CHƯƠNG 2: BIẾN NGẪU NHIÊN VÀ CÁC ĐẶC TRƯNG CỦA CHÚNG

2.1 ĐỊNH NGHĨA VÀ PHÂN LOẠI BIẾN NGẪU NHIÊN

Gieo một con xúc xắc 6 mặt. Ký hiệu A_1, A_2, A_3, A_4, A_5, A_6 lần lượt là biến cố "mặt 1 chấm xuất hiện", "mặt 2 chấm xuất hiện", …, "mặt 6 chấm xuất hiện".

Nếu X là đại lượng biểu diễn số chấm xuất hiện khi gieo con xúc xắc thì X có thể nhận các giá trị 1, 2, 3, 4, 5, 6 một cách ngẫu nhiên và X nhận giá trị k là biến cố A_k, nghĩa là $\{X = k\} = A_k$, với $k = 1, 2, ..., 6$.

Ta gọi X là một biến ngẫu nhiên có miền giá trị $R_X = \{1, 2, ..., 6\}$.

Một cách tổng quát ta có khái niệm biến ngẫu nhiên như sau.

2.1.1 Định nghĩa biến ngẫu nhiên

Định nghĩa 2.1: *Biến ngẫu nhiên X là đại lượng nhận các giá trị nào đó phụ thuộc vào các yếu tố ngẫu nhiên, nghĩa là với mọi giá trị thực $x \in \mathbb{R}$ thì " X nhận giá trị nhỏ hơn bằng x ", ký hiệu $\{X \le x\}$, là một biến cố ngẫu nhiên.*

Đối với biến ngẫu nhiên người ta chỉ quan tâm xem nó nhận một giá trị nào đó hoặc nhận giá trị trong một khoảng nào đó với một xác suất bao nhiêu.

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

Ví dụ 2.1: Gieo đồng thời hai con xúc xắc. Ký hiệu A_k, $k = 2,3,...,12$ là biến cố tổng số chấm xuất hiện của hai con xúc xắc là k.

Nếu gọi X là tổng số chấm xuất hiện khi gieo hai con xúc xắc thì X là một biến ngẫu nhiên có miền giá trị $R_X = \{2,3,...,12\}$ và $\{X = k\} = A_k$ với $k = 2,3,...,12$.

Ví dụ 2.2: Các đại lượng sau là biến ngẫu nhiên

- Tuổi thọ của một thiết bị đang hoạt động.
- Số khách hàng vào một điểm phục vụ trong một khoảng thời gian nào đó.
- Số cuộc gọi đến một tổng đài trong một khoảng thời gian nào đó.
- Sai số khi đo lường một đại lượng vật lý …

Định nghĩa 2.2: *Hai biến ngẫu nhiên X, Y là độc lập nếu X nhận các giá trị nào đó không phụ thuộc Y và ngược lại. Nói cách khác với mọi số thực x, y hai biến cố sau là độc lập*

$$\{X \le x\}, \{Y \le y\}.$$

Trong chương 3 ta sẽ đưa ra các tiêu chuẩn để nhận biết tính chất độc lập của hai biến ngẫu nhiên.

2.1.2 Hàm phân bố xác suất

Các biến ngẫu nhiên được xét trong các phép thử khác nhau (tương ứng với các không gian xác suất khác nhau) nhưng các phân bố xác suất của chúng có thể như nhau. Chẳng hạn xác suất bắn trúng bia của một xạ thủ là 0,8. Xạ thủ này bắn 10 viên, gọi X là số viên bắn trúng bia thì xác suất để xạ thủ bắn trúng k viên là

$$P\{X=k\}=C_{10}^{k}(0,6)^{k}(0,4)^{10-k},\ 0\le k\le 10$$

(xem dãy phép thử Bernoulli mục 1.4 chương 1 và phân bố nhị thức, mục 2.2.2.2).

Tương tự, giả sử tỷ lệ chính phẩm của lô hàng là 0,8. Chọn 10 sản phẩm kiểm tra, gọi Y là số chính phẩm phát hiện được thì xác suất chọn được k chính phẩm là

$$P\{Y=k\}=C_{10}^{k}(0,6)^{k}(0,4)^{10-k},\ 0\le k\le 10.$$

Vậy $P\{X=k\}=P\{Y=k\}$, $0\le k\le 10$. Nói cách khác phân bố xác suất của X và Y như nhau, mặc dù X và Y là hai biến ngẫu nhiên được xét trong hai phép thử khác nhau.

Phân bố xác suất được nghiên cứu thông qua hàm phân bố xác suất định nghĩa như sau.

Định nghĩa 2.3: *Hàm phân bố xác suất (cumulative distribution function, viết tắt CDF) của biến ngẫu nhiên X là hàm số $F_X(x)$ xác định với mọi $x\in\mathbb{R}$ bởi công thức*:

$$F_X(x)=P\{X\le x\};\ -\infty<x<\infty \tag{2.1}$$

trong đó $\{X\le x\}$ là ký hiệu biến cố "biến ngẫu nhiên X nhận giá trị nhỏ hơn hay bằng x".

Hàm phân bố có các tính chất sau:

a) $0\le F_X(x)\le 1$ với mọi $x\in\mathbb{R}$, **(2.2)**

b) $F_X(x)$ là hàm không giảm, liên tục bên phải. Nghĩa là:

$$\text{Với mọi } a,b\in\mathbb{R}:\ a<b\ \Rightarrow F_X(a)\le F_X(b)$$

Với mọi $a\in\mathbb{R}$, ta có

$$F_X(a^+)=F_X(a)\ \text{ với }\ F_X(a^+)=\lim_{x>a,x\to a}F_X(x) \tag{2.3}$$

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

c) $F_X(-\infty)=\lim\limits_{x\to-\infty}F_X(x)=0;\ F_X(+\infty)=\lim\limits_{x\to+\infty}F_X(x)=1$, **(2.4)**

d) $P\{a<X\le b\}=F_X(b)-F_X(a)$. **(2.5)**

e) $P\{X>a\}=1-F_X(a)$; $P\{X<a\}=F_X(a^-)$ với $F_X(a^-)=\lim\limits_{x<a,x\to a}F_X(x)$ **(2.6)**

Nhận xét 2.1: Một số tài liệu coi

$$G_X(x)=P\{X<x\};\ -\infty<x<\infty$$

là hàm phân bố của biến ngẫu nhiên X.

Có một số khác biệt giữa hai định nghĩa này. Chẳng hạn tính chất liên tục phải của $F_X(x)$ được thay bằng liên tục trái của $G_X(x)$, công thức (2.5) sẽ là

$$P\{a \le X < b\} = G_X(b) - G_X(a).$$

Mỗi cách định nghĩa có thuận lợi riêng, tuy nhiên trong giáo trình này ta sử dụng hàm phân bố $F_X(x)$ theo công thức (2.1).

Ví dụ 2.3: Một nguồn thông tin sinh ra các ký hiệu ngẫu nhiên từ bốn ký tự $\{a,b,c,d\}$ với xác suất $P(a)=1/2$, $P(b)=1/4$ và $P(c)=P(d)=1/8$. Mã hóa các ký hiệu này theo các mã nhị phân sau

a	0
b	10
c	110
d	111

Đặt X là biến ngẫu nhiên ký hiệu độ dài của mã, đó là số các bit.

a) Tìm miền giá trị của X .

b) Giả sử các ký hiệu được sinh độc lập. Tính các xác suất

$$P\{X=1\},\ P\{X=2\} \text{ và } P\{X=3\}.$$

c) Tìm hàm phân bố $F_X(x)$ và vẽ đồ thị của $F_X(x)$.

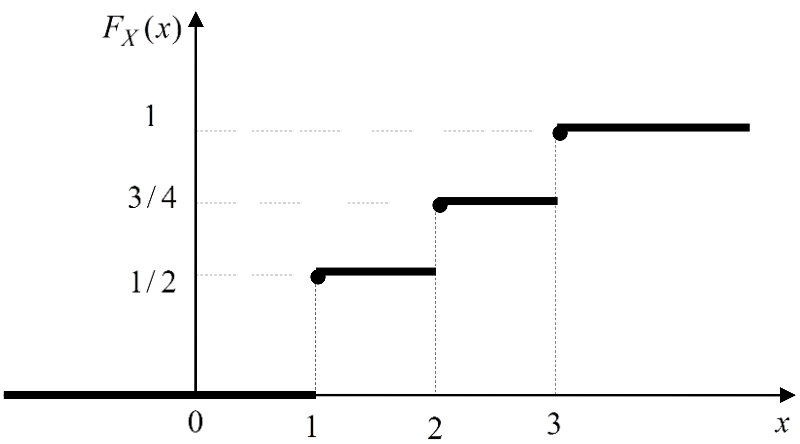
Giải: a) Miền giá trị $R_X = \{1,2,3\}$.

b) $P\{X=1\} = P(a) = 1/2$, $P\{X=2\} = P(b) = 1/4$,

$P\{X=3\} = P\{c,d\} = P(c)+P(d) = 1/4$.

c) $$F_X(x) = \begin{cases} 0 & x<1 \\ 1/2 & 1\le x<2 \\ 3/4 & 2\le x<3 \\ 1 & x\ge 3 \end{cases}$$

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



Hình 2.1

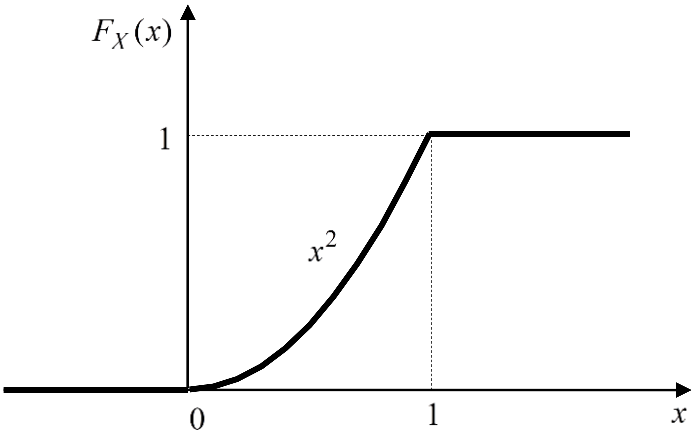
Ví dụ 2.4: Xét phép thử ném phi tiêu vào một đĩa tròn có bán kính bằng 1 (xem ví dụ 1.18). Ký hiệu X là biến ngẫu nhiên đo khoảng cách từ điểm mũi phi tiêu cắm vào đĩa đến tâm của đĩa. Giả sử mũi phi tiêu luôn cắm vào đĩa và đồng khả năng tại mọi điểm của đĩa.

a) Tìm miền giá trị của X .

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

Giải: a) Miền giá trị của X là $R_X = \{x \in \mathbb{R} \mid 0 \le x < 1\}$.

b) $P\{X \le x\} = \dfrac{\pi . x^2}{\pi . 1^2} = x^2$ nếu $0 \le x \le 1$.



Hình 2.2

Vậy hàm phân bố $$F_X(x) = \begin{cases} 0 & x \le 0 \\ x^2 & 0 < x \le 1 \\ 1 & x > 1 \end{cases}$$

Ví dụ 2.5: Xét hàm số $$F(x) = \begin{cases} 0 & x < 0 \\ x + 1/2 & 0 \le x < 1/2 \\ 1 & x \ge 1/2 \end{cases}$$

a) Vẽ đồ thị của $F(x)$ và chứng minh rằng $F(x)$ thỏa mãn các tính chất (2.2)-(2.4).

b) Nếu X là biến ngẫu nhiên có hàm phân bố $F(x)$. Tính

$$P\{X \le 1/4\},\ P\{0 < X \le 1/4\},\ P\{X = 0\} \text{ và } P\{0 \le X \le 1/4\}.$$

Giải: a) Từ đồ thị của $F(x)$ (hình 2.3) có thể suy ra các tính chất (2.2)-(2.4) của hàm phân bố.

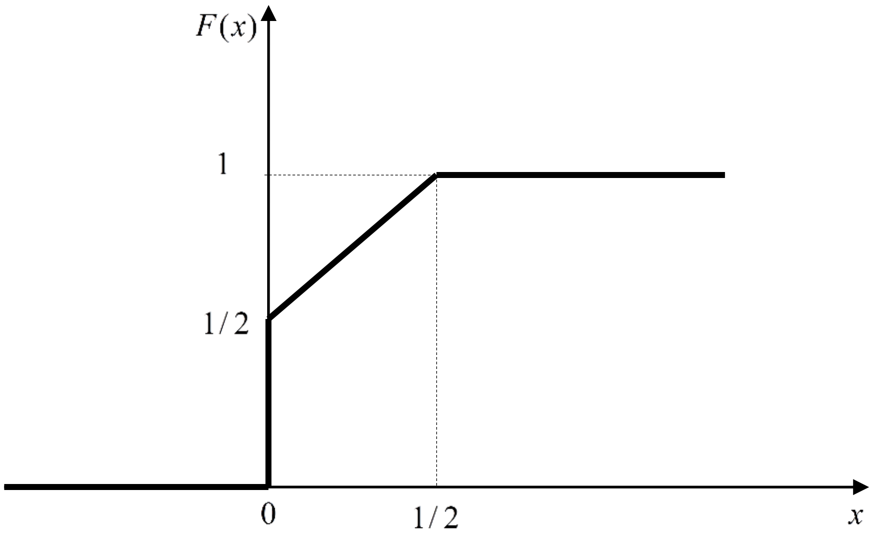
b) $$P\left\{X \le \frac{1}{4}\right\} = F\left(\frac{1}{4}\right) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4};$$

$$P\left\{0 < X \le \frac{1}{4}\right\} = F\left(\frac{1}{4}\right) - F(0) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$$

Áp dụng công thức (2.6) ta được

$$P\{X = 0\} = P\{X \le 0\} - P\{X < 0\} = F(0) - F(0^-) = \frac{1}{2} - 0 = \frac{1}{2}$$

$$P\{0 \le X \le 1/4\} = P\{X = 0\} + P\{0 < X \le 1/4\} = \frac{1}{2} + \frac{3}{4} - \frac{1}{2} = \frac{3}{4}.$$



Hình 2.3

2.1.3 Phân loại

Hàm phân bố của biến ngẫu nhiên là hàm số không giảm và bị chặn do đó hàm phân bố chỉ có thể gián đoạn tại một số không quá đếm được các điểm. Dựa vào dạng hàm phân bố của biến ngẫu nhiên ta có thể chia biến ngẫu nhiên thành ba loại: Biến ngẫu nhiên rời rạc, biến ngẫu nhiên liên tục và loại hỗn hợp.

❖ ***Biến ngẫu nhiên rời rạc***

Biến ngẫu nhiên X là rời rạc nếu miền giá trị gồm một số hữu hạn hoặc vô hạn đếm được các giá trị, nghĩa là có thể liệt kê các giá trị của miền giá trị R_X thành một dãy $x_1, x_2, ...$. Do đó hàm phân bố có đồ thị dạng hình thang (Hình 2.1, ví dụ 2.3)

❖ ***Biến ngẫu nhiên liên tục***

X là biến ngẫu nhiên liên tục nếu miền giá trị của nó có thể lấp đầy một hoặc một số các khoảng hữu hạn hoặc vô hạn và xác suất biến ngẫu nhiên nhận giá trị tại từng điểm đều bằng 0 (nghĩa là $P\{X = a\} = 0$ với mọi a). Do đó hàm phân bố là hàm số liên tục (Hình 2.2, ví dụ 2.4).

