Chương 3: BIẾN NGẪU NHIÊN RỜI RẠC VÀ PHÂN PHỐI XÁC SUẤT

3.1. BIẾN NGẪU NHIÊN

Biến ngẫu nhiên (BNN): là ánh xạ mà qua đó mỗi kết quả của phép thử xác định tương ứng cho nó một giá trị thực.

$$\begin{aligned} X:\Omega &\to \mathbb{R} \\ \omega &\mapsto X(\omega)=x \end{aligned}$$

BNN cũng được gọi là Đại lượng ngẫu nhiên (ĐLNN).

BNN thường được đặt tên là $X, Y, Z, \ldots$

Trong phần lý thuyết, chúng ta sẽ kí hiệu các giá trị mà một BNN nhận tương ứng với các kết quả của phép thử bằng chữ cái giống tên của BNN đó nhưng viết thường, chẳng hạn các giá trị của BNN X là $x_1, x_2, \ldots, x_n$.

Phân loại biến ngẫu nhiên:

Có hai loại BNN:

1) ***BNN rời rạc:*** là BNN mà tập giá trị có số lượng phần tử **hữu hạn** hoặc **vô hạn đếm được**.

2) ***BNN liên tục:*** là BNN mà tập giá trị có số lượng phần tử **vô hạn không đếm được**, tức là tập hợp chứa vô số giá trị thực liên tiếp nhau lấp đầy ít nhất một khoảng nào đó trên trục số thực $\mathbb{R}$.

Ví dụ 1: X là số lượng học sinh giỏi trong 5 học sinh được chọn ngẫu nhiên từ lớp có 9 học sinh giỏi và 28 học sinh khá.
Tập hợp các giá trị của X là {0, 1, 2, 3, 4, 5}.
Do đó X là BNN rời rạc.

Ví dụ 2: Y là chiều cao của một học sinh nữ được chọn ngẫu nhiên từ các học sinh nữ cao từ 1,5m đến 1,6m ở khu vực K.
Tập hợp các giá trị của Y (đơn vị: m) là đoạn [1,5 ; 1,6].
Do đó Y là BNN liên tục.

3.2. PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN RỜI RẠC

Giả sử BNN rời rạc X có tập giá trị $\{x_1, x_2, \ldots, x_n\}$ và $P\{X = x_i\} = p_i$, $(i = 1, 2, \ldots, n)$ khi đó bảng phân phối xác suất của X có dạng:

X	x_1	x_2	…	x_n
$p_i = P\{X = x_i\}$	p_1	p_2	…	p_n

Lưu ý:

- $x_1, x_2, \ldots, x_n$ là các giá trị thực khác nhau từng đôi một và nên được liệt kê vào dòng thứ nhất của bảng phân phối xác suất theo thứ tự tăng dần từ trái sang phải.
- $p_i > 0, \forall i = 1, 2, \ldots, n$
- $p_1 + p_2 + \ldots + p_n = 1$

Ví dụ: Hộp có 10 viên bi đỏ và 4 bi xanh. Lấy ra 3 bi. Đặt X là số bi đỏ trong 3 bi lấy ra. Lập bảng phân phối xác suất cho X.

X	0	1	2	3
$p_i = P\{X = x_i\}$	1/91	15/91	45/91	30/91

$P\{X=0\} = \dfrac{C_4^3}{C_{14}^3} = \dfrac{1}{91}$ $\qquad P\{X=1\} = \dfrac{C_{10}^1 C_4^2}{C_{14}^3} = \dfrac{15}{91}$

$P\{X=2\} = \dfrac{C_{10}^2 C_4^1}{C_{14}^3} = \dfrac{45}{91}$ $\qquad P\{X=3\} = \dfrac{C_{10}^3}{C_{14}^3} = \dfrac{30}{91}$

Hàm xác suất của BNN rời rạc X, thường được kí hiệu $p_X(u)$, là hàm số một biến thực: $p_X(u) = P(X = u) \quad ,\forall u \in \mathbb{R}$

