Chương 2: BIẾN NGẪU NHIÊN RỜI RẠC

Vũ Thị Hương

2nd May 2021

Mục lục

2.1 Biến ngẫu nhiên rời rạc

Khái niệm

Trong nhiều phép thử ngẫu nhiên ta không chỉ quan tâm đến các biến cố của phép thử mà ta còn quan tâm đến giá trị đo các thuộc tính của của các biến cố đó. Với mỗi kết quả của một phép thử chúng ta gán với một số đặc trưng cho kết quả đó. Sau đây ta xem xét một số ví dụ mở đầu về biến ngẫu nhiên rời rạc.

Ví dụ 2.1

Một người bỏ ra 2 đồng để tham gia một trò chơi: anh ta gieo một đồng xu 3 lần, gọi X là biến chỉ số lần gieo được mặt sấp. Nếu $X = 2$ thì anh ta được nhận lại 1 đồng, $X = 3$ thì anh ta được nhận 8 đồng, trong các trường hợp còn lại anh ta không được nhận đồng nào. Gọi Y là biến ngẫu nhiên chỉ số tiền anh ta thu được sau một lần chơi.

Ký hiệu S và N tương ứng là biến cố gieo được mặt sấp và ngửa trong một lần gieo.

Ta có mối liên hệ giữa các biến cố của không gian mẫu và giá trị của biến X và Y như sau:

Không gian mẫu	SSS	SSN	SNS	NSS	SNN	NSN	NNS	NNN
X	3	2	2	2	1	1	1	0
Y	6	-1	-1	-1	-2	-2	-2	-2

Qua ví dụ này chúng ta thấy có thể có nhiều biến ngẫu nhiên cùng xác định trên một không gian mẫu. Trong trường hợp này ta thấy biến ngẫu nhiên Y là một hàm của biến ngẫu nhiên X.

Ví dụ 2.2

Xét phép thử: gieo 1 con xúc xắc cân đối, đồng chất. Gọi X là biến ngẫu nhiên chỉ số chấm thu được khi gieo. Tập các giá trị bnn X có thê nhận là $S = \{1, 2, 3, 4, 5, 6\}$.

Ví dụ 2.3

Một sinh viên ném bóng vào rổ cho đến khi nào ném trúng rổ thì dừng lại. Gọi Z là biến ngẫu nhiên chỉ số lần sinh viên này thực hiện ném bóng. Khi đó tập các giá trị có thể nhận của biến Z là tập vô hạn nhưng đếm được

$$\{1, 2, 3, ..., n, ...\}.$$

Định nghĩa 2.1

Ta gọi biến ngẫu nhiên có tập giá trị có thể nhận $\mathcal{S}$ là tập hữu hạn phần tử $\mathcal{S} = \{x_1, x_2, ..., x_n\}$ hoặc là tập vô hạn phần tử nhưng đếm được $\mathcal{S} = \{x_1, x_2, ..., x_n, ...\}$ là biến ngẫu nhiên rời rạc.

Một số ký hiệu:

- $(X = a)$ là biến cố "bnn X nhận giá trị a"
- $(X < a)$ là biến cố "bnn X nhận giá trị nhỏ hơn a"
- Các biến cố $(X \leqslant a), (a < X < b), (|X| < a)$ v.v. được hiểu tương tự.

Ví dụ

Ký hiệu X là đại lượng ngẫu nhiên chỉ số chấm xuất hiện khi gieo con xúc xắc. Tương ứng ta có các phân tích:

$$(X < 2,3) = (X = 1) + (X = 2),$$
$$(X \geqslant 4) = (X = 4) + (X = 5) + (X = 6).$$

2.2 Bảng phân phối xác suất và hàm khối xác suất

Luật phân phối xác suất được hiểu là cách đưa thông tin về một đại lượng ngẫu nhiên X sao cho ta có thể xác định được hoàn toàn các giá trị xác suất $\mathbb{P}(X = a), \mathbb{P}(X < a), \mathbb{P}(a < X < b)$ v.v..

Hai cách đưa ra luật phân phối xác suất cho đại lượng ngẫu nhiên rời rạc:

- Sử dụng bảng phân phối xác suất hoặc hàm khối xác suất.
- Sử dụng hàm phân phối xác suất.

Các luật phân phối rời rạc được sử dụng trong thực tế

- Luật phân phối đều rời rạc.
- Luật phân phối Bernoulli, luật nhị thức.
- Luật phân phối siêu hình học.
- Luật phân phối hình học.
- Luật phân phối nhị thức âm.
- Luật phân phối Poisson.

Giả sử X là biến ngẫu nhiên rời rạc có tập giá trị có thể nhận là $\mathcal{S} = \{x_1, x_2, ..., x_n\}$. Khi đó bảng phân phối xác suất của X cho ta biết xác suất để biến ngẫu nhiên nhận các giá trị tương ứng. Bảng phân phối xác suất thường có dạng như sau:

X	x_1	x_2	...	x_n
$\mathbb{P}$	p_1	p_2	...	p_n

trong đó

1. $p_i = \mathbb{P}(X = x_i)$ với mọi $i = 1, 2, ..., n$,
2. $\sum_{i=1}^{n} p_i = 1$.

Ví dụ 2.4

Trong Ví dụ 2.1 ta thấy

$$\mathbb{P}(X=0)=\mathbb{P}(X=3)=\frac{1}{8};\ \mathbb{P}(X=1)=\mathbb{P}(X=2)=\frac{3}{8};$$
$$\mathbb{P}(Y=-2)=\frac{1}{2};\ \mathbb{P}(Y=-1)=\frac{3}{8};\ \mathbb{P}(Y=6)=\frac{1}{8}.$$

Vậy bảng phân phối xác suất của biến ngẫu nhiên X là

X	0	1	2	3
$\mathbb{P}$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Ví dụ 2.5

Một tổ vận tải có hai xe tải để vận chuyển hàng hóa cùng hoạt động. Xác suất trong một ngày làm việc hai xe đó gặp sự cố lần lượt là $0,1$ và $0,15$. Gọi X là biến ngẫu nhiên chỉ số xe gặp sự cố trong một ngày làm việc. Lập bảng phân phối xác suất của biến ngẫu nhiên X.

Giải: Tập giá trị có thể nhận của biến X là $\mathcal{S} = \{0, 1, 2\}$. Gọi A_1 và A_2 lần lượt là biến cố xe thứ nhất và xe thứ hai gặp sự cố trong một ngày làm việc. Vì 2 xe hoạt động độc lập nên hai biến cố A_1 và A_2 là độc lập với nhau với $\mathbb{P}(A_1) = 0,1$; $\mathbb{P}(A_2) = 0,15$.
Biến cố $(X = 0) = \overline{A}_1.\overline{A}_2$ là biến cố không có xe nào gặp sự cố trong 1 ngày làm việc. Áp dụng công thức nhân xác suất ta có $\mathbb{P}(X = 0) = \mathbb{P}(\overline{A}_1.\overline{A}_2) = \mathbb{P}(\overline{A}_1)\mathbb{P}(\overline{A}_2) = (1-0,1).(1-0,15) = 0,765.$

2.2 Bảng phân phối xác suất và hàm khối xác suất

Bảng phân phối xác suất

Tương tự, $(X=2) = A_1 \cap A_2$ là biến cố có 2 xe gặp sự cố trong 1 ngày làm việc. Ta có

$$\mathbb{P}(X=2) = \mathbb{P}(A_1.A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2) = 0,1.0,15 = 0,015.$$