Biến ngẫu nhiên có hàm phân bố ở ví dụ 2.5 thuộc loại hỗn hợp, không rời rạc vì miền giá trị chứa khoảng [0;1/2] và không liên tục vì $P\{X = 0\} = 1/2 \ne 0$.

Tuy nhiên chúng ta thường chỉ gặp hai loại biến ngẫu nhiên rời rạc hoặc liên tục. Tập bài giảng này cũng chỉ xét hai loại đó.

Ví dụ 2.6:

- Gọi X là số chấm xuất hiện khi gieo một con xúc xắc thì X là biến ngẫu nhiên rời rạc nhận các giá trị $1, 2, 3, 4, 5, 6$.
- Gọi Y là tuổi thọ của một thiết bị đang hoạt động thì Y là biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.
- Gọi Z là số khách hàng vào một điểm phục vụ trong 1 đơn vị thời gian, Z là biến ngẫu nhiên rời rạc nhận các giá trị $0, 1, 2, ...$

- Số cuộc gọi đến một tổng đài là biến ngẫu nhiên rời rạc nhận các giá trị $0,1,2,...$
- Sai số Y khi đo lường một đại lượng vật lý nào đó là biến ngẫu nhiên liên tục nhận giá trị trong một khoảng.

2.2 BIẾN NGẪU NHIÊN RỜI RẠC

2.2.1 Hàm khối lượng xác suất và bảng phân bố xác suất của biến ngẫu nhiên rời rạc

Biến ngẫu nhiên rời rạc có miền giá trị là một tập hữu hạn hoặc vô hạn đếm được. Các xác suất chỉ tập trung tại các giá trị này.

Giả sử biến ngẫu nhiên rời rạc X có đồ thị của $F_X(x)$ là hàm bậc thang có bước nhảy tại $x_1, x_2, ...$ thì

$$F_X(x_k) - F_X(x_k^-) = P\{X \le x_k\} - P\{X < x_k\} = P\{X = x_k\} \tag{2.7}$$

Đặt

$$p_X(x) = P\{X = x\} \tag{2.8}$$

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

Tính chất của hàm khối lượng xác suất $p_X(x)$:

1. $p_X(x_k) > 0$, với mọi $x_k \in R_X$ **(2.9)**
2. $\sum\limits_{x_k \in R_X} p_X(x_k) = 1$ **(2.10)**
3. $p_X(x) = 0$ với mọi $x \notin R_X$ **(2.11)**

Hàm phân bố của X

$$F_X(x) = P\{X \le x\} = \sum_{x_k \le x;\, x_k \in R_X} p_X(x_k) \tag{2.12}$$

- Nếu biến ngẫu nhiên rời rạc X nhận vô hạn các giá trị $x_1, x_2, ...$ thì hàm phân bố xác suất có dạng:

$$F_X(x) = \begin{cases} 0 & \text{nếu} \quad x < x_1 \\ p_X(x_1) + \cdots + p_X(x_{k-1}) & \text{nếu} \quad x_{k-1} \le x < x_k,\ \forall k > 1 \end{cases} \tag{2.13}$$

- Nếu X chỉ nhận một số hữu hạn các giá trị $x_1, x_2, ..., x_n$ thì các biến cố

$$\{X = x_1\}, \{X = x_2\}, ..., \{X = x_n\} \tag{2.14}$$

lập thành hệ đầy đủ các biến cố.

Hàm phân bố xác suất có dạng:

$$F_X(x) = \begin{cases} 0 & \text{nếu} \quad x < x_1 \\ p_X(x_1) + \cdots + p_X(x_{k-1}) & \text{nếu} \quad x_{k-1} \le x < x_k \\ 1 & \text{nếu} \quad x \ge x_n \end{cases} \qquad \textbf{(2.15)}$$

Chúng ta cũng có thể biểu diễn phân bố của biến ngẫu nhiên rời rạc thông qua *bảng phân bố xác suất*. Đó là bảng có hai hàng, hàng trên ghi các giá trị mà biến ngẫu nhiên nhận được, hàng dưới là giá trị của hàm khối lượng xác suất tương ứng.

Bảng phân bố xác suất của X có dạng sau:

X	x_1	x_2	…
P	$p_X(x_1)$	$p_X(x_2)$	…

Ví dụ 2.7: Xét phép thử tung đồng thời 2 đồng xu (Ví dụ 1.1). Không gian mẫu của phép thử là $\Omega = \{(S,S),(S,N),(N,S),(N,N)\}$ gồm 4 kết quả đồng khả năng. Gọi X là số mặt sấp xuất hiện, khi đó X là một biến ngẫu nhiên rời rạc.

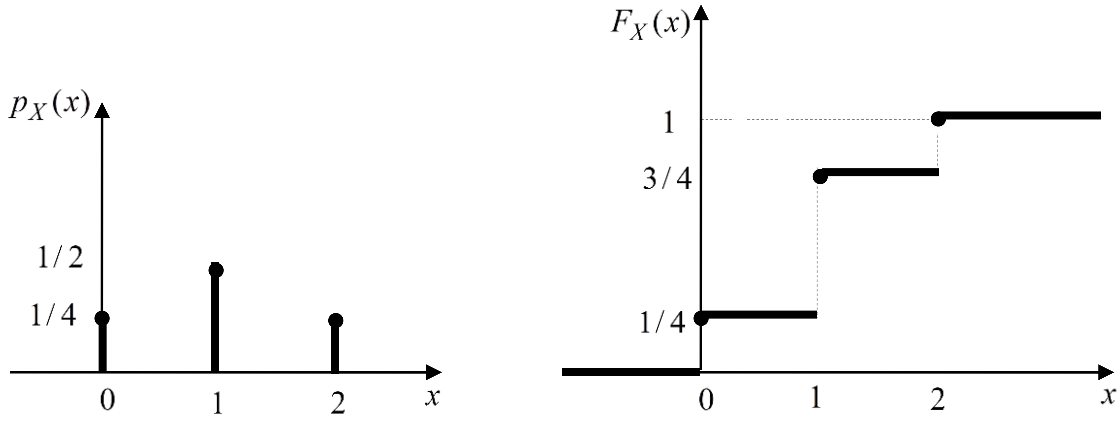
Bảng phân bố xác suất

X	0	1	2
P	$1/4$	$2/4$	$1/4$

Hàm phân bố xác suất

$$F_X(x) = \begin{cases} 0 & \text{nếu} \quad x < 0 \\ 1/4 & \text{nếu} \quad 0 \le x < 1 \\ 3/4 & \text{nếu} \quad 1 \le x < 2 \\ 1 & \text{nếu} \quad x \ge 2 \end{cases}$$

Đồ thị của hàm khối lượng xác suất và hàm phân bố



(a) (b)

Hình 2.4

Ví dụ 2.8: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng. Gọi X là số bi trắng trong 3 bi vừa chọn thì X là một biến ngẫu nhiên rời rạc. Tìm bảng phân bố xác suất và hàm phân bố xác suất.

Giải: $P\{X=0\}=\dfrac{C_6^3}{C_{10}^3}=\dfrac{5}{30}$, $P\{X=1\}=\dfrac{C_6^2C_4^1}{C_{10}^3}=\dfrac{15}{30}$,

$$P\{X=2\}=\frac{C_6^1C_4^2}{C_{10}^3}=\frac{9}{30},\ P\{X=3\}=\frac{C_4^3}{C_{10}^3}=\frac{1}{30}.$$

Hàm khối lượng xác suất

$$p_X(0)=\frac{5}{30},\ p_X(1)=\frac{15}{30},\ p_X(2)=\frac{9}{30},\ p_X(3)=\frac{1}{30};\ \ p_X(x)=0 \text{ với mọi } x \text{ khác } 0,1,2,3.$$

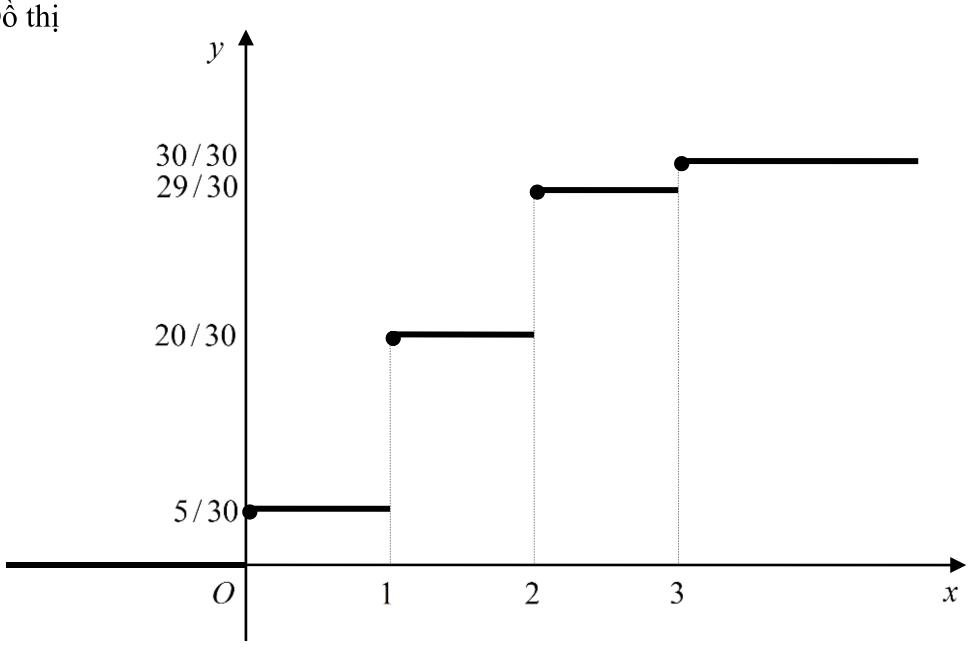
Bảng phân bố xác suất

X	0	1	2	3
P	$5/30$	$15/30$	$9/30$	$1/30$

Hàm phân bố xác suất

$$F_X(x)=\begin{cases} 0 & \text{nếu} \quad x<0 \\ 5/30 & \text{nếu} \quad 0\le x<1 \\ 20/30 & \text{nếu} \quad 1\le x<2 \\ 29/30 & \text{nếu} \quad 2\le x<3 \\ 1 & \text{nếu} \quad x\ge 3 \end{cases}$$

Đồ thị



Hình 2.5

Ví dụ 2.9: Xác suất để xạ thủ bắn trúng bia là p, $0<p<1$. Xạ thủ được phát từng viên đạn để bắn cho đến khi trúng bia. Hãy xây dựng bảng phân bố xác suất của số viên đạn được phát.

Giải: Gọi X là "số viên đạn được phát". X là biến ngẫu nhiên rời rạc với các giá trị có thể có là $1,2,...,k,...$ Gọi A_k là biến cố "xạ thủ bắn trúng bia ở viên đạn thứ k". Các biến cố A_k độc lập nhau và $P(A_k)=p$.

$$P\{X=1\}=P(A_1)=p$$

$$P\{X=2\}=P\left(\overline{A_1}A_2\right)=P\left(\overline{A_1}\right)P\left(A_2\right)=qp\,,\ q=1-p$$

Tổng quát $$P\{X=k\}=P\left(\overline{A_1}...\overline{A_{k-1}}A_k\right)=P\left(\overline{A_1}\right)...P\left(\overline{A_{k-1}}\right)P\left(A_k\right)=q^{k-1}p$$

Hàm khối lượng xác suất

$$p_X(x)=\begin{cases} q^{k-1}p & x=k \\ 0 & x\neq k \end{cases}\ ;\ k=1,2,...;\ 0<p<1 \qquad \textbf{(2.16)}$$

Bảng phân bố xác suất có dạng:

X	1	2	…	k	…
P	p	qp	…	$q^{k-1}p$	…

Biến ngẫu nhiên với phân bố rời rạc có dạng trên được gọi là *biến ngẫu nhiên cấp số nhân* (geometric random variable), hàm khối lượng xác suất lập thành cấp số nhân có công bội là q.

2.2.2 Các phân bố rời rạc thường gặp

2.2.2.1 Phân bố Bernoulli

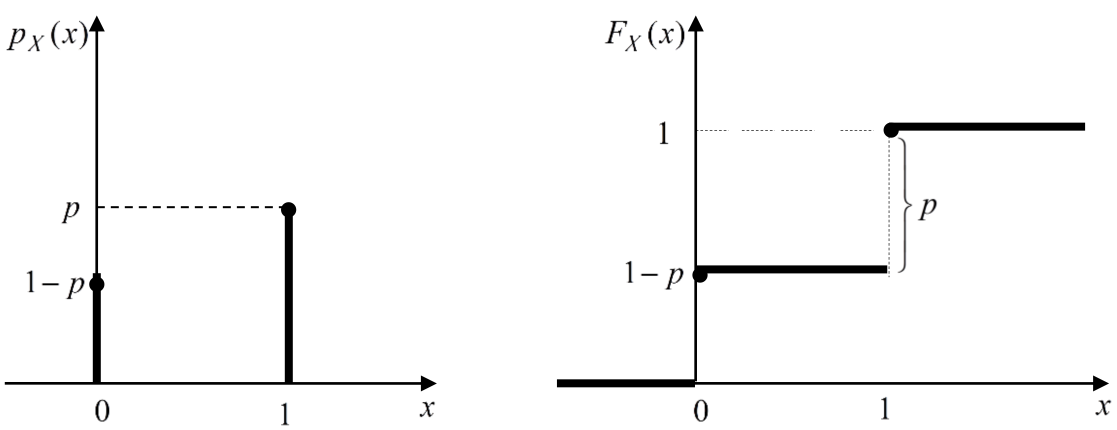
Định nghĩa 2.4: *Biến ngẫu nhiên rời rạc X nhận hai giá trị $0, 1$ với xác suất tương ứng*

$$p_X(k)=P\{X=k\}=p^kq^{1-k};\ k=0,1 \qquad \textbf{(2.17)}$$

trong đó $0<p<1,\ q=1-p$, được gọi là có phân bố Bernoulli tham số p.

Xét phép thử Bernoulli với sự thành công của phép thử là sự xuất hiện của biến cố A và giả sử xác suất xuất hiện của A trong mỗi lần thử là p. Gọi X là số lần thành công trong một lần thử thì X là biến ngẫu nhiên rời rạc có phân bố Bernoulli tham số p. Biến ngẫu nhiên X còn được gọi là phân bố không – một $A(p)$.

Hình 2.5 minh họa phân bố Bernoulli tham số p.



(a) (b)

Hình 2.6: Đồ thị hàm khối lượng xác suất và hàm phân bố Bernoulli tham số p

Trong lý thuyết mẫu người ta thường sử dụng biến ngẫu nhiên có phân bố Bernoulli để biểu diễn cho dấu hiệu định tính (xem chương 5).

2.2.2.2 Phân bố nhị thức $\mathscr{B}(n;p)$

Định nghĩa 2.5: *Biến ngẫu nhiên rời rạc X nhận các giá trị $0, 1, ..., n$ với xác suất tương ứng*

$$p_X(k) = P\{X = k\} = C_n^k p^k q^{n-k};\ k = 0,1,...,n \tag{2.18}$$

trong đó n là số tự nhiên và $0 < p < 1$, được gọi là có phân bố nhị thức tham số n, p, ký hiệu $X \sim \mathscr{B}(n;p)$.

