HCMUS
Viet Nam National University
Ho Chi Minh City
University of Science

Khoa Toán - Tin học
Fac. of Math. & Computer Science

BỔ SUNG KIẾN THỨC PHÂN PHỐI XÁC SUẤT

Lưu Giang Nam

Bộ môn Ứng dụng Tin học
Khoa Toán - Tin học
Trường Đại học KHTN, ĐHQG TPHCM

10/2025

Mục lục

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Phần 1

Phân phối xác suất

Phần 1
Phân phối xác suất

Mục 1: Biến ngẫu nhiên

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Biến ngẫu nhiên

Biến ngẫu nhiên là một biến mà mỗi giá trị của nó tương ứng với một con số đại diện cho mỗi kết quả có thể xảy ra của một phép thử ngẫu nhiên.

- **Phép thử ngẫu nhiên**: Là quá trình hoặc thí nghiệm mà kết quả không thể dự đoán trước một cách chắc chắn, ví dụ: gieo một con xúc xắc, tung một đồng xu.
- **Ký hiệu biến ngẫu nhiên**: Thường được biểu diễn bằng các chữ cái in hoa như **X**, **Y**, **Z**.
- Ví dụ: Xét phép thử tung một xúc xắc. Không gian mẫu của phép thử này là: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Tức là mỗi lần tung xúc xắc sẽ có một trong sáu kết quả có thể xảy ra, tương ứng với các mặt của xúc xắc.
  Gọi biến ngẫu nhiên X là nút xuất hiện khi tung xúc xắc. Biến ngẫu nhiên này có thể nhận các giá trị là các số nguyên từ 1 đến 6, tức là: $X \in \{1, 2, 3, 4, 5, 6\}$.

Ví dụ: Phép thử tung hai đồng xu

Xét phép thử tung hai đồng xu cân đối đồng chất. Không gian mẫu của phép thử này là:

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

Trong đó: - S đại diện cho mặt sấp của đồng xu. - N đại diện cho mặt ngửa của đồng xu.
Gọi biến ngẫu nhiên Y là số mặt ngửa xuất hiện. Biến ngẫu nhiên Y có thể nhận các giá trị trong tập hợp:

$$Y \in \{0, 1, 2\}$$

Tức là, $Y = 0$ khi không có mặt ngửa nào xuất hiện, $Y = 1$ khi có một mặt ngửa xuất hiện và $Y = 2$ khi cả hai mặt đều ngửa.

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Phân loại biến ngẫu nhiên

- Biến ngẫu nhiên được gọi là rời rạc nếu các giá trị mà nó nhận có thể đếm được. Điều này có nghĩa là số lượng các giá trị mà biến ngẫu nhiên có thể nhận là hữu hạn hoặc đếm được vô hạn.
- Biến ngẫu nhiên được gọi là liên tục nếu các giá trị mà nó nhận là các giá trị liên tục, có thể lấp đầy một khoảng trên trục số thực.

Ví dụ:

- Số bài kiểm tra trong một học kỳ -> Rời rạc
- Số người tham gia một cuộc khảo sát -> Rời rạc
- Thu nhập sau 5 năm đi làm -> Liên tục
- Thời gian một chiếc xe di chuyển từ điểm A đến điểm B -> Liên tục

Phần 1

Phân phối xác suất

Mục 2: Biến ngẫu nhiên rời rạc

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. Hàm xác suất $f(x)$ cho biến ngẫu nhiên X được định nghĩa như sau:

$$P(X = x) = f(x) \tag{1}$$

Ví dụ: Giả sử X là biến ngẫu nhiên biểu diễn số mặt ngửa khi tung một đồng xu. Nếu X có thể nhận giá trị 0 hoặc 1, hàm xác suất $f(x)$ của X có thể được thể hiện như sau:

$$P(X = 0) = f(0) = 0.5, \quad P(X = 1) = f(1) = 0.5 \tag{2}$$

Còn $P(X = 2) = f(2) = 0$ là xác suất của biến ngẫu nhiên X nhận giá trị 2.

Hàm xác suất của biến ngẫu nhiên rời rạc

Tính chất:

- $0 \leq f(x) \leq 1$, với $\forall x$ là giá trị mà biến ngẫu nhiên X có thể nhận.
- $\sum_{x} f(x) = 1$, với $f(x) = P(X = x)$, tức là tổng các xác suất của tất cả các giá trị có thể xảy ra của biến ngẫu nhiên X phải bằng 1.

Ví dụ: Gọi X là biến ngẫu nhiên khi tung một xúc xắc. Ta có:

$$f(1) = f(2) = f(3) = f(4) = f(5) = f(6) = \frac{1}{6}$$

Điều này có nghĩa là xác suất để xúc xắc hiện ra bất kỳ số nào từ 1 đến 6 đều là $\frac{1}{6}$.

Phần 1

Phân phối xác suất

Mục 3: Biến ngẫu nhiên liên tục

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.
- Tuy nhiên, khác với biến ngẫu nhiên rời rạc, biến ngẫu nhiên liên tục có thể nhận vô số giá trị trong một khoảng liên tục.
- **Ký hiệu:** Hàm mật độ xác suất được ký hiệu là $f(x)$ hoặc $P(X = x)$, với:
  - X là biến ngẫu nhiên.
  - x là các giá trị mà biến X có thể nhận được.

