngẫu nhiên các phương án trong mỗi câu trả lời. Nếu trả lời đúng 1 câu thì sinh viên B được 0,5 điểm và nếu trả lời sai 1 câu thì được 0 điểm. Tính xác suất sinh viên được đúng 5,5 điểm.

Ví dụ 2.5.9

Một đề thi môn xác suất thống kê có 20 câu trắc nghiệm (4 phương án, có 1 phương án đúng). Sinh viên B chọn ngẫu nhiên các phương án trong mỗi câu trả lời. Nếu trả lời đúng 1 câu thì sinh viên B được 0,5 điểm và nếu trả lời sai 1 câu thì được 0 điểm. Tính xác suất sinh viên được đúng 5,5 điểm.

Giải. Để được đúng 5,5 điểm, sinh viên B phải chọn đúng 11 câu trả lời đúng và 9 câu trả lời sai.

Ví dụ 2.5.9

Một đề thi môn xác suất thống kê có 20 câu trắc nghiệm (4 phương án, có 1 phương án đúng). Sinh viên B chọn ngẫu nhiên các phương án trong mỗi câu trả lời. Nếu trả lời đúng 1 câu thì sinh viên B được 0,5 điểm và nếu trả lời sai 1 câu thì được 0 điểm. Tính xác suất sinh viên được đúng 5,5 điểm.

Giải. Để được đúng 5,5 điểm, sinh viên B phải chọn đúng 11 câu trả lời đúng và 9 câu trả lời sai. Xác suất B được 5,5 điểm là

$$C_{20}^{11} \left(\frac{1}{4}\right)^{11} \left(\frac{3}{4}\right)^9 = 0,003.$$

2.5.3 Phân phối siêu bội (Hypergeometric distribution)

Định nghĩa 2.5.10 (Phân phối siêu bội)

Xét một tập có N phần tử trong đó có M phần tử có tính chất \mathcal{P} và $N - M$ phần tử không có tính chất \mathcal{P} . Chọn ra n phần tử từ N phần tử. Gọi X là số phần tử có tính chất \mathcal{P} trong n phần tử đã chọn. Khi đó X là một biến ngẫu nhiên và ta nói X có **phân phối siêu bội** với ba tham số N, M, n . Ký hiệu $X \sim H(N, M, n)$

2.5.3 Phân phối siêu bội (Hypergeometric distribution)

Định nghĩa 2.5.10 (Phân phối siêu bội)

Xét một tập có N phần tử trong đó có M phần tử có tính chất \mathcal{P} và $N - M$ phần tử không có tính chất \mathcal{P} . Chọn ra n phần tử từ N phần tử. Gọi X là số phần tử có tính chất \mathcal{P} trong n phần tử đã chọn. Khi đó X là một biến ngẫu nhiên và ta nói X có **phân phối siêu bội** với ba tham số N, M, n . Ký hiệu $X \sim H(N, M, n)$

Xác suất trong n phần tử có k phần tử có tính chất \mathcal{P} là

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

trong đó $0 \leq k \leq n$.

Ví dụ 2.5.11 Công ty Z đã nhận được 20 mẫu chào hàng cung cấp với máy in, trong đó 8 máy in PH và 12 máy in Nonac. Chọn ngẫu nhiên 5 mẫu chào hàng. Tính xác suất trong 5 mẫu chào hàng có k ($k = 0, 1, 2, 3, 4$, hoặc 5) máy in Nonac?

Ví dụ 2.5.11 Công ty Z đã nhận được 20 mẫu chào hàng cung cấp với máy in, trong đó 8 máy in PH và 12 máy in Nonac. Chọn ngẫu nhiên 5 mẫu chào hàng. Tính xác suất trong 5 mẫu chào hàng có k ($k = 0, 1, 2, 3, 4$, hoặc 5) máy in Nonac?

Giải. Ta có một tập gồm $N = 20$ phần tử, trong đó có $M = 12$ máy in Nonac và $N - M = 8$ máy in PH. Chọn ra 5 mẫu chào hàng, tức là $n = 5$. Đặt X là số máy in Nonac trong 5 mẫu chào hàng.

Ví dụ 2.5.11 Công ty Z đã nhận được 20 mẫu chào hàng cung cấp với máy in, trong đó 8 máy in PH và 12 máy in Nonac. Chọn ngẫu nhiên 5 mẫu chào hàng. Tính xác suất trong 5 mẫu chào hàng có k ($k = 0, 1, 2, 3, 4$, hoặc 5) máy in Nonac?

Giải. Ta có một tập gồm $N = 20$ phần tử, trong đó có $M = 12$ máy in Nonac và $N - M = 8$ máy in PH. Chọn ra 5 mẫu chào hàng, tức là $n = 5$. Đặt X là số máy in Nonac trong 5 mẫu chào hàng. Khi đó $X \sim H(20, 12, 5)$ và

Ví dụ 2.5.11 Công ty Z đã nhận được 20 mẫu chào hàng cung cấp với máy in, trong đó 8 máy in PH và 12 máy in Nonac. Chọn ngẫu nhiên 5 mẫu chào hàng. Tính xác suất trong 5 mẫu chào hàng có k ($k = 0, 1, 2, 3, 4$, hoặc 5) máy in Nonac?