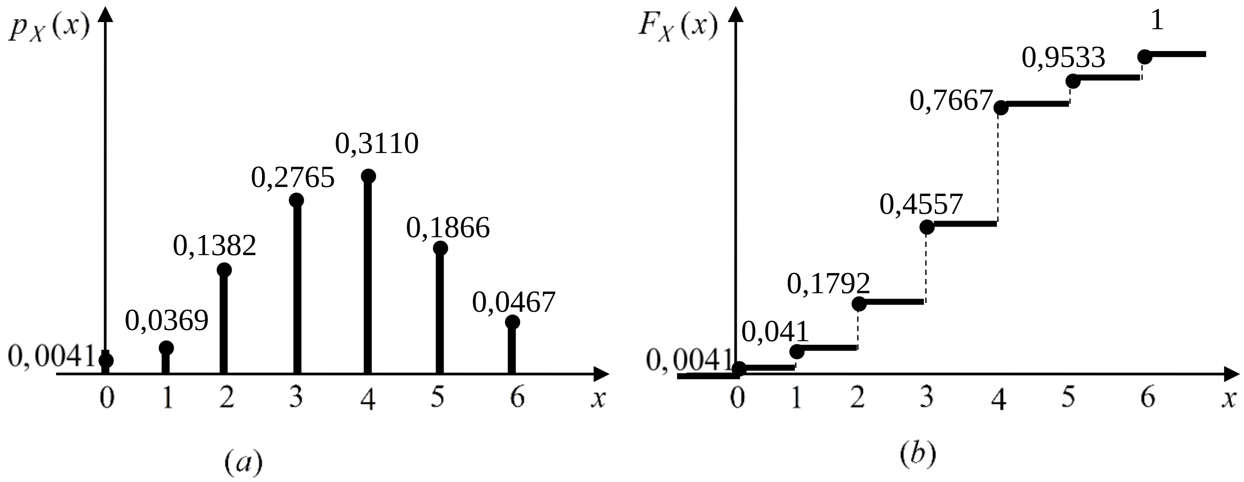
Bảng phân bố xác suất của biến ngẫu nhiên có phân bố nhị thức $\mathscr{B}(n;p)$

X	0	1	…	k	…	n
P	$C_n^0 p^0 q^n$	$C_n^1 p^1 q^{n-1}$	…	$C_n^k p^k q^{n-k}$	…	$C_n^n p^n q^0$

Trong đó $q = 1 - p$.

Hàm phân bố

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < 0 \\ \sum_{k=0}^{m} C_n^k p^k q^{n-k} & \text{nếu } m \le x < m+1,\ 0 \le m \le n-1, \\ 1 & \text{nếu } x \ge n \end{cases} \tag{2.19}$$



Hình 2.7: phân bố nhị thức với $n = 6$ và $p = 0,6$

Nhận xét 2.2:

1. Thực hiện n phép thử Bernoulli với xác suất thành công của biến cố A trong mỗi lần thử là p.

Với mỗi $i = 1, 2, ..., n$; nếu ở lần thử thứ i biến cố A xuất hiện ta cho X_i nhận giá trị 1, nếu biến cố A không xuất hiện ta cho X_i nhận giá trị 0. Như vậy X_i là biến ngẫu nhiên rời rạc có phân bố Bernoulli tham số p:

Gọi X là số thành công trong n phép thử Bernoulli này thì

$$X = X_1 + X_2 + \cdots + X_n \sim \mathscr{B}(n;p) \tag{2.20}$$

2. Từ (2.20) suy ra rằng nếu $X \sim \mathscr{B}(n_1; p)$, $Y \sim \mathscr{B}(n_2; p)$ và X, Y độc lập thì

$$X + Y \sim \mathscr{B}(n_1 + n_2; p) \qquad \textbf{(2.21)}$$

Ví dụ 2.10: Tỉ lệ phế phẩm của lô hàng là 4%. Chọn ngẫu nhiên 20 sản phẩm để kiểm tra. Gọi X là số phế phẩm phát hiện được.

a) Gọi tên luật phân bố xác suất của X.

b) Tính xác suất có đúng 5 phế phẩm phát hiện được.

c) Lô hàng được xem là đạt tiêu chuẩn nếu số phế phẩm phát hiện được không nhiều hơn 2. Tính xác suất để lô hàng đạt tiêu chuẩn.

Giải: Có thể xem kiểm tra chất lượng mỗi sản phẩm là thực hiện một phép thử Bernoulli với sự thành công của phép thử là phát hiện ra phế phẩm. Theo giả thiết xác suất thành công của mỗi lần thử là 0,04. Kiểm tra 20 sản phẩm là thực hiện 20 phép thử.

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

b) Xác suất phát hiện đúng 5 phế phẩm là $P\{X = 5\} = C_{20}^{5}(0,04)^5(0,96)^{15} = 0,0008$.

c) Xác suất để lô hàng đạt tiêu chuẩn là $P\{X \le 2\} = 0,956$.

Ví dụ 2.11: Một nguồn nhị phân phát ra hai ký số (digit) 1 và 0 một cách ngẫu nhiên với xác suất tương ứng 0,6 và 0,4.

a) Tính xác suất có đúng hai ký số 1 và ba ký số 0 trong dãy có năm ký số.

b) Tính xác suất có ít nhất ba ký số 1 trong dãy có năm ký số.

Giải: Gọi X là số các ký số 1 trong dãy có năm ký số. Theo giả thiết chỉ có hai khả năng đầu vào có thể là 1 và 0, xác suất xuất hiện của 1 không đổi và bằng 0,6. Vậy có thể xem mỗi ký số xuất hiện trong dãy tương ứng với một phép thử Bernoulli mà sự thành công của mỗi lần thử là xuất hiện ký số 1. Mỗi dãy có năm ký số tương ứng với 5 phép thử Bernoulli, vậy X có phân bố nhị thức $X \sim \mathscr{B}(n; p)$, với $n = 5,\ p = 0,6$.

a) Xác suất để dãy có đúng hai ký số 1 và ba ký số 0 trong dãy có năm ký số là

$$P\{X = 2\} = C_5^2(0,6)^2(0,4)^3 = 0,23.$$

b) Xác suất có ít nhất ba ký số 1 trong dãy có năm ký số là

$$P\{X \ge 3\} = 1 - P\{X \le 2\}$$

Trong đó
$$P\{X \le 2\} = \sum_{k=0}^{2} C_5^k (0,6)^k (0,4)^{5-k} = 0,317$$

Vậy
$$P\{X \ge 3\} = 1 - 0,317 = 0,683.$$

2.2.2.3 Phân bố Poisson

Định nghĩa 2.6: *Biến ngẫu nhiên X nhận các giá trị $k = 0, 1, 2, \ldots$ với hàm khối lượng xác suất*

$$p_X(k) = P\{X = k\} = e^{-\lambda}\frac{\lambda^k}{k!}; \lambda > 0;\ k = 0,1,2,... \qquad \textbf{(2.22)}$$

gọi là có phân bố Poisson tham số $\lambda > 0$*, ký hiệu* $X \sim \mathscr{P}(\lambda)$.

Hàm phân bố

$$F_X(x) = e^{-\lambda}\sum_{k=0}^{n}\frac{\lambda^k}{k!},\quad n \le x < n+1 \qquad \textbf{(2.23)}$$

Trong thực tế với một số giả thiết thích hợp thì các biến ngẫu nhiên là các quá trình đếm sau:

1) Số cuộc gọi đến một tổng đài,

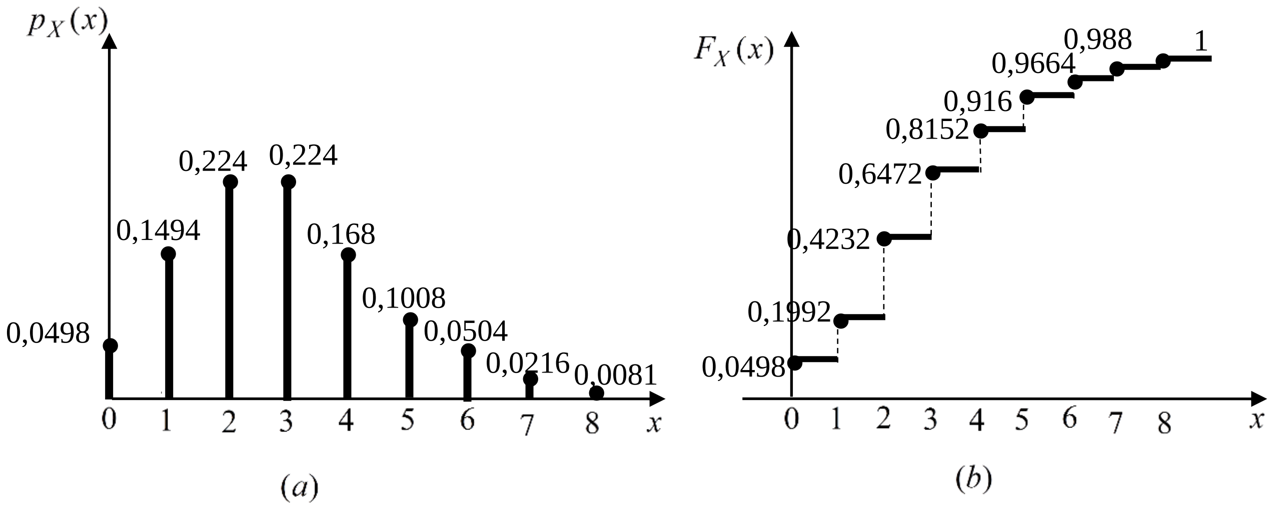
2) Số khách hàng đến 1 điểm phục vụ,

3) Số xe cộ qua 1 ngã tư,

4) Số tai nạn (xe cộ); số các sự cố xảy ra ở một địa điểm …

trong một khoảng thời gian xác định nào đó sẽ có phân bố Poisson với tham số λ, trong đó λ là tốc độ trung bình diễn ra trong khoảng thời gian này.

Hình 2.7 minh họa phân bố Poisson với tham số $\lambda = 3$



Hình 2.8 Phân bố Poisson tham số $\lambda = 3$

Ví dụ 2.12: Ở một tổng đài điện thoại các cuộc gọi đến một cách ngẫu nhiên, độc lập. Ký hiệu $X(t)$ là số cuộc gọi đến tổng đài trong khoảng thời gian t phút, có thể chứng minh được $X(t)$ có phân bố Poisson tham số λt, trong đó λ là số cuộc gọi trung bình trong 1 phút. Giả sử trung bình có 2 cuộc gọi trong 1 phút. Tìm xác suất để:

a) Có đúng 5 cuộc gọi đến trong 2 phút (biến cố A).

b) Không có một cuộc gọi nào trong 30 giây (biến cố B).

c) Có ít nhất 1 cuộc gọi trong 10 giây (biến cố C).

Giải: Theo giả thiết $\lambda = 2$, vậy ta có

a) $X(2) \sim \mathscr{P}(4)$, do đó $P(A) = P\{X(2) = 5\} = e^{-4}\dfrac{4^5}{5!} \approx 0{,}156$.

b) $X(1/2) \sim \mathscr{P}(1)$, do đó $P(B) = P\{X(1/2) = 0\} = e^{-1} \approx 0{,}3679$.

c) $X(1/6) \sim \mathscr{P}(1/3)$, do đó

$$P(C) = P\{X(1/6) \geq 1\} = 1 - P\{X(1/6) = 0\} = 1 - e^{-1/3} \approx 0,2835.$$

Phân bố Poisson được ứng dụng rộng rãi trong nhiều lĩnh vực thực tế như kiểm tra chất lượng sản phẩm, lý thuyết sắp hàng, các hệ phục vụ đám đông, các bài toán chuyển mạch trong tổng đài …

Có thể chứng minh được: nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố Poisson tham số lần lượt λ_1, λ_2 thì $X_1 + X_2$ cũng có phân bố Poisson tham số $\lambda_1 + \lambda_2$ (ví dụ 3.10 chương 3)

$$X_1 + X_2 \sim \mathscr{P}(\lambda_1 + \lambda_2). \qquad \textbf{(2.24)}$$

2.3 BIẾN NGẪU NHIÊN LIÊN TỤC

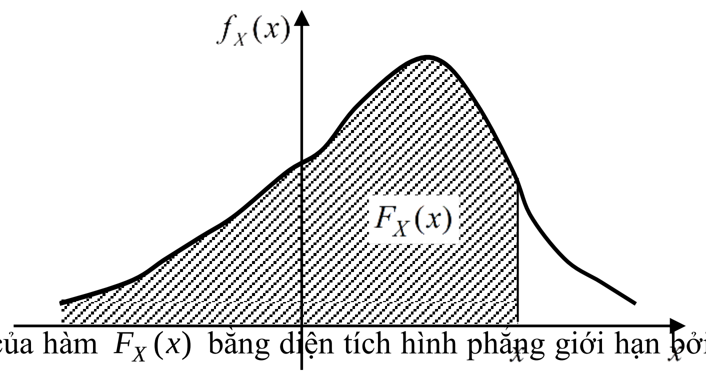
2.3.1 Hàm mật độ xác suất của biến ngẫu nhiên liên tục

Công thức (2.12) cho thấy hàm phân bố xác suất của biến ngẫu nhiên rời rạc có thể được xác định qua hàm khối lượng xác suất. Mục này sẽ chỉ ra rằng hàm phân bố xác suất của biến ngẫu nhiên liên tục có thể được xác định bởi hàm mật độ xác suất tương ứng.

Định nghĩa 2.7: *Giả sử* X *là một biến ngẫu nhiên liên tục có hàm phân bố xác suất* $F_X(x)$, *nếu tồn tại hàm* $f_X(x)$ *sao cho*

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \text{ với mọi } x \in \mathbb{R} \qquad \textbf{(2.25)}$$

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



Hình 2.9

Như vậy giá trị của hàm $F_X(x)$ bằng diện tích hình phẳng giới hạn bởi đồ thị hàm mật độ xác suất $f_X(x)$, trục hoành và đường thẳng song song với trục tung có hoành độ là x.

Tính chất của hàm mật độ xác suất $f_X(x)$

1. $F_X'(x) = f_X(x)$ tại các điểm x mà $f_X(x)$ liên tục. **(2.26)**
2. $f_X(x) \geq 0$ với mọi $x \in \mathbb{R}$, **(2.27)**
3. $\int_{-\infty}^{\infty} f_X(x)dx = 1$, **(2.28)**

Vì xác suất tại từng điểm của biến ngẫu nhiên liên tục bằng 0, do đó ta có

4. $P\{a < X < b\} = P\{a \le X \le b\} = P\{a \le X < b\} = P\{a < X \le b\} = \int_a^b f_X(x)dx$. **(2.29)**

Vậy hàm phân bố của biến ngẫu nhiên liên tục là một nguyên hàm của hàm mật độ, và hàm mật độ là đạo hàm của hàm phân bố.

Ví dụ 2.13: Hàm phân bố xác suất của biến ngẫu nhiên liên tục X có dạng

$$F_X(x) = \begin{cases} 0 & \text{với } x < 0 \\ kx^2 & \text{với } 0 \le x < 1 \\ 1 & \text{với } x \ge 1 \end{cases}$$

a) Xác định hệ số k;

b) Tìm hàm mật độ xác suất $f_X(x)$.

