BÀI 2: BIẾN NGẪU NHIÊN VÀ QUY LUẬT PHÂN BỐ XÁC XUẤT

Mục tiêu

Thông qua các công cụ giải tích, bài này giới thiệu với học viên khái niệm về biến ngẫu nhiên, phân loại các biến ngẫu nhiên, các quy luật phân phối xác suất của biến ngẫu nhiên, các tham số đặc trưng của biến ngẫu nhiên và ý nghĩa của chúng. Hai nội dung quan trọng nhất của chương là quy luật phân phối xác suất và các tham số đặc trưng của một biến ngẫu nhiên.

Thời lượng

- 8 tiết

Các kiến thức cần có

- Định nghĩa và phân loại biến ngẫu nhiên;
- Định nghĩa biến ngẫu nhiên;
- Phân loại biến ngẫu nhiên;
- Quy luật phân phối xác suất của biến ngẫu nhiên;
- Bảng phân phối xác suất;
- Hàm phân phối xác suất;
- Hàm mật độ xác suất;
- Các tham số đặc trưng của biến ngẫu nhiên;
- Kỳ vọng (giá trị trung bình);
- Trung vị;
- Mốt (Mode);
- Phương sai và độ lệch chuẩn;
- Giá trị tới hạn (critical value);
- Mômen trung tâm bậc cao;
- Biến ngẫu nhiên nhiều chiều;
- Biễn nhẫu nhiên k chiều;
- Bảng phân phối xác suất của biễn ngẫu nhiên hai chiều;
- Bảng phân phối xác suất có điều kiện của hai biến ngẫu nhiên;
- Bảng phân phối xác suất có điều kiện của hai biến ngẫu nhiên.

TÌNH HUỐNG KHỞI ĐỘNG BÀI

Tình huống

Một công ty bảo hiểm bán thẻ bảo hiểm với giá 100000đ/1 người/1 năm. Nếu người tham gia bảo hiểm gặp rủi ro trong năm đó thì nhận được số tiền bồi thường là 1 triệu đồng. Theo thống kê biết rằng tỷ lệ người tham gia bảo hiểm bị rủi ro trong năm là 005, hãy tính tiền lãi trung bình khi bán mỗi thẻ bảo hiểm. Nếu bán bảo hiểm được cho 10000 khách hàng thì số tiền lãi trung bình thu về được là bao nhiêu?

Câu hỏi

1. Biểu diễn bảng phân phối xác suất giữa tiền lãi bảo hiểm và khả năng nhận được lãi?
2. Số tiền lãi trung bình là bao nhiêu?
3. Nếu bán bảo hiểm được cho 10000 khách hàng thì số tiền lãi trung bình thu về được là bao nhiêu?

2.1. Định nghĩa và phân loại biến ngẫu nhiên

2.1.1. Định nghĩa biến ngẫu nhiên

Trong thực tế người ta thường gặp rất nhiều đại lượng nhận các giá trị một cách ngẫu nhiên. Ta hãy bắt đầu làm quen với khái niệm biến ngẫu nhiên qua các ví dụ.

Ví dụ 1.1:

Gọi X là số chấm xuất hiện khi gieo một con xúc sắc thì X có thể nhận một trong các giá trị 1, 2, 3, 4, 5 và 6.

Ví dụ 1.2:

Bắn 3 viên đạn một cách độc lập vào mục tiêu, xác suất trúng bia của mỗi viên đạn đều bằng 0,8. Gọi Y là số viên đạn trúng bia. Lúc đó Y có thể nhận các giá trị 0, 1, 2 hoặc 3.

Ví dụ 1.3:

Một hộp có m sản phẩm tốt, n sản phẩm xấu. Lấy ngẫu nhiên từ hộp đó ra 2 sản phẩm. Nếu ký hiệu Z là số sản phẩm tốt lấy ra được thì Z có thể nhận các giá trị 0, 1 hoặc 2.

Ví dụ 1.4:

Bắn 1 viên đạn vào bia có bán kính là 20cm và giả sử viên đạn trúng vào bia. Gọi W là khoảng cách từ tâm bia tới điểm bia trúng đạn thì W có thể nhận các giá trị thuộc nữa đoạn [0; 20).

Các đại lượng X, Y, Z, W trong những ví dụ trên nhận mỗi giá trị có thể có của mình một cách ngẫu nhiên, tương ứng với một xác suất nào đó. Chúng được gọi là biến ngẫu nhiên hay đại lượng ngẫu nhiên.