Biến cố $(X=1) = A_1.\overline{A}_2 + \overline{A}_1.A_2$ là biến cố có 1 xe gặp sự cố trong 1 ngày làm việc. Ta có

$$\begin{aligned}\mathbb{P}(X=1) &= \mathbb{P}(A_1.\overline{A}_2) + \mathbb{P}(A_1.\overline{A}_2)\\ &= \mathbb{P}(A_1)\mathbb{P}(\overline{A}_2) + \mathbb{P}(\overline{A}_1)\mathbb{P}(A_2)\\ &= 0,1.(1-0,15) + (1-0,1).0,15\\ &= 0,22.\end{aligned}$$

Ta cũng có thể tính nhanh $\mathbb{P}(X=1)$ theo công thức $\mathbb{P}(X=1) = 1-\mathbb{P}(X=0)-\mathbb{P}(X=2) = 1-0,765-0,015 = 0,22.$

Khi đó bảng phân phối xác suất của biến ngẫu nhiên X là

X	0	1	2
$\mathbb{P}$	0,765	0,22	0,015

Để thuận tiện cho việc biểu diễn quy luật phân phối xác suất của những biến ngẫu nhiên rời rạc, trong nhiều trường hợp người ta thường dùng một hàm số để mô tả cho luật phân phối xác suất của biến. Hàm số đó được gọi là hàm khối xác suất.

Định nghĩa 2.2

Giả sử X là biến ngẫu nhiên rời rạc có tập giá trị có thể nhận là $\mathcal{S}$. Hàm số $f(x)$ xác định trên tập $\mathcal{S}$ được gọi là hàm khối xác xuất của X nếu

1. $f(x_i) \geqslant 0$, với mọi $x_i \in \mathcal{S}$,
2. $\sum_{x_i \in \mathcal{S}} f(x_i) = 1$,
3. $f(x_i) = \mathbb{P}(X = x_i)$ với mọi $x_i \in \mathcal{S}$.

Ví dụ 2.6

Trở lại ví dụ 2.5, bnn X nhận các giá trị: 0, 1, 2. Hàm khối xác suất của biến ngẫu nhiên X xác định trên tập $S = \{0, 1, 2\}$, cụ thể

$$f(0) = 0,765; \quad f(1) = 0,22; \quad f(2) = 0,015.$$

Nhận xét: biết bảng phân phối xác suất thì biết được hàm khối xác suất và ngược lại.

Ví dụ 2.7

Cho biến ngẫu nhiên rời rạc X có hàm khối xác suất xác định bởi công thức

$$f(x) = P(X = x) = \begin{cases} kx, & \text{nếu } x = 2, 4, 6, \\ k(x-2), & \text{nếu } x = 8, \\ 0, & \text{trường hợp khác} . \end{cases}$$

Tìm hằng số k và lập bảng phân phối xác suất của biến ngẫu nhiên X.

Giải: Từ giả thiết, tập giá trị của biến ngẫu nhiên X là $\mathcal{S} = \{2, 4, 6, 8\}$.

Theo định nghĩa của hàm khối xác suất

$$\sum_{x \in \mathcal{S}} f(x) = 1 \Leftrightarrow 2k + 4k + 6k + 6k = 1 \Leftrightarrow k = \frac{1}{18}.$$

Đồng thời $k = 1/18$ thỏa mãn điều kiện $f(x) \geq 0$ với mọi x. Vậy $k = 1/18$. Dẫn tới

$$P(X=2) = f(2) = 2k = 1/9, \quad P(X=4) = f(4) = 4k = 2/9,$$
$$P(X=6) = f(6) = 6k = 1/3, \quad P(X=8) = f(8) = 6k = 1/3.$$

Vậy bảng phân phối xác suất của X là

X	2	4	6	8
P	1/9	2/9	1/3	1/3

2.3 Hàm phân phối xác suất

Định nghĩa hàm phân phối xác suất

Định nghĩa 2.3

Giả sử X là biến ngẫu nhiên rời rạc có tập giá trị có thể nhận là $\mathcal{S}$ và có hàm khối xác suất là $f(x)$. Hàm số $F(x)$ xác định bởi:

$$F(x) = \mathbb{P}(X \leqslant x) = \sum_{x_i \leqslant x} f(x_i), \text{ với } -\infty < x < \infty, \tag{2.2}$$

được gọi là hàm phân phối xác xuất hay hàm phân phối tích lũy của biến ngẫu nhiên X.

Chú ý hàm phân phối xác suất $F(x)$ xác định với mọi giá trị $x \in \mathbb{R}$.

2.3 Hàm phân phối xác suất

Tính chất của hàm phân phối xác suất

Giả sử biến ngẫu nhiên X có tập giá trị là $\mathcal{S} = \{x_1, x_2, \cdots, x_n\}$, trong đó các giá trị được sắp xếp tăng dần $x_1 < x_2 < \cdots < x_n$ với hàm khối xác suất $f(x_i), i = 1, 2, \ldots, n$.

Hàm phân phối xác suất là hàm bậc thang được xác định như sau

$$F(x) = \begin{cases} 0 & \text{nếu } x < x_1 \\ f(x_1) & \text{nếu } x_1 \leqslant x < x_2 \\ f(x_1) + f(x_2) & \text{nếu } x_2 \leqslant x < x_3 \\ \vdots & \vdots \\ f(x_1) + f(x_2) + \ldots + f(x_k) & \text{nếu } x_k \leqslant x < x_{k+1} \\ \vdots & \vdots \\ 1 & \text{nếu } x \geqslant x_n \end{cases}$$

Ví dụ 2.8

Hàm phân phối xác suất của biến ngẫu nhiên X trong ví dụ 2.7 là

$$F(x) = \begin{cases} 0, & \text{nếu } x < 2 \\ 1/9, & \text{nếu } 2 \le x < 4 \\ 1/9 + 2/9, & \text{nếu } 4 \le x < 6 \\ 1/9 + 2/9 + 1/3, & \text{nếu } 6 \le x < 8 \\ 1/9 + 2/9 + 1/3 + 1/3, & \text{nếu } x \ge 8 \end{cases}$$

$$= \begin{cases} 0, & \text{nếu } x < 2 \\ 1/9, & \text{nếu } 2 \le x < 4 \\ 1/3, & \text{nếu } 4 \le x < 6 \\ 2/3, & \text{nếu } 6 \le x < 8 \\ 1, & \text{nếu } x \ge 8. \end{cases}$$

Hàm phân phối xác suất có các tính chất sau:

- $0 \leqslant F(x) \leqslant 1$ với mọi $-\infty < x < \infty$.
- Nếu $x < y$ thì $F(x) \leqslant F(y)$.
- $\mathbb{P}(a < X \leqslant b) = F(b) - F(a)$.
- $F(x)$ là hàm liên tục phải và có giới hạn trái tại mỗi điểm.

Cho hàm phân phối xác suất $F(x)$ có dạng bậc thang với các bước nhẩy tại các điểm x_k thì ta dễ dàng tìm được hàm khối xác suất theo công thức sau:

$$\begin{aligned}
f(x_1) &= F(x_1), \\
f(x_2) &= F(x_2) - F(x_1), \\
\dots & \qquad \dots \\
f(x_k) &= F(x_k) - F(x_{k-1}), \\
\dots & \qquad \dots \\
f(x_n) &= F(x_n) - F(x_{n-1}).
\end{aligned}$$

Ví dụ 2.9

Cho hàm phân phối xác suất của biến ngẫu nhiên X như sau:

$$F(x) = \begin{cases} 0 & x < -2 \\ 0,2 & -2 \leqslant x < 0 \\ 0,7 & 0 \leqslant x < 2 \\ 1 & 2 \leqslant x. \end{cases}$$

Tìm hàm khối xác suất của X.