Giải: a) Xét tại $x = 1$ và áp dụng tính chất liên tục của hàm phân bố xác suất $F_X(x)$ ta có:

$$\begin{cases} \lim\limits_{x\to 1^-} F_X(x) = \lim\limits_{x\to 1^-} kx^2 = k \\ F_X(1) = 1 \end{cases} \Rightarrow k = 1.$$

b) Theo tính chất (2.26) của hàm mật độ xác suất ta có

$$f_X(x) = \begin{cases} 0 & \text{với } x \le 0 \\ 2x & \text{với } 0 < x < 1 \\ 0 & \text{với } x \ge 1 \end{cases}$$

Ví dụ 2.14: Biến ngẫu nhiên liên tục X với hàm mật độ xác suất có dạng

$$f_X(x) = \begin{cases} 0 & \text{với } x < 1 \\ \dfrac{k}{x^2} & \text{với } x \ge 1 \end{cases}$$

Hãy xác định:

a) Hệ số k;

b) Hàm phân bố xác suất $F_X(x)$;

c) Xác suất $P\{2 < X < 3\}$;

d) Xác suất để trong 4 phép thử độc lập biến ngẫu nhiên X không lấy giá trị trong khoảng $(2,3)$.

Giải: a) Theo tính chất (2.28) ta có $1 = \int_{-\infty}^{\infty} f_X(x)dx = \int_1^{\infty} \frac{k}{x^2}dx = -\lim\limits_{a\to\infty}\left(\frac{k}{x}\Big|_1^a\right) = k$, từ đó $k = 1$.

b) Từ công thức (2.25) ta có

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0 & \text{với } x < 1 \\ 1 - \dfrac{1}{x} & \text{với } x \ge 1 \end{cases}$$

c) Từ công thức (2.29) ta có

$$P\{2 < X < 3\} = F_X(3) - F_X(2) = \left(1 - \frac{1}{3}\right) - \left(1 - \frac{1}{2}\right) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.$$

d) Theo kết quả ở trên xác suất để X không lấy giá trị trong khoảng $(2;3)$ trong một phép thử bằng $1 - \frac{1}{6} = \frac{5}{6}$.

Vậy xác suất để trong 4 phép thử độc lập biến ngẫu nhiên X đều không lấy giá trị trong khoảng $(2;3)$ bằng $\left(\frac{5}{6}\right)^4 \approx 0{,}48$.

2.3.2 Các phân bố liên tục thường gặp

2.3.2.1 Phân bố đều $\mathbf{U}(a,b)$

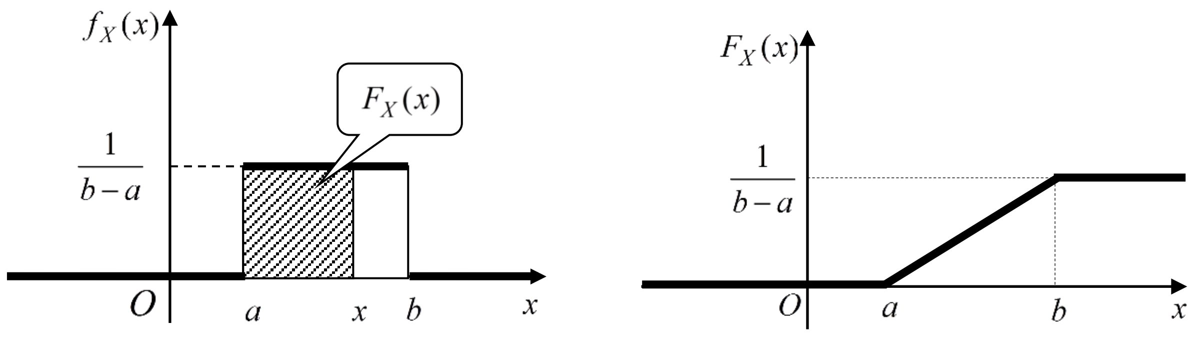
Định nghĩa 2.8: *Biến ngẫu nhiên X được gọi là có phân bố đều trong khoảng $(a;b)$ nếu hàm mật độ xác suất của nó xác định bởi:*

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{nếu } a < x < b \\ 0 & \text{nếu ngược lại} \end{cases} \tag{2.30}$$

Hàm phân bố xác suất tương ứng

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 0 & \text{nếu } x \le a \\ \dfrac{x-a}{b-a} & \text{nếu } a < x < b \\ 1 & \text{nếu } x \ge b \end{cases} \tag{2.31}$$

Vậy X có khả năng nhận giá trị trong khoảng $(a;b)$ là "đều nhau" và không nhận giá trị ngoài $(a;b)$.



Hình 2.10 Phân bố đều $\mathbf{U}(a,b)$

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

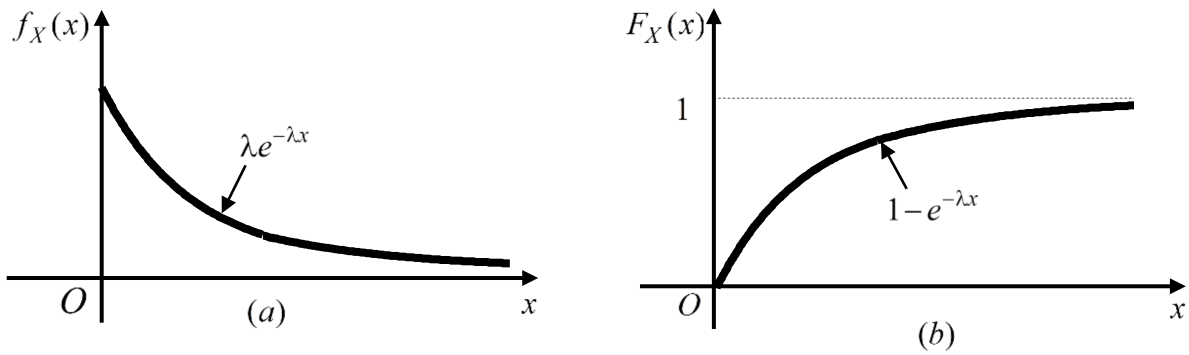
2.3.2.2 Phân bố mũ

Định nghĩa 2.9: *Biến ngẫu nhiên liên tục* X *được gọi là có phân bố mũ tham số* $\lambda > 0$ *nếu hàm mật độ xác suất xác định như sau:*

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \le 0 \end{cases} \tag{2.32}$$

Hàm phân bố xác suất

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 1 - e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \le 0 \end{cases} \tag{2.33}$$



Hình 2.11 Phân bố mũ tham số $\lambda = 1/7$

Phân bố mũ thường [illegible] ống của một loài sinh vật, tuổi thọ của thiết bị … hoặc khoảng thời gian giữa hai lần xuất hiện của một biến cố E nào đó mà số lần xuất hiện của E tuân theo luật phân bố Poisson.

Ví dụ 2.15: Tuổi thọ của một mạch điện tử trong máy tính là một biến ngẫu nhiên có phân bố mũ tham số $\lambda > 0$. Giả sử tuổi thọ trung bình của mạch điện tử này là $\frac{1}{\lambda} = 6,25$(năm) (xem kỳ vọng của phân bố mũ ở mục 2.5.3). Thời gian bảo hành là 2 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Giải: Gọi X là tuổi thọ của mạch điện tử. Xác suất để mạch điện tử bị hỏng trong thời gian bảo hành là:

$$P\{X \le 2\} = 1 - e^{-2\lambda} = 1 - e^{\frac{-2}{6,25}} = 1 - e^{-0,32} = 1 - 0,726 = 0,274\,.$$

Vậy có khoảng 27,4% số mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

Nhận xét 2.3: Tính chất thú vị nhất của phân bố mũ là tính "không nhớ".

Biến ngẫu nhiên X *được gọi là "không nhớ" (memoryless)* nếu

$$P\{X > x + t \mid X > t\} = P\{X > x\};\ \forall\, x, t > 0$$

Điều này tương đương với

$$P\{X > x + t\} = P\{X > x\}P\{X > t\};\ \forall\, x, t > 0 \tag{2.34}$$

Gọi $F(x)$ là hàm phân bố xác suất của X, đặt $G(x) = P\{X > x\} = 1 - F(x)$.

Điều kiện (2.34) có thể viết lại

$$G(x+t)=G(x)G(t) \tag{2.35}$$

Giải phương trình (2.35) (phương trình hàm Cauchy) với điều kiện $G(x)=1,\ \forall x<0$ và $G(+\infty)=0$ ta được $G(x)=e^{-\lambda x}$.

Vậy biến ngẫu nhiên X "không nhớ" khi và chỉ khi X có phân bố mũ. Chính vì thế phân bố mũ là phân bố có tính chất "không nhớ" và được gọi là *phân bố Markov*.

Ví dụ 2.16: Giả sử thời gian mỗi cuộc gọi điện thoại (tính theo phút) là một biến ngẫu nhiên X với phân bố mũ tham số $\lambda=1/10$. Một bốt điện thoại chỉ phục vụ từng người và giả sử A vào bốt điện thoại trước khi B đến. Tính xác suất B phải chờ đến lượt mình trong khoảng thời gian:

a) Ít hơn 5 phút

b) Trong khoảng từ 5 đến 10 phút.

Giải: a) Vì tính chất "không nhớ" của phân bố mũ do đó thời gian chờ của B bằng thời gian A tiếp tục hoàn thành cuộc gọi tính từ lúc B đến và không phụ thuộc A đã gọi trong thời gian bao lâu. Theo công thức (2.33) ta được xác suất B phải chờ ít hơn 5 phút là

$$P\{X<5\}=P\{X\le 5\}=F_X(5)=1-e^{-0,5}\approx 0,394\,.$$

b) Lập luận tương tự trên ta được xác suất B phải chờ trong khoảng từ 5 đến 10 phút là

$$P\{5<X<10\}=F_X(10)-F_X(5)=(1-e^{-1})-(1-e^{-0,5})=e^{-0,5}-e^{-1}\approx 0,239\,.$$

2.3.2.3 Phân bố Erlang

Định nghĩa 2.10: *Biến ngẫu nhiên X có phân bố Erlang tham số $(k;\lambda)$; k là số tự nhiên và $\lambda>0$, nếu hàm mật độ xác suất có dạng:*

$$f_X(x)=\begin{cases}\dfrac{(\lambda x)^{k-1}}{(k-1)!}\lambda e^{-\lambda x} & \text{nếu } \ x>0\\ 0 & \text{nếu } \ x\le 0\end{cases} \tag{2.36}$$

Có thể chứng minh được rằng nếu $X_1,X_2,...,X_k$ là k biến ngẫu nhiên độc lập cùng có phân bố mũ tham số $\lambda>0$ thì

$$X=X_1+X_2+\cdots+X_k$$

có phân bố Erlang tham số $(k;\lambda)$.

Phân bố Erlang tham số $(1;\lambda)$ là phân bố mũ tham số λ.

2.3.2.4 Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$

2.3.2.4.1 Định nghĩa

Định nghĩa 2.11: *Biến ngẫu nhiên liên tục X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$, ký hiệu $X\sim\mathbf{N}(\mu;\sigma^2)$, nếu hàm mật độ xác suất có dạng*

$$f_X(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\ ;\ \forall x\in\mathbb{R} \tag{2.37}$$

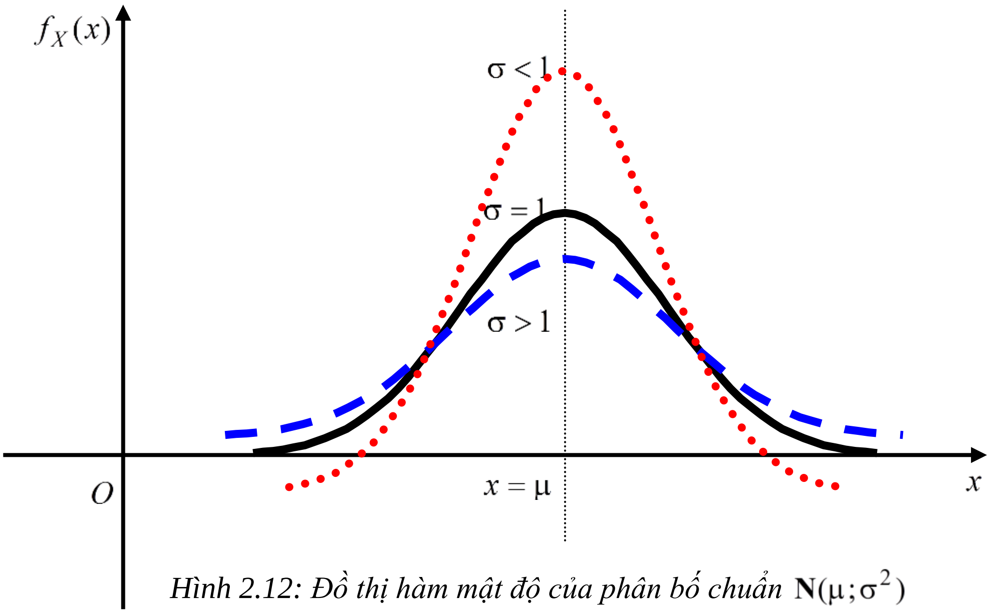
Phân bố chuẩn được Gauss tìm ra năm 1809 nên nó còn được gọi là *phân bố Gauss*. Phân bố chuẩn thường được sử dụng trong các bài toán đo đạc các đại lượng vật lý, thiên văn ...

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

2.3.2.4.2 Tính chất đồ thị của hàm mật độ xác suất của phân bố chuẩn

Từ công thức xác định hàm mật độ xác suất (2.26) ta suy ra các tính chất sau của đồ thị:

- Nhận trục $x=\mu$ làm trục đối xứng.
- Tiệm cận với trục hoành khi $x \to \pm\infty$.
- Diện tích giới hạn bởi đồ thị và trục hoành bằng 1.
- Đạt cực đại tại $x=\mu$ và có giá trị cực đại bằng $\dfrac{1}{\sigma\sqrt{2\pi}}$.
- Có 2 điểm uốn tại $x=\mu\pm\sigma$.
- Do đó khi μ tăng lên thì đồ thị dịch sang phải, còn khi μ giảm đồ thị dịch sang trái.
- Khi σ tăng lên thì đồ thị sẽ thấp xuống, còn khi σ giảm đồ thị cao lên và nhọn hơn.



Hình 2.12: Đồ thị hàm mật độ của phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$

Nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố chuẩn $X_1 \sim \mathbf{N}(\mu_1;\sigma_1^2)$ và $X_2 \sim \mathbf{N}(\mu_2;\sigma_2^2)$ thì tổ hợp tuyến tính bất kỳ của X_1, X_2 cũng có phân bố chuẩn, đặc biệt

$$X_1 + X_2 \sim \mathbf{N}(\mu_1+\mu_2;\sigma_1^2+\sigma_2^2) \qquad \textbf{(2.38)}$$

2.3.2.4.3 Phân bố chuẩn tắc

Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ với $\mu=0$, $\sigma^2=1$ gọi là *phân bố chuẩn tắc* $\mathbf{N}(0;1)$.