Định nghĩa 1.1:

Biến ngẫu nhiên là đại lượng mà việc nó có thể nhận một giá trị cụ thể nào đó, hoặc một giá trị nằm trong một khoảng nào đó thuộc miền các khoảng giá trị có thể có của nó, là một biến cố ngẫu nhiên nếu như phép thử chưa được thực hiện.

CHÚ Ý

Sau khi phép thử được thực hiện, biến ngẫu nhiên sẽ chỉ nhận một và chỉ một giá trị trong các giá trị có thể có của biến ngẫu nhiên đó.

Ta thường ký hiệu biến ngẫu nhiên bởi các chữ in hoa: X, Y, Z, ... hoặc X_1, X_2, … , Y_1, Y_2, ... và các giá trị của chúng bởi các chữ thường $x_1, x_2, \ldots, y_1, y_2, \ldots z_1, z_2 \ldots$

Hình 2.1: Kết quả tung đồng xu chỉ có thể nhận được một trong hai giá trị: sấp và ngửa

CHÚ Ý

Để đơn giản, ta kí hiệu $(X = x)$ thay cho biến cố "biến ngẫu nhiên X nhận giá trị bằng x" và viết $(X < x)$ thay cho biến cố "biến ngẫu nhiên X nhận giá trị nhỏ hơn x".

Nếu biến ngẫu nhiên X chỉ nhận các giá trị: $x_1, x_2, ... x_n$ thì các biến cố $(X = x_1)$, $(X = x_2), ..., (X = x_n)$ tạo nên một hệ đầy đủ biến cố trong phép thử.

2.1.2. Phân loại biến ngẫu nhiên

Người ta thường chia các biến cố ngẫu nhiên làm hai loại: Biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục.

- Biến ngẫu nhiên được gọi là rời rạc khi các giá trị có thể có của nó xếp thành dãy hữu hạn hoặc vô hạn đếm được $x_1, x_2, ..., x_j, ..., x_k$. Nói cách khác, ta có thể liệt kê tất cả các giá trị của biến ngẫu nhiên đó. Các biến ngẫu nhiên X, Y, Z tương ứng trong các ví dụ 1.1, 1.2, 1.3 là các biến ngẫu nhiên rời rạc.

- Biến ngẫu nhiên được gọi là *liên tục* trong một khoảng giá trị nếu như các giá trị có thể có của nó lấp đầy khoảng giá trị đó. Biến ngẫu nhiên W trong Ví dụ 1.4 là một biến ngẫu nhiên liên tục.

Hình 2.2: Số lượng cá câu được là một biến ngẫu nhiên rời rạc

2.2. Quy luật phân phối xác suất của biến ngẫu nhiên

Như đã trình bày ở trên, biến ngẫu nhiên nhận mỗi giá trị của nó tương ứng với một biến cố ngẫu nhiên nào đó và do vậy tương ứng với một xác suất của biến cố đó. Quy luật phân phối xác suất của biến ngẫu nhiên là cách biểu diễn mối quan hệ giữa giá trị có thể có của biến ngẫu nhiên và các xác suất tương ứng để biến ngẫu nhiên nhận các giá trị đó.

Các phương pháp được sử dụng phổ biến để mô tả quy luật phân phối xác suất của biến ngẫu nhiên bao gồm:

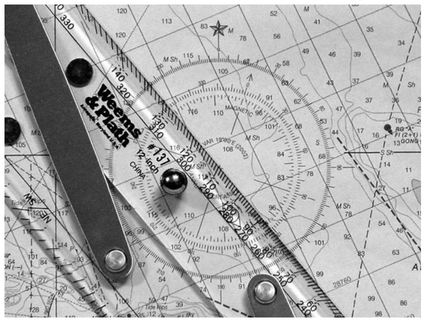
- Bảng phân phối xác suất (áp dụng cho biến ngẫu nhiên rời rạc)
- Hàm phân phối xác suất (áp dụng cho cả hai loại biến ngẫu nhiên rời rạc và liên liên tục)
- Hàm mật độ xác suất (áp dụng cho biến ngẫu nhiên liên tục)

Hình 2.3: Chiều cao của người là một biến ngẫu nhiên liên tục

2.2.1. Bảng phân phối xác suất

Giả sử biến ngẫu nhiên rời rạc X nhận các giá trị $x_1, x_2, \ldots x_n$ với các xác suất tương ứng $p_i = P(X = x_i), i = 1 \div n$. Khi đó bảng phân phối xác suất của biến ngẫu nhiên X được trình bày như sau:

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



Hình 2.4: Quy luật phân phối của biến ngẫu nhiên

Trong đó:

$$\begin{cases} 0 \le p_i \le 1 \\ \sum_{i=1}^{n} p_i = 1 \end{cases}$$ (khi X nhận vô hạn đếm được các giá trị thì $\sum_{i=1}^{\infty} p_i = 1$).

