HỌC VIỆN CÔNG NGHỆ BƯU CHÍNH VIỄN THÔNG

o0o

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

XÁC SUẤT THỐNG KÊ

Bộ môn Toán, Khoa Cơ bản 1

Hà Nội - 2023

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

2.1.1 Định nghĩa biến ngẫu nhiên

Định nghĩa

Biến ngẫu nhiên (Bnn) là biến nhận giá trị là các số thực tùy thuộc vào kết quả ngẫu nhiên của phép thử. Người ta thường dùng các chữ in hoa $X, Y, Z, \ldots$ để chỉ các biến ngẫu nhiên.

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

- Bnn được gọi là rời rạc nếu nó nhận một số hữu hạn hay vô hạn đếm được các giá trị.
- Bnn được gọi là liên tục nếu các giá trị của nó là một khoảng hoặc một số khoảng thực nào đó.

Ví dụ 1. Gieo 1 con súc sắc. Gọi X là số chấm xuất hiện $\Rightarrow X$ là một Bnn rời rạc có thể nhận các giá trị $1, 2, 3, 4, 5, 6$.

Ví dụ 2. Một xạ thủ bắn vào bia cho tới khi nào trúng thì dừng. Gọi Y là số viên đạn cần dùng $\Rightarrow Y$ là một Bnn rời rạc có thể nhận các giá trị $1, 2, \ldots$.

Ví dụ 3. Thắp sáng liên tục một bóng đèn điện cho tới khi dây tóc của bóng đèn bị cháy. Gọi Z là thời gian bóng đèn sáng $\Rightarrow Z$ là một Bnn liên tục.

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

Hàm khối lượng xác suất

Hàm khối lượng xác suất của Bnn rời rạc X, ký hiệu $p_X(x)$, là hàm số được xác định bởi

$$p_X(x) = P(X = x), x \in \mathbb{R}$$

Tính chất của hàm khối lượng xác suất:

1) $p_X(x) > 0, \forall x \in R_X$
2) $\sum\limits_{x \in R_X} p_X(x) = 1$
3) $p_X(x) = 0, \forall x \notin R_X$

Bảng phân bố xác suất

Bảng phân bố xác suất của Bnn rời rạc là 1 bảng gồm 2 dòng: dòng trên ghi các giá trị có thể có của Bnn, dòng dưới ghi các xác suất tương ứng

X	x_1	x_2	...
P	$p_X(x_1)$	$p_X(x_2)$	...

Ví dụ 4. Một công ty cần tuyển 3 nhân viên. Có 10 người nộp đơn trong đó có 6 nam và 4 nữ. Khả năng được tuyển của mỗi người là như nhau. Gọi X là số nữ được chọn. Lập bảng phân bố xác suất của X.

2.1.3 Hàm phân bố xác suất

Định nghĩa

Hàm phân bố xác suất của Bnn X, ký hiệu $F_X(x)$, là hàm số được xác định bởi

$$F_X(x) = P(X \leq x), x \in \mathbb{R}$$

Nhận xét: Nếu X chỉ nhận một số hữu hạn các giá trị $x_1, x_2, \ldots, x_n$ thì

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < x_1, \\ p_X(x_1) & \text{nếu } x_1 \leq x < x_2, \\ \ldots & \\ p_X(x_1) + \ldots + p_X(x_{i-1}) & \text{nếu } x_{i-1} \leq x < x_i, \\ \ldots & \\ 1 & \text{nếu } x \geq x_n \end{cases}$$

Ví dụ 5. Tìm hàm phân bố xác suất của Bnn X cho trong Ví dụ 4.

Tính chất của hàm phân bố xác suất:

1) $0 \leq F_X(x) \leq 1$
2) $F_X(-\infty) = \lim\limits_{x \to -\infty} F_X(x) = 0,$
   $F_X(+\infty) = \lim\limits_{x \to +\infty} F_X(x) = 1$
3) $P(a < X \leq b) = F_X(b) - F_X(a)$
4) $F_X(x)$ là hàm không giảm.
5) $F_X(x)$ là hàm liên tục phải.