Hàm xác suất của biến ngẫu nhiên liên tục

- $P(X = x) = 0$, tức là xác suất để X nhận một giá trị cụ thể là không thể xảy ra.
- Để tính xác suất trong một khoảng $[a, b]$, ta sử dụng tích phân của hàm mật độ xác suất:

$$P(a \leq X \leq b) = \int_a^b f(x)\,dx \tag{3}$$

- $f(x) \geq 0$ và $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$
- $P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$: Vì hàm mật độ xác suất là liên tục, nên các dạng dấu bất đẳng thức trên đều có giá trị xác suất như nhau.

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Phần 1

Phân phối xác suất

Mục 4: Phân phối xác suất

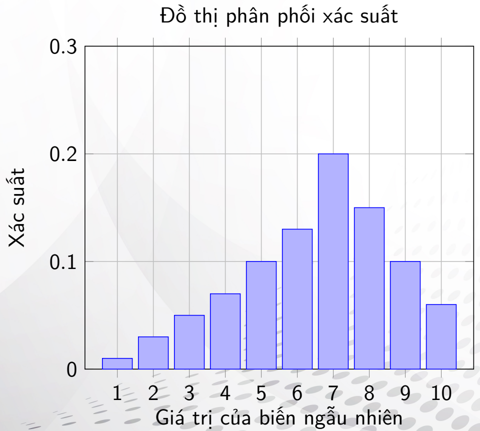
Phân phối xác suất của biến ngẫu nhiên

- 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.
- Đối với biến ngẫu nhiên rời rạc, ta có thể biểu diễn phân phối xác suất thông qua bảng phân phối xác suất.
- Bảng phân phối xác suất là một bảng liệt kê tất cả các giá trị mà biến ngẫu nhiên có thể nhận được cùng với xác suất tương ứng của từng giá trị.

Giá trị	**Xác suất**
x_1	$P(X = x_1)$
x_2	$P(X = x_2)$
$\vdots$	$\vdots$
x_n	$P(X = x_n)$

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 và trục tung là giá trị xác suất.



Hàm phân phối tích lũy

- Hàm phân phối tích lũy (CDF - Cumulative Distribution Function) là hàm $F(x)$ được định nghĩa như sau:

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

- Hàm phân phối tích lũy cho biết xác suất mà biến ngẫu nhiên X nhận giá trị nhỏ hơn hoặc bằng x.
- Hàm phân phối tích lũy là một hàm không giảm và có giá trị trong khoảng từ 0 đến 1.

Ví dụ: Biến ngẫu nhiên khi tung xúc xắc

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

Giá trị x	1	2	3	4	5	6
Xác suất $P(X = x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- Hàm phân phối tích lũy $F(x)$ là xác suất $P(X \leq x)$, tức là:

$$F(x) = \begin{cases} 0 & \text{nếu } x < 1 \\ \frac{1}{6} & \text{nếu } x < 2 \\ \frac{2}{6} & \text{nếu } x < 3 \\ \frac{3}{6} & \text{nếu } x < 4 \\ \frac{4}{6} & \text{nếu } x < 5 \\ \frac{5}{6} & \text{nếu } x < 6 \\ 1 & \text{nếu } x \leq 6 \end{cases}$$

Phần 1

Phân phối xác suất

Mục 5: ĐẶC TRƯNG CỦA BIẾN NGẪU NHIÊN

Kỳ vọng

- **Kỳ vọng (Expectation)** 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:

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

  - Trường hợp biến ngẫu nhiên liên tục:

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

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, $\mathrm{var}(X)$, $V(X)$

Công thức:

- Trường hợp biến ngẫu nhiên rời rạc:

$$\sigma^2 = \mathrm{var}(X) = \sum_x (x - E(X))^2 f(x)$$

- Trường hợp biến ngẫu nhiên liên tục:

$$\sigma^2 = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\, dx$$

Phần 2

Phân phối rời rạc

Phần 2

Phân phối rời rạc

Mục 1: Phân phối Nhị thức

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. Xác suất để mặt sấp xuất hiện trong mỗi lần tung là 0.5. Gọi X là biến ngẫu nhiên thể hiện số lần mặt sấp xuất hiện trong 10 lần tung. X có phân phối nhị thức:

$$X \sim B(10, 0.5)$$

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 $X \sim B(n, p)$:

$$P(X = k) = C_n^k \cdot p^k \cdot q^{n-k} \tag{4}$$

trong đó:

- p là xác suất thành công trong một lần thử nghiệm.
- $q = 1 - p$ là xác suất thất bại.

Các tham số đặc trưng

Nếu $X \sim B(n, p)$ (hay X có phân phối nhị thức), ta có các tham số đặc trưng sau:

- **Kỳ vọng:**

$$\mu = n \cdot p \tag{5}$$

- **Phương sai:**

$$\sigma^2 = n \cdot p \cdot q \quad \text{với} \quad q = 1 - p \tag{6}$$

- **Độ lệch chuẩn:**

$$\sigma = \sqrt{n \cdot p \cdot q} \tag{7}$$

Các tham số này giúp mô tả đặc điểm phân phối nhị thức trong các tình huống thử nghiệm với hai kết quả có xác suất thành công cố định.

