Chương

05

PHÂN PHỐI XÁC SUẤT RỜI RẠC

DẪN NHẬP

- Để có một cái nhìn tổng thể về phép thử, ta sẽ xem xét khả năng xảy ra của tất cả các kết quả của phép thử: làm sao biết các kết quả nào hay có khả năng xảy ra? Làm sao biết kết quả nào là không thường xảy ra? Trong chương này chúng ta sẽ tìm hiểu khái niệm *phân phối xác suất* để giải quyết vấn đề này.

- Chúng ta sẽ dùng các kiến thức về *thống kê mô tả* ở chương 2, chương 3 và một vài kiến thức về *xác suất* ở chương 4 để có thể mô tả và phân tích *phân phối xác suất*

DẪN NHẬP

- Để có thể nắm vững được khái niệm phân phối xác xuất, trước hết ta sẽ tìm hiểu khái niệm về *biến ngẫu nhiên*. Sau đó, chúng ta cần phân biệt được hai loại *biến ngẫu nhiên rời rạc* và *biến ngẫu nhiên liên tục*.
- Trong nội dung chương này chúng ta sẽ tìm hiểu phân phối của biến ngẫu nhiên rời rạc thông qua hai loại phân phối là *phân phối nhị thức* và *phân phối Poisson*
- Trong chương sau chúng ta sẽ tìm hiểu về phân phối của biến ngẫu nhiên liên tục thông qua *phân phối chuẩn*

NỘI DUNG

- Biến ngẫu nhiên
- Phân phối xác suất
- Đặc trưng của biến ngẫu nhiên
- Một số phân phối xác xuất rời rạc
  - Phân phối Nhị thức
  - Phân phối Poisson

NỘI DUNG

- **Biến ngẫu nhiên**
- Phân phối xác suất
- Đặc trưng của biến ngẫu nhiên
- Một số phân phối xác xuất rời rạc
  - Phân phối Nhị thức
  - Phân phối Poisson

BIẾN NGẪU NHIÊN

- *Biến ngẫu nhiên* là biến mà mỗi giá trị nó mang là một con số đại diện cho mỗi kết quả của phép thử.

- Ký hiệu: thường dùng các chữ cái in hoa như X, Y, Z…

BIẾN NGẪU NHIÊN

Ví dụ: Thực hiện phép thử tung một xúc xắc. Không gian mẫu của phép thử của phép thử:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Gọi biến ngẫu nhiên X là nút xuất hiện.

X có thể nhận các giá trị: {1, 2, 3, 4, 5, 6}

BIẾN NGẪU NHIÊN

Ví dụ: Thực hiện phép thử tung hai đồng xu cân đối đồng chất. Ta có, không gian mẫu của phép thử:

$$\Omega = \{SS, SN, NS, NN\}$$

Gọi biến ngẫu nhiên Y là số mặt ngửa xuất hiện.

Y có thể nhận các giá trị: $\{0, 1, 2\}$

PHÂN LOẠI BIẾN NGẪU NHIÊN

- Biến ngẫu nhiên được gọi là *rời rạc* nếu các giá trị mà nó nhận có thể đếm được.

Ví dụ: số bài kiểm tra trong một học kỳ? Số tin nhắn nhận được trong một ngày?...

- Biến ngẫu nhiên được gọi là *liên tục* nếu các giá trị mà nó nhận là giá trị liên tục, có thể lấp đầy một khoảng trên trục số

Ví dụ: chiều cao của sinh viên CNTT? Thu nhập sau 5 năm đi làm?...

NỘI DUNG

- Biến ngẫu nhiên
- **Phân phối xác suất**
- Đặc trưng của biến ngẫu nhiên
- Một số phân phối xác xuất rời rạc
  - Phân phối Nhị thức
  - Phân phối Poisson

HÀM XÁC SUẤT CỦA BIẾN NGẪU NHIÊN RỜI RẠC

- Đối với biến ngẫu nhiên rời rạc: **hàm độ lớn xác suất** (**pmf** – probability mass function) của một biến ngẫu nhiên rời rạc là hàm thể hiện xác suất của biến ngẫu nhiên đó ứng với một giá trị cụ thể.

Ký hiệu: **f(x)** hay P(X = x) với X là biến ngẫu nhiên, x các giá trị mà biến X có thể nhận được.

Ví dụ:

- P (X = 2) = f(2): xác suất của biến ngẫu nhiên X nhận giá trị 2

HÀM XÁC SUẤT CỦA BIẾN NGẪU NHIÊN RỜI RẠC

- Tính chất:
  - $0 \leq f(x) \leq 1$, $\forall x$ là giá trị X có thể nhận.
  - $\sum f(x) = 1$ Với $f(x) = P(X = x)$

- ***Ví dụ:*** gọi X là biến ngẫu nhiên khi tung xúc sắc. Ta có $f(1) = f(2) = f(3) = \ldots = f(6) = 1/6$

HÀM XÁC SUẤT CỦA BIẾN NGẪU NHIÊN LIÊN TỤC

- Đối với biến ngẫu nhiên liên tục: **hàm mật độ xác suất** (**pdf** – probability density function) của một biến ngẫu nhiên liên tục là hàm thể hiện xác suất của biến ngẫu nhiên đó ứng với một giá trị cụ thể.

