

BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC ĐỒNG THÁP

TaiLieu.vn

THỐNG KÊ TRONG KHOA HỌC XÃ HỘI
Dùng cho các lớp thuộc ngành xã hội

ĐỒNG THÁP 2014-2015

MỞ ĐẦU

"Giáo trình Lý thuyết xác suất thống kê" là tài liệu được biên soạn cho các sinh viên ngành Khoa học Xã hội như: Công tác Xã hội, Việt Nam học, Thư viện Thông tin, Giáo dục Thể chất,...

Mục đích của bài giảng là trang bị cho các sinh viên kiến thức về thống kê trong khoa học xã hội từ đó nghiên cứu, thu thập và xử lý thông tin kinh tế - xã hội...

Bài giảng bao gồm 4 chương. Chương 1: Khái quát những khái niệm cơ bản về lý thuyết xác suất để làm nền tảng cho việc nghiên cứu phần thống kê. Bao gồm: xác suất cổ điển, xác suất theo quan điểm thống kê, tính chất của xác suất, các biến ngẫu nhiên, hàm phân phối và một số phân phối quan trọng. Chương 2: Mẫu ngẫu nhiên và ước lượng tham số. Chương này mục đích đưa ra các khái niệm về mẫu ngẫu nhiên, các đặc trưng mẫu và các ước lượng tham số. Chương 3: Kiểm định giả thiết. Chương này trình bày một số bài toán kiểm định giả thiết như: kiểm định trung bình, kiểm định tỷ lệ, kiểm định phương sai, kiểm định tính độc lập, quy luật phân phối và các bài toán so sánh. Chương 4 trình bày về tương quan và hồi quy tuyến tính.

Trong tất cả các chương đưa ra đều có những ví dụ minh họa cụ thể cho từng dạng bài toán, sau cuối của mỗi chương đều có hệ thống bài tập khá đa dạng và phong phú.

Vì nhiều lý do, chắc chắn bài giảng không tránh khỏi những sai sót. Chúng tôi mong được sự đóng góp của đồng nghiệp và các bạn sinh viên.

Tác giả

MỤC LỤC

	Trang
MỞ ĐẦU	2
MỤC LỤC	3
Chương 1. Các khái niệm cơ bản về xác suất	6
1.1. Bổ túc về giải tích tổ hợp	6
1.1.1. Các nguyên lý cơ bản	6
1.1.2. Hoán vị	6
1.1.3. Chỉnh hợp	6
1.1.4. Chỉnh hợp lặp	7
1.1.5. Tổ hợp	7
1.1.6. Công thức nhị thức Newton	7
1.2. Phép thử ngẫu nhiên và biến cố	7
1.2.1. Phép thử ngẫu nhiên	7
1.2.2. Biến cố	8
1.2.3. Quan hệ và phép toán giữa các biến cố	8
1.3. Các định nghĩa về xác suất	9
1.3.1. Định nghĩa xác suất theo cổ điển	9
1.3.2. Định nghĩa xác suất theo tần suất và thống kê	9
1.3.3. Tính chất của xác suất	10
1.4. Các công thức xác suất	10
1.5. Biến ngẫu nhiên và hàm phân phối	12
1.5.1. Khái niệm biến ngẫu nhiên và hàm phân phối	12
1.5.2. Biến ngẫu nhiên rời rạc	13
1.5.3. Biến ngẫu nhiên liên tục	14
1.6. Các đặc trưng của biến ngẫu nhiên	16
1.6.1. Kỳ vọng	16
1.6.2. Phương sai	17
1.6.3. Mod	17
1.6.4. Median	17
1.7. Một số phân phối thường gặp	18
1.7.1. Phân phối nhị thức	18