Giải: Hàm khối xác suất xác định tại các điểm mà làm thay đổi giá trị của hàm phân phối xác suất tại các điểm đó. Ta có hàm khối xác suất của biến ngẫu nhiên X là

$$\begin{aligned} f(-2) &= 0,2 - 0 = 0,2 \\ f(0) &= 0,7 - 0,2 = 0,5 \\ f(2) &= 1 - 0,7 = 0,3. \end{aligned}$$

Định nghĩa 2.4

Cho X là biến ngẫu nhiên rời rạc có tập giá trị có thể là $\mathcal{S}$ và có hàm khối xác suất là $f(x), x \in \mathcal{S}$.

- Giá trị trung bình hay còn gọi là kỳ vọng của X được ký hiệu là μ hoặc $\mathbb{E}[X]$ và được xác định bởi
$$\mu = \mathbb{E}[X] = \sum_{x \in \mathcal{S}} x f(x). \tag{2.3}$$
- Phương sai của X được ký hiệu là σ^2 hoặc $\mathbb{V}[X]$ và được xác định bởi
$$\sigma^2 = \mathbb{V}[X] = \mathbb{E}[X-\mu]^2 = \sum_{x \in \mathcal{S}} (x-\mu)^2 f(x). \tag{2.4}$$
- Độ lệch chuẩn của X là
$$\sigma = \sqrt{\sigma^2} \tag{2.5}$$

Nếu ta khai triển công thức tính phương sai (2.4), ta nhận được

$$\begin{aligned}\mathbb{V}[X] = \sum_{x\in\mathcal{S}}(x-\mu)^2 f(x) &= \sum_{x\in\mathcal{S}} x^2 f(x) - 2\mu\sum_{x\in\mathcal{S}} x f(x) + \mu^2\sum_{x\in\mathcal{S}} f(x)\\ &= \sum_{x\in\mathcal{S}} x^2 f(x) - 2\mu.\mu + \mu^2\\ &= \sum_{x\in\mathcal{S}} x^2 f(x) - \mu^2 = \mathbb{E}[X^2] - \mu^2.\end{aligned}$$

Công thức tính phương sai

$$\sigma^2 = \mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \sum_{x\in\mathcal{S}} x^2 f(x) - \mu^2. \tag{2.8}$$

Nếu X có bảng phân phối xác suất

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

thì ta viết

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i p_i, \quad \mathbb{E}[X^2] = \sum_{i=1}^{n} x_i^2 p_i.$$

☞ $\mathbb{E}[X]$ là giá trị trung bình của các giá trị có thể nhận của biến X theo trọng số xác suất.

☞ Trong cơ học, nếu $f(x)$ là trọng lực đặt lên một thanh dài, mỏng tại ví trí x thì $\mathbb{E}[X]$ chính là điểm cân bằng của thanh. Như vậy $\mathbb{E}[X]$ là một đặc trưng cho số đo tâm của X.

☞ Phương sai $\mathbb{V}[X]$ của biến X là số đo mức độ phân tán của các giá trị có thể của biến X xung quanh giá trị trung bình.

☞ Hai biến ngẫu nhiên có thể có cùng giá trị trung bình nhưng phương sai khác nhau. Đối với hai biến như thế biến ngẫu nhiên có phương sai nhỏ hơn được nhìn nhận là ổn định hơn và dễ kiểm soát hơn.

Ví dụ minh họa cho $\mathbb{E}X$

Một lớp có 40 sinh viên. Sau khi hoàn thành một môn học thì có 10 sinh viên đạt 5 điểm, 14 sinh viên đạt 6 điểm, 8 sinh viên đạt 7 điểm, 5 sinh viên đạt 8 điểm, 3 sinh viên đạt 9 điểm.

1. Điểm trung bình của cả lớp là:

$$\begin{aligned} \text{Điểm TB} &= \frac{\text{tổng điểm}}{\text{số sinh viên}} \\ &= \frac{5.10 + 6.14 + 7.8 + 8.5 + 9.3}{40} = \frac{257}{40} = 6,425 \end{aligned}$$

2. Chọn ngẫu nhiên một sinh viên của lớp. Gọi X là điểm thi của sinh viên được chọn. Ta có bảng phân phối xác suất

2.4 Kỳ vọng và phương sai

Định nghĩa của kỳ vọng và phương sai

X	5	6	7	8	9
P	$\frac{10}{40}$	$\frac{14}{40}$	$\frac{8}{40}$	$\frac{5}{40}$	$\frac{3}{40}$

Kỳ vọng của X

$$\begin{aligned}\mathbb{E}[X] &= \sum_{x_i \in S} x_i f(x_i) \\ &= 5 \times \frac{10}{40} + 6 \times \frac{14}{40} + 7 \times \frac{8}{40} + 8 \times \frac{5}{40} + 9 \times \frac{3}{40} = 6,425\end{aligned}$$

Ví dụ

Gọi X là đại lượng ngẫu nhiên chỉ mức thu nhập hàng tháng của một kỹ sư được chọn ngẫu nhiên. Tương ứng $\mathbb{E}X$ được hiểu là giá trị trung bình của mức thu nhập hàng tháng trong tập hợp tất cả các kỹ sư.

- Trong trường hợp tổng quát nếu X là biến ngẫu nhiên rời rạc có hàm khối xác suất là $f(x), x \in \mathcal{S}$ và $h(x)$ là hàm số xác định trên tập $\mathcal{S}$ thì $Y = h(X)$ cũng là một biến ngẫu nhiên rời rạc.
- Giá trị trung bình của $h(X)$ được tính theo công thức

$$\mathbb{E}[h(X)] = \sum_{x \in \mathcal{S}} h(x)f(x). \tag{2.6}$$

- Nói riêng, ta có thể tính giá trị trung bình của biến X^2 theo công thức

$$\mathbb{E}[X^2] = \sum_{x \in \mathcal{S}} x^2 f(x). \tag{2.7}$$

Ví dụ 2.10

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

X	0	1	2
P	0,765	0,220	0,015

Hãy tìm $\mathbb{E}[X]$ và $\mathbb{V}[X]$.

Giải:

$$\mathbb{E}[X] = 0.0,765 + 1.0,220 + 2.0,015 = 0,25;$$
$$\mathbb{E}[X^2] = 0^2.0,765 + 1^2.0,220 + 2^2.0,015 = 0,28;$$
$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = 0,28 - 0,25^2 = 0,2175.$$

Cho X là biến ngẫu nhiên, a, b, c là các hằng số. Khi đó ta có

1. $\mathbb{E}c = c$,
2. $\mathbb{V}c = 0$,
3. $\mathbb{E}aX = a\mathbb{E}X$,
4. $\mathbb{V}aX = a^2\mathbb{V}X$,
5. $\mathbb{E}aX + b = a\mathbb{E}X + b$,
6. $\mathbb{V}aX + b = a^2\mathbb{V}X$.

Biến ngẫu nhiên có phân phối đều rời rạc là biến có luật phân phối đơn giản nhất. Đó là những biến có tập giá trị có thể nhận là tập hữu hạn với xác suất để biến nhận các giá trị đó đều bằng nhau.