- Ký hiệu: f(x) hay P(X = x) với X là biến ngẫu nhiên, x các giá trị mà biến X có thể nhận được.

HÀM XÁC SUẤT CỦA BIẾN NGẪU NHIÊN LIÊN TỤC

- Tính chất:
  - $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
  - $P(a \leq X \leq b) = \int_a^b f(x)dx$: diện tích vùng bên dưới đường cong f(x) với x chạy từ a đến b
- Sử dụng để tính xác suất:
  - $P(X=a) = 0$
  - $P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$

PHÂN PHỐI XÁC SUẤT CỦA BIẾN NGẪU NHIÊN

- *Phân phối xác suất* của biến ngẫu nhiên là một mô tả đầy đủ về giá trị có thể nhận được của biến ngẫu nhiên với xác suất tương ứng với giá trị đó.

- Nói cách khác, phân phối xác suất cho biến ngẫu nhiên X là tất cả các giá trị x mà X có thể nhận được và xác suất f(x) hay P(X = x) tương ứng của nó

- Phân phối xác suất của biến ngẫu nhiên có thể biểu diễn bằng cách lập bảng, vẽ đồ thị hoặc biểu diễn bằng công thức toán học.

BẢNG PHÂN PHỐI XÁC SUẤT

- Đối với biến ngẫu nhiên rời rạc, ta có thể biểu diễn phân phối xác suất thông qua bảng phân phối xác suất.
- Bản phân phối xác suất

Ký hiệu:

$X = x_i$: BNN X nhận giá trị x_i.

$p_i = P(X = x_i)$: Xác suất để X nhận giá trị x_i.

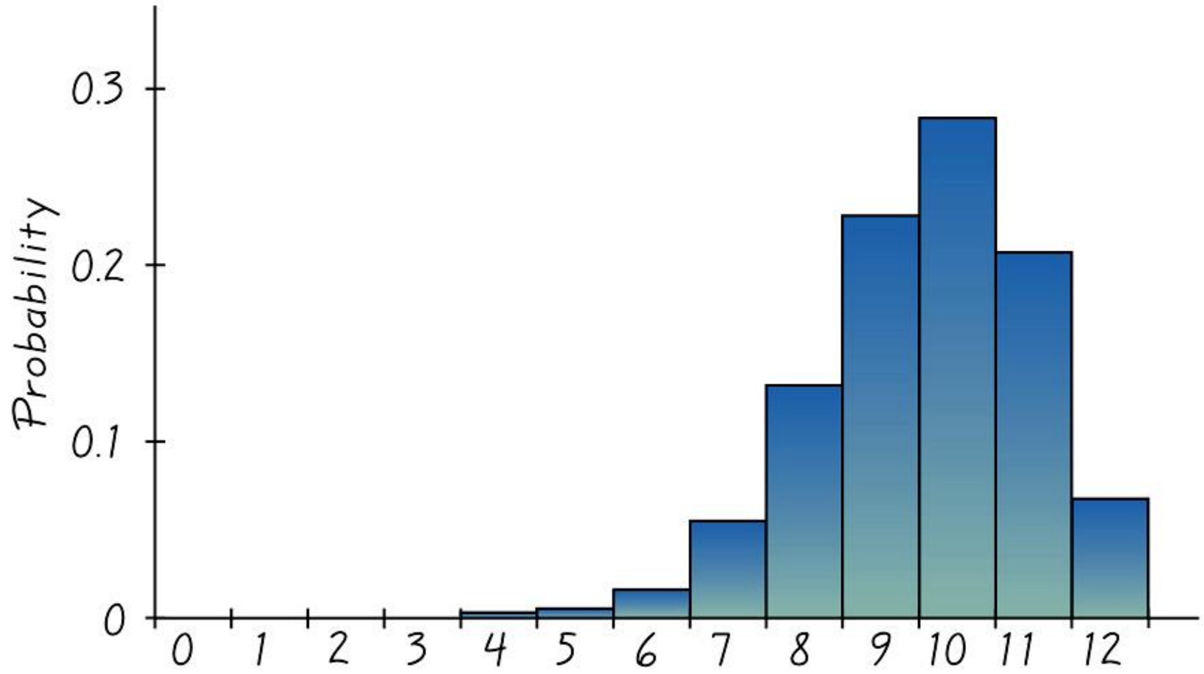
Giả sử $X = \{x_1, x_2, ..., x_n\} \quad (x_1 < x_2 < ... < x_n)$.