Tính chất: $p_X(u) \geq 0, \forall u \in \mathbb{R}$ và $\sum_{u \in \mathbb{R}} p_X(u) = 1$

Nếu X có tập giá trị $\{x_1, x_2, \dots, x_n\}$ và $P\{X = x_i\} = p_i$ thì:

$(i = 1, 2, \dots, n)$

$$p_X(u) = \begin{cases} p_1 & khi\ u = x_1 \\ p_2 & khi\ u = x_2 \\ \vdots & \\ p_n & khi\ u = x_n \\ 0 & khi\ u \notin \{x_1, x_2, \dots, x_n\} \end{cases}$$

3.3. KỲ VỌNG VÀ PHƯƠNG SAI CỦA BIẾN NGẪU NHIÊN RỜI RẠC

Định nghĩa: Kỳ vọng (còn gọi là "Trung bình (theo xác suất)") của BNN rời rạc X được kí hiệu là E(X) và được xác định như sau:

$$E(X) = \sum_{x_i \in \text{ tập giá trị của } X} x_i . P\{X = x_i\}$$

Tính chất:

1/ $E(c) = c \quad khi \quad c = const$

2/ $E(X+Y) = E(X) + E(Y)$

$E(c+X) = c + E(X) \quad khi \quad c = const$

3/ $E(c.X) = c.E(X) \quad khi\ c = const$

4/ Chỉ với X và Y là các BNN độc lập thì có được $E(XY) = E(X).E(Y)$

Chú ý: với X là BNN trong một phép thử cụ thể thì $E(X)$ là hằng số.

Qui tắc LOTUS (law of the unconscious statistician): Với X là một BNN rời rạc và hàm số một biến g(u) có giá trị thực tương ứng với mỗi u thuộc tập giá trị của X thì Y = g(X) cũng là BNN rời rạc và có trung bình là:

$$E[Y] = E[g(X)] = \sum_{x_i \in \text{ tập giá trị của } X} g(x_i).P\{X = x_i\}$$

Áp dụng LOTUS, ta có:

$$E[X^2] = \sum_{x_i \in \text{ tập giá trị của } X} x_i^2 .P\{X = x_i\}$$

Định nghĩa: Phương sai của BNN X được kí hiệu là Var(X) (hay D(X) hay V(X)) và được định nghĩa như sau:

$$Var(X) = E\left[(X - E(X))^2\right]$$

Công thức tương đương để tính phương sai là:

$$Var(X) = E\left(X^2\right) - E^2\left(X\right)$$

Tính chất:

(i) $Var(X) \geq 0$

(ii) $Var(c) = 0 \quad (c = const)$

(iii) $Var(cX) = c^2 Var(X)$.

(iv) Chỉ với X và Y là các BNN độc lập thì $Var(X+Y) = Var(X) + Var(Y)$

Chú ý: với X là BNN trong một phép thử cụ thể thì $Var(X)$ là hằng số.

Định nghĩa: Độ lệch tiêu chuẩn của BNN X là $\sigma(X) = \sqrt{var(X)}$

Ví dụ: Tính trung bình, phương sai, độ lệch chuẩn của BNN rời rạc X trong ví dụ về bảng phân phối xác suất.

3.4. Phân phối Nhị thức B(n ; p)

3.4. 1. Phân phối Bernoulli

Định nghĩa: Phân phối xác suất của BNN rời rạc X có tập giá trị là {0 ; 1} và P{X=1}= p với 0 < p < 1 được gọi là phân phối Bernoulli tham số p.

Tính chất: Khi BNN X có phân phối Bernoulli tham số p thì E[X] = p và Var(X) = p(1 – p)

Ví dụ: Xét phép thử có xác suất xảy ra biến cố A là P(A) = p (0 < p < 1). Đặt:

$$X = \begin{cases} 0 & \text{khi A không xảy ra;} \\ 1 & \text{khi A xảy ra.} \end{cases}$$

thì X có phân phối Bernoulli tham số p.