CHÚ Ý

Nếu biến ngẫu nhiên X có bảng phân phối xác suất như trên thì

$$p(a < X < b) = \sum_{a < x_i < b} P(X = x_i) = \sum_{a < x_i < b} p_i \quad (2.1)$$

Ví dụ 2.1:

Tung một con xúc xắc cân đối và đồng chất. Gọi X là "số chấm của mặt trên cùng". Khi ấy X là một biến ngẫu nhiên, ta có bảng phân phối xác suất của X như sau:

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

Ví dụ 2.2:

Với biến ngẫu nhiên Y trong Ví dụ 1.2, ta có:

$$P(Y = 0) = C_3^0 \times 0,8^0 \times 0,2^3 = 0,008$$

$$P(Y=1)=C_3^1\times 0,8^1\times 0,2^2=0,096$$
$$P(Y=2)=C_3^2\times 0,8^2\times 0,2^1=0,384$$
$$P(Y=3)=C_3^3\times 0,8^3\times 0,2^0=0,512.$$

Từ đó suy ra bảng phân phối xác suất của biến ngẫu nhiên Y có dạng:

Y	0	1	2	3
P	0,008	0,096	0,384	0,512

Ví dụ 2.3:

Với biến ngẫu nhiên Z trong ví dụ 1.3, ta có bảng phân phối xác suất của biến ngẫu nhiên Z như sau:

Z	0	1	2
P	$\frac{C_m^0\times C_n^2}{C_{m+n}^2}$	$\frac{C_m^1\times C_n^1}{C_{m+n}^2}$	$\frac{C_m^2\times C_n^0}{C_{m+n}^2}$

Ví dụ 2.4:

Một người phải tiến hành thí nghiệm cho tới khi thành công thì dừng. Lập bảng phân phối xác suất của số lần tiến hành thí nghiệm. Biết rằng các lần tiến hành thí nghiệm là độc lập với nhau và xác suất thành công mỗi lần là $p\ (0<p<1)$.

Giải:

Gọi X là số lần phải tiến hành thí nghiệm. Các giá trị có thể có của X là 0, 1, 2, …, ,n …

Gọi A_i là biến cố ở lần thí nghiệm thứ i thì thành công $(i=1,2,\ldots)$. Ta có:

$$P(X=1)=P(A_1)=p.$$

Biến cố (X = 2) tương đương với biến cố $\bar{A}_1A_2$. Từ đó ta có:

$$P(X=2)=P(\bar{A}_1A_2)=(1-p)\times p.$$

Tương tự, ta có bảng phân phối xác suất của biến ngẫu nhiên X như sau:

X	1	2	3	…	n	…
P	p	$(1-p)\times p$	$(1-p)^2\times p$	…	$(1-p)^{n-1}\times p$	…

Ví dụ 2.5:

Một người được phát 3 viên đạn và lần lượt bắn một tấm bia đến khi nào trúng thì dừng. Lập bảng phân phối xác suất số viên đạn phải bắn, biết rằng các lần bắn độc lập với nhau và xác suất trúng đích của mỗi lần bắn là 0,7.

Giải:

Ký hiệu X là số viên đạn phải bắn, các giá trị mà X có thể nhận là 1, 2 và 3. Gọi A_i là biến cố viên đạn thứ i trúng bia $(i=1,2,3)$. Ta có $P(X=1)=P(A_1)=0,7$. Mặt khác ta thấy biến cố (X = 2) tương đương với biến cố $\bar{A}_1A_2$.

Do vậy:

$$P(X=2)=P(\overline{A}_1A_2)=0,3\times 0,7=0,21$$

Đồng thời, biến cố (X = 3) tương đương với biến cố $\overline{A}_1\overline{A}_2A_3+\overline{A}_1\overline{A}_2\overline{A}_3$.