Định nghĩa 2.5

Biến ngẫu nhiên rời rạc X được gọi là có phân phối đều rời rạc trên tập $\{x_1, x_2, ..., x_n\}$ nếu hàm khối xác suất của nó thỏa mãn

$$f(x_i) = \frac{1}{n}, \quad \text{với mọi } i = 1, 2, ..., n.$$

Khi đó, giá trị kỳ vọng và phương sai của biến ngẫu nhiên X được xác định như sau:

$$\mu = \mathbb{E}[X] = \frac{x_1 + x_2 + \cdots + x_n}{n}, \tag{2.9}$$

$$\sigma^2 = \mathbb{V}[X] = \frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n} - \mu^2. \tag{2.10}$$

Nếu ta giả sử tập giá trị có thể nhận của biến là các số nguyên liên tiếp $a, a+1, a+2, ..., b,\ a < b$ thì giá trị trung bình và phương sai của X được tính như sau:

$$\mu = \mathbb{E}[X] = \frac{a+b}{2} \text{ và } \mathbb{V}[X] = \frac{(b-a+1)^2 - 1}{12}. \tag{2.11}$$

Ví dụ 2.12

Bước sóng của một bức xạ quang hợp có phân phối đều trên dải phổ đỏ $\{675, 676, ..., 700\}$ nm (nano mét). Hãy tìm hàm khối xác suất, giá trị trung bình và phương sai của biến ngẫu nhiên chỉ bước sóng của bức xạ này.

Giải: Hàm khối xác suất là hàm $f(x)$ thỏa mãn

$$f(675) = f(676) = \cdots = f(700) = \frac{1}{26}.$$

Giá trị trung bình của bước sóng là

$$\mathbb{E}[X] = \frac{675 + 700}{2} = 687,5 \ (nm).$$

Phương sai của bước sóng là

$$\mathbb{V}[X] = \frac{(700 - 675 + 1)^2 - 1}{12} = 56,25(nm^2).$$

2.5 Một số phân phối xác suất rời rạc

Phân phối nhị thức

Trước tiên chúng ta xét các mô hình về dãy các phép thử độc lập và biến ngẫu nhiên rời rạc gắn với dãy phép thử đó.

Ví dụ 2.13

1. Gieo một đồng xu 20 lần. Gọi X là biến ngẫu nhiên chỉ số lần gieo được mặt sấp.
2. Các bit mã hóa được truyền qua một kênh thông tin kỹ thuật số. Xác suất truyền lỗi của kênh là 5%. Gọi X là biến ngẫu nhiên chỉ số bit bị truyền lỗi trong số 4 bit được truyền qua kênh thông tin.
3. Một dây truyền sản xuất hàng loạt một loại bóng đèn, xác suất bóng bị lỗi của dây truyền đó là 1%. Gọi X là biến ngẫu nhiên chỉ số bóng bị lỗi trong số 200 bóng được dây truyền đó sản xuất.

2.5 Một số phân phối xác suất rời rạc

Phân phối nhị thức

Mỗi ví dụ trên là một mô hình mà ở đó có dãy phép thử độc lập: gieo 1 đồng xu 20 lần, truyền 4 bit qua một kênh thông tin ... Còn X là biến ngẫu nhiên đếm số phép thử có kết quả thỏa mãn một tiêu chí nào đó, chẳng hạn tiêu chí: gieo được mặt sấp, một bit bị lỗi,một bóng đèn bị lỗi. Ta chia kết quả có thể của mỗi phép thử thành hai loại:

☞ "thành công" nếu kết quả của phép thử đó thỏa mãn tiêu chí được đặt ra,

☞ "thất bại" nếu kết quả của phép thử đó không thỏa mãn tiêu chí được đặt ra.

Người ta gọi những phép thử có tập kết quả có thể nhận được phân thành hai loại như vậy là những phép thử Bernoulli. Hơn nữa trong mỗi phép thử Bernoulli, xác suất của kết quả "thành công" là hằng số không đổi p. Chẳng hạn trong các ví dụ trên $p = 0,5$; $p = 0,05$; $p = 0,01$.

Định nghĩa phân phối nhị thức

- Thực hiện n lần cùng một phép thử Bernoulli.
- Các phép thử là độc lập.
- Mỗi phép thử có hai kết quả có thể là "thành công" và "thất bại".
- Xác suất để một phép thử có kết quả "thành công" là hằng số p, với $0 < p < 1$.

Khi đó biến ngẫu nhiên rời rạc X chỉ số phép thử Bernoulli có kết quả "thành công" là biến ngẫu nhiên có phân phối nhị thức với hai tham số n và p. Hàm khối xác suất của X (theo công thức Becnoulli- Chương 1) là

$$f(x) = \mathbb{P}(X = x) = C_n^x\, p^x (1-p)^{n-x}, \; x = 0, 1, ..., n. \tag{2.12}$$

Ta kí hiệu biến X có phân phối nhị thức là $X \sim \mathscr{B}(n, p)$.

Giá trị kỳ vọng của phân phối nhị thức được tính như sau:

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{n} x f(x) = \sum_{x=0}^{n} x C_n^x p^x q^{n-x} \\
&= \sum_{x=1}^{n} x C_n^x p^x q^{n-x} = p \sum_{x=1}^{n} x C_n^x p^{x-1} q^{n-x} \\
&= p \frac{\partial}{\partial p} \left(\sum_{x=0}^{n} C_n^x p^x q^{n-x} \right) = p \frac{\partial}{\partial p} (p+q)^n = pn(p+q)^{n-1} = np.
\end{aligned}
$$

Tương tự như trên ta có

$$\begin{aligned}
\mathbb{E}[X(X-1)] &= \sum_{x=0}^{n} x(x-1)f(x) = \sum_{x=0}^{n} x(x-1)C_n^x p^x q^{n-x} \\
&= \sum_{x=2}^{n} x(x-1)C_n^x p^x q^{n-x} = p^2 \sum_{x=2}^{n} x(x-1)C_n^x p^{x-2} q^{n-x} \\
&= p \frac{\partial^2}{\partial p^2}\left(\sum_{x=0}^{n} C_n^x p^x q^{n-x}\right) = p \frac{\partial^2}{\partial p^2}(p+q)^n \\
&= pn(n-1)(p+q)^{n-2} = n(n-1)p^2.
\end{aligned}$$

Từ hệ thức trên ta có $\mathbb{E}[X^2] - \mathbb{E}[X] = n^2p^2 - np^2$ hay $\mathbb{E}[X^2] = np + n^2p^2 - np^2$. Vậy phương sai của phân phối nhị thức là

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = np + n^2p^2 - np^2 - (np)^2 = np(1-p).$$

Công thức kỳ vọng phương sai

Nếu X là biến ngẫu nhiên có phân phối nhị thức $B(n, p)$ với hai tham số n và p thì

$$\mathbb{E}[X] = np \text{ và } \mathbb{V}[X] = np(1-p). \tag{2.13}$$

Ví dụ 2.15

Gieo một đồng xu 20 lần. Gọi X là biến ngẫu nhiên chỉ số lần gieo được mặt sấp. Khi đó X có phân phối nhị thức $B(n = 20, p = 0,5)$.