Giải. Ta có một tập gồm $N = 20$ phần tử, trong đó có $M = 12$ máy in Nonac và $N - M = 8$ máy in PH. Chọn ra 5 mẫu chào hàng, tức là $n = 5$. Đặt X là số máy in Nonac trong 5 mẫu chào hàng. Khi đó $X \sim H(20, 12, 5)$ và

$$P(X = k) = \frac{C_{12}^k C_8^{5-k}}{C_{20}^5}.$$

Mệnh đề 2.5.12

Kỳ vọng và phương sai của một phân phối siêu bội $X \sim H(N, M, n)$ được xác định như sau:

$$E(X) = n \cdot \frac{M}{N} \text{ và } V(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

Mệnh đề 2.5.12

Kỳ vọng và phương sai của một phân phối siêu bội $X \sim H(N, M, n)$ được xác định như sau:

$$E(X) = n \cdot \frac{M}{N} \text{ và } V(X) = \frac{N-n}{N-1} \cdot n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right)$$

Ví dụ 2.5.13 Trong 100 laptop trưng bày tại cửa hàng POTPAL có 70 laptop hiệu PH. Chọn ngẫu nhiên 40 laptop trong số 100 laptop. Gọi X là số laptop hiệu PH trong số 40 laptop được chọn.

- Tính xác suất để chọn được từ 27 đến 29 laptop PH.
 - Tính trung bình số laptop PH được chọn và $V(X)$.
-
.....
.....
.....
.....
.....
.....
.....
.....
.....

2.5.4 Phân phối Poisson (Poisson distribution)

Phân phối Poisson là một phân phối xác suất của biến ngẫu nhiên rời rạc. Thông tin mà ta biết **không phải** là **xác suất của một biến cố** trong một lần thử như trong phân phối Bernoulli hay là **số lần xảy ra biến cố** đó trong n lần thử như trong phân phối nhị thức, mà **ta biết trung bình số lần xảy ra của một biến cố trong một khoảng thời gian nhất định**. Giá trị trung bình này được kí hiệu là μ . Phân phối Poisson còn được dùng cho khoảng mà đơn vị khác thời gian như: khoảng cách, diện tích hay thể tích.



Siméon Denis Poisson (1781 - 1840) là nhà toán học, nhà vật lý người Pháp.Ảnh nguồn: Internet.

Ví dụ 2.5.14 Một bệnh viện có trung bình 10 ca sinh mỗi giờ. Đặt X là số ca sinh trong một giờ. Khi đó X là một biến ngẫu nhiên có phân phối Poisson.

Chúng ta có thể sử dụng phân phối Poisson để trả lời các câu hỏi chẳng hạn như:

- Xác suất để có nhiều hơn 12 lần sinh xảy ra trong một giờ nhất định là bao nhiêu?
- Xác suất để xảy ra ít hơn 5 ca sinh trong một giờ nhất định?
- Xác suất để từ 8 đến 11 ca sinh xảy ra trong một giờ nhất định là bao nhiêu?

Dịnh nghĩa 2.5.15 (Phân phối Poisson)

Một biến ngẫu nhiên rời rạc X được gọi là có **phân phối Poisson** với tham số $\lambda (\lambda > 0)$, ký hiệu $X \sim P(\lambda)$, nếu nó có phân phối xác suất là

$$P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, x = 0, 1, 2, \dots$$

trong đó λ là trung bình số lần xuất hiện biến cố và x là số lần xuất hiện biến cố mà ta quan tâm.

Dịnh nghĩa 2.5.15 (Phân phối Poisson)

Một biến ngẫu nhiên rời rạc X được gọi là có **phân phối Poisson** với tham số $\lambda (\lambda > 0)$, ký hiệu $X \sim P(\lambda)$, nếu nó có phân phối xác suất là

$$P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, x = 0, 1, 2, \dots$$

trong đó λ là trung bình số lần xuất hiện biến cố và x là số lần xuất hiện biến cố mà ta quan tâm.

Mệnh đề 2.5.16

Cho X là một biến ngẫu nhiên có phân phối Poisson $X \sim P(\lambda)$. Khi đó

$$E(X) = V(X) = \lambda.$$

Dịnh nghĩa 2.5.15 (Phân phối Poisson)

Một biến ngẫu nhiên rời rạc X được gọi là có **phân phối Poisson** với tham số $\lambda (\lambda > 0)$, ký hiệu $X \sim P(\lambda)$, nếu nó có phân phối xác suất là

$$P(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, x = 0, 1, 2, \dots$$

trong đó λ là trung bình số lần xuất hiện biến cố và x là số lần xuất hiện biến cố mà ta quan tâm.

Mệnh đề 2.5.16

Cho X là một biến ngẫu nhiên có phân phối Poisson $X \sim P(\lambda)$. Khi đó

$$E(X) = V(X) = \lambda.$$

Giá trị của hàm phân phối xác suất $F(X) = P(X \leq x)$ của biến ngẫu nhiên có phân phối Poisson có thể tìm trong bảng A3.



Ví dụ 2.5.17 Số lượng cơn bão lớn trung bình trong thành phố T là 2 cơn bão mỗi năm. Xác suất chính xác 3 cơn bão sẽ đổ bộ vào thành phố T trong năm tới là bao nhiêu?

.....
.....
.....
.....
.....
.....
.....
.....
.....

Ví dụ 2.5.18

Giả sử có 200 lỗi đánh máy được phân bố ngẫu nhiên trong bản thảo 500 trang sách. Tìm xác suất một trang đã cho có đúng 3 lỗi.

