Bài 5:
Phân bố xác suất

Nội dung

- Biến ngẫu nhiên (nhắc lại)
  - Khái niệm
  - Tính xác suất
  - Phân phối xác suất
  - Kỳ vọng - Phương sai - Độ lệch chuẩn
- Phân phối đều rời rạc
  - Khái niệm
  - Đặc trưng
- Phân phối nhị thức
  - Khái niệm
  - Đặc trưng

Biến ngẫu nhiên

- Khái niệm Biến ngẫu nhiên: là ánh xạ từ một tập hợp, xây dựng trên nền không gian mẫu S, vào tập các xác suất có thể xảy ra.
  - Biến ngẫu nhiên rời rạc: nếu nó chỉ có hữu hạn, hoặc vô hạn đếm được các giá trị
    - Ví dụ: X_1 = Điểm đạt được khi tung xúc xắc

    X_1 có thể nhận 1 trong các giá trị {1,2,3,4,5,6}
  - Biến ngẫu nhiên liên tục
    - Ví dụ: X_2 = Chiều cao của 1 người Việt Nam

    X_2 có thể có giá trị 1.6m, 1.61m, 1.612m,…

Tính xác suất – Hàm xác suất

- Đối với biến ngẫu nhiên rời rạc:
  - Gọi là hàm độ lớn xác suất (pms – probability mass function): là hàm gán xác suất cho từng giá trị ***x*** của ***X***, ký hiệu **f(x).**
    - $0 \leq f(x) \leq 1$, $\forall x$ là giá trị X có thể nhận.
    - $\sum_x f(x) = 1$
    - $f(x) = Pr(X=x)$

Tính xác suất – Hàm xác suất

- Sử dụng để tính xác suất

a) $Pr(X = a) = f(a)$

b) $Pr(X = a \text{ hay } X = b) = \ldots$

c) $Pr(X > a) = \ldots$

d) $Pr(X \geq a) = \ldots$

e) $Pr(X < a) = \ldots$

f) $Pr(X \leq a) = \ldots$

g) $Pr(a \leq X \leq b) = \ldots$

h) $Pr(a < X \leq b) = \ldots$

i) $Pr(a \leq X < b) = \ldots$

j) $Pr(a \leq X < b) = \ldots$

Ví dụ

- 1 X = Điểm đạt được khi tung xúc xắc
- Pr(X = 1)= f(1) = xác suất tung xúc xắc được 1 điểm = …

Tương tự Pr(X = 2)=f(2) …,

Pr(X = 3)= f(3)=…, Pr(X = 4)= f(4)=…,

Pr(X = 5)= f(5)=…, Pr(X = 6)= f(6)=…

Hàm độ lớn xác suất:

X= x	1	2	3	4	5	6
Pr(X=x) = f(s)						

Pr(X =3 hay X =5) = …

Ví dụ (tt)

- Ví dụ trên
- $Pr(1<X<5) = f(2) + f(3) + f(4) = 3/6$
- $Pr(X<4)=$...

Tính xác suất – Hàm xác suất

- Đối với biến ngẫu nhiên liên tục:
  - Gọi là hàm mật độ xác suất (probability density function)

    (1) $f(x) \geq 0$

    (2) $\int_{-\infty}^{\infty} f(x)dx = 1$

    (3) $\Pr(a \leq X \leq b) = \int_a^b f(x)dx$: diện tích vùng bên dưới đường cong f(x) với x chạy từ a đến b

    - Sử dụng để tính xác suất

    Pr(X=a) = 0

    $Pr(a \leq X \leq b) = Pr(a < X \leq b)$
    $= Pr(a \leq X < b) = Pr(a < X < b)$

Tính xác suất – Hàm phân phối tích lũy

- Hàm phân phối tích lũy (cdf – cumulative distribution function)
- Trường hợp rời rạc
  - Định nghĩa $$F(a) = \Pr(X \le a) = \sum_{x \le a} f(x)$$
  - Sử dụng để tính xác suất
    a) *Pr(X > a) = 1-Pr(X ≤ a) =1-F(a) ;*
    b) *Pr(X ≥ a) =…*
    c) *Pr(a ≤ x ≤ b) =...*
    d) *Pr(a < x ≤ b) =…*
    e) *Pr(a ≤ x < b) = ...*