Hàm mật độ xác suất của phân bố chuẩn tắc $\mathbf{N}(0;1)$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\ ;\ \forall x\in\mathbb{R} \qquad \textbf{(2.39)}$$

Hàm phân bố xác suất của $\mathbf{N}(0;1)$

$$\Phi(x)=\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt\ ;\ \forall x\in\mathbb{R} \qquad \textbf{(2.40)}$$

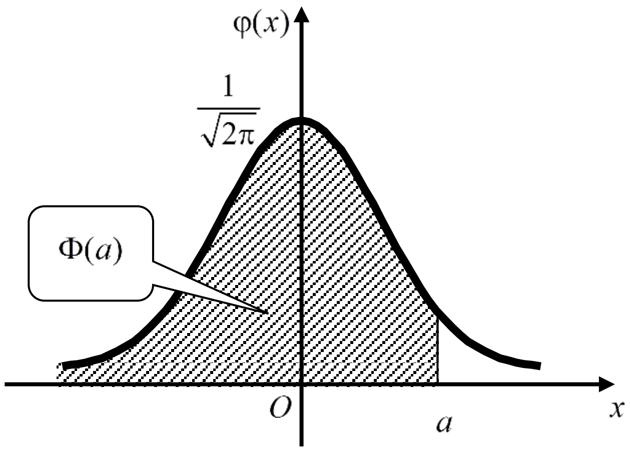
Có bảng tính sẵn các giá trị của $\varphi(x)$ và $\Phi(x)$ (xem Phụ lục I và Phụ lục II).

Cần chú ý rằng một số tài liệu cho bảng tính $\Phi_0(x)=\int_0^x\varphi(t)dt=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ với $x\ge 0$.

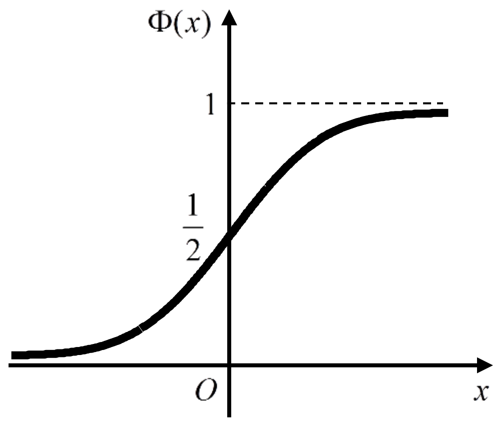
Công thức liên hệ là

$$\Phi(x)=\Phi_0(x)+0,5 \text{ với mọi } x\ge 0.$$

Đồ thị của hàm mật độ xác suất $\varphi(x)$ và hàm phân bố xác suất $\Phi(x)$



Hình 2.13a: Đồ thị hàm mật độ của phân bố chuẩn tắc $\mathbf{N}(0;1)$



Hình 2.13b: Đồ thị hàm phân bố của phân bố chuẩn tắc $\mathbf{N}(0;1)$

C it $\Phi(x)$:

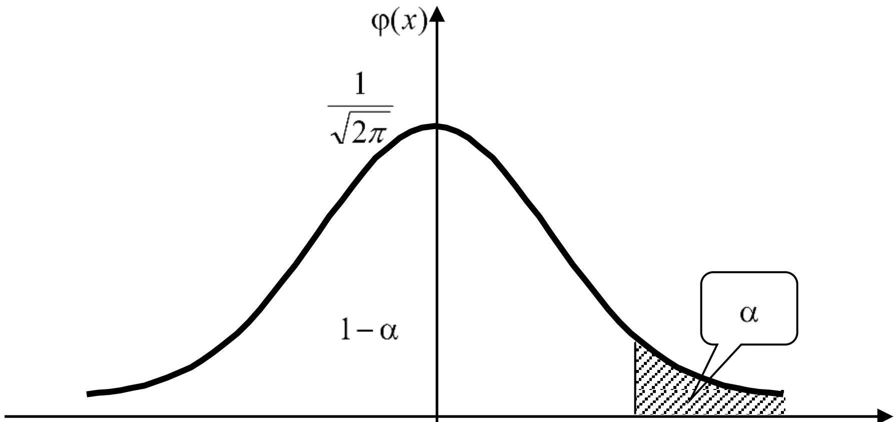
1) $\Phi(x)+\Phi(-x)=1$, $\Phi(-x)=1-\Phi(x)$.

2) Nếu $X\sim\mathbf{N}(0;1)$ thì

$$\forall a>0,\ P\{|X|<a\}=2\Phi(a)-1,\ P\{|X|>a\}=2(1-\Phi(a)). \qquad \textbf{(2.41)}$$

Định nghĩa 2.12: *Giá trị U_α gọi là giá trị tới hạn mức α của phân bố chuẩn tắc nếu*

$$\Phi^{-1}(1-\alpha)=U_\alpha. \qquad \textbf{(2.42)}$$



Từ (2.41)-(2.42) ta có: nếu $X \sim \mathbf{N}(0;1)$ thì

$$P\{X > U_\alpha\} = \alpha \; ; \; P\left\{|X| > U_{\frac{\alpha}{2}}\right\} = \alpha \; ; \; P\left\{|X| < U_{\frac{\alpha}{2}}\right\} = 1 - \alpha \, . \qquad \textbf{(2.43)}$$

Ta có thể chứng minh được (xem ví dụ 3.15 chương 3):

$$\text{Nếu } X \sim \mathbf{N}(\mu;\sigma^2) \text{ thì } \frac{X-\mu}{\sigma} \sim \mathbf{N}(0;1) \, . \qquad \textbf{(2.44)}$$

Từ đó ta có

$$F_X(x) = P\{X \le x\} = P\left\{\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right). \qquad \textbf{(2.45)}$$

$$P\{a < X < b\} = P\{a < X \le b\} = P\left\{\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \qquad \textbf{(2.46)}$$

Ví dụ 2.17: Giả sử $X \sim \mathbf{N}(\mu;\sigma^2)$, $\mu = 2100, \sigma = 200$. Hãy tìm:

a) $P\{X < 2400\}$.

b) $P\{1700 < X < 2200\}$.

c) Xác định a để $P\{X > a\} = 0,03$.

Giải: Áp dụng công thức (2.44), (2.45), (2.46) ta có:

a) $P\{X < 2400\} = P\{X \le 2400\} = \Phi\left(\dfrac{2400-2100}{200}\right) = \Phi(1,5) = 0,9332$.

b) $P\{1700 < X < 2200\} = \Phi\left(\dfrac{2200-2100}{200}\right) - \Phi\left(\dfrac{1700-2100}{200}\right)$

$$= \Phi(0,5) - \Phi(-2) = 0,6688 \, .$$

c) $P\{X > a\} = 1 - \Phi\left(\dfrac{a-2100}{200}\right) = 0,03 \;\Rightarrow\; \Phi\left(\dfrac{a-2100}{200}\right) = 0,97$

Tra bảng ta được $0,97 = \Phi(1,88) \;\Rightarrow\; \dfrac{a-2100}{200} = 1,88 \;\Rightarrow\; a = 2476$.

2.3.2.4.4 Quy tắc hai xích ma và ba xích ma

Áp dụng công thức (2.41), (2.44) ta có thể tính xác suất của sự sai lệch giữa biến ngẫu nhiên có phân bố chuẩn $X \sim \mathbf{N}(\mu;\sigma^2)$ và tham số μ của nó theo công thức

$$P\{|X-\mu|<\varepsilon\} = 2\Phi\left(\frac{\varepsilon}{\sigma}\right)-1 \qquad \textbf{(2.47)}$$

Nếu trong công thức (2.47) ta đặt $\varepsilon = 2\sigma$ tức là bằng hai lần độ lệch chuẩn của X (xem mục 2.42) thì $P\{|X-\mu|<2\sigma\} = 2\Phi(2)-1 = 0,9546$. Vậy

$$P\{\mu-2\sigma < X < \mu+2\sigma\} = 0,9546 \qquad \textbf{(2.48)}$$

Tương tự thay $\varepsilon = 3\sigma$ ta được

$$P\{\mu-3\sigma < X < \mu+3\sigma\} = 0,9974 \qquad \textbf{(2.49)}$$

Hai công thức trên là cơ sở của quy tắc hai xích ma và ba xích ma:

Nếu X có phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì có đến 95,46% giá trị của X nằm trong khoảng $(\mu-2\sigma;\ \mu+2\sigma)$ và hầu như toàn bộ giá trị của X nằm trong khoảng $(\mu-3\sigma;\ \mu+3\sigma)$.

2.3.2.5 Phân bố "khi bình phương"

Định nghĩa 2.13: *Biến ngẫu nhiên liên tục X có phân bố "khi bình phương" n bậc tự do, ký hiệu $X \sim \chi_n^2$ nếu hàm mật độ xác suất có dạng*

$$f_X(x) = \begin{cases} \dfrac{\left(\dfrac{x}{2}\right)^{n/2-1}}{\Gamma(n/2)} \dfrac{1}{2} e^{-\frac{x}{2}} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \le 0 \end{cases} \qquad \textbf{(2.50)}$$

trong đó $\Gamma(x) = \int\limits_0^{+\infty} t^{x-1}e^{-t}dt$, $x > 0$ là hàm Gamma.

Phân bố χ^2 do Karl Pearson đưa ra vào năm 1900.

Có thể chứng minh được rằng nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập có cùng phân bố chuẩn tắc $\mathbf{N}(0,1)$ thì

$$\sum_{i=1}^{n} X_i^2 = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi_n^2 \qquad \textbf{(2.51)}$$

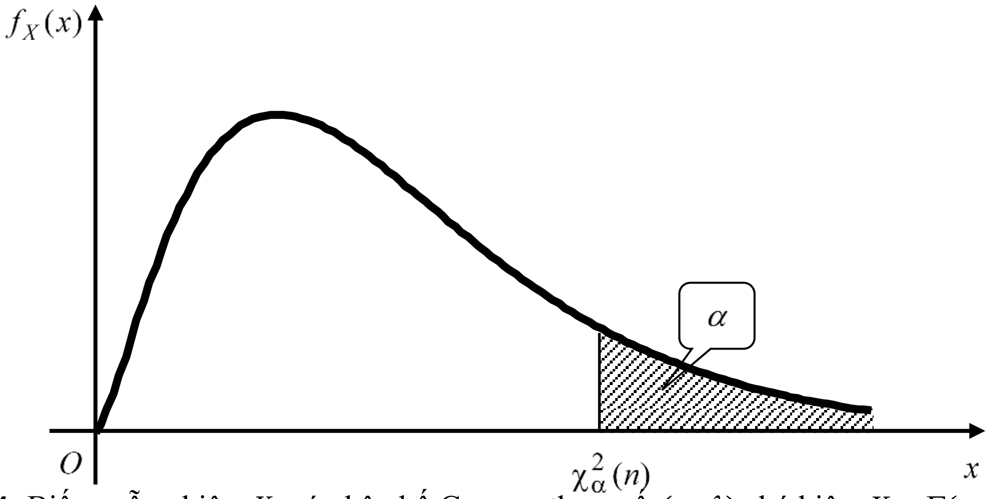
Từ (2.51) suy ra rằng nếu X_1, X_2 là hai biến ngẫu nhiên độc lập có phân bố "khi bình phương" lần lượt n_1 và n_2 bậc tự do thì $X_1 + X_2$ là biến ngẫu nhiên có phân bố "khi bình phương" $n_1 + n_2$ bậc tự do

$$X_1 + X_2 \sim \chi_{n_1+n_2}^2 \qquad \textbf{(2.52)}$$

Giá trị tới hạn "khi bình phương" n bậc tự do mức α, ký hiệu $\chi_\alpha^2(n)$, được định nghĩa như sau:

$$P\{\chi^2 > \chi_\alpha^2(n)\} = \alpha\,. \qquad \textbf{(2.53)}$$

Bảng các giá trị tới hạn $\chi^2_\alpha(n)$ được tính sẵn trong bảng ở Phụ lục IV.



Hình 2.15: *Giá trị tới hạn của phân bố "khi bình phương"*

Nhận xét 2.4: Biến ngẫu nhiên X có phân bố Gamma tham số (α,λ), ký hiệu $X \sim \Gamma(\alpha;\lambda)$, nếu hàm mật độ có dạng

$$f_X(x)=\begin{cases}\dfrac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}\lambda e^{-\lambda x} & \text{nếu } \ x>0\\ \quad 0 & \text{nếu } \ x\le 0\end{cases} \qquad \textbf{(2.54)}$$

Trong đó α, $\lambda>0$.

- Khi $\alpha=k$ là một số nguyên dương thì phân bố $\Gamma(k;\lambda)$ là phân bố Erlang tham số $(k;\lambda)$ có hàm mật độ xác định bởi công thức (2.36).
- Khi $\alpha=1$ thì $\Gamma(1;\lambda)$ là phân bố mũ tham số λ.
- Khi $\lambda=1/2$, $\alpha=n/2$ thì phân bố $\Gamma(n/2,1/2)$ là phân bố "khi bình phương" n bậc tự do.

2.3.2.6 Phân bố Student $\mathbf{T}(n)$

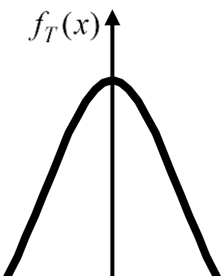
Định nghĩa 2.14: *Biến ngẫu nhiên liên tục T có phân bố Student n bậc tự do, ký hiệu $T\sim\mathbf{T}(n)$, nếu hàm mật độ xác suất có dạng*:

$$f_T(x)=\frac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(n/2\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty<x<\infty. \qquad \textbf{(2.55)}$$

trong đó $\Gamma(x)$ là hàm Gamma.

Người ta chứng minh được rằng nếu $Z\sim\mathbf{N}(0;1)$, $V\sim\chi^2_n$; Z và V độc lập thì

$$T=\frac{Z}{\sqrt{V/n}}\sim\mathbf{T}(n) \qquad \textbf{(2.56)}$$



Giá trị tới hạn mức α của phân bố Student n bậc tự do ký hiệu $t_{\alpha}(n)$ thỏa mãn:

$$P\{T > t_{\alpha}(n)\} = \alpha \,. \tag{2.57}$$

Bảng tính các giá trị tới hạn $t_{\alpha}(n)$ cho trong Phụ lục III.

Hàm mật độ xác suất (2.54) là hàm chẵn nên đồ thị đối xứng qua trục tung. Khi số bậc tự do tăng lên, phân bố Student hội tụ rất nhanh về phân bố chuẩn tắc $\mathbf{N}(0,1)$. Do đó khi n đủ lớn ($n \ge 30$) có thể dùng phân bố chuẩn tắc thay cho phân bố Student. Tuy nhiên khi n nhỏ ($n < 30$) việc thay thế như trên sẽ gặp sai số lớn.

2.4 CÁC THAM SỐ ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

2.4.1 Kỳ vọng toán

2.4.1.1 Định nghĩa

Với mọi biến ngẫu nhiên X ta ký hiệu $\mathrm{E}\,X$ hoặc $\mathrm{E}(X)$ hoặc $\mathrm{E}[X]$ và xác định như sau:

(i) Trường hợp biến ngẫu nhiên X rời rạc có miền giá trị R_X với hàm khối lượng xác suất $p_X(x_i)$, ta ký hiệu

$$\mathrm{E}\,X = \sum_{x_i \in R_X} x_i p_X(x_i) \tag{2.58}$$

(ii) Trường hợp biến ngẫu nhiên X liên tục có hàm mật độ xác suất $f_X(x)$, ta ký hiệu

$$\mathrm{E}\,X = \int_{-\infty}^{\infty} x f_X(x)dx \tag{2.59}$$

Nếu chuỗi (2.58) hội tụ tuyệt đối (trường hợp X rời rạc) hoặc tích phân (2.59) hội tụ tuyệt đối (trường hợp X liên tục) thì ta gọi $\mathrm{E}\,X$ *là kỳ vọng của biến ngẫu nhiên* X (expected value). Trường hợp ngược lại ta nói X không tồn tại kỳ vọng.