Ví dụ 2.5.18

Giả sử có 200 lỗi đánh máy được phân bố ngẫu nhiên trong bản thảo 500 trang sách. Tìm xác suất một trang đã cho có đúng 3 lỗi.

Giải. Đầu tiên, ta sẽ tìm λ trung bình số lỗi của 1 trang. Ta có

$$\lambda = \dots\dots\dots$$

Xác suất để một trang có 3 lỗi là

$$P(X = 3) = \frac{e^{-\lambda} \cdot (\lambda)^3}{3!} = 0,0072.$$

Bài tập Giả sử rằng số lần truy cập trung bình vào một trang web là 2 lần trong 1 giây.

- a. Tính xác suất có 5 lần truy cập trong 1 giây.
- b. Tính xác suất có nhiều nhất 5 lần truy cập trong 1 giây.
- c. Tính xác suất có nhiều hơn 5 lần truy cập trong 2 giây.

Giải. a. Đặt X là số lần truy cập trong 1 giây. Khi đó X có phân phối Poisson với tham số $\lambda = 2$. Suy ra

$$P(X = 5) = \frac{e^{-2} \cdot 2^5}{5!} = 0,0361.$$

b. Dùng bảng A3, ta có

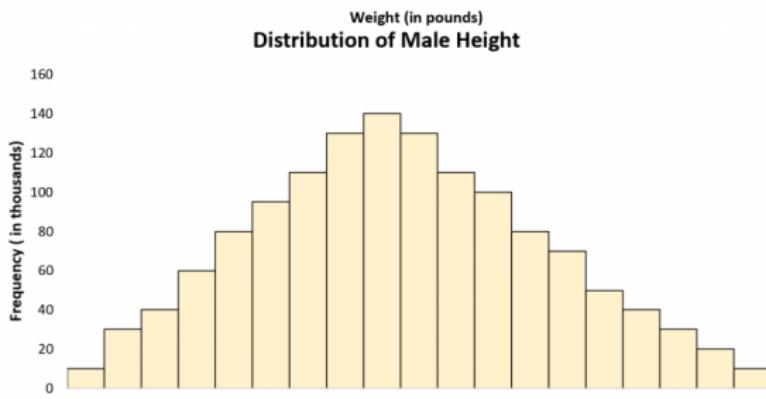
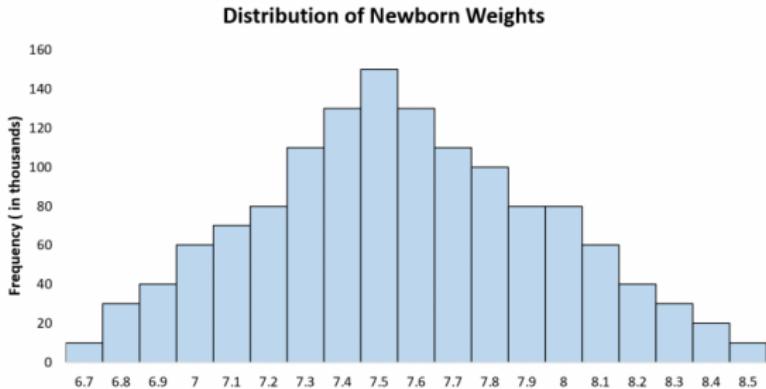
$$P(X \leq 5) = 0,983$$

c. Trung bình số lần truy cập trong 2 giây là 4. Đặt Y là số lần truy cập vào trang web trong 2 giây. Khi đó $Y \sim P(4)$ và do đó

$$P(Y > 5) = 1 - P(Y \leq 5) = 1 - 0,785 = 0,215$$

(Dùng bảng A3, ta có $P(Y \leq 5) = 0,785$)

2.5.5 Phân phối chuẩn



Phân phối chuẩn (Normal distribution) là phân phối quan trọng nhất trong xác suất và thống kê. Phân phối chuẩn được nêu ra bởi một người Anh gốc Pháp tên là Abraham de Moivre (1733). Sau đó, Gauss, một nhà toán học người Đức, đã dùng phân phối chuẩn để nghiên cứu các dữ liệu về thiên văn học (1809) và do vậy nó cũng được gọi là **phân phối Gauss**.



Johann Carl Friedrich Gauss (1777-1855) và Abraham de Moivre (1667-1754).

Nguồn: [Wikipeadia.org](#)

Định nghĩa 2.5.19 (Phân phối chuẩn-Normal distribution)

Một biến ngẫu nhiên liên tục X được gọi là có **phân phối chuẩn** nếu hàm mật độ xác suất của nó có dạng

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

Ký hiệu $X \sim N(\mu; \sigma^2)$.