3.4.2 Phân phối Nhị thức:

Định nghĩa: Giả sử có **n** phép thử độc lập mà xác suất *biến cố A xảy ra* trong mỗi phép thử đều bằng **p** (với 0 < p <1). Gọi X là số lượng phép thử có biến cố A xảy ra trong **n** phép thử độc lập này, thì X là BNN rời rạc có ***phân phối nhị thức*** với tham số **n** và **p**. Kí hiệu: X ~ B(**n** ; **p**).

Tính chất:

i) X tuân theo phân phối Nhị thức tham số n và p thì có tập giá trị là {0, 1, 2,…, n} và:

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

$$E[X] = np$$

ii) $Var(X) = np(1-p)$

$$np + p - 1 \le Mod(X) \le np + p$$

với $Mod(X)$ là kí hiệu cho giá trị mà X có xác suất nhận trội nhất trong tập giá trị của X.

iii) Tổng các phân phối nhị thức độc lập: khi $X_i \sim B(n_i, p)$, $i = 1, 2, ..., m$ là các BNN độc lập thì $X = \sum_{i=1}^{m} X_i \sim B(n; p)$ với $n = \sum_{i=1}^{m} n_i$

13

Ví dụ 1: Một xạ thủ bắn 3 lần độc lập vào một mục tiêu, xác suất trúng mỗi lần bắn đều là 0,8. Gọi X là số lượng lần bắn trúng mục tiêu. Xác định luật phân phối xác suất của X và lập bảng phân phối xác suất của X.

Ví dụ 2: Mua ngẫu nhiên 100 tờ vé số trong một đợt phát hành, biết xác suất trúng với mỗi tờ đều là 0,0005.

a) Tính xác suất trúng được 8 tờ.
b) Gọi X là số lượng tờ vé trúng trong 100 tờ vé số chọn mua, tính E[X], Var(X).
c) Phải mua ít nhất bao nhiêu tờ vé số thì xác suất "có mua được tờ vé trúng" mới không dưới 90%?

14

3.5. PHÂN PHỐI SIÊU BỘI VÀ PHÂN PHỐI NHỊ THỨC ÂM

3.5.1. Phân phối Siêu bội:

Giả sử có ***N*** phần tử trong đó có ***M*** phần tử có tính chất A.

Lấy ra ***n*** phần tử, *X* là số phần tử có tính chất A trong ***n*** phần tử lấy ra. Khi đó X có ***phân phối siêu bội tham số n; M; N.***

Kí hiệu: $X \sim H(n; M; N)$

$\forall \boldsymbol{k} \in \mathbb{N}$ ***sao cho*** $\boldsymbol{Max[0\,;\, n - N + M] \le k \le Min[n; M]}$, **thì:**

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

15

Tính chất của BNN $\boldsymbol{X \sim H(n\,;\, M\,;\, N)}$:

$$E(X) = n.\frac{M}{N}$$

$$Var(X) = n.\frac{M}{N}(1 - \frac{M}{N}).\frac{N-n}{N-1}$$

<u>*Ví dụ 1:*</u> Một hộp có 50 viên bi trong đó có 10 bi màu xanh. Lấy ra 20 viên bi, gọi X là số bi xanh. Tính E(X) ; Var(X).

16

<u>*Ví dụ 2:*</u> Một lô hàng có 10000 sản phẩm. Số sản phẩm loại I có trong lô là 2000. Người mua hàng chọn ngẫu nhiên 10 sản phẩm từ lô.

a/ Gọi X là số sp loại I trong 10 sp được chọn, tính E(X), Var(X).

b/ Nếu trong 10 sản phẩm được chọn có hơn 2 sản phẩm loại I thì lô hàng này sẽ được mua. Tính xác suất lô hàng này được mua?