2.1.4 Hàm mật độ xác suất

Định nghĩa

Giả sử X là biến ngẫu nhiên liên tục có hàm phân bố xác suất $F_X(x)$. Nếu tồn tại hàm không âm $f_X(x)$ sao cho

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt, \ \forall x \in \mathbb{R}$$

thì $f_X(x)$ được gọi là hàm mật độ xác suất của biến ngẫu nhiên X.

Tính chất của hàm mật độ xác suất:

1) $f_X(x) \geq 0, \forall x \in \mathbb{R}$
2) $f_X(x) = F'_X(x)$ tại các điểm x mà $f_X(x)$ liên tục.
3) $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b)$
   $= P(a < X < b) = \int_a^b f_X(x)dx$
4) $\int_{-\infty}^{+\infty} f_X(x)dx = 1$

Ví dụ 6. Biến ngẫu nhiên X có hàm mật độ xác suất sau:

$$f_X(x) = \begin{cases} 0 & \text{nếu } x \leq 2 \\ \dfrac{k}{x^2} & \text{nếu } x > 2 \end{cases}$$

a) Xác định hệ số k

b) Tìm hàm phân bố $F_X(x)$

c) Tính $P(1 < X \leq 10|X > 4)$

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

2.2.1 Kỳ vọng

Định nghĩa

Kỳ vọng của Bnn X, ký hiệu $E(X)$, được xác định như sau:

- Nếu X là Bnn rời rạc có miền giá trị $R_X = \{x_1, x_2, \dots, x_n\}$ thì

$$E(X) = \sum_{i=1}^{n} x_i p_X(x_i)$$

- Nếu X là Bnn rời rạc có miền giá trị $R_X = \{x_1, x_2, \dots\}$ thì

$$E(X) = \sum_{i=1}^{\infty} x_i p_X(x_i)$$

với điều kiện chuỗi $\sum_{i=1}^{\infty} |x_i| p_X(x_i)$ hội tụ.

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

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

với điều kiện $\int_{-\infty}^{\infty} |x| f_X(x)dx$ hội tụ.

Ví dụ 7. Một cuộc khảo sát về những người mua hàng trên website Amazon.com cho thấy phân bố xác suất của số lượng sách X được mua mỗi lần khách hàng truy cập như sau:

X	0	1	2	3	4	5	6
P	0,35	0,25	0,2	0,08	0,06	0,03	0,03

Tính kỳ vọng của X.

Ví dụ 8. Thời gian xếp hàng chờ mua hàng của khách hàng là Bnn liên tục X có hàm mật độ xác suất như sau (đơn vị: phút)

$$f_X(x) = \begin{cases} \dfrac{4}{81}x^3 & \text{nếu } x \in (0,3) \\ 0 & \text{nếu } x \notin (0,3) \end{cases}$$

Tìm thời gian xếp hàng trung bình của khách hàng.

Tính chất của kỳ vọng:

1) $E(C) = C$, với C là một hằng số
2) $E(kX) = kE(X)$, với k là hằng số
3) $E(X+Y) = E(X) + E(Y)$

   Hệ quả:

   - $E(X-Y) = E(X) - E(Y)$
   - $E(a+bX) = a + bE(X)$ với a, b là các hằng số
   - $E(k_1X_1 + k_2X_2 + \ldots + k_nX_n) = k_1E(X_1) + k_2E(X_2) + \ldots + k_nE(X_n)$

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

$$E[g(X)] = \begin{cases} \displaystyle\sum_{x \in R_X} g(x)p_X(x) & \text{nếu } X \text{ là Bnn rời rạc} \\ \displaystyle\int_{-\infty}^{\infty} g(x)f_X(x)dx & \text{nếu } X \text{ là Bnn liên tục} \end{cases}$$

Hai biến ngẫu nhiên độc lập

Hai Bnn X và Y được gọi là độc lập với nhau nếu quy luật phân bố xác suất của Bnn này không phụ thuộc vào việc Bnn kia lấy giá trị nào.

5) Nếu X, Y là 2 Bnn độc lập thì $E(XY) = E(X)E(Y)$.