Định nghĩa 2.5.19 (Phân phối chuẩn-Normal distribution)

Một biến ngẫu nhiên liên tục X được gọi là có **phân phối chuẩn** nếu hàm mật độ xác suất của nó có dạng

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

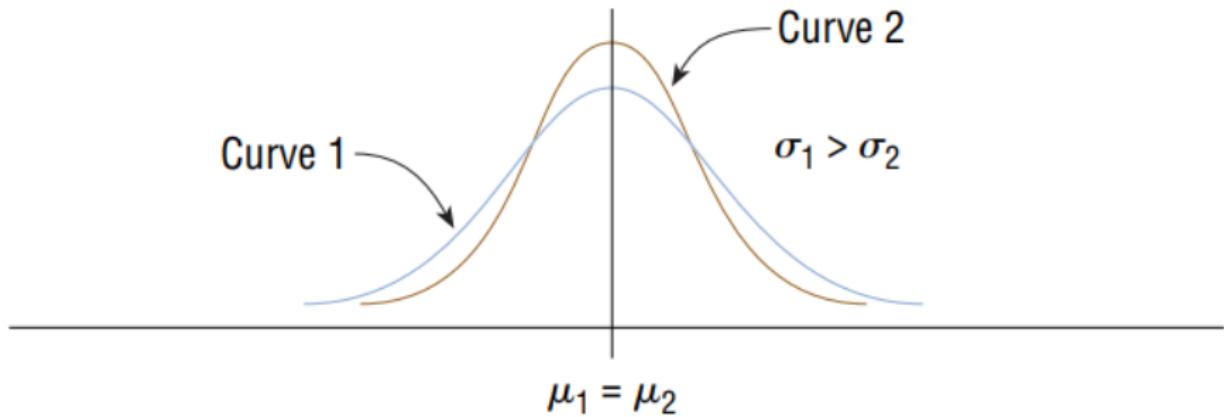
Ký hiệu $X \sim N(\mu; \sigma^2)$.

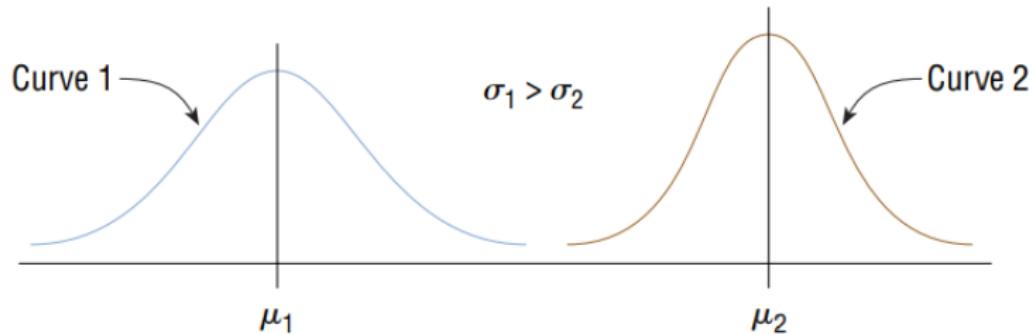
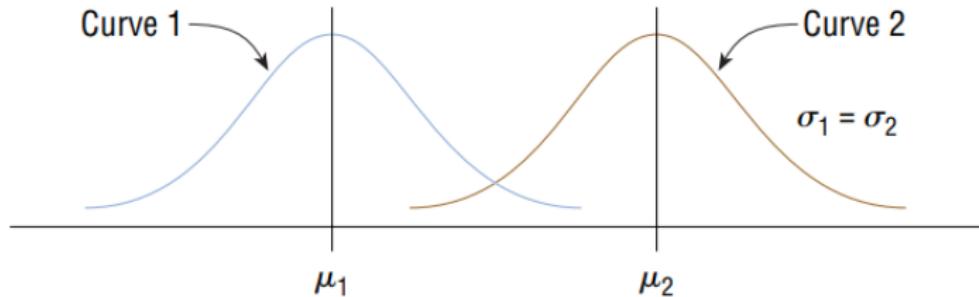
Mệnh đề 2.5.20

Cho $X \sim N(\mu; \sigma^2)$. Khi đó

$$E(X) = \mu \text{ và } V(X) = \sigma^2.$$

Hình dạng và vị trí của đường cong phân phối chuẩn (đồ thị của hàm mật độ xác suất) phụ thuộc vào hai tham số μ và σ .





Một số tính chất của phân phối chuẩn

Một số tính chất của phân phối chuẩn

- ① Kỳ vọng, trung vị bằng nhau và nằm ở trung điểm của phân phối.

Một số tính chất của phân phối chuẩn

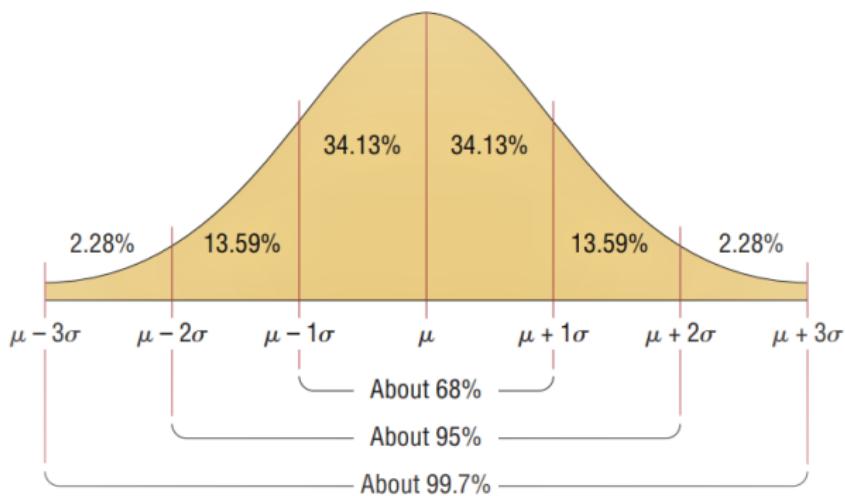
- ① Kỳ vọng, trung vị bằng nhau và nằm ở trung điểm của phân phối.
- ② Đồ thị của hàm mật độ xác suất có hình chuông, đối xứng qua đường thẳng đứng đi qua kỳ vọng.