17

3.5.2. Phân phối Nhị thức âm:

Định nghĩa: Giả sử có thể làm lần lượt nhiều phép thử độc lập mà xác suất biến cố A xảy ra trong mỗi phép thử đều bằng p (với $0 < p < 1$).

Gọi T_r là số lượng phép thử không có biến cố A xảy ra sẽ được làm cho đến khi ***có vừa đủ r phép thử mà biến cố A xảy ra*** (với $r \in \mathbb{N}^*$), thì T_r là BNN rời rạc có ***phân phối Nhị thức âm*** với tham số r và p. Kí hiệu: $T_r \sim NB(r\,;\, p)$.

18

Tính chất: Cho T_r là BNN có phân phối Nhị thức âm tham số r và p, ta có:

i) T_r có tập giá trị là $\mathbb{N}$, và:

$$P\{T_r = k\} = C_{r+k-1}^{r-1} . p^r (1-p)^k \ , \ \forall k \in \mathbb{N} \ ,$$

ii) $$E[T_r] = \frac{r(1-p)}{p} \quad ; \quad Var(T_r) = \frac{r(1-p)}{p^2}$$

19

Ví dụ: Người ta dự định kiểm tra lần lượt (không hoàn lại) từng sản phẩm được lấy ngẫu nhiên từ rất nhiều sản phẩm thuộc cùng một lô của xưởng F. Số liệu thống kê sản xuất cho thấy tỉ lệ phế phẩm trong mỗi lô sản phẩm của xưởng F là 8%.

a) Nếu việc kiểm tra chỉ được tiến hành vừa đủ để phát hiện (1) phế phẩm thì dừng ngay, hãy tính trung bình của số lượng sản phẩm được lấy kiểm tra và xác suất mà số lượng sản phẩm được lấy kiểm tra thấp hơn một nửa mức trung bình này.

b) Nếu việc kiểm tra được tiến hành để phát hiện vừa đủ 8 phế phẩm thì dừng ngay, hãy tính trung bình của số lượng sản phẩm được lấy kiểm tra và xác suất mà số lượng sản phẩm được lấy kiểm tra là 20.

20

3.6. Phân phối Poisson

Định nghĩa: BNN rời rạc X tuân theo phân phối Poisson tham số λ ($\lambda > 0$) khi X có tập giá trị là $\mathbb{N}$ và:

$$\boxed{P(X=k) = \frac{e^{-\lambda}.\lambda^k}{k!} \qquad , \forall k \in \mathbb{N}}$$

Kí hiệu : $X \sim P(\lambda)$

Mệnh đề: (Xấp xỉ phân phối Nhị thức bởi phân phối Poisson) Nếu $X \sim B(n,p)$ với n *lớn* và p *nhỏ* ($n \geq 1000$ và $n.p \leq 5$) thì có thể xem như $X \sim P(\lambda)$ với tham số $\lambda = n.p$.

21

Tính chất:

i) $$E[X] = Var(X) = \lambda$$ $$\lambda - 1 \leq Mod(X) \leq \lambda$$

ii) Tổng các phân phối Poisson độc lập: khi $X_i \sim P(\lambda_i)$, $i = 1, 2, ..., m$, là các BNN độc lập thì

$$X = \sum_{i=1}^{m} X_i \sim P(\sum_{i=1}^{m} \lambda_i)$$

22

Ví dụ 1: Biết số lượng lỗi in trong mỗi trang của một cuốn sách tuân theo phân phối Poisson và trung bình trong mỗi trang có 0,5 lỗi in. Hãy tính xác suất trong 3 trang nào đó của cuốn sách này có tổng cộng 3 lỗi.

Ví dụ 2: Tỉ lệ phế phẩm trong ***rất nhiều*** sản phẩm mà một máy sản xuất ra được cho biết là 0,002. Mỗi lô hàng gồm 1000 sản phẩm do máy này sản xuất có trung bình bao nhiêu phế phẩm? Tính xác suất trong một lô hàng như vậy có 50 phế phẩm.

23