Tương ứng hàm khối xác suất của biến X là

$$f(x) = C_{20}^{x}\ .0,5^{x}.0,5^{20-x} = C_{20}^{x}\ 0,5^{20},\ x = 0,1,...,20.$$

Giả sử ta có một lô hàng có N sản phẩm trong đó có M sản phẩm tốt (chính phẩm) và $N - M$ sản phẩm xấu (phế phẩm). Chọn ngẫu nhiên một mẫu gồm có n sản phẩm (chọn không hoàn lại) từ N sản phẩm trên với $n \leqslant N$. Gọi X là biến ngẫu nhiên chỉ số chính phẩm được chọn trong số n sản phẩm lấy ra.

Hàm khối xác suất của X có công thức

$$f(x) = \mathbb{P}(X = x) = \frac{C_M^x C_{N-M}^{n-x}}{C_N^n} \tag{2.14}$$

với $\max\{0, n + M - N\} \leqslant x \leqslant \min\{M, n\}$. Ta gọi X là biến ngẫu nhiên có phân phối siêu hình học và ký hiệu là $X \sim \mathscr{H}(n, M, N)$.

Tập các giá trị x của phân phối siêu hình học không vượt quá số sản phẩm được lấy ra n và tổng số sản phẩm tốt M nên $x \leqslant \min\{M, n\}$. Mặt khác, số sản phẩm tốt lấy ra hiển nhiên là không âm nên $x \geq 0$ và số phế phẩm lấy ra không vượt quá số phế phẩm trong lô hàng $n - x \leqslant N - M$ hay $x \geqslant n + M - N$. Vậy các giá trị của phân phối siêu hình học X có giá trị nguyên từ $\max\{0, n + M - N\}$ đến $\min\{M, n\}$.

Ví dụ 2.16

Một lô hàng có 100 sản phẩm loại A và 200 sản phẩm loại B. Rút ngẫu nhiên 4 sản phẩm để kiểm tra. Gọi X là số sản phẩm loại A. Lập bảng phân phối xác suất của X.

Giải: Trong trường hợp này biến ngẫu nhiên X có phân phối siêu hình học và X nhận các giá trị $\{0, 1, 2, 3, 4\}$.

Ta tính hàm khối xác suất của X như sau:

$$\mathbb{P}(X=0)=\frac{C_{100}^{0}C_{200}^{4}}{C_{300}^{4}}=0,1955;\quad \mathbb{P}(X=1)=\frac{C_{100}^{1}C_{200}^{3}}{C_{300}^{4}}=0,3970;$$

$$\mathbb{P}(X=2)=\frac{C_{100}^{2}C_{200}^{2}}{C_{300}^{4}}=0,2978;\quad \mathbb{P}(X=3)=\frac{C_{100}^{3}C_{200}^{1}}{C_{300}^{4}}=0,0978;$$

$$\mathbb{P}(X=4)=\frac{C_{100}^{4}C_{200}^{0}}{C_{300}^{4}}=0,0119.$$

Vậy bảng phân phối xác suất của X là

X	0	1	2	3	4
$\mathbb{P}$	0,1955	0,3970	0,2978	0,0978	0,0119

Tiếp theo ta tính kỳ vọng và phương sai của phân phối siêu hình học. Trước hết do $f(x)$ là hàm khối xác suất nên $\sum_x f(x) = 1$ hay

$$\sum_x C_M^x C_{N-M}^{n-x} = C_N^n. \tag{2.15}$$

Sử dụng công thức (2.15) ta có

$$\sum_x x C_M^x C_{N-M}^{n-x} = M \sum_x C_{M-1}^{x-1} C_{N-M}^{n-x} = M C_{N-1}^{n-1}.$$

Vậy kỳ vọng của phân phối siêu hình học là

$$\mathbb{E}[X] = \sum_x \frac{x C_M^x C_{N-M}^{n-x}}{C_N^n} = \frac{M C_{N-1}^{n-1}}{C_N^n} = \frac{nM}{N}.$$

Tiếp tục áp dụng công thức (2.15) ta có

$$\sum_x x(x-1) C_M^x C_{N-M}^{n-x} = M(M-1) \sum_x C_{M-2}^{x-2} C_{N-M}^{n-x} = M(M-1) C_{N-2}^{n-2}.$$

Từ đó ta có

$$\mathbb{E}[X(X-1)] = \sum_x \frac{x(x-1)C_M^x C_{N-M}^{n-x}}{C_N^n}$$
$$= \frac{M(M-1)C_{N-2}^{n-2}}{C_N^n} = \frac{n(n-1)M(M-1)}{N(N-1)}.$$

Khi đó phương sai của phân phối của phân phối siêu hình học là

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X(X-1)] + \mathbb{E}[X] - (\mathbb{E}[X])^2$$
$$= \frac{n(n-1)M(M-1)}{N(N-1)} + \frac{nM}{N} - \frac{n^2M^2}{N^2}$$
$$= \frac{nM}{N}\left(\frac{(n-1)(M-1)}{N-1} + 1 - \frac{nM}{N}\right)$$
$$= \frac{nM}{N}\left(\frac{(N-M)(N-n)}{N(N-1)}\right) = n\frac{M}{N}\left(1-\frac{M}{N}\right)\frac{N-n}{N-1}.$$

Công thức kỳ vọng phương sai

Cho X có phân phối siêu hình học với các tham số N, M, và n thì

$$\mathbb{E}X = np, \mathbb{V}X = np(1-p)\frac{N-n}{N-1} \tag{2.16}$$

với $p = \dfrac{M}{N}$.

Tham số p ở trên được hiểu là tỷ lệ thành công. Đối với phân phối siêu hình học thì giá trị kỳ vọng bằng giá trị kỳ vọng của phân phối nhị thức. Tuy nhiên, giá trị phương sai của hai biến ngẫu nhiên này lại khác nhau, bởi vì các phép thử trong phân phối nhị thức là độc lập còn các phép thử trong phân phối siêu hình học không độc lập.

Thừa số $\dfrac{N-n}{N-1}$ được gọi là hệ số hiệu chỉnh cho mẫu hữu hạn.

Trong phần này chúng ta tiếp tục xét dãy phép thử độc lập Bernoulli (dãy phép thử độc lập và xác suất thành công trong một phép thử là hằng số p). Tuy nhiên chúng ta sẽ thực hiện liên tiếp các phép thử Bernoulli cho đến khi phép thử đầu tiên có được kết quả "thành công" thì dừng lại.

Ví dụ 2.17

Xác suất để một bit được truyền qua một kênh thông tin bị lỗi là $0,1$. Giả sử các bit được truyền độc lập và gọi X là biến ngẫu nhiên chỉ số bit được truyền qua kênh thông tin cho đến khi bit đầu tiên bị lỗi xuất hiện. Tìm hàm khối xác suất của X.

Giải: Hàm khối xác suất được tìm tương tự như trong Ví dụ 2.7. Khi truyền bit thứ i qua kênh thông tin, gọi T_i là biến cố "bit đó được truyền thành công" và L_i là biến cố "bit đó bị truyền lỗi".

Ta có các biến cố $T_i, i \geqslant 1$ độc lập và có xác suất bằng nhau và bằng $p = \mathbb{P}(T_i) = 0,9$.
Biến cố $(X = k)$ là biến cố "$k-1$ bit đầu tiên được truyền thành công và bit thứ k bị truyền lỗi". Vậy

$$(X = k) = T_1 T_2 ... T_{k-1} L_k, \ k = 1, 2, 3, ...$$

Do các bit được truyền độc lập nên ta có hàm khối xác suất của biến X là

$$f(k) = \mathbb{P}(X = k) = \mathbb{P}(T_1)\mathbb{P}(T_2)...\mathbb{P}(T_{k-1})P(L_k) = 0,1.0,9^{k-1}, \ k = 1, 2, 3,$$

Tập các giá trị của biến ngẫu nhiên X có dạng vô hạn đếm được.