Tính xác suất –
Hàm phân phối tích lũy

- Trường hợp liên tục
  - Định nghĩa $$F(a) = \int_{-\infty}^{a} \Pr(a)$$
  - Sử dụng để tính xác suất

  a) *Pr(X < a) = Pr(X ≤ a) = F(a)*

  b) *Pr(X > a) = Pr(X ≥ a)=1-F(a) ;*

  c) *Pr(a ≤ X ≤ b) = Pr(a<x ≤b)*

  = Pr(a ≤ x<b) = F(b)-F(a)

  - Lưu ý: Với bài toán mà biến nhận nhiều hơn hai giá trị, việc dùng hàm phân phối tích lũy sẽ hiệu quả hơn dùng hàm xác suất

Ví dụ

- X = Điểm đạt được khi tung xúc xắc

Hàm độ lớn xác suất:

X=x	1	2	3	4	5	6
f(x)	1/6	1/6	1/6	1/6	1/6	1/6

Hàm phân phối tích lũy:

X=x	1	2	3	4	5	6
F(a) = sum(f(x),x<=a)	F(1) =1/6	F(2) =2/6	F(3)= 3/6	F(4) =4/6	F(5) =5/6	F(6)= 1

Bài tập

- X = Điểm đạt được khi tung xúc xắc

a) $Pr(X>3)=\ldots$

b) $Pr(X\geq 3)=\ldots$

c) $Pr(X<3)=\ldots$

d) $Pr(X\leq 3)=\ldots$

e) $Pr(2<X \leq 5) =\ldots$

f) $Pr(2 \leq X \leq 5)=\ldots$

g) $Pr(2 \leq X<5) =\ldots$

Phân phối xác suất

- Khái niệm Phân phối xác suất cho X:

  Là tất cả các giá trị *x* mà *X* có thể nhận và xác suất *f(x)* tương ứng của chúng.

  - Phân phối rời rạc
  - Phân phối liên tục
- Đặc trưng bởi :
  - Hàm xác suất, hàm xác suất tích lũy
  - Kỳ vọng - Phương sai - Độ lệch chuẩn
- Mô hình xác suất = {biến ngẫu nhiên ; phân phối xác suất}

Kỳ vọng-Phương sai-Độ lệch chuẩn

- Kỳ vọng
  - Khái niệm: là giá trị trung bình sau khi lặp lại một thí nghiệm vô số lần.
  - Ký hiệu: E(X) hay μ
  - Định nghĩa:
    - Trường hợp biến ngẫu nhiên rời rạc

$$\mu = E(X) = \sum_{x} x f(x)$$

    - Trường hợp biến ngẫu nhiên liên tục

$$\mu = E(X) = \int_{x} x f(x)$$

Ví dụ

- Cho biến ngẫu nhiên X rời rạc có bảng mật độ xác suất như sau:

X=x	1	3	6
f(x)	1/6	2/6	3/6

- Kỳ vọng của X:

$$\mu = \sum x f(x) = 1\times 1/6 + 3\times 2/6 + 6\times 3/6 = 25/6$$

Kỳ vọng-Phương sai-Độ lệch chuẩn

- Phương sai
  - Khái niệm: là trung bình bình phương độ lệch so với kỳ vọng sau khi lặp lại một thí nghiệm vô số lần. Ký hiệu: σ^2 ,var(x), V(x)
  - Định nghĩa:
    - Trường hợp rời rạc

$$\sigma^2 = E[(X-\mu)^2] = \sum_x f(x)\times(x-\mu)^2$$
$$= E\left(X^2\right) - \mu^2 = \sum_x x^2 f(x) - \mu^2$$