Bảng phân phối xác suất của X:

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

ĐỒ THỊ PHÂN PHỐI XÁC SUẤT

- Đồ thị phân phối xác suất giống với đồ thị phân phối tần số tương đối (với trục hoành là giá trị của biến ngẫu nhiên, trục tung là giá trị xác suất)



Probability Histogram for Number of Mexican-American Jurors Among 12

VÍ DỤ

- Gọi X là biến ngẫu nhiên được khi tung xúc sắc
- Bảng phân phối xác suất của X:

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(x) = P(X <= x)	F(1)= 1/6	F(2)= 2/6	F(3)= 3/6	F(4)= 4/6	F(5)= 5/6	F(6)=1

HÀM PHÂN PHỐI TÍCH LŨY

- Hàm phân phối tích lũy (**cdf** – culmulative densitve function) là hàm F(x) được định nghĩa:

$$F(x) = P(X \le x)$$

NỘI DUNG

- Biến ngẫu nhiên
- Phân phối xác suất
- **Đặc trưng của biến ngẫu nhiên**
- Một số phân phối xác xuất rời rạc
  - Phân phối Nhị thức
  - Phân phối Poisson

KỲ VỌNG

- **Kỳ vọng** (*expected value*): là giá trị trung bình sau khi lặp lại một thí nghiệm **vô số lần**
- Ký hiệu: E(X) hoặc μ
- Công thức:
  - 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

- Hãy tính giá trị kỳ vọng của X?

PHƯƠNG SAI

- **Phương sai** (*variance*): là trung bình của tổng bình phương độ lệch của tất cả giá trị của biến ngẫu nhiên so với giá trị kỳ vọng
- Ký hiệu: σ^2 ,var(x), V(x)
- Công thức:
  - Trường hợp biến ngẫu nhiên 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 biến ngẫu nhiên 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

- **Độ lệch chuẩn** (*standard variation*): là căn bậc 2 của giá trị phương sai
- Ký hiệu: σ hoặc SD(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

- Hãy tính giá trị phương sai và độ lệch chuẩn của X?

NỘI DUNG

- Biến ngẫu nhiên
- Phân phối xác suất
- Đặc trưng của biến ngẫu nhiên
- **Một số phân phối xác xuất rời rạc**
  - Phân phối Nhị thức
  - Phân phối Poisson

NỘI DUNG

- Biến ngẫu nhiên
- Phân phối xác suất
- Đặc trưng của biến ngẫu nhiên
- **Một số phân phối xác xuất rời rạc**
  - **Phân phối Nhị thức**
  - Phân phối Poisson

PHÂN PHỐI NHỊ THỨC

- Xét biến cố A liên quan đến một phép thử.
- Giả sử xác suất để A xảy ra là: $P(A) = p$
- Thực hiện phép thử trên n lần *độc lập nhau*.
- Gọi X là biến ngẫu nhiên thể hiện số lần A xảy ra.
  $(X = 0, 1,2, \ldots, n)$
- Khi đó, ta nói X có phân phối nhị thức (Binomial Distribution)
- Ký hiệu: **$X \sim B(n, p)$**

VÍ DỤ

- Tung một đồng xu cân đối đồng chất 10 lần độc lập nhau. Hỏi khả năng xảy ra mặt sấp 3 lần là bao nhiêu?
- Ta có:
  - Đồng xu cân đối, đồng chất nên khả năng xuất hiện mặt sấp trong 1 lần tung đồng xu là 0.5
  - Nếu gọi X là biến ngẫu nhiên thể hiện số lần mặt sấp xuất hiện thì vời điều kiện 10 lần tung là độc lập nhau ta có thể nói X có phân phối nhị thức hay X ~ B(10, 0.5)

ĐIỀU KIỆN ĐỂ CÓ PHÂN PHỐI NHỊ THỨC

- Một phân phối là phân phối nhị thức khi:
  - Số lần thực hiện thử nghiệm ngẫu nhiên là hữu hạn
  - Kết quả của thử nghiệm được phân thành hai lớp (ví dụ: thành công hoặc thất bại)
  - Xác suất thành công trong mọi lần thử nghiệm là như nhau
  - Các thử nghiệm đều độc lập nhau

HÀM XÁC SUẤT CỦA PHÂN PHỐI NHỊ THỨC

- Nếu X ~ B(n, p). Ta có:

$$\boxed{P\left(X=k\right)=C_n^k p^k q^{n-k}}$$

(với q = 1 – p)

VÍ DỤ

- Giả sử xác suất một người trưởng thành biết đến Twitter là 0.85 . Hỏi nếu tìm ngẫu nhiên 5 người thì xác suất tìm được 3 người biết đến Twitter là bao nhiêu? Theo bạn số người biết đến Twitter có phải có phân phối Nhị Thức hay không?