Một số tính chất của phân phối chuẩn

- ① Kỳ vọng, trung vị bằng nhau và nằm ở trung điểm của phân phối.
- ② Đồ thị của hàm mật độ xác suất có hình chuông, đối xứng qua đường thẳng đứng đi qua kỳ vọng.
- ③ Đường cong liên tục và không chạm vào trục x .

Một số tính chất của phân phối chuẩn

- ① Kỳ vọng, trung vị bằng nhau và nằm ở trung điểm của phân phối.
- ② Đồ thị của hàm mật độ xác suất có hình chuông, đối xứng qua đường thẳng đứng đi qua kỳ vọng.
- ③ Đường cong liên tục và không chạm vào trục x .
- ④ Diện tích dưới phần của đường cong phân phối chuẩn trong đoạn $[\mu - \sigma; \mu + \sigma]$ là khoảng 0,68; trong đoạn $[\mu - 2\sigma; \mu + 2\sigma]$ là 0,95; và trong khoảng $[\mu - 3\sigma; \mu + 3\sigma]$ là khoảng 0,997.



Định nghĩa 2.5.21

Phân phối chuẩn $N(0; 1)$ được gọi là **phân phối chuẩn chuẩn tắc** (standard normal distribution).

Định nghĩa 2.5.21

Phân phối chuẩn $N(0; 1)$ được gọi là **phân phối chuẩn chuẩn tắc** (standard normal distribution).

Các phân phối chuẩn có thể được chuyển thành phân phối chuẩn chuẩn tắc bằng cách đổi biến

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

Định nghĩa 2.5.21

Phân phối chuẩn $N(0; 1)$ được gọi là **phân phối chuẩn chuẩn tắc** (standard normal distribution).

Các phân phối chuẩn có thể được chuyển thành phân phối chuẩn chuẩn tắc bằng cách đổi biến

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

Biến ngẫu nhiên Z có hàm mật độ xác suất

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

Định nghĩa 2.5.21

Phân phối chuẩn $N(0; 1)$ được gọi là **phân phối chuẩn chuẩn tắc** (standard normal distribution).

Các phân phối chuẩn có thể được chuyển thành phân phối chuẩn chuẩn tắc bằng cách đổi biến

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

Biến ngẫu nhiên Z có hàm mật độ xác suất

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

và hàm phân phối xác suất

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

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

① Đổi biến $Z = \frac{X - \mu}{\sigma} \Rightarrow Z \sim N(0; 1)$.

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

① Đổi biến $Z = \frac{X - \mu}{\sigma} \Rightarrow Z \sim N(0; 1)$.

② $P(X \leq a) = P(Z \leq \frac{a - \mu}{\sigma}) = \Phi(\frac{a - \mu}{\sigma})$

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

① Đổi biến $Z = \frac{X - \mu}{\sigma} \Rightarrow Z \sim N(0; 1)$.

② $P(X \leq a) = P(Z \leq \frac{a - \mu}{\sigma}) = \Phi(\frac{a - \mu}{\sigma})$

③ Tìm $\Phi(z) = P(Z \leq z)$ từ Bảng A4

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

① Đổi biến $Z = \frac{X - \mu}{\sigma} \Rightarrow Z \sim N(0; 1)$.

② $P(X \leq a) = P(Z \leq \frac{a - \mu}{\sigma}) = \Phi(\frac{a - \mu}{\sigma})$

③ Tìm $\Phi(z) = P(Z \leq z)$ từ Bảng A4

④ Xác định giá trị của z như sau: cột đầu tiên bên trái là hai chữ số đầu tiên của z , hàng ngang đầu tiên là chữ số thứ ba của z .

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

① Đổi biến $Z = \frac{X - \mu}{\sigma} \Rightarrow Z \sim N(0; 1)$.

② $P(X \leq a) = P(Z \leq \frac{a - \mu}{\sigma}) = \Phi(\frac{a - \mu}{\sigma})$

③ Tìm $\Phi(z) = P(Z \leq z)$ từ Bảng A4

④ Xác định giá trị của z như sau: cột đầu tiên bên trái là hai chữ số đầu tiên của z , hàng ngang đầu tiên là chữ số thứ ba của z .

⑤ Giao của cột và hàng này là giá trị $\Phi(z) = P(Z \leq z)$.

Bài toán. Giả sử $X \sim N(\mu; \sigma^2)$. Tính $P(X \leq a)$.

① Đổi biến $Z = \frac{X - \mu}{\sigma} \Rightarrow Z \sim N(0; 1)$.

② $P(X \leq a) = P(Z \leq \frac{a - \mu}{\sigma}) = \Phi(\frac{a - \mu}{\sigma})$

③ Tìm $\Phi(z) = P(Z \leq z)$ từ Bảng A4

④ Xác định giá trị của z như sau: cột đầu tiên bên trái là hai chữ số đầu tiên của z , hàng ngang đầu tiên là chữ số thứ ba của z .

⑤ Giao của cột và hàng này là giá trị $\Phi(z) = P(Z \leq z)$.