    - Trường hợp liên tục

$$\sigma^2 = E[(X-\mu)^2] = \int_x f(x)\times(x-\mu)^2$$
$$= E\left(X^2\right) - \mu^2 = \int_x x^2 f(x) - \mu^2$$

- Độ lệch chuẩn

$$SD(X), \sigma$$

Ví dụ

- Cho biến ngẫu nhiên X rời rạc có bảng mật độ xác suất như sau:

X=x	1	3	6
f(x)	1/6	2/6	3/6

- Kỳ vọng của X: $\mu= 25/6$
- Phương sai của X: $\sigma^2=E(X^2)- \mu= \sum x^2 f(x)- \mu=1^2\times 1/6+3^2\times 2/6+6^2\times 3/6-(25/6)^2=137/36$
- Độ lệch chuẩn của X: $\sigma=sqrt(137/36)$

Bài tập

- Trong buổi họp chi đoàn, mọi người bỏ phiếu ủng hộ/ không ủng hộ An làm bí thư. Thực hiện thí nghiệm đếm số phiếu ủng hộ. Biết số đoàn viên tham gia bỏ phiếu là 5.

  X = số phiếu ủng hộ

  Xác định hàm xác suất (pmf), hàm xác suất tích lũy (cdf)

  Tính xác suất An không được ai ủng hộ.

  Tính xác suất An được 1 người ủng hộ.

  Tính xác suất An được ≤ 2 người ủng hộ.

Bài tập

- Trong kỳ thi tuyển sinh đại học năm 2010, khối A. Giả sử thang điểm làm tròn đến 1.

  X = tổng điểm thi đại học môn toán, lý khối A

  - X là biến ngẫu nhiên rời rạc hay liên tục?
  - Xác định Pr(1)
  - Tính F(3)
  - Tính xác suất 1 học sinh đạt tổng điểm 2 môn thuộc khoảng [1,3]

Phân phối đều rời rạc

- Định nghĩa: Biến X có phân phối đều rời rạc (discrete uniform distribution) nếu nó thỏa hai điều kiện sau:
  - X có thể nhận các giá trị nguyên trong đoạn [a, b].
  - Các giá trị mà X có thể nhận có xác suất bằng nhau.

Ví dụ

- Thảy xúc xắc. X=điểm đạt được khi thảy xúc xắc.
- $X \in \{1,2,3,4,5,6\}$ hay $X \in [1,6]$
- $Pr(a)=1/6$, với mọi $a \in [1,6]$
- X có phân phối đều rời rạc

Phân phối đều rời rạc

- Đặc trưng
  - Hàm xác suất (Pmf)

$$f(x) = \frac{1}{b-a+1}, \ (a \le x \le b).$$

  - Hàm phân phối tích lũy (Cdf)

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a+1}{b-a+1}, & a \le x < b. \\ 1, & x \ge b \end{cases}$$

  - Kỳ vọng $\mu = \dfrac{b+a}{2}$
  - Phương sai $\sigma^2 = \dfrac{(b-a+2)(b-a)}{12}.$
  - Độ lệch chuẩn $\sigma = \sqrt{\dfrac{(b-a+2)(b-a)}{12}}.$

Ví dụ trên

- Tung xúc xắc. X=điểm đạt được khi thảy xúc xắc.
- X có phân phối đều rời rạc

$f(3)=1/(6-1+1)=1/6$

$F(3)=(3-1+1)/(6-1+1)=3/6$

$\mu=(1+6)/2=3.5$

$\sigma^2=(6-1+2)(6-1)/12=35/12$

Bài tập

- Xác định phân phối xác suất trong ví dụ tung đồng xu với X = kết quả tung được.
  - Hàm xác suất
  - Hàm phân phối tích lũy
  - Kỳ vọng
  - Phương sai
  - Độ lệch chuẩn

Ví dụ (tt)