Khái niệm phân phối hình học

Trong dãy phép thử độc lập Bernoulli, các phép thử có cùng xác suất thành công là hằng số p. Gọi X là biến ngẫu nhiên chỉ số phép thử được thực hiện cho đến phép thử đầu tiên thành công. Khi đó biến X tuân theo phân phối hình học với tham số $0 < p < 1$ và hàm khối xác suất của X là

$$f(x) = p(1-p)^{x-1},\ x = 1, 2, \ldots$$

Ký hiệu biến ngẫu nhiên X có phân phối hình học với tham số p là $X \sim \mathcal{G}(p)$.

Giá trị trung bình của biến ngẫu nhiên X tuân theo luật phân phối hình học với tham số p là

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = p \sum_{k=1}^{\infty} kq^{k-1}$$

với $q = 1 - p$. Ta sử dụng công thức tổng của chuỗi cấp số nhân theo biến q

$$\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}. \tag{2.17}$$

Tương ứng ta có tính toán sau

$$\mathbb{E}[X] = p \sum_{k=1}^{\infty} kq^{k-1} = p \frac{\partial}{\partial q} \left(\sum_{k=0}^{\infty} q^k \right) = p \frac{\partial}{\partial q} \left(\frac{1}{1-q} \right)$$
$$= \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}.$$

Để tính phương sai của biến X trước tiên chúng ta cần tính $\mathbb{E}[X(X-1)]$. Cụ thể là

$$\mathbb{E}[X(X-1)] = \sum_{k=1}^{\infty} k(k-1)pq^{k-1} = pq\sum_{k=2}^{\infty} k(k-1)q^{k-2}$$

$$= pq\frac{\partial^2}{\partial q^2}\Big(\sum_{k=0}^{\infty} q^k\Big) = pq\frac{\partial^2}{\partial q^2}\Big(\frac{1}{1-q}\Big)$$

$$= pq\frac{2}{(1-q)^3} = \frac{2q}{p^2} = \frac{2(1-p)}{p^2}.$$

Tương ứng ta nhận được

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \frac{2(1-p)}{p^2} + \frac{1}{p} = \frac{2-p}{p^2}$$

và

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1-p}{p^2}.$$

Như vậy, ta có kết luận sau đây.

Công thức kỳ vọng phương sai

Nếu X là biến ngẫu nhiên có luật phân phối hình học với tham số p thì

$$\mathbb{E}[X] = \frac{1}{p} \text{ và } \mathbb{V}[X] = \frac{1-p}{p^2}. \tag{2.18}$$

Tính chất không nhớ. Ta gọi hàm đối của hàm phân phối xác suất $F(x)$ là hàm $\overline{F}(x)$ được xác định như sau:

$$\overline{F}(x) = 1 - F(x) = 1 - P(X \leqslant x) = P(X > x).$$

Hàm đối xác suất còn gọi là hàm sống sót (survival function). Đối với biến ngẫu nhiên X có phân phối hình học, ta có

$$\begin{aligned}\overline{F}(x) &= 1 - F(x) = 1 - P(X \leqslant x)\\ &= 1 - \sum_{k=1}^{x} p(1-p)^{k-1} = 1 - p\sum_{k=1}^{x} q^{k-1}\\ &= 1 - p\frac{1-q^x}{1-q} = 1 - (1-q^x) = q^x = (1-p)^x.\end{aligned}$$

Khái niệm

Biến ngẫu nhiên rời rạc X gọi là không nhớ nếu

$$P(X > s + t | X > s) = P(X > t), \text{ với } s, t \text{ là các số thực.} \quad (2.19)$$

Biến ngẫu nhiên X có phân phối hình học là biến ngẫu nhiên không nhớ. Thật vậy, ta có

$$\begin{aligned} P(X > s + t | X > s) &= \frac{P(X > s + t, X > s)}{P(X > s)} = \frac{P(X > s + t)}{P(X > s)} \\ &= \frac{\overline{F}(s + t)}{\overline{F}(s)} = \frac{(1 - p)^{s+t}}{(1 - p)^s} \\ &= (1 - p)^t = \overline{F}(t) = P(X > t). \end{aligned}$$

Biến ngẫu nhiên có phân phối hình học được định nghĩa là số lần thực hiện phép thử đến khi thành công thì dừng. Do các phép thử là độc lập với nhau, nên việc đếm số lần phép thử cho đến khi thành công sẽ không phụ thuộc vào điểm bắt đầu. Nói cách khác, ta có thể đếm tại bất kỳ thời điểm nào của dãy phép thử cũng không gây ảnh hưởng tới luật phân phối của biến ngẫu nhiên mô tả số lần đếm. Đây là cách nhìn nhận trực giác về tính chất không nhớ của biến hình học.

Ví dụ 2.18

Một công ty bán hàng qua mạng, công ty đó có 20 mặt hàng trong đó có máy tính Lenovo E430 và hàng ngày hệ thống mạng sẽ lựa chọn ngẫu nhiên một mặt hàng để giảm giá. Ngày hôm nay bạn bắt đầu theo dõi trang mạng bán hàng đó và bạn muốn mua một máy tính Lenovo E430 của công ty đó.

a) Bạn phải chờ đợi trung bình bao nhiêu ngày nữa mới đến ngày giảm giá sản phẩm máy tính Lenovo E430 đó?

b) Tính xác suất để đúng 20 ngày sau sẽ có đợt giảm giá máy tính Lenovo E430.

c) Bạn đã đợi 5 ngày mà chưa đến đợt giảm giá máy tính Lenovo E430. Hãy tính xác suất để bạn đợi thêm 15 ngày nữa sẽ đến đợt giảm giá loại máy tính đó.

d) Tính xác suất để máy tính đó được giảm giá trong vòng 3 ngày tới.

Giải: Gọi X là biến ngẫu nhiên chỉ số ngày phải chờ đến ngày giảm giá của máy tính Lenovo E430. Khi đó X có phân phối hình học. Do phân phối hình học là phân phối *không nhớ* tức là ta có thể chọn một thời điểm bất kỳ để bắt đầu quá trình đếm số ngày cho đến ngày đầu tiên sản phẩm máy tính Lenovo E430 được giảm giá mà không làm thay đổi phân phối xác suất của biến ngẫu nhiên X. Như vậy ta chọn ngày hôm nay là ngày bắt đầu đếm và gọi X là biến ngẫu nhiên chỉ số ngày kể từ ngày hôm nay cho đến ngày đầu tiên giảm giá máy tính Lenovo E430. Khi đó X tuân theo luật phân phối hình học với tham số $p = \dfrac{1}{20} = 0,05$. Hàm khối xác suất của X là

$$f(x) = 0,05.0,95^{x-1},\ x = 1,2,3,...$$

a) Bạn phải chờ đợi trung bình $\mathbb{E}[X] = \dfrac{1}{p} = 20$ ngày thì mới đến ngày đầu tiên giảm giá máy tính Lenovo E430.

b) $\mathbb{P}(X = 20) = f(20) = 0,05.0,95^{19} = 0,0189.$
c) Vì phân phối hình học là phân phối không nhớ nên xác suất của biến cố "bạn đã chờ được 5 ngày và phải chờ thêm 15 ngày nữa để đến đợt giảm giá" bằng xác suất của biến cố "đợi 15 ngày để đến đợt giảm giá". Vì vậy xác suất cần tính là

$$\mathbb{P}(X = 15) = f(15) = 0,05.0,95^{14} = 0,0244.$$

d) Xác suất cần tính là

$$\begin{aligned}\mathbb{P}(X \leqslant 3) &= P(X = 3) + P(X = 2) + P(X = 1)\\ &= f(3) + f(2) + f(1)\\ &= 0,05(0,95^2 + 0,95^1 + 0,95^0)\\ &= 0,1426.\end{aligned}$$

Trong phần này chúng ta sẽ xét đến phân phối xác suất tổng quát hơn phân phối hình học. Đó là phân phối của biến ngẫu nhiên chỉ số phép thử Bernoulli cần thực hiện cho đến khi có r phép thử có kết quả thành công. Ta sẽ làm quen với luật phân phối của biến này qua một ví dụ cụ thể dưới đây.