Từ đó:

$$P(X=3)=P(\overline{A}_1\overline{A}_2A_3)+P(\overline{A}_1\overline{A}_2\overline{A}_3)=0,3\times 0,3\times 0,7+0,3\times 0,3\times 0,3=0,09$$

Tổng hợp các kết quả trên, ta lập được bảng phân phối xác suất của X như sau:

X	1	2	3
P	0,7	0,21	0,09

Ví dụ 2.6:

Một người bắn một viên đạn vào bia với xác suất trúng bia là 0,7. Thử lập bảng phân phối xác suất của khoảng cách từ điểm bia trúng đạn tới tâm bia, biết bia có bán kính là 20cm.

Chúng ta dễ dàng thấy việc lập bảng phân phối xác xuất với một biến ngẫu nhiên liên tục như trong ví dụ này không thể thực hiện được. Vì vậy cần sử dụng công cụ thứ hai mô tả quy luật phân phối xác suất của các biến ngẫu nhiên, đó là hàm phân phối xác suất.

2.2.2. Hàm phân phối xác suất

2.2.2.1. Định nghĩa hàm phân phối xác suất

Hình 2.5: Hàm phân bố xác suất

Cho biến ngẫu nhiên X. Với mỗi số thực x, xác định duy nhất một biến cố (X < x) và do đó có tương ứng một và chỉ một xác suất $P(X<x)$. Quan hệ tương ứng này cho ta một hàm số xác định trên $\mathbb{R}$, hàm số này được ký hiệu là F(x).

Định nghĩa 2.1:

Hàm số $F(x)=P(X<x)$, $x\in\mathbb{R}$, được gọi là hàm phân phối (hàm phân bố) xác suất của biến ngẫu nhiên X.

Nếu X là một biến ngẫu nhiên rời rạc có bảng phân phối xác suất ở mục 2.1 thì hàm phân phối xác suất của X xác định như sau:

$$F(x)=\sum_{x_i<x}p_i\,,\qquad x\in\mathbb{R} \tag{2.2}$$

Ví dụ 2.7:

Cho biết ngẫu nhiên X có bảng phân phối xác suất

X	0	1	2
P	$\frac{1}{5}$	$\frac{3}{10}$	$\frac{1}{2}$

Tìm hàm phân phối xác suất của X .

Giải:

Ta có

$$F(x)=\begin{cases} 0 & ; x \le 0 \\ 1/5 & ; 0 < x \le 1 \\ 1/2 & ; 1 < x \le 2 \\ 1 & ; x > 2 \end{cases}$$

2.2.2.2. Tính chất của hàm phân phối xác suất

Hình 2.6: Tính chất của hàm Hàm phân phối xắc xuất

Từ định nghĩa, ta có thể chứng minh được hàm phân phối xác suất của một biến ngẫu nhiên có một số tính cơ bản sau:

Tính chất 1:

$$0 \le F(x) \le 1, \ \forall x.$$

Tính chất 2:

Nếu a là giá trị nhỏ nhất có thể có của X và b là giá trị lớn nhất có thể có của X thì:

$$F(x)=0 \qquad \text{với mọi } x \le a$$

$$F(x)=1 \qquad \text{với mọi } x > b$$

Chứng minh:

Vì a là giá trị nhỏ nhất của X nên với $x \le a$ thì biến cố $X < a$ là biến cố không thể có. Do vậy $F(x) = P(X < x) = P(V) = 0$.

Tương tự, vì b là giá trị lớn nhất có thể có của X nên với $x > b$ thì $(X < x) = U$. Từ đó $F(x) = P(X < x) = P(U) = 1$.

Tính chất 3:

Hàm phân phối xác suất của biến ngẫu nhiên là một hàm không giảm.

Thật vậy, giả sử $x_1, x_2 \in \mathbb{R}$ và $x_1 < x_2$.

Ta có:

$$F(x_1)=P(X<x_1), \qquad F(x_2)=P(X<x_2).$$

Vì biến cố $(X < x_2)$ có thể tách thành hai biến cố xung khắc $(X < x_1)$ và $(x_1 \le X < x_2)$ nên

$$P(X<x_2)=P(X<x_1)+P(x_1 \le X \le x_2),$$

Do đó $F(x_2)=F(x_1)+P(x_1 \le X < x_2) \ge F(x_1)$.

Vì vậy F(x) là hàm không giảm.