Ví Dụ: Xác suất lọ hỏng

Tỷ lệ hỏng trong một lô thuốc là $p = 0.2$. Chọn ngẫu nhiên 5 lọ. Ta cần thực hiện các phần sau:

a) Tính xác suất chọn được 1 lọ hỏng.

b) Lập bảng phân phối xác suất đối với số lọ hỏng.

c) Tính kỳ vọng và độ lệch chuẩn đối với số lọ hỏng lấy được.

Giả sử X là số lọ hỏng.

Phần 2

Phân phối rời rạc

Mục 2: 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).
- Ký hiệu: $X \sim P(\lambda)$, với λ là số lần xuất hiện của biến cố trong khoảng thời gian hoặc không gian cho trước.
- Nếu $X \sim P(\lambda)$, ta có hàm xác suất:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots \tag{8}$$

với $e \approx 2.71828$ là cơ số tự nhiên.

Ví Dụ

Trong một nhà máy dệt, 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 sợi bị đứt.

- Ta có $\lambda = 4$ và $k = 3$.
- Áp dụng công thức phân phối Poisson:

$$P(X = 3) = \frac{\lambda^3 e^{-\lambda}}{3!} = \frac{4^3 e^{-4}}{3!}$$

- Tính giá trị:

$$P(X = 3) = \frac{64e^{-4}}{6} \approx \frac{64 \times 0.0183}{6} \approx 0.197$$

- Vậy, xác suất trong 1 giờ có 3 ống sợi bị đứt là khoảng 0.197.

Các Tham Số Đặc Trưng Của Phân Phối Poisson

- Nếu $X \sim P(\lambda)$, với λ là tham số của phân phối Poisson, thì các tham số đặc trưng của phân phối này là:
  - Kỳ vọng (mean):
  $$\mu = \lambda \tag{9}$$
  - Phương sai (variance):
  $$\sigma^2 = \lambda \tag{10}$$
  - Độ lệch chuẩn (standard deviation):
  $$\sigma = \sqrt{\lambda} \tag{11}$$
- Lưu ý: Với phân phối Poisson, kỳ vọng và phương sai đều bằng λ, và độ lệch chuẩn là căn bậc hai của λ.

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Phân Phối Nhị Thức và Phân Phối Poisson

- **Phân phối nhị thức (Binomial Distribution):**
  - Bị ảnh hưởng bởi hai tham số: n (số lần thử) và p (xác suất thành công trong mỗi lần thử).
  - Giá trị tối đa mà biến ngẫu nhiên có thể nhận là n, tức là không thể có quá n lần thành công.
- **Phân phối Poisson:**
  - Bị ảnh hưởng bởi một tham số duy nhất: λ (số lần xuất hiện của biến cố trong một khoảng thời gian, không gian,...).
  - Không có giới hạn trên cho giá trị của biến ngẫu nhiên, có thể có bất kỳ số lần xuất hiện nào.
- **Xấp xỉ phân phối nhị thức bằng phân phối Poisson:**
  - Khi n rất lớn, việc tính toán hàm xác suất của phân phối nhị thức gặp nhiều khó khăn. Lúc này, ta có thể sử dụng phân phối Poisson để xấp xỉ.
  - Nếu $X \sim B(n, p)$ và $n \cdot p \geq 10$, ta có thể xấp xỉ $X \sim P(\lambda)$ với $\lambda = n \cdot p$ (điều kiện là p gần 0).

Ví dụ 1: Nhị Thức vs Poisson

Trong một lô thuốc, tỉ lệ thuốc hỏng là 0.003. Kiểm tra 1000 ống, tính xác suất gặp 4 ống hỏng.

- **Cách 1: Sử dụng phân phối Nhị thức**

$$P(X=4) = C_{1000}^{4}(0.003)^4(1-0.003)^{996} \approx 0.171$$

- **Cách 2: Sử dụng phân phối Poisson**
  Khi n lớn và np nhỏ, ta có thể sử dụng phân phối Poisson với $\lambda = np = 1000 \times 0.003 = 3$.

$$P(X=4) = \frac{3^4 e^{-3}}{4!} \approx \frac{81 \times e^{-3}}{24} \approx 0.167$$

- **So sánh**
  Cả hai cách đều cho ra kết quả gần nhau, nhưng cách sử dụng phân phối Poisson đơn giản và nhanh hơn khi n rất lớn.

Ví dụ 2: Nhị Thức vs Poisson

Trong một lớp học có 30 học sinh, tỉ lệ học sinh đạt điểm cao là 0.2. Tính xác suất có 6 học sinh đạt điểm cao khi chọn ngẫu nhiên 10 học sinh.

- **Cách 1: Sử dụng phân phối nhị thức**

$$P(X = 6) = C_{10}^{6}(0.2)^6(1 - 0.2)^{10-6} \approx 0.024$$

- **Cách 2: Sử dụng phân phối Poisson**
  Với $n = 10$, $p = 0.2$, ta có $\lambda = n \times p = 10 \times 0.2 = 2$. Hàm xác suất phân phối Poisson:

$$P(X = 6) = \frac{2^6 e^{-2}}{6!} \approx 0.0025$$

Bài tập Phân phối rời rạc

1. Một nhà máy sản xuất có tỉ lệ sản phẩm đạt tiêu chuẩn là 0.95. Chọn ngẫu nhiên 20 sản phẩm.
   a. Tính xác suất có đúng 18 sản phẩm đạt tiêu chuẩn.
   b. Tính xác suất có ít nhất 17 sản phẩm đạt tiêu chuẩn.
   c. Sử dụng phân phối Poisson để ước lượng xác suất cho câu 1 (nếu áp dụng được).
2. Số lượng cuộc gọi đến một tổng đài trung bình là 4 cuộc mỗi phút.
   a. Tính xác suất trong một phút có đúng 6 cuộc gọi đến tổng đài.
   b. Tính xác suất trong 5 phút có ít hơn 15 cuộc gọi đến tổng đài.
   c. Xác định tham số của phân phối Poisson cho mỗi câu.
3. Trong một bài kiểm tra trắc nghiệm, xác suất một học sinh trả lời đúng một câu hỏi là 0.7. Chọn ngẫu nhiên 10 câu hỏi.
   a. Tính xác suất học sinh trả lời đúng 7 câu.
   b. Tính xác suất học sinh trả lời đúng từ 6 đến 8 câu.
   c. Sử dụng phân phối Poisson để xấp xỉ kết quả trong câu 1.

Bài tập Phân phối rời rạc

4. Trung bình có 3 sự cố xảy ra tại một trạm điện mỗi tháng.
   a) Tính xác suất trong một tháng có đúng 2 sự cố.
   b) Tính xác suất trong 3 tháng có tổng cộng 5 sự cố.
   c) Giả sử số sự cố xảy ra theo phân phối Poisson, hãy xác định tham số của phân phối cho mỗi trường hợp.
5. Trong một trận đấu bóng rổ, xác suất cầu thủ thực hiện thành công một cú ném là 0.6. Cầu thủ thực hiện 15 cú ném.
   a) Tính xác suất cầu thủ thực hiện thành công đúng 10 cú ném.
   b) Tính xác suất cầu thủ thực hiện thành công ít nhất 12 cú ném.
   c) Kiểm tra xem có thể sử dụng phân phối Poisson để xấp xỉ kết quả trong câu 1 hay không.

Phần 3

Phân phối liên tục

Phần 3

Phân phối liên tục

Mục 1: Phân phối chuẩn

Phân Phối Chuẩn

- Phân phối chuẩn là một mô hình xác suất phổ biến, đặc trưng bởi hai tham số:
  - Trung bình μ (Mean): xác định tâm của phân phối.
  - Phương sai σ^2 (Variance): xác định độ rộng của phân phối.
- Ký hiệu:

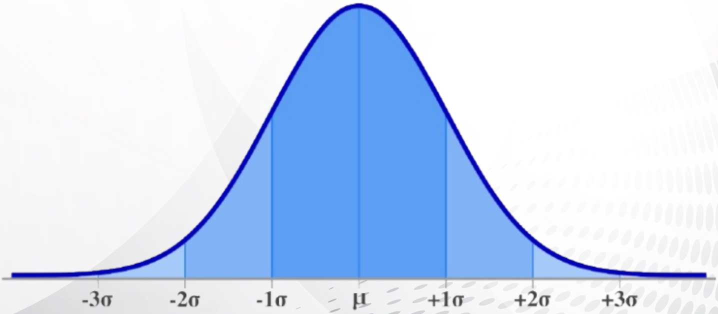
$$X \sim N(\mu, \sigma^2) \quad \text{hoặc} \quad N(\mu, \sigma). \tag{12}$$

- Hàm mật độ xác suất (*pdf*):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty. \tag{13}$$

Phân Phối Chuẩn

- Đặc điểm của đồ thị:
  - Hình dạng giống như một chiếc chuông (Bell-shaped curve).
  - Trung tâm tại μ, và độ rộng phụ thuộc vào σ.



Phân phối chuẩn chính tắc

- **Phân phối Z** là một trường hợp đặc biệt của phân phối chuẩn với:
  - Trung bình $\mu = 0$.
  - Phương sai $\sigma^2 = 1$.
- Phân phối Z còn được gọi là **Phân phối chuẩn chính tắc**.
- Hàm mật độ xác suất (*pdf*):

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, \quad -\infty < z < \infty. \tag{14}$$

Lưu ý

Phân phối Z thường được sử dụng để chuẩn hóa dữ liệu, giúp so sánh giá trị từ các phân phối khác nhau bằng cách chuyển chúng về dạng chuẩn.

Ký hiệu thường dùng

Ký hiệu thường dùng trong phân phối Z:

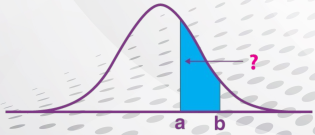
- $P(a \leq Z \leq b)$: biểu thị xác suất biến ngẫu nhiên Z nằm giữa a và b.
- $P(Z > a)$: biểu thị xác suất biến ngẫu nhiên Z lớn hơn a.
- $P(Z < a)$: biểu thị xác suất biến ngẫu nhiên Z nhỏ hơn a.

Lưu ý:

- Trong phân phối Z, các xác suất này được tính dựa trên bảng phân phối chuẩn hoặc bằng cách tích phân hàm mật độ xác suất $f(z)$.
- Tổng xác suất của tất cả các giá trị Z luôn bằng 1.