Ví dụ 2.19

Xác suất để một bit được truyền qua một kênh thông tin bị lỗi là $0,1$. Giả sử các bit được truyền độc lập. Gọi X là biến ngẫu nhiên chỉ số bit được truyền cho đến khi có 4 bit bị truyền lỗi. Người ta gọi X là biến ngẫu nhiên tuân theo luật phân phối nhị thức âm với tham số $r = 4$. Hãy tính $\mathbb{P}(X = 12)$.

Giải: Gọi A_k là biến cố "bit thứ 12 bị truyền lỗi", và gọi B_k là biến cố "có 3 bit bị truyền lỗi trong k bit đầu tiên". Ta thấy rằng $(X = 12) = B_{11}A_{12}$. Vì các phép thử là độc lập nên theo phân phối nhị thức xác suất để trong 11 phép thử đầu tiên có đúng 3 bit bị lỗi là

$$\mathbb{P}(B_{11}) = C_{11}^3.0,1^3.0,9^8$$

và xác suất để phép thử thứ 12 bị lỗi là $\mathbb{P}(A_{12}) = 0,1$. Vậy ta có

$$\mathbb{P}(X = 12) = C_{11}^3.0,1^3.0,9^8.0,1 = C_{11}^3.0,1^4.0,9^8.$$

Khái niệm phân phối nhị thức âm

Trong dãy phép thử Bernoulli độc lập với xác suất thành công của mỗi phép thử là hằng số p $(0 < p < 1)$, gọi X là biến ngẫu nhiên chỉ số phép thử được thực hiện cho đến khi có r $(r = 1, 2, 3\cdots)$ phép thử thành công. Khi đó ta nói X là biến ngẫu nhiên có phân phối nhị thức âm với hai tham số p và r với hàm khối xác suất của biến X là

$$f(x) = C_{x-1}^{r-1}(1-p)^{x-r}p^r, \quad x = r, r+1, r+2, ... \tag{2.20}$$

Ta ký hiệu biến ngẫu nhiên X có phân phối nhị thức âm với các tham số r và p là $X \sim \mathscr{N}\mathscr{B}(r,p)$.

Ta thấy tập giá trị của biến X là $\mathcal{S} = \{r, r+1, r+2, ...\}$. Trường hợp $r = 1$ thì X chính là biến hình học với tham số p.

Ta có

$$\sum_{x=r}^{+\infty} C_{x-1}^{r-1}(1-p)^{x-r}p^r = 1. \tag{2.21}$$

Sử dụng công thức trên ta tính kỳ vọng và phương sai của phân phối nhị thức âm.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=r}^{+\infty} xC_{x-1}^{r-1}(1-p)^{x-r}p^r \\
&= \sum_{x=r}^{+\infty} rC_x^r(1-p)^{x-r}p^r \\
&= \frac{r}{p}\sum_{x=r}^{+\infty} C_{(x+1)-1}^{(r+1)-1}(1-p)^{(x+1)-(r+1)}p^{r+1} \\
&= \frac{r}{p}.
\end{aligned}$$

Để tính phương sai, trước hết ta tính $\mathbb{E}[X(X+1)]$. Ta có

$$\begin{aligned}\mathbb{E}[X(X+1)] &= \sum_{x=r}^{+\infty} x(x+1)C_{x-1}^{r-1}(1-p)^{x-r}p^r \\ &= \frac{r(r+1)}{p^2}\sum_{x=r}^{+\infty} C_{x+1}^{r+1}(1-p)^{x-r}p^{r+2} \\ &= \frac{r(r+1)}{p^2}.\end{aligned}$$

Vậy ta có phương sai của phân phối nhị thức âm là

$$\mathbb{V}[X] = \mathbb{E}[X(X+1)] - \mathbb{E}[X] - (\mathbb{E}[X])^2 = \frac{r(r+1)}{p^2} - \frac{r}{p} - \frac{r^2}{p^2} = \frac{r(1-p)}{p^2}.$$

Công thức kỳ vọng phương sai

Giả sử X là biến ngẫu nhiên có phân phối nhị thức âm với hai tham số p và r với $0 < p < 1$ và $r = 1, 2, \dots$ ta có

$$\mathbb{E}[X] = \frac{r}{p} \text{ và } \mathbb{V}[X] = \frac{r(1-p)}{p^2}. \tag{2.22}$$

Biến ngẫu nhiên có phân phối nhị thức đếm số lần thực hiện thành công trong n lần thực hiện phép thử, nghĩa là số phép thử đã được xác định trước và số lần thành công là ngẫu nhiên. Còn đối với biến ngẫu nhiên có phân phối nhị thức âm số lần thành công r được xác định trước và số phép thử là ngẫu nhiên. Do đó biến ngẫu nhiên nhị thức âm là trường hợp ngược lại của biến ngẫu nhiên nhị thức nên có tên gọi là nhị thức âm.

Trong phần này chúng ta xét đến những phân phối nhị thức với dãy phép thử Bernoulli độc lập sao cho khi số phép thử n tăng lên vô hạn mà giá trị trung bình np vẫn không đổi. Khi đó ta nghiên cứu đến phân phối giới hạn của phân phối nhị thức này.

Ví dụ 2.20

Xét bài toán truyền n bit qua một kênh thông tin, giả thiết các bit được truyền độc lập và xác suất 1 bit bị truyền lỗi là hằng số p.

Gọi X là biến ngẫu nhiên chỉ số bit bị truyền lỗi trong số n bit cần truyền thì X là biến nhị thức với hai tham số n và p. Theo công thức Bernoulli, ta có

$$\mathbb{P}(X = x) = C_n^x p^x (1-p)^{n-x} = C_n^x \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}, \qquad (2.23)$$

với $\lambda = \mathbb{E}[X] = np$.

Giả sử rằng khi số bit truyền n tăng lên thì xác suất p giảm đi sao cho $\mathbb{E}[X] = \lambda = np$ là hằng số không đổi thì

$$\lim_{n\to\infty} C_n^x \frac{1}{n^x} = \frac{1}{x!}, \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-x} = 1, \quad \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} = e^{-\lambda}.$$

Do đó

$$\lim_{n\to\infty} \mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0, 1, 2, \ldots \tag{2.24}$$

Việc tính xác suất giới hạn theo công thức (2.24) rõ ràng thuận lợi hơn rất nhiều so với việc tính xác suất theo phân phối nhị thức (2.23).