Hình 2.7: Hàm phân phối xắc suất của một biến ngẫu nhiên liên tục bên trái

Tính chất 4:

Hàm phân phối xác suất của một biến ngẫu nhiên liên tục bên trái.

Từ các tính chất trên, ta có các hệ quả sau:

Hệ quả 2.1:

$$F(-\infty)=\lim_{x\to-\infty}P(X<x)=0 \qquad (2.3)$$

$$F(+\infty)=\lim_{x\to+\infty}P(X<x)=1 \qquad 2.4)$$

Hệ quả 2.2:

$$P(a\le X<b)=F(b)-F(a) \qquad (2.5)$$

Hệ quả 2.3: Nếu X là một biến ngẫu nhiên liên tục thì $P(X=x)=0$ với mọi $x\in\mathbb{R}$.

Ý nghĩa của hệ quả này là trong quá trình nghiên cứu biến ngẫu nhiên liên tục ta không cần quan tâm đến xác suất để biến ngẫu nhiên đó nhận một giá trị cụ thể nào, mà cần quan tâm đến xác suất để nó nhận giá trị trong một khoảng giá trị nào đó.

Hệ quả 2.4:

Nếu X là biến ngẫu nhiên liên tục thì ta có:

$$P(x_1<X<x_2)=P(x_1<X\le x_2)=P(x_1\le X<x_2)=P(x_1\le X\le x_2)$$

Ý nghĩa của hệ quả này là với biến ngẫu nhiên liên tục ta không cần phân biệt xác xuất để nó nhận giá trị trong đoạn hay trong khoảng giá trị nào đó của nó.

CHÚ Ý

Nếu hàm F(x) có các tính chất 1, 2, 3 thì nó là hàm phân phối xác suất của một biến ngẫu nhiên nào đó.

Hàm F(x) cho biết tỷ lệ phần trăm giá trị của X nằm về bên trái của số thực x.

Ví dụ 2.8:

Biến ngẫu nhiên rời rạc X có phân phối xác suất

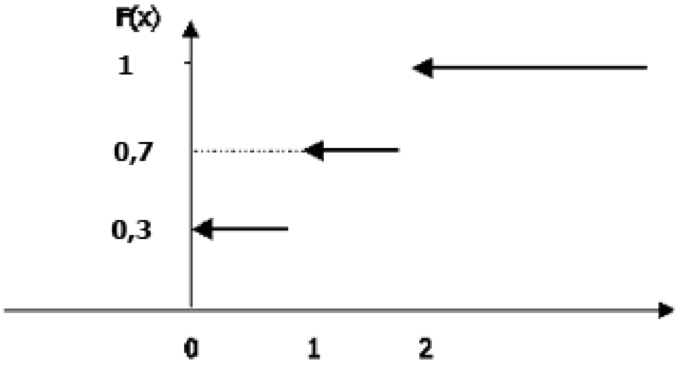
X	0	1	2
P	0,3	0,4	0,3

- Lập hàm phân phối xác suất của X .
- Tính $p(0<X\le 2)$ và $P(1<X<5)$

Giải:

- Ta có:

$$F(x)=\begin{cases}0 & x\le 0\\ 0,3 & 0<x\le 1\\ 0,7 & 1<x\le 2\\ 1 & 2<x\end{cases}$$



Hình 2.8: Đồ thị hàm phân phối F(x) của biến ngẫu nhiên rời rạc

- Để tính $P(0<X\le 2)$ ta có thể sử dụng hai cách:

Cách 1: Tính thông qua hàm phân phối:

$$\begin{aligned} P(0<X\le 2) &= P(X<2)+P(X=2)-(P(X<0)+P(X=0)) \\ &= F(2)+P(X=2)-F(0)-P(X=0) \\ &= 0,7+0,3-0-0,3=0,7 \end{aligned}$$

Cách 2: Tính trực tiếp:

$$\begin{aligned} P(0<X\le 2) &= P(X=1)+P(X=2) \\ &= 0,4+0,3=0,7 \end{aligned}$$

Tương tự ta tính được:

$$\begin{aligned} P(1<X<5) &= P(X<5)-P(X<1)-P(X=1) \\ &= F(5)-F(1)-P(X=1)=1-0,3-0,4=0,3 \end{aligned}$$

hoặc bằng cách khác: $P(1<X<5)=P(X=2)=0,3$

2.2.3. Hàm mật độ xác suất

Hình 2.9: Mật độ xe tại nút giao thông