Tìm Giá Trị Xác Suất

- **Cách tính xác suất**:
  - Có thể tính diện tích (xác suất) của các vùng khác nhau của phân phối chuẩn chính tắc:
    - Sử dụng các công cụ có sẵn (phần mềm thống kê, máy tính).
    - Sử dụng **Bảng Z**.
  - Tra bảng Z khi cần tính xác suất tích lũy của phân phối chuẩn chính tắc, tức là $P(Z < a)$.
- **Xử lý các trường hợp khác**:
  - Nếu không phải xác suất tích lũy $P(Z < a)$, ta cần chuyển về xác suất tích lũy trước khi tra bảng Z:
    - $P(Z > a) = 1 - P(Z < a)$
    - $P(a < Z < b) = P(Z < b) - P(Z < a)$



Ví dụ: Xác suất mật độ Canxi trong xương

Giả sử z là kết quả mật độ Canxi trong xương của một người lớn được chọn ngẫu nhiên. z có phân phối chuẩn với trung bình $\mu = 0$ và độ lệch chuẩn $\sigma = 1$, tức là $z \sim N(0, 1)$. Tính xác suất của một người lớn có mật độ Canxi trong xương dưới 1.27.

- Ta cần tính $P(z < 1.27)$.
- Sử dụng bảng Z hoặc công cụ tính toán để tìm giá trị xác suất này.
- Tra bảng Z cho $z = 1.27$, ta có:

$$P(z < 1.27) \approx 0.8980$$

- Kết luận: Xác suất là khoảng **0.8980** hay **89.80%**.

Tra bảng Z

z	.00	.01	.02	.03	.04	.05	.06	.07
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292

Ví dụ: Tìm z khi biết xác suất

Bài toán: Sử dụng thử nghiệm mật độ Canxi trong xương, hãy tìm giá trị z sao cho:

$$P(z > Z_1) = 0.25 \quad \text{và} \quad P(z < Z_2) = 0.95$$

Hướng dẫn:

- Sử dụng bảng Z hoặc công cụ tính toán để tra giá trị z ứng với xác suất tích lũy.
- Đối với xác suất $P(z > Z_1) = 0.25$:
  - Diện tích bên trái của Z_1 là $1 - 0.25 = 0.75$.
  - Từ bảng Z, $P(z < Z_1) = 0.75$ tương ứng với $Z_1 = 0.674$.
- Đối với xác suất $P(z < Z_2) = 0.95$:
  - Từ bảng Z, $P(z < Z_2) = 0.95$ tương ứng với $Z_2 = 1.645$.

Kết luận:

- Giá trị $Z_1 = 0.674$ và $Z_2 = 1.645$.
- Xác suất $25\% < z < 95\%$ là vùng giữa Z_1 và Z_2.

Critical Value (z_α)

Định nghĩa:

- Đối với phân phối chuẩn, **Critical Value** là giá trị z mà tại đó phần diện tích bên phải z bằng một giá trị α.
- Ký hiệu: z_α.

Ví dụ:

- Nếu $\alpha = 0.025$, thì $z_{0.025} = 1.96$.
- Điều này có nghĩa là:

$$P(z > z_{0.025}) = 0.025$$

- Tương ứng, phần diện tích bên trái $z_{0.025}$ là:

$$P(z < z_{0.025}) = 1 - 0.025 = 0.975$$

Phần 3

Phân phối liên tục

Mục 2: Chuẩn hóa phân phối chuẩn

Chuẩn hóa Phân phối Chuẩn

Định nghĩa:

- Chuẩn hóa phân phối chuẩn $N(\mu, \sigma)$ là quá trình biến đổi phân phối chuẩn đã cho sang phân phối Z với:

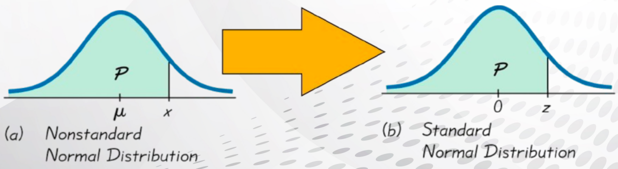
$$Z \sim N(0, 1), \quad Z = \frac{X - \mu}{\sigma}$$

  - X: Biến ngẫu nhiên ban đầu.
  - μ: Giá trị trung bình của X.
  - σ: Độ lệch chuẩn của X.
- **Lưu ý:**
  - Việc chuẩn hóa giúp sử dụng được bảng phân phối Z mà không ảnh hưởng đến các xác suất cần tính.
  - Chuẩn hóa không thay đổi bản chất của bài toán gốc.
  - Mọi xác suất tính được từ phân phối $N(\mu, \sigma)$ có thể tìm thông qua phân phối Z.

Ý nghĩa của Chuẩn hóa Phân phối Chuẩn

Ý nghĩa:

- Chuẩn hóa giúp đưa phân phối chuẩn bất kỳ $N(\mu, \sigma)$ về phân phối chuẩn chính tắc $Z \sim N(0,1)$, tạo sự thống nhất khi tính toán xác suất.
- Giúp tra bảng phân phối Z một cách dễ dàng để tính các xác suất liên quan.
- Đơn giản hóa việc so sánh các giá trị từ các phân phối khác nhau bằng cách đưa chúng về cùng một thang đo.



