§ 6 BIẾN NGẪU NHIÊN RỜI RẠC

1. Khái niệm biến ngẫu nhiên rời rạc

Ví dụ 1. Gieo đồng xu 5 lần liên tiếp. Kí hiệu X là số lần xuất hiện mặt ngửa. Đại lượng X có các đặc điểm sau :

– Giá trị của X là một số thuộc tập $\{0, 1, 2, 3, 4, 5\}$;

– Giá trị của X là ngẫu nhiên, không đoán trước được.

Ta nói X là một biến ngẫu nhiên rời rạc.

Một cách khái quát :

> Đại lượng X được gọi là một **biến ngẫu nhiên rời rạc** nếu nó nhận giá trị bằng số thuộc một tập hữu hạn nào đó và giá trị ấy là ngẫu nhiên, không dự đoán trước được.

2. Phân bố xác suất của biến ngẫu nhiên rời rạc

Giả sử X là một biến ngẫu nhiên rời rạc nhận các giá trị $\{x_1, x_2, ..., x_n\}$. Để hiểu rõ hơn về X, ta thường quan tâm đến xác suất để X nhận giá trị x_k tức là các số $P(X = x_k) = p_k$ với $k = 1, 2, ..., n$.

Các thông tin về X như vậy được trình bày dưới dạng bảng sau đây :

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

Bảng 1

Bảng 1 được gọi là **bảng phân bố xác suất của biến ngẫu nhiên rời rạc** X.

Người ta chứng minh được rằng trong bảng 1, tổng các số ở dòng thứ hai bằng $p_1 + p_2 + ... + p_n = 1$.

Ví dụ 2. Số vụ vi phạm luật giao thông trên đoạn đường A vào tối thứ bảy hàng tuần là một biến ngẫu nhiên rời rạc X. Giả sử X có bảng phân bố xác suất như sau :

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

Bảng 2

Nhờ bảng 2 ta biết được chẳng hạn xác suất để tối thứ bảy trên đoạn đường A không có vụ vi phạm luật giao thông nào là 0,1 và xác suất để xảy ra nhiều nhất một vụ vi phạm luật giao thông là 0,1 + 0,2 = 0,3.

H1 *Tính xác suất để tối thứ bảy trên đoạn đường A :*

a) Có hai vụ vi phạm luật giao thông ;

b) Có nhiều hơn ba vụ vi phạm luật giao thông.

Ví dụ 3. Một túi đựng 6 viên bi đỏ và 4 viên bi xanh. Chọn ngẫu nhiên 3 viên bi. Gọi X là số viên bi xanh trong 3 viên bi được chọn ra. Rõ ràng X là biến ngẫu nhiên rời rạc nhận giá trị trong tập $\{0, 1, 2, 3\}$.

Để lập bảng phân bố xác suất của X ta phải tính các xác suất $P(X=0)$, $P(X=1)$, $P(X=2)$ và $P(X=3)$.

Số trường hợp có thể là $C_{10}^3 = 120$.

Ta có $P(X=0)$ là xác suất chọn được cả 3 viên bi đỏ. Số cách chọn 3 viên bi đỏ là $C_6^3 = 20$. Vậy $P(X=0) = \frac{20}{120} = \frac{1}{6}$.

Ta có $P(X=1)$ là xác suất để chọn được 1 viên bi xanh và 2 viên bi đỏ. Ta có $C_4^1 = 4$ cách chọn 1 viên bi xanh và $C_6^2 = 15$ cách chọn 2 viên bi đỏ. Theo quy tắc nhân, ta có $4.15 = 60$ cách chọn 1 viên bi xanh và 2 viên bi đỏ. Vậy $P(X=1) = \frac{60}{120} = \frac{1}{2}$. □

H2 *Hãy tính* $P(X=2)$ *và* $P(X=3)$ *rồi lập bảng phân bố xác suất của X.*

3. Kì vọng

ĐỊNH NGHĨA

Cho X là biến ngẫu nhiên rời rạc với tập giá trị là $\{x_1, x_2, ..., x_n\}$. **Kì vọng** của X, kí hiệu là $E(X)$, là một số được tính theo công thức

$$E(X) = x_1p_1 + x_2p_2 + \cdots + x_np_n = \sum_{i=1}^{n} x_ip_i,$$

ở đó $p_i = P(X = x_i)$, $(i = 1, 2, ..., n)$.

Ý nghĩa : $E(X)$ là một số cho ta một ý niệm về độ lớn trung bình của X. Vì thế kì vọng $E(X)$ còn được gọi là giá trị trung bình của X.

Nhận xét. Kì vọng của X không nhất thiết thuộc tập các giá trị của X.

Ví dụ 4. Gọi X là số vụ vi phạm luật giao thông trong đêm thứ bảy ở đoạn đường A nói trong ví dụ 2. Tính $E(X)$.