Ví dụ 2.5.22

Cho biến ngẫu nhiên X có phân phối chuẩn chuẩn tắc $X \sim N(0; 1)$.

a. Tính $P(X \leq 1,35)$.

b. Tính $P(-1,37 \leq X \leq 1,68)$.

Giải. a. Tra bảng A4, ta thấy

Table A4, continued. Standard Normal distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

Giải. a. Tra bảng A4, ta thấy

Table A4, continued. Standard Normal distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

$$P(X \leq 1,35) = \Phi(1,35) = 0,9115.$$

Giải. a. Tra bảng A4, ta thấy

Table A4, continued. Standard Normal distribution

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545

$$P(X \leq 1,35) = \Phi(1,35) = 0,9115.$$

b. Ta thấy

$$P(-1,37 \leq X \leq 1,68) = \Phi(1,68) - \Phi(-1,37) = 0,9535 - 0,0853 = 0,8682$$

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.
Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.

Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình.

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.

Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình. Theo giả thiết

$$X \sim N(900; 200^2) \text{ với } \mu = 900 \text{ và } \sigma = 200.$$

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.

Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình. Theo giả thiết

$X \sim N(900; 200^2)$ với $\mu = 900$ và $\sigma = 200$. Đặt $Z = \frac{X - 900}{200}$, khi đó

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.

Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình. Theo giả thiết

$X \sim N(900; 200^2)$ với $\mu = 900$ và $\sigma = 200$. Đặt $Z = \frac{X - 900}{200}$, khi đó

$$P(600 \leq X \leq 1200)$$

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.

Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình. Theo giả thiết

$X \sim N(900; 200^2)$ với $\mu = 900$ và $\sigma = 200$. Đặt $Z = \frac{X - 900}{200}$, khi đó

$$P(600 \leq X \leq 1200) = P\left(\frac{600 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{1200 - \mu}{\sigma}\right)$$

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.

Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình. Theo giả thiết

$X \sim N(900; 200^2)$ với $\mu = 900$ và $\sigma = 200$. Đặt $Z = \frac{X - 900}{200}$, khi đó

$$\begin{aligned} P(600 \leq X \leq 1200) &= P\left(\frac{600 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{1200 - \mu}{\sigma}\right) \\ &= P(-1,5 \leq Z \leq 1,5) \end{aligned}$$

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.

Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình. Theo giả thiết

$X \sim N(900; 200^2)$ với $\mu = 900$ và $\sigma = 200$. Đặt $Z = \frac{X - 900}{200}$, khi đó

$$\begin{aligned}P(600 \leq X \leq 1200) &= P\left(\frac{600 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{1200 - \mu}{\sigma}\right) \\&= P(-1,5 \leq Z \leq 1,5) \\&= \Phi(1,5) - \Phi(-1,5)\end{aligned}$$

Ví dụ 2.5.23 Giả sử thu nhập của các gia đình trong một tháng của một quốc gia có phân phối chuẩn với trung bình là 900\$ và độ lệch chuẩn là 200\$. Gặp ngẫu nhiên một gia đình.

- a. Tính xác suất gia đình này có thu nhập từ 600\$ đến 1200\$.
- b. Chính phủ sẽ trợ cấp cho 0,3% gia đình có mức thu nhập thấp nhất.

Hỏi các gia đình này có thu nhập nhiều nhất là bao nhiêu đôla?

Giải. a. Đặt X là thu nhập của một gia đình. Theo giả thiết

$X \sim N(900; 200^2)$ với $\mu = 900$ và $\sigma = 200$. Đặt $Z = \frac{X - 900}{200}$, khi đó

$$\begin{aligned}P(600 \leq X \leq 1200) &= P\left(\frac{600 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{1200 - \mu}{\sigma}\right) \\&= P(-1,5 \leq Z \leq 1,5) \\&= \Phi(1,5) - \Phi(-1,5) \\&= 0,9332 - 0,0668 = 0,8664.\end{aligned}$$

Xác suất gia đình này có thu nhập từ 600\$ đến 1200\$ là 86,64%.

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \leq t) = P(Z \leq \frac{t - \mu}{\sigma}) = \Phi(\frac{t - \mu}{\sigma}) = 0,003$$

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \leq t) = P(Z \leq \frac{t - \mu}{\sigma}) = \Phi(\frac{t - \mu}{\sigma}) = 0,003$$

Tra bảng A4,

z	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	-0.00
-(3.9+)	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.7	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.6	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
-3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
-3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007
-3.1	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010
-3.0	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
-2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
-2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
-2.7	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
-2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \leq t) = P(Z \leq \frac{t - \mu}{\sigma}) = \Phi(\frac{t - \mu}{\sigma}) = 0,003$$

Tra bảng A4,

z	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	-0.00
-(3.9+)	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000	.0000
-3.8	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.7	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
-3.6	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0002	.0002
-3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
-3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
-3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
-3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007
-3.1	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010
-3.0	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
-2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
-2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
-2.7	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
-2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \leq t) = P(Z \leq \frac{t - \mu}{\sigma}) = \Phi(\frac{t - \mu}{\sigma}) = 0,003$$

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \leq t) = P(Z \leq \frac{t - \mu}{\sigma}) = \Phi(\frac{t - \mu}{\sigma}) = 0,003$$

Ta thấy $z = -2,75$. Suy ra

$$\frac{t - 900}{200} = -2,75$$

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \leq t) = P(Z \leq \frac{t - \mu}{\sigma}) = \Phi(\frac{t - \mu}{\sigma}) = 0,003$$

Ta thấy $z = -2,75$. Suy ra

$$\frac{t - 900}{200} = -2,75$$

hay $t = 900 - 2,75.200 = 350$.

b. Gọi t là thu nhập (tính bằng \$) của các gia đình thỏa mãn

$$P(X \leq t) = 0,3\% = 0,003.$$

Như vậy

$$P(X \leq t) = P(Z \leq \frac{t - \mu}{\sigma}) = \Phi(\frac{t - \mu}{\sigma}) = 0,003$$

Ta thấy $z = -2,75$. Suy ra

$$\frac{t - 900}{200} = -2,75$$

hay $t = 900 - 2,75 \cdot 200 = 350$. Như vậy, các gia đình này có thu nhập không quá 350\$.