Ví dụ 2.21

Giả sử X là biến ngẫu nhiên đếm số vết nứt trên một đoạn bất kỳ có độ dài là L (mm) của sợi dây đồng và số trung bình các vết nứt trong một đoạn L (mm) bằng λ. Việc tìm phân phối của biến X được tiến hành tương tự Ví dụ 2.20.

Bây giờ ta chia nhỏ đoạn dây L (mm) thành n đoạn con, mỗi đoạn có độ dài 1 đơn vị độ dài, chẳng hạn 1 (μm). Sợi dây được chia nhỏ sao cho ta có thể giả thiết

- xác suất để trong 1 đoạn con có nhiều hơn 1 vết nứt là không đáng kể,
- hơn nữa ta có thể giả thiết xác suất để trong mỗi đoạn con có 1 vết nứt là bằng nhau và bằng p,
- giả thiết việc các vết nứt xuất hiện trong mỗi đoạn con là độc lập với nhau.

Khi đó ta có thể xấp xỉ phân phối của biến X thành phân phối nhị thức. Mặt khác $\mathbb{E}[X] = \lambda = np$ hay $p = \lambda/n$, nên ta có xác suất để trong một đoạn con có chứa 1 vết nứt là $p = \lambda/n$. Vì vậy ta có thể tìm phân phối của biến X theo công thức (2.23) hoặc (2.24) trong Ví dụ 2.20.

☞ Ta có thể lập luận tương tự với các với các mô hình liên quan đến một đơn vị thời gian, một đơn vị diện tích hoặc một đơn vị thể tích. Vì vậy phương pháp trên được áp dụng rộng rãi để mô hình hóa cho nhiều ứng dụng trong kỹ thuật, chẳng hạn như

- đếm số hạt bị ô nhiễm trong sản suất chất bán dẫn,
- đếm số vết xước trong một cuộn vải sản suất công nghiệp,
- đếm số cuộc gọi đến một tổng đài,
- đếm số hạt nguyên tử giải phóng ra từ một mẫu,
- đếm số vụ tai nạn giao thông trên một đoạn đường cao tốc v.v..

Một cách tổng quát ta xét một khoảng số thực T và chia khoảng này thành các khoảng nhỏ có độ dài Δt đủ nhỏ sao cho:

- xác suất có nhiều hơn 1 sự kiện xuất hiện trong 1 đoạn nhỏ Δt bằng không,
- xác suất có 1 sự kiện xuất hiện trong 1 đoạn nhỏ Δt bằng $\lambda \Delta t / T$,
- các sự kiện xuất hiện trong mỗi đoạn nhỏ là độc lập với nhau.

Những mô hình mà thỏa mãn các điều kiện này được gọi là những quá trình Poisson. Với những giả thiết như vậy ta có thể xét dãy $n = T/\Delta t$ phép thử độc lập Bernoulli với xác suất thành công trong một phép thử là $p = \lambda \Delta t / T$, ở đây $\lambda = np$ giả thiết là hằng số không đổi. Ta có định nghĩa sau

Định nghĩa

Biến ngẫu nhiên X đếm số sự kiện xuất hiện trong một khoảng đơn vị (thời gian, độ dài, v.v.) được gọi là biến ngẫu nhiên tuân theo luật Poisson với tham số $\lambda > 0$ và có hàm khối xác suất là

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots \tag{2.25}$$

Ta ký hiệu biến ngẫu nhiên X có phân phối Poisson với tham số λ là $X \sim \mathscr{P}(\lambda)$.

Ta nhớ rằng, hàm e^λ có thể khai triển Taylor thành chuỗi theo công thức

$$e^\lambda = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}. \tag{2.26}$$

Do đó, $\sum\limits_{x=0}^{\infty} f(x) = e^{-\lambda} \sum\limits_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$

Ví dụ 2.22

Giả sử số vết nứt dọc theo một sợi dây đồng là biến Poisson với trung bình 2,3 vết trên 1mm.
a) Tính xác suất để trong 1mm sợi dây có 3 vết nứt.
b) Tính xác suất để trong 5mm sợi dây có 15 vết nứt.
c) Tính xác suất để có ít nhất 1 vết nứt trong 2mm của sợi dây.

Giải: a) Gọi X là biến ngẫu nhiên chỉ số vết nứt trên 1mm của sợi dây đồng. Biến ngẫu nhiên X có phân phối Poisson với tham số $\lambda = 2,3$. Ta có xác suất để trong 1 mm sợi dây có 3 vết nứt là

$$\mathbb{P}(X = 3) = \frac{e^{-2,3}(2,3)^3}{3!} = 0,2033.$$

b) Gọi Y là biến ngẫu nhiên chỉ số vết nứt trong mỗi 5mm của sợi dây, khi đó Y cũng là biến Poisson với tham số $\lambda = 5.2,3 = 11,5$.

Bởi vậy xác suất để trong 5mm sợi dây có 15 vết nứt là

$$P(Y = 15) = \frac{e^{-11,5} 11,5^{15}}{15!} = 0,0630.$$

c) Gọi Z là biến ngẫu nhiên chỉ số vết nứt trong mỗi 2mm của sợi dây thì Z là biến Poisson với tham số $\lambda = 2.2,3 = 4,6$. Xác suất để có ít nhất 1 vết nứt trong 2mm của sợi dây đồng là

$$\mathbb{P}(Z \geqslant 1) = 1 - \mathbb{P}(Z = 0) = 1 - \frac{e^{-4,6} 4,6^{0}}{0!} = 0,9899.$$

Trong Ví dụ này khi ta cần tính xác suất liên quan đến 2mm của sợi dây đồng thì tham số của biến Poisson được điều chỉnh thành $\lambda = 4,6$; tức là trung bình trong 2mm của sợi dây có 4,6 vết nứt. Với sự điều chỉnh về tham số λ như vậy thì ta có thể tính xác suất trong những đoạn có độ dài bất kỳ của sợi dây.

Bây giờ áp dụng khai triển (2.26) ta chứng minh được

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} \frac{ke^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{ke^{-\lambda}\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda.$$

Để tính phương sai của X ta tính, $\mathbb{E}[X(X-1)]$. Ta có

$$\mathbb{E}[X(X-1)] = \sum_{k=0}^{\infty} \frac{k(k-1)e^{-\lambda}\lambda^k}{k!} = \sum_{k=2}^{\infty} \frac{k(k-1)e^{-\lambda}\lambda^k}{k!}$$

$$= \lambda^2 \sum_{k=2}^{\infty} \frac{e^{-\lambda}\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda} e^{\lambda} = \lambda^2.$$

Ta tính được

$$\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda.$$

Từ đó ta có

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Vậy ta có kết luận:

Công thức kỳ vọng phương sai

Giả sử X là biến ngẫu nhiên Poisson với tham số $\lambda > 0$ thì

$$\mathbb{E}[X] = \lambda \text{ và } \mathbb{V}[X] = \lambda. \tag{2.27}$$

Ta thấy rằng giá trị trung bình và phương sai của biến Poisson là bằng nhau. Chẳng hạn trong Ví dụ 2.22 có trung bình 2,3 vết nứt trên 1mm sợi dây và có phương sai cũng là $\mathbb{V}[X] = 2,3$. Tương ứng, độ lệch chuẩn là $\sigma = \sqrt{2,3} = 1,5166$ vết nứt/1mm.

☞ Như vậy nếu một mẫu dữ liệu có phương sai mẫu lớn hơn giá trị trung bình thì ta không nên dùng biến Poisson để mô hình hóa cho nguồn cung cấp mẫu dữ liệu đó.