Bài tập: Chiều dài cá theo phân phối chuẩn

Bài toán: Chiều dài cá được mô hình hóa bằng phân phối chuẩn $N(\mu = 16 \text{ (cm)}, \sigma = 4 \text{ (cm)})$. Trả lời các câu hỏi sau:

1. Xác suất bắt được con cá nhỏ (nhỏ hơn 8 cm)?
2. Giả sử, ai bắt được con cá lớn (lớn hơn 24 cm) sẽ được thưởng. Hỏi xác suất được thưởng là bao nhiêu?
3. Xác suất bắt được con cá vừa (trong khoảng từ 16 cm đến 24 cm)?

Phương pháp giải:

1. Chuẩn hóa dữ liệu về phân phối chuẩn tắc $Z = \frac{X-\mu}{\sigma}$.
2. Tính các giá trị Z tương ứng cho mỗi câu hỏi.
3. Sử dụng bảng Z hoặc công cụ tính toán để tìm xác suất.

Bài tập: Thời gian hoàn thành dự án

Bài toán: Thời gian hoàn thành một dự án tuân theo phân phối chuẩn $N(\mu = 50 \text{ (ngày)}, \sigma = 5 \text{ (ngày)})$.

1. Xác suất hoàn thành dự án trong vòng dưới 45 ngày là bao nhiêu?
2. Xác suất hoàn thành dự án trong khoảng từ 45 đến 55 ngày là bao nhiêu?
3. Bao nhiêu phần trăm các dự án cần nhiều hơn 60 ngày để hoàn thành?

Bài tập: Điểm thi của sinh viên

Bài toán: Điểm thi của sinh viên trong một kỳ thi tuân theo phân phối chuẩn $N(\mu = 70, \sigma = 10)$.

1. Xác suất một sinh viên đạt điểm dưới 50 là bao nhiêu?
2. Bao nhiêu phần trăm sinh viên đạt điểm trong khoảng từ 60 đến 80?
3. Điểm số tối thiểu để một sinh viên nằm trong top 10% cao nhất là bao nhiêu?

Phần 3

Phân phối liên tục

Mục 3: Phân phối mẫu và công cụ ước lượng

Định nghĩa:

- Phân phối mẫu (*sampling distribution*) là phân phối của một giá trị thống kê (ví dụ: trung bình mẫu, tỷ lệ mẫu, phương sai mẫu) khi ta lấy tất cả các mẫu có thể có với kích thước n từ một quần thể.
- Phân phối mẫu cung cấp thông tin về hành vi của giá trị thống kê từ mẫu, giúp ta hiểu được tính biến động của thống kê khi mẫu thay đổi.

Ví dụ phổ biến:

- Phân phối mẫu của trung bình mẫu.
- Phân phối mẫu của tỷ lệ mẫu.

Phân phối mẫu của trung bình mẫu

Tính chất: Nếu quần thể có trung bình μ và độ lệch chuẩn σ, thì:

- Trung bình của phân phối mẫu: $\mu_{\bar{X}} = \mu$.
- Độ lệch chuẩn của phân phối mẫu (*Standard Error*):

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- Nếu quần thể tuân theo phân phối chuẩn hoặc n đủ lớn ($n \geq 30$), phân phối mẫu của trung bình mẫu cũng tuân theo phân phối chuẩn (**Định lý giới hạn trung tâm**).

Ý nghĩa:

- Giúp ta ước lượng trung bình quần thể từ mẫu.
- Giảm độ biến động của trung bình khi kích thước mẫu tăng.

Phân phối mẫu của tỷ lệ mẫu

Tính chất: Nếu tỷ lệ của quần thể là p, thì:

- Trung bình của phân phối mẫu: $\mu_{\hat{p}} = p$.
- Độ lệch chuẩn của phân phối mẫu (*Standard Error*):

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

- Nếu n đủ lớn, phân phối mẫu của tỷ lệ mẫu gần với phân phối chuẩn.

Ý nghĩa:

- Hỗ trợ việc kiểm định giả thuyết và xây dựng khoảng tin cậy cho tỷ lệ.

Bài tập 1

- Tuổi trung bình của học sinh trong một trường là 16.5 tuổi, với độ lệch chuẩn là 0.5 tuổi.
- Một mẫu ngẫu nhiên gồm 50 học sinh được chọn.
- Hỏi xác suất để tuổi trung bình mẫu nhỏ hơn 16.4 là bao nhiêu?

Bài tập 2

- Một nhà máy sản xuất có thời gian trung bình để hoàn thành một sản phẩm là 5.2 giờ, với độ lệch chuẩn 0.4 giờ.
- Một mẫu ngẫu nhiên gồm 36 sản phẩm được chọn.
- Hỏi xác suất để thời gian trung bình mẫu lớn hơn 5.4 giờ là bao nhiêu?

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Phần 3

Phân phối liên tục

Mục 4: XẤP XỈ PHÂN PHỐI NHỊ THỨC BẰNG PHÂN PHỐI CHUẨN

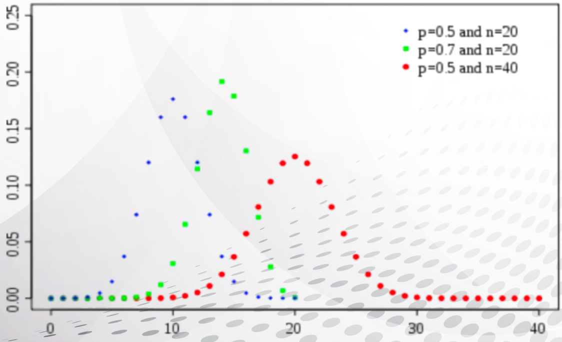
Xấp Xỉ Phân Phối Nhị Thức Bằng Phân Phối Chuẩn