Bài tập 1 Một thành

phố lắp đặt đèn điện để chiếu sáng đường. Loại đèn này có tuổi thọ trung bình là 1000 giờ với độ lệch chuẩn là 200 giờ. Giả sử tuổi thọ của loại đèn này có phân phối chuẩn.

- Tính xác suất đèn này bị hỏng trong 700 giờ.
- Tính xác suất đèn sử dụng được từ 900 giờ đến 1300 giờ.
- Sau bao lâu thì còn 10% đèn sử dụng được?



Bài tập 2 Các nghiên cứu cho thấy quãng đường các loại xe ôtô cỡ nhỏ chạy được với 1 lít xăng có phân phối chuẩn với trung bình là 15,1 km/lít và độ lệch chuẩn là 1,92 km/lít.

- Có bao nhiêu phần trăm xe cỡ nhỏ chạy được nhiều hơn 17 km/lít?

b. Các xe tiết kiệm xăng nhất chiếm 7,15%. Quãng đường mà những xe này chạy với 1 lít xăng ít nhất là bao nhiêu?



Bài tập 1 Một thành

phố lắp đặt đèn điện để chiếu sáng đường. Loại đèn này có tuổi thọ trung bình là 1000 giờ với độ lệch chuẩn là 200 giờ. Giả sử tuổi thọ của loại đèn này có phân phối chuẩn.

- Tính xác suất đèn này bị hỏng trong 700 giờ.
- Tính xác suất đèn sử dụng được từ 900 giờ đến 1300 giờ.
- Sau bao lâu thì còn 10% đèn sử dụng được?



Bài tập 2 Các nghiên cứu cho thấy quãng đường các loại xe ôtô cỡ nhỏ chạy được với 1 lít xăng có phân phối chuẩn với trung bình là 15,1 km/lít và độ lệch chuẩn là 1,92 km/lít.

- Có bao nhiêu phần trăm xe cỡ nhỏ chạy được nhiều hơn 17 km/lít?

b. Các xe tiết kiệm xăng nhất chiếm 7,15%. Quãng đường mà những xe này chạy với 1 lít xăng ít nhất là bao nhiêu?



2.5.6 Phân phối χ^2 (Chi-Squared Distribution)

Nếu X là một biến ngẫu nhiên có phân phối chuẩn thì biến ngẫu nhiên $Y = X^2$ được gọi là có phân phối χ^2 với bậc tự do là 1.

2.5.6 Phân phối χ^2 (Chi-Squared Distribution)

Nếu X là một biến ngẫu nhiên có phân phối chuẩn thì biến ngẫu nhiên $Y = X^2$ được gọi là có phân phối χ^2 với bậc tự do là 1. Nếu các biến ngẫu nhiên $X_i \sim N(0; 1)$ ($1 \leq i \leq n$) độc lập thì biến ngẫu nhiên

$$Y = X_1^2 + X_2^2 + \cdots + X_n^2$$

được gọi là có phân phối **chi-bình phương** với bậc tự do là n .

2.5.6 Phân phối χ^2 (Chi-Squared Distribution)

Nếu X là một biến ngẫu nhiên có phân phối chuẩn thì biến ngẫu nhiên $Y = X^2$ được gọi là có phân phối χ^2 với bậc tự do là 1. Nếu các biến ngẫu nhiên $X_i \sim N(0; 1)$ ($1 \leq i \leq n$) độc lập thì biến ngẫu nhiên

$$Y = X_1^2 + X_2^2 + \cdots + X_n^2$$

được gọi là có phân phối **chi-bình phương** với bậc tự do là n .

Định nghĩa 2.5.24 (Hàm gamma)

Cho $\alpha > 0$, hàm gamma $\Gamma(\alpha)$ được xác định bởi

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$$

Mệnh đề 2.5.25

- ① Với $\alpha > 1$, ta có $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.

Mệnh đề 2.5.25

- ① VỚI $\alpha > 1$, ta có $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
- ② VỚI n là một số nguyên dương, ta có $\Gamma(n) = (n - 1)!$

Mệnh đề 2.5.25

- ① VỚI $\alpha > 1$, ta có $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
- ② VỚI n là một số nguyên dương, ta có $\Gamma(n) = (n - 1)!$
- ③ $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

Mệnh đề 2.5.25

- ① Với $\alpha > 1$, ta có $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$.
- ② Với n là một số nguyên dương, ta có $\Gamma(n) = (n - 1)!$
- ③ $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Định nghĩa 2.5.26 (Phân phối χ^2)

Cho ν là một số nguyên dương. Một biến ngẫu nhiên X được gọi là có phân phối χ^2 (chi bình phương) với tham số ν , ký hiệu $X \sim \chi_{\nu}^2$, nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \begin{cases} \frac{1}{2^{\nu/2}\Gamma(\nu/2)}x^{(\nu/2)-1}e^{-x/2}, & x \geq 0 \\ 0, & x < 0. \end{cases}$$

Tham số ν được gọi là **bậc tự do** của X .