Ví dụ 9. Số sản phẩm cửa hàng A bán được trong ngày là Bnn X có bảng phân bố xác suất như sau:

X	0	1	2	3	4	5
P	0,07	0,13	0,3	0,3	0,15	0,05

a) Tính xác suất để trong một tuần (7 ngày) có 4 ngày cửa hàng bán được nhiều hơn 2 sản phẩm.

b) Tính số sản phẩm bán được trung bình trong một ngày. Để lợi nhuận trung bình mỗi ngày của cửa hàng đạt 500 ngàn đồng thì cần quy định lợi nhuận của mỗi sản phẩm là bao nhiêu?

2.2.2 Phương sai

Định nghĩa

Phương sai của Bnn X, ký hiệu $D(X)$, là 1 số được xác định bởi

$$D(X) = E[(X - E(X))^2].$$

Chú ý:

- Phương sai là đại lượng đặc trưng cho mức độ phân tán của Bnn quanh giá trị trung bình lý thuyết của nó.
- 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.
- $D(X) = E(X^2) - [E(X)]^2$.

Ví dụ 10. Giả sử Bnn X có bảng phân bố xác suất như sau:

X	0	1	2	3
P	0,24	0,46	0,26	0,04

Tính phương sai của X.

Tính chất của phương sai:

1) $D(C) = 0$, với C là một hằng số.
2) $D(kX) = k^2 D(X)$, với k là hằng số.
3) Nếu X, Y là 2 Bnn độc lập thì

$$D(X + Y) = D(X) + D(Y)$$

Hệ quả:

- $D(a + bX) = b^2 D(X)$ với a, b là các hằng số.
- Nếu $X_1, X_2, \ldots, X_n$ là các Bnn độc lập thì

$$D(k_1X_1 + k_2X_2 + \ldots + k_nX_n) = k_1^2 D(X_1) + k_2^2 D(X_2) + \ldots + k_n^2 D(X_n)$$

2.2.3 Độ lệch chuẩn

Định nghĩa

Độ lệch chuẩn của Bnn X là số

$$\sigma(X) = \sqrt{D(X)}.$$

Chú ý: Đơn vị đo của phương sai bằng bình phương đơn vị đo của biến ngẫu nhiên. Vì vậy khi cần đánh giá mức độ phân tán của Bnn theo đơn vị đo của nó người ta thường tính độ lệch chuẩn.

2.2.4 Mode

Định nghĩa

Mode của Bnn X, ký hiệu Mod$X$, là giá trị của Bnn tương ứng với:

- xác suất lớn nhất nếu X là Bnn rời rạc,
- cực đại của hàm mật độ nếu X là Bnn liên tục.

Ví dụ 11.

a) Tìm mode của Bnn X trong Ví dụ 10.

b) Tìm mode của Bnn X có hàm mật độ xác suất là

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Chương 2: Biến ngẫu nhiên và các đặc trưng của chúng

2.3.1 Phân bố nhị thức

Định nghĩa

Bnn X được gọi là có phân bố nhị thức với các tham số n, p, ký hiệu $X \sim B(n; p)$, nếu X có thể nhận một trong các giá trị: $0, 1, 2, \ldots, n$ và

$$P(X = k) = C_n^k p^k (1-p)^{n-k};\ k = 0, 1, \ldots, n,$$

trong đó $n \in \mathbb{N}^*$, $0 < p < 1$.

Chú ý: Nếu X là số lần xuất hiện biến cố A trong một dãy phép thử Bernoulli thì $X \sim B(n; p)$, trong đó n là số phép thử, p là xác suất xuất hiện biến cố A trong một phép thử.

Ví dụ 12. Một nhân viên tại một trung tâm trò chơi điện tử phụ trách 20 máy. Xác suất để mỗi máy cần sự giúp đỡ của nhân viên trong một giờ là 0,1. Gọi X là số máy cần sự giúp đỡ của nhân viên trong một giờ, ta có $X \sim B(20; 0,1)$.

Ví dụ 13. Một phân xưởng có 5 máy hoạt động độc lập. Xác suất để trong một ngày mỗi máy bị hỏng đều bằng 0,05. Gọi X là số máy hỏng trong ngày, ta có $X \sim B(5; 0,05)$.

Kỳ vọng và phương sai

Nếu $X \sim B(n;p)$ thì

$$E(X) = np, \;\; D(X) = np(1-p)$$

Ví dụ 14. 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.