Giải

Ta có $E(X) = 0 . 0{,}1 + 1 . 0{,}2 + 2 . 0{,}3 + 3 . 0{,}2 + 4 . 0{,}1 + 5 . 0{,}1 = 2{,}3$.

(Như vậy ở đoạn đường A mỗi tối thứ bảy có trung bình 2,3 vụ vi phạm luật giao thông). □

4. Phương sai và độ lệch chuẩn

a) Phương sai

ĐỊNH NGHĨA

Cho X là biến ngẫu nhiên rời rạc với tập giá trị là $\{x_1, x_2, ..., x_n\}$.

Phương sai của X, kí hiệu là $\mathrm{V}(X)$, là một số được tính theo công thức

$$\mathrm{V}(X) = (x_1 - \mu)^2 p_1 + (x_2 - \mu)^2 p_2 + \cdots + (x_n - \mu)^2 p_n$$
$$= \sum_{i=1}^{n} (x_i - \mu)^2 p_i,$$

ở đó $p_i = \mathrm{P}(X = x_i)$ $(i = 1, 2, ..., n)$ và $\mu = \mathrm{E}(X)$.

Ý nghĩa : Phương sai là một số không âm. Nó cho ta một ý niệm về *mức độ* phân tán các giá trị của X xung quanh giá trị trung bình. Phương sai càng lớn thì độ phân tán này càng lớn.

b) Độ lệch chuẩn

ĐỊNH NGHĨA

Căn bậc hai của phương sai, kí hiệu là $\sigma(X)$, được gọi là **độ lệch chuẩn** của X, nghĩa là

$$\sigma(X) = \sqrt{\mathrm{V}(X)}.$$

Ví dụ 5. Gọi X là số vụ vi phạm luật giao thông vào tối thứ bảy nói trong ví dụ 2. Tính phương sai và độ lệch chuẩn của X.

Giải

Từ ví dụ 4 ta có $\mu = \mathrm{E}(X) = 2{,}3$. Từ công thức tính phương sai, ta có

$$\mathrm{V}(X) = (0 - 2{,}3)^2 . 0{,}1 + (1 - 2{,}3)^2 . 0{,}2 + (2 - 2{,}3)^2 . 0{,}3 + (3 - 2{,}3)^2 . 0{,}2$$
$$+ (4 - 2{,}3)^2 . 0{,}1 + (5 - 2{,}3)^2 . 0{,}1 = 2{,}01.$$

Độ lệch chuẩn là $\sigma(X) = \sqrt{2{,}01} \approx 1{,}418$ □

CHÚ Ý

Có thể chứng minh được rằng

$$V(X) = \sum_{i=1}^{n} x_i^2 p_i - \mu^2. \qquad (1)$$

Trong thực hành, ta thường dùng công thức (1) để tính phương sai.

Ví dụ 6. Dùng công thức (1) để tính phương sai của số vụ vi phạm luật giao thông trong ví dụ 2, ta có

$$V(X) = 0^2.0,1 + 1^2.0,2 + 2^2.0,3 + 3^2.0,2 + 4^2.0,1 + 5^2.0,1 - (2,3)^2 = 2,01.$$

Câu hỏi và bài tập

43. Một cuộc điều tra được tiến hành như sau : Chọn ngẫu nhiên một bạn học sinh trên đường và hỏi xem gia đình bạn đó có bao nhiêu người. Gọi X là số người trong gia đình bạn học sinh đó. Hỏi X có phải là biến ngẫu nhiên rời rạc không ? Vì sao ?

44. Chọn ngẫu nhiên một gia đình trong số các gia đình có ba con. Gọi X là số con trai trong gia đình đó. Hãy lập bảng phân bố xác suất của X (giả thiết rằng xác suất sinh con trai là 0,5).

45. Số ca cấp cứu ở một bệnh viện vào tối thứ bảy là một biến ngẫu nhiên rời rạc X có bảng phân bố xác suất như sau :

X	0	1	2	3	4	5
P	0,15	0,2	0,3	0,2	0,1	0,05

Biết rằng, nếu có hơn 2 ca cấp cứu thì phải tăng cường thêm bác sĩ trực.

a) Tính xác suất để phải tăng cường thêm bác sĩ trực vào tối thứ bảy.

b) Tính xác suất để xảy ra ít nhất một ca cấp cứu vào tối thứ bảy.

46. Số cuộc điện thoại gọi đến một tổng đài trong khoảng thời gian 1 phút vào buổi trưa (từ 12 giờ đến 13 giờ) là một biến ngẫu nhiên rời rạc X có bảng phân bố xác suất sau :

X	0	1	2	3	4	5
P	0,3	0,2	0,15	0,15	0,1	0,1

Tính xác suất để trong khoảng thời gian từ 12 giờ 30 phút đến 12 giờ 31 phút có nhiều hơn 2 cuộc gọi.