Kỳ vọng mang ý nghĩa là giá trị trung bình (average, mean value) của biến ngẫu nhiên X .

Ví dụ 2.18: Tính kỳ vọng của biến ngẫu nhiên X cho ở ví dụ 2.8.

Giải: $\mathrm{E}\,X = 0 \times \frac{5}{30} + 1 \times \frac{15}{30} + 2 \times \frac{9}{30} + 3 \times \frac{1}{30} = \frac{6}{5}$.

Ví dụ 2.19: Theo thống kê việc một người Mỹ 25 tuổi sẽ sống thêm trên một năm có xác suất là 0,992, còn xác suất để người đó chết trong vòng một năm tới là 0,008 (xem ví dụ 1.18). Một chương trình bảo hiểm đề nghị người đó bảo hiểm sinh mạng cho 1 năm với số tiền chi trả 1000 đô la, còn tiền đóng là 10 đô la. Hỏi lợi nhuận của công ty bảo hiểm nhận được là bao nhiêu?

Giải: Rõ ràng lợi nhuận là biến ngẫu nhiên X với 2 giá trị là $+10$ đô la (nếu người bảo hiểm không chết) và -990 đô la (nếu người đó chết). Bảng phân bố xác suất tương ứng.

X	–990	+10
P	$0,008$	$0,992$

Do đó kỳ vọng $\mathrm{E}\,X = (-990)\cdot 0,008 + 10\cdot 0,992 = 2$. Ta thấy lợi nhuận trung bình là một số dương vì vậy công ty bảo hiểm có thể làm ăn có lãi.

Ví dụ 2.20: Tuổi thọ của một loại côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất như sau:

$$f_X(x) = \begin{cases} kx^2(4-x) & \text{nếu } 0 \le x \le 4 \\ 0 & \text{nếu ngược lại} \end{cases}$$

Tìm hàm phân bố xác suất và tìm tuổi thọ trung bình của loài côn trùng trên.

Giải: Vì $\int\limits_0^4 x^2(4-x)dx = \frac{64}{3} \Rightarrow k = \frac{3}{64}$. Vậy hàm phân bố xác suất

$$F_X(x) = \int\limits_{-\infty}^{x} f_X(t)dt = \begin{cases} 0 & \text{nếu } x \le 0 \\ \dfrac{3x^3}{64}\left(\dfrac{4}{3} - \dfrac{x}{4}\right) & \text{nếu } 0 < x \le 4 \\ 1 & \text{nếu } x > 4 \end{cases}$$

Tuổi thọ trung bình $\mathrm{E}\,X = \frac{3}{64}\int\limits_0^4 x^3(4-x)dx = \frac{3}{64}\left(x^4 - \frac{x^5}{5}\right)\Bigg|_0^4 = \frac{12}{5}$ (tháng).

2.4.1.2 Ý nghĩa của kỳ vọng

Kỳ vọng mang ý nghĩa là giá trị trung bình mà biến ngẫu nhiên nhận được. Giả sử biến ngẫu nhiên X nhận các giá trị $x_1, x_2, ..., x_m$ với các tần số tương ứng $r_1, r_2, ..., r_m$.

$r_i x_i$ là tổng giá trị X nhận được với cùng giá trị x_i.

Do đó $r_1x_1 + r_2x_2 + \cdots + r_mx_m$ là tổng tất cả các giá trị X nhận được.

$\dfrac{r_1x_1 + r_2x_2 + \cdots + r_mx_m}{n}$ là giá trị trung bình của X, trong đó $r_1 + r_2 + \cdots + r_m = n$.

Đặt $f_i = \dfrac{r_i}{n}$ là tần suất nhận giá trị x_i của X.

$$\frac{r_1x_1 + r_2x_2 + \cdots + r_mx_m}{n} = f_1x_1 + f_2x_2 + \cdots + f_mx_m.$$

Trong trường hợp tổng quát thì tần suất f_i được thay bằng xác suất p_i.

Trường hợp biến ngẫu nhiên liên tục phép tính tổng của giá trị trung bình được thay bằng phép tính tích phân xác định.

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

2.4.1.3 Tính chất

1) $\mathrm{E}(C) = C$ với mọi hằng số C. **(2.60)**

2) $\mathrm{E}(CX) = C\mathrm{E}(X)$ với mọi hằng số C. **(2.61)**

3) $\mathrm{E}(X_1 + \cdots + X_n) = \mathrm{E}(X_1) + \cdots + \mathrm{E}(X_n)$ **(2.62)**

4) Cho hàm số $g(x)$, kỳ vọng của biến ngẫu nhiên $Y = g(X)$ được tính theo công thức

$$\mathrm{E}Y = \begin{cases} \sum\limits_{x_i \in R_X} g(x_i) p_X(x_i) & \text{nếu } X \text{ rời rạc với } p_X(x_i) = P\{X = x_i\} \\ \int\limits_{-\infty}^{\infty} g(x) f_X(x) dx & \text{nếu } X \text{ liê n tục có hàm mật độ } f_X(x) \end{cases} \quad \textbf{(2.63)}$$

Đặc biệt ta có công thức tính kỳ vọng của X^2:

$$\mathrm{E}X^2 = \begin{cases} \sum\limits_{x_i \in R_X} x_i^2 p_X(x_i) & \text{nếu } X \text{ rời rạc} \\ \int\limits_{-\infty}^{\infty} x^2 f_X(x) dx & \text{nếu } X \text{ liên tục có hàm mật độ } f_X(x) \end{cases} \quad \textbf{(2.34)}$$

5) Nếu $X_1, ..., X_n$ độc lập thì $\mathrm{E}(X_1 \cdots X_n) = \mathrm{E}(X_1) \cdots \mathrm{E}(X_n)$. **(2.65)**

Các tính chất (2.62)-(2.65) được chứng minh trong chương 3.

Ví dụ 2.21: Chọn ngẫu nhiên 3 bi từ một túi có 6 bi đen, 4 bi trắng. Xét hai bài toán sau:

a) Nếu chọn được 1 bi trắng sẽ được thưởng 200\$. Gọi Y là số tiền nhận được. Tính kỳ vọng của Y.

b) Nếu chọn được 1 bi trắng sẽ được thưởng 200\$ và chọn được 1 bi đen sẽ được thưởng 300\$. Gọi Z là số tiền nhận được. Tính kỳ vọng của Z.

Giải: a) Gọi X là số bi trắng trong 3 bi vừa chọn (xem ví dụ 2.8) thì $Y = g(X) = 200X$ là một biến ngẫu nhiên rời rạc có bảng phân bố sau:

$Y = g(X)$	0	200	400	600
P	$5/30$	$15/30$	$9/30$	$1/30$

$$\mathrm{E}Y = 0 \times \frac{5}{30} + 200 \times \frac{15}{30} + 400 \times \frac{9}{30} + 600 \times \frac{1}{30} = 240.$$

Mặt khác, theo công thức (2.61) và ví dụ 2.18 ta cũng được $\mathrm{E}Y = 200\mathrm{E}X = 200 \times \frac{6}{5} = 240$.

b) $Z = 200X + 300(3 - X) = 900 - 100X$

$$\Rightarrow \ \mathrm{E}\, Z = \mathrm{E}\big(900 - 100X\big) = 900 - 100\mathrm{E}X = 900 - 100 \times \frac{6}{5} = 780\$.$$

Ví dụ 2.22: Tung con xúc xắc n lần. Tìm kỳ vọng của tổng số chấm thu được.

Giải: Gọi $X_i \ (i = 1,...,n)$ là số chấm thu được ở lần tung thứ i, gọi X là tổng số chấm thu được trong n lần tung. Như vậy $X = \sum_{i=1}^{n} X_i$. Theo công thức (2.62) ta có $\mathrm{E}\, X = \sum_{i=1}^{n} \mathrm{E}\, X_i$.

Các biến ngẫu nhiên X_i đều có bảng phân bố xác suất như sau

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

Do đó $\mathrm{E}\, X_i = \frac{1}{6}\big(1+2+3+4+5+6\big) = \frac{7}{2} \ \Rightarrow \ \mathrm{E}\, X = \frac{7}{2} n.$

Ví dụ 2.23: (Bài toán đố). Ví dụ này khi mở rộng một cách thích hợp sẽ là mô hình về sự lựa chọn thứ tự tối ưu thỏa mãn điều kiện cho trước.

Xét trò chơi trả lời 2 câu hỏi A và B; người chơi có quyền chọn câu hỏi nào để trả lời đầu tiên. Câu hỏi A được trả lời đúng với xác suất 0.8 và khi đó người chơi sẽ được thưởng \$100, câu hỏi B được trả lời đúng với xác suất 0,6 và người chơi được thưởng \$200. Nếu không trả lời đúng lần thứ nhất sẽ không được trả lời tiếp. Vậy người chơi nên chọn câu hỏi nào trả lời đầu tiên để tiền thưởng trung bình nhận được cao hơn.

Gọi X là số tiền thưởng nhận được khi người chơi chọn câu hỏi A trả lời đầu tiên thì

X	\$0	\$100	\$300
P	0,2	0,8.0,4	0,8.0,6

Do đó $\mathrm{E}\, X = 0,8.0,4.100 + 0,8.0,6.300 = \$176$

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

Y	\$0	\$200	\$300
P	0,4	0,6.0,2	0,6.0,8

Do đó

$$\mathrm{E}Y = 0,6.0,2.200 + 0,6.0,8.300 = \$168$$

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

2.4.2 Phương sai

2.4.2.1 Định nghĩa

Phương sai (variance) *hay độ lệch* (deviation) *bình phương trung bình của biến ngẫu nhiên* X *là đại lượng đo sự phân tán bình phương trung bình của* X *xung quanh giá trị trung bình* $\mathrm{E}\, X$. *Nói cách khác phương sai của* X *là kỳ vọng của* $\left(X - \mathrm{E}\, X\right)^2$.

Phương sai của X được ký hiệu $\mathrm{D}X$ hoặc $\mathrm{Var}\, X$.

Vậy $$\mathrm{D}X=\mathrm{E}\left(X-\mathrm{E}\,X\right)^2 \qquad \textbf{(2.66)}$$

Độ lệch chuẩn của X là $\sigma_X=\sqrt{\mathrm{D}X}$.

Khai triển vế phải công thức (2.66) và áp dụng các tính chất của kỳ vọng ta có thể tính phương sai theo công thức sau:

$$\mathrm{D}X=\mathrm{E}\,X^2-\left(\mathrm{E}\,X\right)^2 \qquad \textbf{(2.67)}$$

Sử dụng công thức (2.64), (2.67) ta suy ra công thức thường dùng để tính phương sai của biến ngẫu nhiên rời rạc và phương sai của biến ngẫu nhiên liên tục:

(i). Nếu X rời rạc nhận các giá trị x_i với hàm khối lượng xác suất $p_X(x_i)$ thì

$$\mathrm{D}\,X=\mathrm{E}\,X^2-\left(\mathrm{E}\,X\right)^2=\sum_{x_i\in R_X}x_i^2p_X(x_i)-\left(\sum_{x_i\in R_X}x_ip_X(x_i)\right)^2 \qquad \textbf{(2.68)}$$

(ii). Nếu X liên tục có hàm mật độ xác suất $f_X(x)$ thì

$$\mathrm{D}\,X=\mathrm{E}\,X^2-\left(\mathrm{E}\,X\right)^2=\int_{-\infty}^{\infty}x^2f_X(x)dx-\left(\int_{-\infty}^{\infty}xf_X(x)dx\right)^2 \qquad \textbf{(2.69)}$$

Ví dụ 2.24: Tính phương sai của biến ngẫu nhiên xét trong ví dụ 2.19.

Giải: $\mathrm{E}\,X^2=(-990)^2\cdot 0{,}008+10^2\cdot 0{,}992=7940$

$$\Rightarrow\ \mathrm{D}X=\mathrm{E}X^2-\left(\mathrm{E}X\right)^2=7940-4=7936\ \Rightarrow\ \sigma_X=\sqrt{\mathrm{D}X}=\sqrt{7936}\approx 89{,}08.$$

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

Ví dụ 2.25: Tính phương sai và độ lệch chuẩn của biến ngẫu nhiên xét trong ví dụ 2.20.

Giải: $$\mathrm{E}\,X^2=\frac{3}{64}\int_0^4x^4(4-x)dx=\frac{3}{64}\left(\frac{4x^5}{5}-\frac{x^6}{6}\right)\Bigg|_0^4=\frac{32}{5}$$

$$\Rightarrow\ \mathrm{D}X=\mathrm{E}X^2-\left(\mathrm{E}X\right)^2=\frac{32}{5}-\left(\frac{12}{5}\right)^2=\frac{16}{25}\ \Rightarrow\ \sigma_X=\frac{4}{5}.$$

Phương sai của biến ngẫu nhiên X là độ lệch bình phương trung bình quanh giá trị trung bình $\mathrm{E}\,X$. Trong kỹ thuật phương sai đặc trưng cho mức độ phân tán của các chi tiết gia công hay sai số của thiết bị. Trong quản lý và kinh doanh thì phương sai đặc trưng cho mức độ rủi ro của các quyết định.

Ví dụ 2.19 cho thấy đầu tư bảo hiểm cho những người 25 tuổi là có lãi, nhưng ví dụ 2.24 cho thấy rủi ro của bảo hiểm rất lớn.

2.4.2.2 Tính chất

1) $\mathrm{D}(a)=0$ với mọi hằng số a . **(2.70)**

2) $\mathrm{D}(aX+b)=a^2\mathrm{D}(X)$ với mọi hằng số $a,\,b$. **(2.71)**

3) Nếu $X_1,\,...,X_n$ độc lập và có các phương sai hữu hạn thì

$$\mathrm{D}\left(a_1X_1+\cdots+a_nX_n\right)=a_1^2\mathrm{D}\left(X_1\right)+\cdots+a_n^2\mathrm{D}\left(X_n\right). \qquad \textbf{(2.72)}$$

Nói riêng: Nếu X,Y độc lập và $\mathrm{D}X$, $\mathrm{D}Y$ hữu hạn thì $\mathrm{D}\left(X\pm Y\right)=\mathrm{D}X+\mathrm{D}Y$.

Ví dụ 2.26: Tung con xúc xắc n lần độc lập nhau. Tìm phương sai của tổng số chấm xuất hiện.

Giải: Xét $X=\sum\limits_{i=1}^{n}X_i$ ở ví dụ 2.14. Vì các $X_i\ (i=1,...,n)$ độc lập nhau, do đó theo công thức (2.72) ta có $\mathrm{D}X=\sum\limits_{i=1}^{n}\mathrm{D}X_i$.

Mặt khác $\quad \mathrm{E}\,X_i=\dfrac{7}{2}$; $\quad \mathrm{E}\,X_i^2=\dfrac{1}{6}\left(1^2+2^2+3^2+4^2+5^2+6^2\right)=\dfrac{91}{6}$

Do đó $\quad \mathrm{D}X_i=\dfrac{91}{6}-\dfrac{7^2}{2^2}=\dfrac{35}{12}$. $\quad$ Vậy $\mathrm{D}X=\dfrac{35}{12}n$.