2.3.2 Phân bố Poisson

Định nghĩa

Bnn X được gọi là có phân bố Poisson với tham số $\lambda > 0$, ký hiệu $X \sim P(\lambda)$, nếu X có thể nhận một trong các giá trị $0, 1, 2, \ldots$ và

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \ldots$$

Phân phối Poisson có ứng dụng trong các quá trình liên quan đến số quan sát với một đơn vị thời gian hoặc không gian, chẳng hạn như số cuộc điện thoại gọi đến một tổng đài trong một phút, số lỗi trên một mét vuông vải, số khách hàng đến nhà băng trong 30 phút, ...

Kỳ vọng và phương sai

Nếu $X \sim P(\lambda)$ thì
$$E(X) = D(X) = \lambda$$

Ví dụ 15. 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 đúng 5 cuộc điện thoại trong 2 phút,

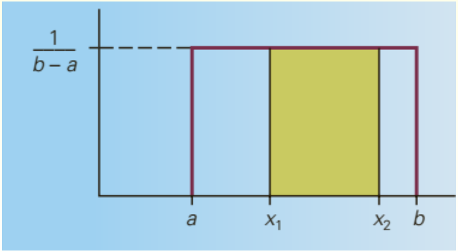
b) có ít nhất một cuộc điện thoại trong khoảng thời gian 30 giây.

2.3.3 Phân bố đều

Định nghĩa

Biến ngẫu nhiên X được gọi là có phân bố đều trên đoạn $[a; b]$, ký hiệu $X \sim U[a; b]$, nếu hàm mật độ xác suất của nó là

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{nếu } x \in [a; b] \\ 0 & \text{nếu } x \notin [a; b] \end{cases}$$



Kỳ vọng và phương sai

Nếu X có phân bố đều trên đoạn $[a;b]$ thì

$$E(X) = \frac{a+b}{2},\ D(X) = \frac{(b-a)^2}{12}$$

Hàm phân bố xác suất

Nếu X có phân bố đều trên đoạn $[a; b]$ thì

$$F_X(x) = \begin{cases} 0 & \text{nếu } x < a \\ \dfrac{x-a}{b-a} & \text{nếu } a \le x \le b \\ 1 & \text{nếu } x > b \end{cases}$$

Ví dụ 16. Một xe buýt xuất hiện tại một trạm xe buýt cứ 15 phút một chuyến. Một hành khách tới trạm vào một thời điểm ngẫu nhiên. Thời gian chờ xe của hành khách đó là biến ngẫu nhiên X có phân bố đều trên khoảng $[0; 15]$.

a) Viết hàm phân bố xác suất của X.

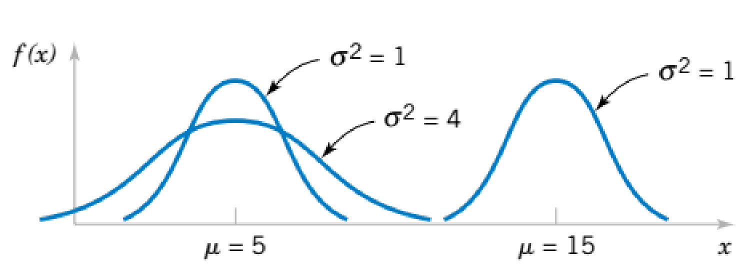
b) Tính xác suất để hành khách đó phải đợi ít hơn 5 phút.

c) Tính xác suất để hành khách đó phải đợi nhiều hơn 10 phút.

Định nghĩa

Bnn X được gọi là có phân bố chuẩn với các tham số μ, σ^2 $(\sigma > 0)$, ký hiệu $X \sim N(\mu, \sigma^2)$, nếu hàm mật độ xác suất của nó là