CÁC THAM SỐ ĐẶC TRƯNG CỦA PP NHỊ THỨC

Nếu X ~ B(n, p) (hay X có phân phối nhị thức). Ta có:

- Kỳ vọng:

$$\boxed{\mu = n \cdot p}$$

- Phương sai:

$$\boxed{\sigma^2 = n \cdot p \cdot q}$$

- Độ lệch chuẩn:

$$\boxed{\sigma = \sqrt{n \cdot p \cdot q}}$$

VÍ DỤ

- Tỷ lệ hỏng trong một lô thuốc là p = 0.2 . Chọn ngẫu nhiên 5 lọ. Hãy thực hiện:
  a) Tính xác suất chọn được 1 lọ hỏng.
  b) Lập bảng phân phối xác xuất đối với số lọ hỏng.
  c) Tính kỳ vọng và độ lệch chuẩn đối với số lọ hỏng lấy được

NỘI DUNG

- Biến ngẫu nhiên
- Phân phối xác suất
- Đặc trưng của biến ngẫu nhiên
- **Một số phân phối xác xuất rời rạc**
  - Phân phối Nhị thức
  - **Phân phối Poisson**

PHÂN PHỐI POISSON

- Phân phối Poisson là một phân phối xác suất của biến ngẫu nhiên rời rạc, thường được dùng để mô tả khả năng xuất hiện của các biến cố hiếm (khả năng xuất hiện nhỏ)
- Một biến ngẫu nhiên X, được gọi là có phân phối Poisson nếu X mô tả số lần xuất hiện của một biến cố trong một khoảng (thời gian, không gian, diện tích, khối lượng hoặc một số đơn vị tương tự)
- Ký hiệu: **$X \sim P(\lambda)$**, với λ là số lần xuất hiện của biến cố
- *Ví dụ:*
  – Số cuộc gọi đến tổng đài trong thời gian 1 phút.
  – Số lỗi in sai trên một trang sách.

ĐIỀU KIỆN ĐỂ CÓ PHÂN PHỐI POISSON

- Một biến ngẫu nhiên X mô tả số lần xuất hiện của một biến cố, được gọi là có phân phối Poisson nếu:
  - Số lần xuất hiện của biến cố trong cùng một khoảng (thời gian, không gian…) là không đổi
  - Sự xuất hiện của biến cố trong khoảng (thời gian, không gian…) này không ảnh hưởng đến sự xuất hiện của biến cố đó trong khoảng (thời gian, không gian…) sau.

HÀM XÁC SUẤT CỦA PHÂN PHỐI POISSON

➢ Nếu **X ~ P(λ)**. Ta có:

$$P(X = k) = e^{-\lambda}.\frac{\lambda^k}{k!}$$

(với e = 2.71828)

VÍ DỤ

- Trong một nhà máy dệt, biết số ống sợi bị đứt trong 1 giờ có phân phối Poisson với trung bình là 4. Tính xác suất trong 1 giờ có 3 ống sợ bị đứt.

CÁC THAM SỐ ĐẶC TRƯNG CỦA PP POISSON

Nếu X ~ P(λ) (hay X có phân phối Poisson). Ta có:

$$\mu = \sigma^2 = \lambda$$

VÍ DỤ

- Theo thống kê trong khoảng thời gian 100 năm, có 530 cơn bão đến từ Đại Tây Dương. Giả sử số cơn bão trong một năm có phân phối Poisson. Hãy tính
  a) Số cơn bão trung bình mỗi năm.
  b) Tính xác suất một năm có hai cơn bão đến từ Đại Tây Dương

SỰ KHÁC BIỆT GIỮA PP NHỊ THỨC VÀ PP POISSON

- PP Nhị thức ảnh hưởng bởi hai tham số là n và p. Trong khi, phân phối Poisson chỉ bị ảnh hưởng bởi tham số λ

- Trong phân phối nhị thức, giá trị tối đa mà biến ngẫu nhiên có thể nhận là n. Còn đối với pp Poisson là không có giới hạn trên.

XẤP XỈ PP NHỊ THỨC BẰNG PP POISSON

- Đối với phân phối nhị thức, khi n rất lớn việc tính toán hàm xác suất gặp nhiều khó khăn. Khi đó, ta có thể dùng phân phối Poisson để tính.
- Nếu $X \sim B(n, p)$ và $n.p \geq 100$ và $n.p \leq 10$. Ta có thể xấp xỉ $X \sim P(\lambda)$ với $\lambda = n.p$

VÍ DỤ

- Trong một lô thuốc, tỉ lệ thuốc hỏng là 0.003 . Kiểm tra 1000 ống:
  a) Tính xác suất gặp 4 ống hỏng?
  b) Tính xác suất gặp 60 ống hỏng?

HỎI ĐÁP