2.4.3 Phân vị, Trung vị

2.4.3.1 Phân vị

Phân vị mức α của biến ngẫu nhiên X, ký hiệu v_α, là giá trị phân chia miền giá trị của X thỏa mãn

$$P\{X<v_\alpha\}\le\alpha\le P\{X\le v_\alpha\}$$

Nghĩa là

$$F_X(v_\alpha-)\le\alpha\le F_X(v_\alpha) \qquad \textbf{(2.73)}$$

- Trường hợp biến ngẫu nhiên X liên tục thì phân vị v_α là điểm phân chia miền giá trị của X thành 2 miền với xác suất tương ứng là α và $1-\alpha$. Vậy v_α là nghiệm duy nhất của phương trình $F_X(x)=\alpha$.

$$v_\alpha=F_X^{-1}(\alpha) \qquad \textbf{(2.74)}$$

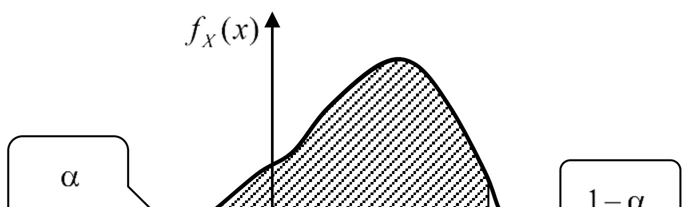
Giá trị tới hạn mức α là phân vị mức $1-\alpha$. **(2.75)**

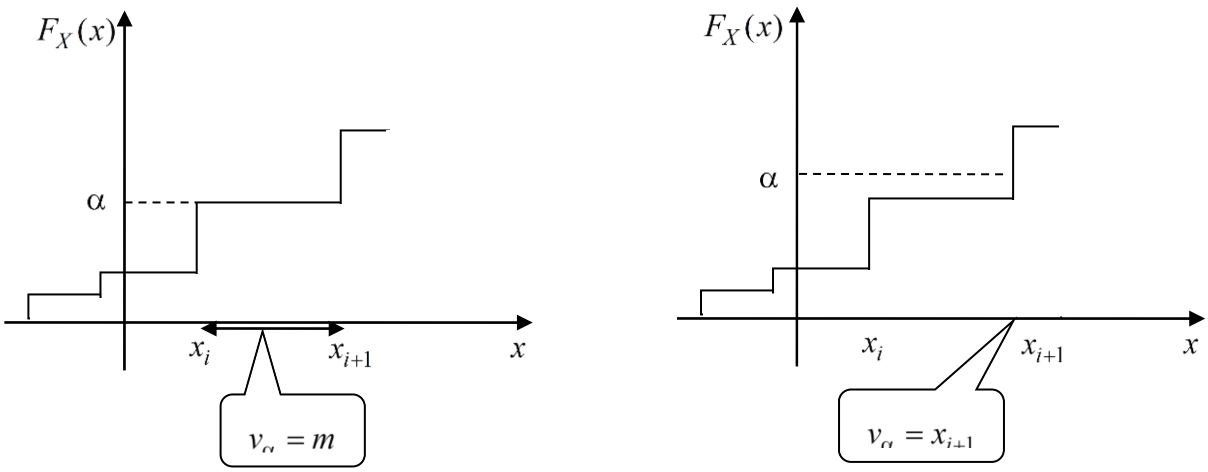
- Trường hợp biến ngẫu nhiên X rời rạc có phân bố:

X	x_1	x_2	…
P	$p_X(x_1)$	$p_X(x_1)$	…

Ta có: $F_X(x_i)=p_X(x_1)+\cdots+p_X(x_i)$ thì

$$v_\alpha=\begin{cases} m,\ \ \forall m\in\left[x_i,x_{i+1}\right) & \text{nếu } F_X(x_i)=\alpha<F_X(x_{i+1})\\ x_{i+1} & \text{nếu } F_X(x_i)<\alpha<F_X(x_{i+1})\end{cases} \qquad \textbf{(2.76)}$$





Hình 2.18: Phân vị mức α *của biến ngẫu nhiên rời rạc*

2.4.3.2 Trung vị

Phân vị mức 1/2 được gọi là *median* hay *trung vị* của X , ký hiệu $\operatorname{Med} X$. Như vậy trung vị là điểm phân chia phân bố xác suất thành hai phần bằng nhau.

2.4.4 Mốt

Mốt (Mode) của biến ngẫu nhiên X là giá trị mà biến ngẫu nhiên X nhận với xác suất lớn nhất. Một biến ngẫu nhiên có thể có nhiều Mốt.

- Mốt của biến ngẫu nhiên rời rạc X với bảng phân bố:

X	x_1	x_2	…
P	$p_X(x_1)$	$p_X(x_1)$	…

được xác định như sau:

$$x_{i_0} = \operatorname{Mod} X \iff p_X(x_{i_0}) = \max\{p_X(x_1), p_X(x_2), ...\} \quad \textbf{(2.77)}$$

- Nếu X liên tục có hàm mật độ xác suất $f_X(x)$ thì

$$c = \operatorname{Mod} X \iff f_X(c) = \max\{f_X(x),\ x \in \mathbb{R}\}. \quad \textbf{(2.78)}$$

Ví dụ 2.27: Biến ngẫu nhiên X ở ví dụ 2.8 có Mốt và trung vị $\operatorname{Mod} X = \operatorname{Med} X = 1$.

Ví dụ 2.28: Tìm trung vị và Mốt của biến ngẫu nhiên rời rạc có bảng phân bố xác suất

X	20	21	22	23	24
P	0,3	0,25	0,18	0,14	0,13

Giải: Dễ thấy rằng $\operatorname{Mod} X = 20$.

Hàm phân bố xác suất của X

$$F_X(x) = \begin{cases} 0 & \text{nếu} \quad x < 20 \\ 0,3 & \text{nếu} \quad 20 \le x < 21 \\ 0,55 & \text{nếu} \quad 21 \le x < 22 \\ 0,73 & \text{nếu} \quad 22 \le x < 23 \\ 0,87 & \text{nếu} \quad 23 \le x < 24 \\ 1 & \text{nếu} \quad x \ge 24 \end{cases}$$

Từ đó suy ra $\operatorname{Med} X = 21$; các giá trị $x \in [22; 23)$ là phân vị mức 0,73 của X.

Ví dụ 2.29: Tìm $\operatorname{Med} X$ và $\operatorname{Mod} X$ của biến ngẫu nhiên liên tục X xét trong ví dụ 2.6

Giải: $\operatorname{Med} X$ là nghiệm của phương trình $F_X(x) = x^2 = \dfrac{1}{2} \Rightarrow \operatorname{Med} X = \dfrac{1}{\sqrt{2}}$.

Hàm mật độ xác suất $f_X(x) = \begin{cases} 0 & \text{với} \quad x \le 0 \\ 2x & \text{với} \quad 0 < x \le 1 \\ 0 & \text{với} \quad x > 1 \end{cases}$ đạt cực đại tại $x = 1$, vậy $\operatorname{Mod} X = 1$.

Ví dụ 2.30: Tìm $\operatorname{Med} X$ và $\operatorname{Mod} X$ của biến ngẫu nhiên liên tục X có hàm mật độ xác suất xác định như sau

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

Giải: Hàm phân bố xác suất

$$F_X(x) = \begin{cases} 0 & \text{với} \quad x \le 0 \\ \dfrac{3}{4}\left(x^2 - \dfrac{x^3}{3}\right) & \text{với} \quad 0 < x \le 2 \\ 1 & \text{với} \quad x > 2 \end{cases}$$

$\operatorname{Med} X$ là nghiệm của phương trình $F_X(x) = \dfrac{1}{2} \Leftrightarrow \begin{cases} x^3 - 3x^2 + 2 = 0 \\ 0 < x \le 2 \end{cases}$. Từ đó $\operatorname{Med} X = 1$.

Hàm mật độ xác suất $f_X(x)$ có đạo hàm $f_X'(x) = \begin{cases} \dfrac{3}{2}(1-x) & \text{với} \quad 0 < x < 2 \\ 0 & \text{nếu trái lại} \end{cases}$

đổi dấu từ dương sang âm khi đi qua $x = 1$, do đó đạt cực đại tại điểm này. Vậy $\operatorname{Mod} X = 1$.

- Mốt của phân bố nhị thức (xem công thức (1.20))

$$X \sim \mathscr{B}(n; p) \text{ thì } \operatorname{mod} X = [(n+1)p] \qquad \textbf{(2.79)}$$

- Mốt của phân bố Poisson tham số $\lambda > 0$

$$X \sim \mathscr{P}(\lambda) \text{ thì } \operatorname{mod} X = [\lambda]. \quad \textbf{(2.80)}$$

2.4.5 Moment, hệ số bất đối xứng, hệ số nhọn (*)

1) Moment cấp k $\quad m_k = \mathrm{E}X^k$; $k = 1, 2, \ldots$ **(2.81)**

2) Moment quy tâm cấp k $\quad \mu_k = \mathrm{E}\left(X - \mathrm{E}X\right)^k$; $k = 1, 2, \ldots$ **(2.82)**

3) Hệ số bất đối xứng $\quad \alpha_3 = \dfrac{\mu_3}{\sigma^3}$ với $\sigma = \sqrt{\mathrm{D}X}$. **(2.83)**

4) Hệ số nhọn $\quad \alpha_4 = \dfrac{\mu_4}{\sigma^4}$. **(2.84)**

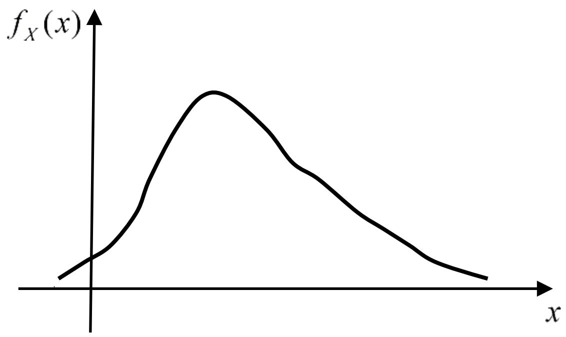
Nhận xét 2.5:

- $m_1 = \mathrm{E}X$, $\mu_1 = 0$, $\mu_2 = \mathrm{D}X$.
- α_3 đo mức độ bất đối xứng của luật phân bố :

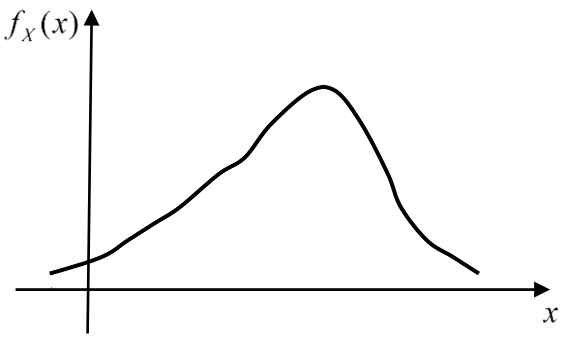
Nếu $\alpha_3 < 0$ thì phân bố xác suất và đồ thị của hàm mật độ xác suất sẽ lệch về bên trái hơn.

$\alpha_3 = 0$ thì phân bố xác suất và đồ thị của hàm mật độ xác suất đối xứng.

$\alpha_3 > 0$ thì phân bố xác suất và đồ thị của hàm mật độ xác suất sẽ lệch về bên phải hơn.



Hình 2.19a: Hệ số bất đối xứng $\alpha_3 < 0$



Hình 2.19b: Hệ số bất đối xứng $\alpha_3 > 0$

- Hệ số nhọn α_4 đặc trưng cho độ nhọn của đồ thị hàm mật độ xác suất so với đồ thị hàm mật độ xác suất của phân bố chuẩn.

Với biến ngẫu nhiên có phân bố chuẩn thì $\alpha_4 = 3$.

$\alpha_4 > 3$ thì đồ thị hàm mật độ xác suất sẽ nhọn hơn so với đồ thị hàm mật độ xác suất chuẩn.

$\alpha_4 < 3$ thì đồ thị hàm mật độ xác suất sẽ tù hơn so với đồ thị hàm mật độ xác suất chuẩn.

- Khi phân bố của X đối xứng hoặc gần đối xứng thì dùng kỳ vọng để định vị là tốt nhất, song nếu phân bố của X quá lệch thì nên dùng Median và Mode để định vị.

2.4.6 Kỳ vọng và phương sai của các biến ngẫu nhiên có phân bố xác suất thường gặp

Phân bố xác suất của X	**Hàm khối lượng $p_X(x_k)$ hoặc mật độ $f_X(x)$**	**Kỳ vọng** $\mathrm{E}X$	**Phương sai** $\mathrm{D}X$
Bernoulli tham số p	$p_X(k) = p^k q^{1-k}$; $k = 0,1$	p	pq

Nhị thức $\mathscr{B}(n;p)$	$p_X(k)=C_n^k p^k q^{n-k};\ k=0,1,\ldots,n$	np	npq
Poisson $\mathscr{P}(\lambda)$	$p_X(k)=e^{-\lambda}\dfrac{\lambda^k}{k!};\lambda>0;\ k=0,1,2,\ldots$	λ	λ
Đều $\mathbf{U}(a,b)$	$\dfrac{1}{b-a},\ a<x<b$	$(a+b)/2$	$(b-a)^2/12$
Phân bố mũ tham số $\lambda>0$	$\lambda e^{-\lambda x},\ x>0$	$1/\lambda$	$1/\lambda^2$
Phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$	$f_X(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}};\ \forall x\in\mathbb{R}$	μ	σ^2
Phân bố Erlang tham số $(k;\lambda)$	$\dfrac{(\lambda x)^{k-1}}{(k-1)!}\lambda e^{-\lambda x},\ x>0$	k/λ	$\dfrac{k}{\lambda^2}$
"Khi bình phương" n bậc tự do	$\dfrac{(x/2)^{n/2-1}}{\Gamma(n/2)}\dfrac{1}{2}e^{-(x/2)},\ x>0$	n	$2n$
Student n bậc tự do, $n>2$	$\dfrac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}\left(1+\dfrac{x^2}{n}\right)^{-(n+1)/2},\ x\in\mathbb{R}$	0	$\dfrac{n}{n-2}$
Gamma $\Gamma(\alpha,\lambda)$	$\dfrac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}\lambda e^{-\lambda x},\ x>0$	$\dfrac{\alpha}{\lambda}$	$\dfrac{\alpha}{\lambda^2}$

TÓM TẮT

Biến ngẫu nhiên X được nghiên cứu qua hàm phân bố xác suất $F_X(x)=P\{X\le x\}$.

- Nếu biến ngẫu nhiên X rời rạc với hàm khối lượng xác suất $p_X(x_k),\ x_k\in R_X$

  Hàm phân bố được xác định theo công thức $F_X(x)=\sum\limits_{x_k\le x;x_k\in R_X} p_X(x_k)$

  Ngược lại $p_X(x_k)=F_X(x_k)-F_X(x_{k-1}^{\ -}),\ x_k\in R_X$.