47. Tính kì vọng, phương sai và độ lệch chuẩn của biến ngẫu nhiên rời rạc X trong bài tập 44 (tính chính xác đến hàng phần trăm).

48. Tính kì vọng, phương sai và độ lệch chuẩn của biến ngẫu nhiên rời rạc X trong bài tập 45 (tính chính xác đến hàng phần trăm).

49. Tính kì vọng, phương sai và độ lệch chuẩn của biến ngẫu nhiên rời rạc X trong bài tập 46 (tính chính xác đến hàng phần trăm).

Bài đọc thêm

LIÊN HỆ GIỮA BIẾN NGẪU NHIÊN RỜI RẠC VÀ THỐNG KÊ

Xét dấu hiệu X với tập giá trị hữu hạn $\{x_1, x_2, ..., x_n\}$. Giả sử trên một mẫu điều tra kích thước N về dấu hiệu X, ta thấy có n_i số liệu có giá trị x_i $(i = 1, 2, ..., n)$, tức là giá trị x_i có tần số n_i. Tần suất của giá trị x_i là $f_i = \frac{n_i}{N}$.

Khi đó bảng phân bố tần suất của mẫu số liệu trên là

Giá trị x	x_1	x_2	...	x_n
Tần suất f	f_1	f_2	...	f_n

Có thể coi dấu hiệu X nói trên là một biến ngẫu nhiên rời rạc với tập giá trị $\{x_1, x_2, ..., x_n\}$. Giả sử phân bố xác suất của X là

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

Chúng ta đã biết rằng tần suất f_i là một giá trị gần đúng của xác suất p_i. Do đó *bảng phân bố tần suất của mẫu số liệu cho ta một "hình ảnh" gần đúng về bảng phân bố xác suất của X.*

Số trung bình của mẫu số liệu là $\overline{x} = \dfrac{\sum_{i=1}^{n} n_i x_i}{N} = \sum_{i=1}^{n} x_i \dfrac{n_i}{N} = \sum_{i=1}^{n} x_i f_i$.

Vì $f_i \approx p_i$ nên $\overline{x} = \sum_{i=1}^{n} x_i f_i \approx \sum_{i=1}^{n} x_i p_i = \mathrm{E}(X)$.

Như vậy, *số trung bình của mẫu số liệu là một giá trị gần đúng của kì vọng của X.*

Tương tự, phương sai của mẫu số liệu là

$$s^2 = \frac{\sum_{i=1}^{n} n_i (x_i - \overline{x})^2}{N} = \sum_{i=1}^{n} (x_i - \overline{x})^2 f_i \approx \sum_{i=1}^{n} (x_i - \mu)^2 p_i = \mathrm{V}(X)$$

ở đó $\mu = \mathrm{E}(X)$.

Vậy *phương sai và độ lệch chuẩn của mẫu số liệu là các giá trị gần đúng của phương sai và độ lệch chuẩn của X.*

Luyện tập

50. Chọn ngẫu nhiên 3 đứa trẻ từ một nhóm trẻ gồm 6 trai và 4 gái. Gọi X là số bé gái trong số 3 đứa trẻ được chọn. Lập bảng phân bố xác suất của X.

51. Số đơn đặt hàng đến trong một ngày ở một công ti vận tải là một biến ngẫu nhiên rời rạc X có bảng phân bố xác suất như sau :

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

a) Tính xác suất để số đơn đặt hàng thuộc đoạn [1 ; 4].

b) Tính xác suất để có ít nhất 4 đơn đặt hàng đến công ti đó trong một ngày.

c) Tính số đơn đặt hàng trung bình đến công ti đó trong một ngày.

52. Cho biến ngẫu nhiên rời rạc X có bảng phân bố xác suất như sau :

X	0	1	2	3	4	5	6	7	8	9
P	0,01	0,05	0,1	0,14	0,18	0,25	0,15	0,07	0,04	0,01

a) Tính $\mathrm{P}(2 < X < 7)$.

b) Tính $\mathrm{P}(X > 5)$.

53. Cho biến ngẫu nhiên rời rạc X có bảng phân bố xác suất như sau :

X	0	1	2	3
P	$\frac{1}{28}$	$\frac{15}{56}$	$\frac{27}{56}$	$\frac{3}{14}$

Tính E(X), V(X) và σ(X) (tính chính xác đến hàng phần nghìn).

54. Cho biến ngẫu nhiên rời rạc X có bảng phân bố xác suất như sau :

X	15	18	21	24
P	$\frac{3}{14}$	$\frac{27}{56}$	$\frac{15}{56}$	$\frac{1}{28}$

Tính E(X), V(X) và σ(X) (tính chính xác đến hàng phần nghìn).