- Khi số phép thử n rất lớn (ví dụ $n = 100$), việc tính toán xác suất chính xác theo phân phối nhị thức trở nên khó khăn.
- Phân phối chuẩn có thể được dùng để xấp xỉ phân phối nhị thức khi n lớn.
- Điều kiện áp dụng xấp xỉ:
  - $n \cdot p \geq 10$ và $n \cdot (1 - p) \geq 10$
- Khi đó, phân phối nhị thức $B(n, p)$ có thể được xấp xỉ bằng phân phối chuẩn $N(\mu, \sigma^2)$ với:

$$\mu = n \cdot p \quad \text{và} \quad \sigma = \sqrt{n \cdot p \cdot (1 - p)} \tag{15}$$

Xấp Xỉ Phân Phối Nhị Thức Bằng Phân Phối Chuẩn

- Khi số lần thí nghiệm n lớn, phân phối nhị thức $B(n, p)$ có thể xấp xỉ bằng phân phối chuẩn $N(\mu, \sigma^2)$.
- Nếu p gần 0.5, phân phối nhị thức sẽ có dạng chuông đối xứng, giống phân phối chuẩn.
- Nếu p gần 0 hoặc 1, phân phối nhị thức sẽ bị lệch sang trái (khi p gần 0) hoặc lệch sang phải (khi p gần 1).



Xấp Xỉ Phân Phối Nhị Thức Bằng Phân Phối Chuẩn

- Do phân phối nhị thức có dạng gần giống phân phối chuẩn khi n lớn, ta có thể sử dụng phân phối chuẩn để xấp xỉ phân phối nhị thức.
- Ba câu hỏi cần lưu ý:
  1. n thế nào là lớn?
     Thông thường, nếu $n \cdot p \geq 10$ và $n \cdot (1-p) \geq 10$, ta có thể sử dụng phân phối chuẩn để xấp xỉ.
  2. Dùng trung bình và độ lệch chuẩn nào để chuẩn hóa X về Z?
     Trung bình $\mu = n \cdot p$ và độ lệch chuẩn $\sigma = \sqrt{n \cdot p \cdot (1-p)}$.
  3. Phân phối nhị thức là rời rạc, còn phân phối chuẩn là liên tục. Làm thế nào để điều chỉnh cho liên tục?
     Chúng ta sử dụng điều chỉnh liên tục (continuity correction): tính xác suất trong khoảng từ $k-0.5$ đến $k+0.5$ thay vì tại k.

Quy Tắc Hiệu Chỉnh

- Biểu diễn bài toán dưới dạng xác suất, ví dụ:
  - $P(X < 235)$
  - $P(X \leq 235)$
  - $P(X \geq 235)$
  - $P(X > 235)$
  - $P(X = 235)$
- Nếu xác suất nhỏ hơn hoặc bằng, cộng 0.5 vào X trước khi tính xác suất:
  - Ví dụ: Nếu $P(X \leq 235)$, ta chuyển thành $P(X \leq 235.5)$.
- Nếu xác suất lớn hơn hoặc bằng, trừ 0.5 khỏi X:
  - Ví dụ: Nếu $P(X \geq 235)$, ta chuyển thành $P(X \geq 234.5)$.
- Nếu xác suất nhỏ hơn hẳn, chuyển X thành $X - 0.5$:
  - Ví dụ: $P(X < 235)$ chuyển thành $P(X \leq 234.5)$.
- Nếu xác suất lớn hơn hẳn, chuyển X thành $X + 0.5$:
  - Ví dụ: $P(X > 235)$ chuyển thành $P(X \geq 235.5)$.

Quy Tắc Hiệu Chỉnh

- Nếu xác suất là đúng một giá trị cố định, ta nắn X thành khoảng $[X - 0.5, X + 0.5]$:
  - Ví dụ: $P(X = 235)$ chuyển thành $P(234.5 \leq X \leq 235.5)$.
- Nếu xác suất nằm giữa hai giá trị, ví dụ $2 \leq X \leq 5$, ta thực hiện các bước hiệu chỉnh cho từng giá trị:
  - Nắn giá trị $X = 2$ và $X = 5$, sau đó tính $P(1.5 \leq X \leq 5.5)$ theo phân phối chuẩn.

Bài toán Xác suất

Trong 431 trò chơi bóng đá NFL, các đội giành được quyền tung đồng xu tiếp tục chiến thắng 235 trong 431 trò chơi. Nếu phương pháp tung tiền xu là công bằng, các đội thắng cược sẽ thắng khoảng 50% số trận đấu (chúng tôi hy vọng 215.5 thắng trong 431 trận đấu thêm giờ). Giả sử xác suất thắng một ván sau khi thắng tung đồng xu là 0.5, hãy tìm xác suất nhận được ít nhất 235 trận thắng.