2.5.7 Phân phối Student (Student distribution))

Định nghĩa 2.5.27 (Phân phối Student)

Một biến ngẫu nhiên liên tục X được gọi là có phân phối Student với bậc tự do $\nu - 1$, ký hiệu $X \sim St(\nu)$, nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

2.5.7 Phân phối Student (Student distribution))

Dịnh nghĩa 2.5.27 (Phân phối Student)

Một biến ngẫu nhiên liên tục X được gọi là có phân phối Student với bậc tự do $\nu - 1$, ký hiệu $X \sim St(\nu)$, nếu hàm mật độ xác suất của nó có dạng

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

Ta có thể viết $X \sim \frac{N(0; 1)}{\sqrt{\chi_{\nu}^2/\nu}}$.

Phân phối Student sẽ được dùng trong việc ước lượng khoảng của trung bình trong chương 5.



William Sealey Gosset.
Nguồn:
[wikipedia.org](https://en.wikipedia.org)

William Sealey Gosset (1876-1937) học toán và hóa học tại Đại học Oxford ở Anh. Năm 1899, ông chuyển đến Dublin, Ireland và làm nhân viên thống kê cho nhà máy bia Guinness. Trong quá trình nghiên cứu chất lượng lúa mạch và hoa bia, Gosset đề xuất việc sử dụng t -phân phối. Năm 1908, ông đã công bố ý tưởng của mình trong một bài báo khoa học bằng bút danh "Student" vì Guinness cấm nhân viên của mình xuất bản kết quả nghiên cứu của riêng họ. Do đó, t -phân phối thường được gọi là phân phối Student.

2.5.8 Phân phối Fisher (F-distribution)

Dịnh nghĩa 2.5.28 (Phân phối Fisher)

Cho hai biến ngẫu nhiên độc lập có phân phối χ^2 với các bậc tự do lần lượt là ν_1 và ν_2 . Khi đó tỉ số

$$\frac{\chi_{\nu_1}^2 / \nu_1}{\chi_{\nu_2}^2 / \nu_2}$$

được gọi là *F*-phân phối hay phân phối Fisher (*F*-distribution), ký hiệu là F_{ν_1, ν_2} .



Ronald Aylmer Fisher năm 1913. Nguồn: wikipedia.org

Ronald Aylmer Fisher (1890 - 1962) là một nhà toán học, thống kê, nhà di truyền học và học thuật người Anh.

2.6 Hàm đặc trưng

Hàm đặc trưng

Định nghĩa 2.6.1

Hàm đặc trưng của biến ngẫu nhiên X là một hàm số $\phi(t) : \mathbb{R} \rightarrow \mathbb{C}$ xác định bởi

Hàm đặc trưng

Định nghĩa 2.6.1

Hàm đặc trưng của biến ngẫu nhiên X là một hàm số $\phi(t) : \mathbb{R} \rightarrow \mathbb{C}$ xác định bởi

- Nếu X là biến ngẫu nhiên rời rạc thì

$$\phi(t) = \sum_{j \geq 1} e^{itx_j} P(X = x_j)$$

Hàm đặc trưng

Định nghĩa 2.6.1

Hàm đặc trưng của biến ngẫu nhiên X là một hàm số $\phi(t) : \mathbb{R} \rightarrow \mathbb{C}$ xác định bởi

- Nếu X là biến ngẫu nhiên rời rạc thì

$$\phi(t) = \sum_{j \geq 1} e^{itx_j} P(X = x_j)$$

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

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx,$$

Hàm đặc trưng

Định nghĩa 2.6.1

Hàm đặc trưng của biến ngẫu nhiên X là một hàm số $\phi(t) : \mathbb{R} \rightarrow \mathbb{C}$ xác định bởi

- Nếu X là biến ngẫu nhiên rời rạc thì

$$\phi(t) = \sum_{j \geq 1} e^{itx_j} P(X = x_j)$$

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

$$\phi(t) = \int_{-\infty}^{+\infty} e^{itx} f(x) dx,$$

trong đó $e^{itx} = \cos tx + i \sin tx$ và $i^2 = -1$.

Phân phối của một biến ngẫu nhiên có thể được xác định hoàn toàn thông qua hàm đặc trưng của nó. Hàm đặc trưng cũng cho ta một tiêu chuẩn để xác định xem khi nào biến ngẫu nhiên có hàm mật độ.

Biến ngẫu nhiên có phân phối	Tham số	Hàm đặc trưng
Bernoulli	p	$\phi(t) = 1 - p + pe^{it}$
Nhị thức	n, p	$\phi(t) = (1 - p + pe^{it})^n$
Siêu bội	p	$\phi(t) = \frac{p}{1 - e^{it}(1 - p)}$
Poisson	μ	$\phi(t) = e^{\mu(e^{it} - 1)}$
Chuẩn	μ, σ^2	$\phi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$
Chi-bình phương	ν	$\phi(t) = (1 - 2it)^{-\frac{\nu}{2}}$