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

$$F_X(x)=\int_{-\infty}^{x} f_X(x)dx \text{ và } f_X(x)=\frac{d}{dx}F_X(x).$$

Kỳ vọng $\quad \mathrm{E}\,X=\sum\limits_{x_k\in R_X} x_k p_X(x_k)$ (trường hợp rời rạc)

$\mathrm{E}\,X=\int_{-\infty}^{\infty} xf_X(x)dx$ (trường hợp liên tục)

Phương sai $\quad \mathrm{D}\,X=\mathrm{E}(X-\mathrm{E}\,X)^2=\mathrm{E}\,X^2-(\mathrm{E}\,X)^2$, trong đó

$$\mathrm{E}\,X^2 = \sum_{x_k \in R_X} x_k^2 p_X(x_k) \text{ (trường hợp rời rạc)}$$

$$\mathrm{E}\,X^2 = \int_{-\infty}^{\infty} x^2 f_X(x)dx \quad \text{(trường hợp liên tục)}$$

Phân vị mức α của biến ngẫu nhiên X, ký hiệu v_α, là giá trị thỏa mãn $F_X(v_\alpha -) \le \alpha \le F_X(v_\alpha)$.

$$x_{i_0} = \mathrm{Mod}\,X \quad \Leftrightarrow \quad p_X(x_{i_0}) = \max\{p_X(x_1), p_X(x_2), ...\} \text{ (trường hợp rời rạc)}$$

$$c = \mathrm{Mod}\,X \quad \Leftrightarrow \quad f_X(c) = \max\{f_X(x),\, x \in \mathbb{R}\} \quad \text{(trường hợp liên tục)}$$

CÂU HỎI ÔN TẬP VÀ BÀI TẬP

2.1 Biến ngẫu nhiên luôn luôn nhận giá trị dương.

Đúng ☐ Sai ☐.

2.2 Biến ngẫu nhiên rời rạc chỉ nhận một số hữu hạn các giá trị.

Đúng ☐ Sai ☐.

2.3 Nếu biến ngẫu nhiên X rời rạc chỉ nhận các giá trị $x_1, ..., x_n$ thì hệ các biến cố $\{X = x_1\}$, …, $\{X = x_n\}$ lập thành một hệ đầy đủ.

Đúng ☐ Sai ☐.

2.4 Kỳ vọng của biến ngẫu nhiên rời rạc là giá trị nó lấy thường xuyên nhất.

Đúng ☐ Sai ☐.

2.5 Kỳ vọng của tổng hai biến ngẫu nhiên luôn luôn bằng tổng các kỳ vọng của nó.

Đúng ☐ Sai ☐.

2.6 Hai biến ngẫu nhiên có cùng kỳ vọng sẽ có cùng phương sai.

Đúng ☐ Sai ☐.

2.7 Phương sai của tổng hai biến ngẫu nhiên rời rạc luôn luôn bằng tổng phương sai của nó.

Đúng ☐ Sai ☐.

2.8 Biến ngẫu nhiên tồn tại phương sai thì cũng tồn tại kỳ vọng.

Đúng ☐ Sai ☐.

2.9 Hàm mật độ xác suất $f(x)$ của biến ngẫu nhiên liên tục có tính chất $f(x) \ge 0$.

Đúng ☐ Sai ☐.

2.10 Tổng của hai biến ngẫu nhiên phân bố theo phân bố nhị thức bất kỳ luôn luôn là một biến ngẫu nhiên phân bố theo phân bố nhị thức.

Đúng ☐ Sai ☐.

2.11 Biến ngẫu nhiên phân bố theo phân bố Poisson là biến ngẫu nhiên rời rạc nên chỉ nhận một số hữu hạn các giá trị.

Đúng ☐ Sai ☐.

2.12 Nếu X là biến ngẫu nhiên có phân bố Poisson tham số $\lambda > 0$ thì kỳ vọng, phương sai và Mốt của X đều bằng λ.

☐ ☐

Đúng Sai .

2.13 Nếu biến ngẫu nhiên X phân bố theo phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì xác suất sai lệch giữa X và kỳ vọng của nó thỏa mãn $P\{|X-\mu|<\varepsilon\}=2\Phi\left(\frac{\varepsilon}{\sigma}\right)-1$.

Đúng ☐ Sai ☐.

2.14 Nếu biến ngẫu nhiên X phân bố theo phân bố chuẩn $\mathbf{N}(\mu;\sigma^2)$ thì $\frac{X-\mu}{\sigma}$ có phân bố chuẩn tắc $\mathbf{N}(0;1)$.

Đúng ☐ Sai ☐.

2.15 Biến ngẫu nhiên có phân bố Student chỉ nhận những giá trị dương.

Đúng ☐ Sai ☐.

2.16 Xác định các hằng số a, b sao cho hàm số sau là hàm phân bố của biến ngẫu nhiên X nào đó

$$F(x)=\begin{cases}1-ae^{-x/b} & \text{nếu } x\ge 0\\ 0 & \text{nếu } x<0\end{cases}$$

2.17 Biến ngẫu nhiên X có bảng phân bố

X	-5	2	3	4
P	0,4	0,3	0,1	0,2

Tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

2.18 Biến ngẫu nhiên rời rạc X nhận ba giá trị có thể có là x_1, x_2, x_3. Biết $x_1=0,6, x_2=4$ với xác suất tương ứng $p_1=0,3$, $p_2=0,5$ và có kỳ vọng $\mathrm{E}X=8$. Tìm x_3 và p_3.

2.19 Cho X_1 và X_2 là hai biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X_1	2	3	5
P	0,3	0,5	0,2

X_2	1	4
P	0,2	0,8

a) Tính $\mathrm{E}X_1$; $\mathrm{E}X_2$; $\mathrm{D}X_1$; $\mathrm{D}X_2$.

b) Tính $\mathrm{E}(X_1-X_2)$ và $\mathrm{D}(X_1-X_2)$.

2.20 Cho X_1, X_2, X_3 là ba biến ngẫu nhiên độc lập có bảng phân bố xác suất như sau:

X_1	0	2
P	0,6	0,4

X_2	1	2
P	0,4	0.6

X_3	0	2
P	0,8	0.2

Lập bảng phân bố xác suất của $\overline{X}=\frac{X_1+X_2+X_3}{3}$. Tính $\mathrm{E}(\overline{X})$; $\mathrm{D}(\overline{X})$.

2.21 Hai biến ngẫu nhiên X, Y độc lập. Tính $\mathrm{D}(Z)$ với:

a) $Z = 2X + 3Y$. b) $Z = -3X + Y$.

Cho biết $\mathrm{D}(X) = 4$, $\mathrm{D}(Y) = 5$.

2.22 Biến ngẫu nhiên rời rạc X nhận các giá trị có thể có là $x_1 = -1$; $x_2 = 0$; $x_3 = 1$. Tìm các xác suất tương ứng p_1; p_2; p_3 biết rằng $\mathrm{E}(X) = 0{,}1$ và $\mathrm{D}(X) = 0{,}89$.

2.23 Xếp ngẫu nhiên 5 hành khách lên 3 toa tầu I, II, III. Gọi X là số khách lên toa I và Y là số khách lên toa II và III.

a) Tính xác suất để cả 3 toa đều có khách.

b) Lập bảng phân bố xác suất của biến ngẫu nhiên X và biến ngẫu nhiên Y.

2.24 Tính kỳ vọng của biến ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f_X(x) = \begin{cases} \dfrac{\cos x}{2} & \text{nếu } x \in (-\pi/2; \pi/2) \\ 0 & \text{nếu } x \notin (-\pi/2; \pi/2) \end{cases}$$

2.25 Tuổi thọ của một loài côn trùng nào đó là một biến ngẫu nhiên X (đơn vị là tháng) với hàm mật độ xác suất như sau

$$f_X(x) = \begin{cases} kx^2(2-x) & \text{nếu } 0 \le x \le 2 \\ 0 & \text{nếu trái lại} \end{cases}$$

a) Tìm k; Tính xác suất để côn trùng chết trước khi nó được một tháng tuổi;

b) Tìm $\mathrm{E}X$, $\mathrm{D}X$.

2.26 Hai xạ thủ $\mathscr{A}$ và $\mathscr{B}$ tập bắn. Mỗi người bắn hai phát. Xác suất bắn trúng đích của $\mathscr{A}$ trong mỗi lần bắn là 0,4; còn của $\mathscr{B}$ là 0,5.

a) Gọi X là số phát bắn trúng của $\mathscr{A}$ trừ đi số phát bắn trúng của $\mathscr{B}$. Tìm phân bố xác suất của X, kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

b) Tìm phân bố xác suất của $Y = |X|$ và kỳ vọng $\mathrm{E}Y$.

2.27 Một xí nghiệp có hai ôtô vận tải hoạt động. Xác suất trong ngày làm việc các ôtô bị hỏng tương ứng bằng 0,1 và 0,2. Gọi X là số ôtô bị hỏng trong thời gian làm việc. Lập bảng phân bố xác suất, tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$ của X.

2.28 Cho biến ngẫu nhiên rời rạc X có bảng phân bố xác suất

X	1	2	3	4	5	6	7
P	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

a) Xác định k.

b) Tính xác suất $P\{X \ge 5\}$ và $P\{X < 3\}$.

c) Tính kỳ vọng $\mathrm{E}X$.

d) Tính phương sai $\mathrm{D}X$.

2.29 Có 5 sản phẩm trong đó có 4 chính phẩm và 1 phế phẩm. Người ta lấy ra lần lượt 2 sản phẩm (lấy không hoàn lại).

a) Gọi X là "số phế phẩm có thể gặp phải". Lập bảng phân bố xác suất của X .

Tính kỳ vọng $\mathrm{E}X$ và phương sai $\mathrm{D}X$.

b) Gọi Y là "số chính phẩm có thể nhận được". Lập hệ thức cho biết mối quan hệ giữa Y và X . Tính kỳ vọng $\mathrm{E}Y$ và phương sai $\mathrm{D}Y$.

2.30 Một nhóm có 10 người trong đó có 6 nam và 4 nữ. Chọn ngẫu nhiên ra 3 người. Gọi X là số nữ có trong nhóm được chọn. Lập bảng phân bố xác suất của X . Tính kỳ vọng $\mathrm{E}X$.

2.31 Hai kiện tướng bóng bàn ngang sức thi đấu với nhau. Hỏi thắng 2 trong 4 ván dễ hơn hay thắng 3 trong 6 ván dễ hơn.

2.32 Trong một lô hàng có 800 sản phẩm loại 1 và 200 sản phẩm loại 2. Lấy ngẫu nhiên ra 5 sản phẩm theo phương thức có hoàn lại. Gọi X là số sản phẩm loại 1 lấy được.

a) X có phân bố gì? Viết biểu thức tổng quát của phân bố.

b) Tìm kỳ vọng và phương sai của X .

c) Tìm mốt của X và tính khả năng để xảy ra điều đó.

2.33 Xác suất để sản phẩm sản xuất ra bị hỏng bằng 0,1.

a) Tìm xác suất để trong 5 sản phẩm sản xuất ra có không quá 2 sản phẩm hỏng.

b) Tìm số sản phẩm hỏng trung bình trong 5 sản phẩm đó.

c) Tìm số sản phẩm hỏng có khả năng xảy ra nhiều nhất.

2.34 Một bài thi trắc nghiệm gồm có 10 câu hỏi, mỗi câu hỏi có 5 phương án trả lời, trong đó chỉ có một phương án đúng. Giả sử mỗi câu trả lời đúng được 4 điểm và câu trả lời sai bị trừ 2 điểm. Một học sinh kém làm bài bằng cách chọn hú hoạ một phương án cho mỗi câu hỏi. Tính xác suất để:

a) Anh ta được 4 điểm.

b) Anh ta bị điểm âm.

2.35 Tín hiệu thông tin được phát đi 5 lần độc lập nhau. Xác suất thu được tin của mỗi lần phát là 0,7. Tính xác suất:

a) Thu được tín hiệu đúng 2 lần.

b) Thu được tín hiệu nhiều nhất 1 lần.

c) Thu được tin.

2.36 Một cầu thủ nổi tiếng về đá phạt đền, xác suất đá vào gôn là 4/5. Có người cho rằng cứ “sút” 5 quả thì chắc chắn rằng có 4 quả vào lưới. Điều khẳng định đó có đúng không? Tìm xác suất để trong 5 lần sút có đúng 4 lần bóng vào lưới.

2.37 Ở một tổng đài bưu điện các cuộc điện thoại gọi đến xuất hiện một cách ngẫu nhiên, độc lập với nhau và trung bình có 2 cuộc gọi trong một phút. Tính xác suất để:

a) Có ít nhất một cuộc gọi trong khoảng thời gian 10 giây.

b) Trong khoảng thời gian 3 phút có nhiều nhất ba cuộc gọi.

c) Trong khoảng thời gian 3 phút liên tiếp mỗi phút có nhiều nhất một cuộc gọi.

2.38 Giả sử X là biến ngẫu nhiên rời rạc với phân bố cấp số nhân tham số p có hàm khối lượng xác suất $p_X(x)$ thỏa mãn công thức (2.16).

a) Nghiệm lại công thức $\sum_{k=1}^{\infty} p_X(k) = 1$. Tìm hàm phân bố $F_X(x)$.

b) Tính kỳ vọng $\mathrm{E}\,X$, phương sai $\mathrm{D}\,X$.

2.39 Biến ngẫu nhiên X tuân theo phân bố chuẩn với kỳ vọng $\mu = 10$ và phương sai $\sigma^2 = 4$. Tính xác suất để X nhận giá trị trong khoảng (8; 12).

2.40 Biến ngẫu nhiên X tuân theo phân bố chuẩn với kỳ vọng $\mu = 10$. Xác suất để X nhận giá trị trong khoảng (10; 20) là 0,3. Tìm xác suất để X nhận giá trị trong khoảng (0; 10).

2.41 Trọng lượng sản phẩm X do một máy tự động sản xuất là một biến ngẫu nhiên tuân theo phân bố chuẩn với $\mu = 100$ gam và độ lệch chuẩn $\sigma = 1$ gam. Sản phẩm được coi là đạt tiêu chuẩn kỹ thuật nếu trọng lượng của nó đạt từ 98 đến 102gam.

a) Tìm tỷ lệ sản phẩm đạt tiêu chuẩn kỹ thuật của nhà máy.

b) Tìm tỷ lệ phế phẩm của nhà máy.

c) Giải thích bằng đồ thị kết quả tìm được ở phần a).

2.42 Cho X_i $(i = \overline{1,n})$ là các biến ngẫu nhiên độc lập, cùng tuân theo phân bố chuẩn với

$$\mathrm{E}(X_1) = \mathrm{E}(X_2) = ... = \mathrm{E}(X_n) = \mu\,; \quad \mathrm{D}(X_1) = \mathrm{D}(X_2) = ... = \mathrm{D}(X_n) = \sigma^2$$

Lập công thức tính $P\left\{\left|\overline{X} - \mu\right| < \varepsilon\right\}$ biết rằng $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ và cũng tuân theo phân bố chuẩn, $\varepsilon > 0$ tùy ý.

2.43 Biến ngẫu nhiên liên tục X được gọi là *biến ngẫu nhiên Rayleigh* tham số σ nếu hàm mật độ có dạng

$$f_X(x) = \begin{cases} \dfrac{x}{\sigma^2} e^{-x^2/(2\sigma^2)} & x > 0 \\ 0 & x < 0 \end{cases}$$

a) Xác định hàm phân bố $F_X(x)$

b) Tính kỳ vọng $\mathrm{E}\,X$, phương sai $\mathrm{D}\,X$.