Các tham số cần thiết:

- Số lần thử $n = 431$ (số trận đấu). Xác suất thành công (trận thắng) $p = 0.5$. Số trận thắng mong muốn $k = 235$.
- Kiểm tra điều kiện:
  - $np = 431.(0.5) = 215.5 \geq 10$
  - $nq = 431.(0.5) = 215.5 \geq 10$
- Ta có phân phối nhị thức $X \sim B(n, p)$, trong đó:

$$n = 431, \quad p = 0.5$$

Bài giải

Cần tính xác suất để có ít nhất 235 trận thắng, tức là $P(X \geq 235)$. Ta có thể xấp xỉ bằng phân phối chuẩn khi n lớn. Với $X \sim B(431, 0.5)$, ta tính các tham số của phân phối chuẩn:

$$\mu = n \cdot p = 431 \cdot 0.5 = 215.5$$

$$\sigma = \sqrt{n \cdot p \cdot (1 - p)} = \sqrt{431 \cdot 0.5 \cdot 0.5} \approx 10.38027$$

Chúng ta chuẩn hóa giá trị $X = 235$ bằng công thức:

$$Z = \frac{X - \mu}{\sigma} = \frac{235 - 215.5}{13.06} \approx 1.83$$

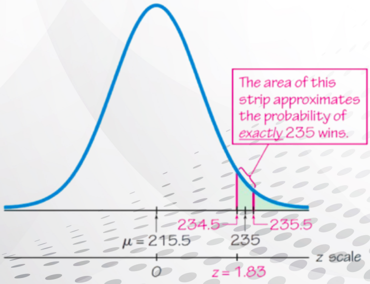
Từ bảng phân phối chuẩn, ta tra cứu giá trị $Z = 1.83$:

$$P(Z \leq 1.83) \approx 0.9664$$

Do đó, xác suất $P(Z \geq 1.83)$ là:

$$P(Z \geq 1.83) = 1 - P(Z \leq 1.83) = 1 - 0.9664 = 0.0336$$

Xác suất để có ít nhất 235 trận thắng trong 431 trận đấu là khoảng 0.0682 hay 6.82%.



Bài tập

1. Trong một cuộc khảo sát, 75% người tham gia cho biết họ ủng hộ chính sách mới của chính phủ. Nếu tiến hành khảo sát 400 người, hãy tính xác suất để có ít nhất 300 người ủng hộ chính sách.
2. Một nhà máy sản xuất có tỷ lệ sản phẩm lỗi là 0.03. Trong một ca sản xuất có 500 sản phẩm, tính xác suất để có ít nhất 20 sản phẩm bị lỗi.
3. Một trạm dịch vụ ô tô có 10% khách hàng cần thay dầu xe. Nếu 100 khách hàng đến trạm trong một ngày, tính xác suất để có ít nhất 15 khách hàng cần thay dầu xe.

Giải Bài tập 1

- Xác suất thành công $p = 0.75$. Số lần thử nghiệm $n = 400$. Ta cần tính xác suất có ít nhất 300 người ủng hộ chính sách: $P(X \geq 300)$. Sử dụng phân phối nhị thức: $X \sim B(400, 0.75)$.
- Để áp dụng phân phối chuẩn, ta tính trung bình và độ lệch chuẩn:

$$\mu = np = 400 \times 0.75 = 300$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{400 \times 0.75 \times 0.25} = \sqrt{75} \approx 8.66$$

- Chuyển sang phân phối chuẩn: $Z = \frac{X-\mu}{\sigma}$.
- Tính $P(X \geq 300)$:

$$P(X \geq 300) \approx P\left(Z \geq \frac{300-300}{8.66}\right) = P(Z \geq 0)$$

$$P(Z \geq 0) = 0.5$$

Vậy xác suất để có ít nhất 300 người ủng hộ chính sách là 0.5.

Giải Bài tập 2

- Xác suất có lỗi trong một sản phẩm: $p = 0.03$. Số sản phẩm: $n = 500$. Ta cần tính xác suất có ít nhất 20 sản phẩm lỗi: $P(X \geq 20)$. Sử dụng phân phối nhị thức: $X \sim B(500, 0.03)$.
- Tính trung bình và độ lệch chuẩn:

$$\mu = np = 500 \times 0.03 = 15$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{500 \times 0.03 \times 0.97} \approx 4.03$$

- Chuyển sang phân phối chuẩn: $Z = \frac{X-\mu}{\sigma}$.
- Tính $P(X \geq 20)$:

$$P(X \geq 20) \approx P\left(Z \geq \frac{20-15}{4.03}\right) = P(Z \geq 1.24)$$

$$P(Z \geq 1.24) \approx 0.107$$

Vậy xác suất để có ít nhất 20 sản phẩm lỗi là khoảng 0.107.

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

Giải Bài tập 3

- Xác suất cần thay dầu xe: $p = 0.1$. Số khách hàng: $n = 100$. Ta cần tính xác suất có ít nhất 15 khách hàng cần thay dầu xe: $P(X \geq 15)$. Sử dụng phân phối nhị thức: $X \sim B(100, 0.1)$.
- Tính trung bình và độ lệch chuẩn:

$$\mu = np = 100 \times 0.1 = 10$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{100 \times 0.1 \times 0.9} \approx 3$$

- Chuyển sang phân phối chuẩn: $Z = \frac{X-\mu}{\sigma}$.
- Tính $P(X \geq 15)$:

$$P(X \geq 15) \approx P\left(Z \geq \frac{15-10}{3}\right) = P(Z \geq 1.67)$$

$$P(Z \geq 1.67) \approx 0.0475$$

Vậy xác suất để có ít nhất 15 khách hàng cần thay dầu xe là khoảng 0.0475.

Good luck!