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}$$



Hình 3.1: Đồ thị hàm mật độ xác suất của phân bố chuẩn

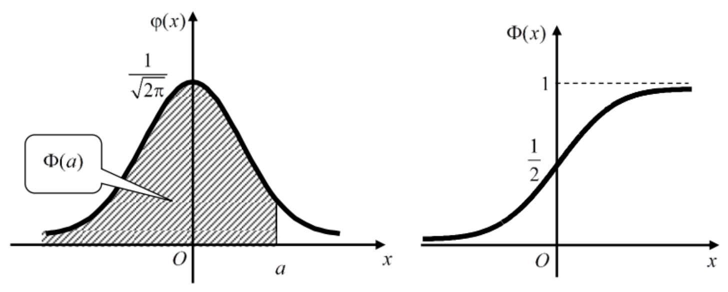
Phân bố chuẩn tắc

Nếu $X \sim N(0,1)$ thì ta nói X có phân bố chuẩn tắc. Khi đó hàm mật độ xác suất của X là

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\dfrac{x^2}{2}}$$

và hàm phân bố xác suất của X là

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\dfrac{t^2}{2}} dt.$$



Hình 3.2: Đồ thị hàm mật độ xác suất $\varphi(x)$ và hàm phân bố xác suất $\Phi(x)$

Tính chất của hàm phân bố xác suất $\Phi(x)$:

1) $\Phi(-x) = 1 - \Phi(x)$
2) Nếu $x \geq 3,5$ thì $\Phi(x) \approx 1$

Giá trị tới hạn

Cho biến ngẫu nhiên $X \sim N(0,1)$. Giá trị tới hạn mức α của X, ký hiệu U_α, là giá trị thỏa mãn

$$P(X > U_\alpha) = \alpha.$$

Cách tìm U_α:
Tính $\Phi(U_\alpha) = 1 - \alpha$, sau đó tra bảng Phụ lục II để tìm U_α.

Một số giá trị thường dùng:

$$U_{0,025} = 1,96;\ U_{0,05} = 1,645;\ U_{0,01} = 2,33; U_{0,005} = 2,575$$

Kỳ vọng và phương sai

Nếu $X \sim N(\mu, \sigma^2)$ thì

$$E(X) = \mu, \; D(X) = \sigma^2$$

Định lý

Nếu $X \sim N(\mu, \sigma^2)$ thì $\dfrac{X-\mu}{\sigma} \sim N(0,1)$

Các công thức tính xác suất

Nếu $X \sim N(\mu, \sigma^2)$ thì

a) $P(X < a) = \Phi\left(\dfrac{a-\mu}{\sigma}\right)$

b) $P(a < X < b) = \Phi\left(\dfrac{b-\mu}{\sigma}\right) - \Phi\left(\dfrac{a-\mu}{\sigma}\right)$

Ví dụ 17. Chỉ số IQ của người trưởng thành là biến ngẫu nhiên có phân bố chuẩn với trung bình là 100 và độ lệch chuẩn là 16.

a) Tính tỷ lệ người trưởng thành có chỉ số IQ trên 120.

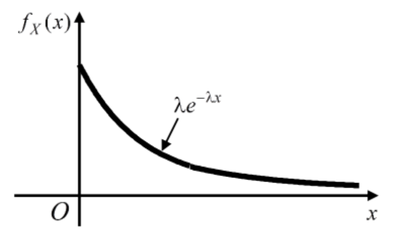
b) Nếu chọn ngẫu nhiên 5 người trưởng thành thì xác suất để có ít nhất 2 người có chỉ số IQ trên 120 là bao nhiêu?

c) Nếu chọn ngẫu nhiên 80 người trưởng thành thì nhiều khả năng nhất có bao nhiêu người có chỉ số IQ từ 100 đến 120?

Định nghĩa

Biến ngẫu nhiên X được gọi là có phân bố mũ với tham số $\lambda > 0$ nếu hàm mật độ xác suất của nó là

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



Hình 3.3: Đồ thị hàm mật độ xác suất của phân bố mũ

Nếu X có phân phối mũ với tham số λ thì hàm phân bố xác suất của X là

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & \text{nếu } x > 0 \\ 0 & \text{nếu } x \leq 0 \end{cases}$$

Kỳ vọng và phương sai

Nếu X có phân bố mũ với tham số λ thì

$$E(X) = \frac{1}{\lambda},\ D(X) = \frac{1}{\lambda^2}$$

Ví dụ 18. Tuổi thọ của một mạch điện tử trong máy tính là một biến ngẫu nhiên có phân bố mũ. Giả sử tuổi thọ trung bình của mạch điện tử này là 6,25 năm và 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?