A: là tên cướp

B: bị máy phát hiện nói dối báo động

$Pr(A) = 2/40$

$Pr(B|A) = 0.85$, $Pr(B|A^c) = 0.08$

Tính $Pr(A|B)$

Phân phối nhị thức

- Định nghĩa: Biến X có phân phối nhị thức nếu nó thỏa các điều kiện sau:
  - Số lần thí nghiệm của tiến trình ngẫu nhiên đang xét là cố định
  - Hậu quả của thí nghiệm chỉ có thể được phân thành 2 lớp (thành công hay thất bại)
  - Xác suất thành công trong mọi lần thí nghiệm là như nhau
  - Các lần thí nghiệm là độc lập nhau
  - X = số lần thí nghiệm thành công trong n lần thí nghiệm

Ví dụ

- Bài kiểm tra: 5 câu trắc nghiệm, mỗi câu có 4 lựa chọn a, b, c, d
- Vì chưa học bài, bạn An chọn ngẫu nhiên kết quả cho từng câu trắc nghiệm.
- Gọi X=số câu bạn An trả lời đúng
- X có phân phối nhị thức không?
  - Số lần thí nghiệm: 5
  - Xác suất thành công mỗi lần: ¼
  - Các lần thí nghiệm: độc lập
  - Kết quả mỗi lần thí nghiệm: đúng/ sai

Phân phối nhị thức

- Đặc trưng
  - Hàm xác suất $$f(x) = C_n^x p^x (1-p)^{n-x}$$
  - Hàm phân phối tích lũy $$F(a) = \sum_{x \le a} f(x)$$
  - Kỳ vọng $$E(X) = n \times p$$
  - Phương sai $$\sigma^2 = np(1-p)$$
  - Độ lệch chuẩn $$\sigma = \sqrt{np(1-p)}$$

Bài tập

- 3 ngã tư đèn xanh đèn đỏ
- xác suất đèn đỏ bật: p=0.7
- Các đèn bật/ tắt độc lập nhau
- Hậu quả: {đèn đỏ bật – thành công, đèn đỏ tắt - thất bại}

Tính xác suất gặp đèn đỏ ít nhất 1 lần?

X=số lần thành công

n =3

$$\Pr(0) = C_3^0 0.7^0 \left(1-0.7\right)^{3-0} = 0.027$$

Pr(x≥1) = 1-P(0)=1-0.027=0.073

Tính kỳ vọng: μ=np=3×0.7=2.1

Tính phương sai: σ^2=np(1-p)=3×0.7× (1-0.7)=0.63

Bài tập: Tính xác suất gặp đèn đỏ từ 1 đến 2 lần?

Bài tập

- Giả sử “70% người bị ung thư phổi là người hút thuốc trong thời gian dài” là đúng
  - Tìm xác suất trong 5 người nhập viện gần đây vì ung thư phổi, có ít hơn 1 nửa là những người hút thuốc lá trong thời gian dài.
- Giả sử xác suất bình phục là 0.8 và các ca hồi phục độc lập nhau.
  - Tìm xác suất 7 trong 10 người sẽ bình phục

Bài tập

- Kiểm tra hàng nhập kho: sẽ trả về nếu như >10% hàng nhập kho bị lỗi. Thực hiện lấy ngẫu nhiên 10 sản phẩm để kiểm tra, xác suất để quyết định trả hàng về là bao nhiêu. Biết xác suất lỗi của 1 sản phẩm là như nhau và bằng 0.1.

Tóm tắt

- Tóm tắt: Biến ngẫu nhiên, Phân phối xác suất, Phân phối đều rời rạc và phân phối nhị thức
- Từ khóa:
  - Biến ngẫu nhiên (random variable), rời rạc (discrete), liên tục (continuous)
  - Hàm độ lớn xác suất (pms – probability mass function), Hàm phân phối tích lũy (cdf – cumulative distribution function)
  - Kỳ vọng (expected value), Phương sai (variance), Độ lệch chuẩn (standard deviation - SD)
  - Phân phối xác suất (probability distribution), Mô hình xác suất (probability model)
  - Phân phối đều rời rạc (uniform distribution), Phân phối nhị thức (binomial distribution)