1.7.2. Phân phối Poisson	18
1.7.3. Phân phối chuẩn.....	19
1.7.4. Tính gần đúng phân phối nhị thức	20
1.8. Véc tơ ngẫu nhiên	22
1.8.1. Biến ngẫu nhiên rời rạc.	22
1.8.2. Biến ngẫu nhiên liên tục.	24
1.8.3. Các đặc trưng của véc tơ ngẫu nhiên.....	27
Bài tập chương 1.....	30
Chương 2. Lý thuyết chọn mẫu và ước lượng tham số.....	37
2.1. Mẫu ngẫu nhiên và hàm phân phối	37
2.1.1. Mẫu ngẫu nhiên	37
2.1.2. Hàm phân phối - Đa giác tần số và tổ chức đồ.....	38
2.1.3. Mẫu ngẫu nhiên hai chiều.....	41
2.1.4. Các đặc trưng mẫu	42
2.2. Ước lượng điểm	45
2.2.1. Ước lượng không chệch.....	46
2.2.2. Ước lượng vững.	46
2.2.3. Ước lượng hiệu quả	47
2.2.4. Ước lượng hợp lý cực đại.	47
2.2.5. Ước lượng điểm cho kỳ vọng.....	49
2.2.6. Ước lượng điểm cho phương sai	49
2.2.4. Ước lượng điểm cho xác suất.....	49
2.3. Ước lượng khoảng	50
2.3.1. Ước lượng khoảng đối với giá trị trung bình.....	50
2.3.2. Ước lượng khoảng đối với giá trị tỷ lệ.....	54
2.3.3. Ước lượng khoảng đối với phương sai.....	57
Bài tập chương 2.....	58
Chương 3. Kiểm định giả thiết.....	65
3.1. Đặt vấn đề.....	65
3.2. Kiểm định giả thiết về giá trị trung bình	66
3.2.1. Trường hợp phương sai σ^2 đã biết.	66
3.2.2. Trường hợp phương sai σ^2 chưa biết $n \geq 30$	68
3.2.3. Trường hợp phương sai σ^2 chưa biết $n < 30$	71
3.3. Kiểm định tỷ lệ	73

3.3.1. Kiểm định hai phía	73
3.3.2. Kiểm định một phía.....	73
3.4. Kiểm định phương sai.....	75
3.4.1. Trường hợp chưa biết μ	75
3.4.2. Trường hợp đã biết μ	76
3.5. Kiểm định về tính độc lập.....	78
3.6. Kiểm định giả thiết về luật phân phối.....	81
3.7. Bài toán so sánh	83
3.7.1. Bài toán so sánh hai giá trị trung bình.....	83
3.7.2. Bài toán so sánh hai giá trị tỷ lệ.....	90
Bài tập chương 3.....	92
Chương 4. Tương quan và hồi quy tuyến tính.....	99
4.1. Tương quan tuyến tính.....	99
4.1.1. Định nghĩa	99
4.1.2. Tính chất	99
4.1.3. Hệ số tương quan mẫu	99
4.1.4. Ý nghĩa của hệ số tương quan mẫu	100
4.2. Hồi quy tuyến tính.....	101
Bài tập chương 4.	104
Các bảng số thông dụng	106
Tài liệu tham khảo.....	111

Chương 1

Các khái niệm cơ bản về xác suất

1.1 Bổ túc về giải tích tổ hợp

1.1.1 Các nguyên lý đếm cơ bản

a) Nguyên lý cộng

Giả sử có k công việc, việc thứ nhất có n_1 cách làm, việc thứ hai có n_2 cách làm, ..., việc thứ k có n_k cách làm, ... các công việc này không làm đồng thời. Khi đó ta có $n_1 + n_2 + \dots + n_k$ cách làm k công việc trên.

b) Nguyên lý nhân.

Giả sử hành động H được thực hiện qua k giai đoạn liên tiếp $H_1, H_2, H_3, \dots, H_k$. Giai đoạn H_1 có n_1 cách làm, ..., H_k có n_k cách làm. Khi đó $n_1 \cdot n_2 \cdot \dots \cdot n_k$ cách làm công việc H .

1.1.2 Hoán vị

Định nghĩa 1.1.1. Cho tập M có n phần tử, mỗi cách sắp xếp n phần tử này theo một thứ tự nhất định gọi là một hoán vị của tập M . Gọi số các hoán vị của tập M là:

$$P_n = n! = 1.2.3 \dots (n-1)n$$

Ví dụ 1. a) Ta có 3 người A, B, C xếp vào 3 chỗ ngồi. Khi đó ta có $3! = 3.2.1 = 6$ cách xếp như sau:

$$ABC, ACB, BAC, BCA, CAB, CBA$$

b) Số cách sắp xếp cho 80 sinh viên vào 80 chỗ ngồi là $P_{80} = 80!$

1.1.3 Chỉnh hợp

Định nghĩa 1.1.2. Cho tập M có n phần tử, $0 \leq k \leq n$, một chỉnh hợp chập k của n phần tử là một bộ sắp thứ tự (phân biệt) lấy từ n phần tử đã cho và được ký hiệu là

$$A_n^k = \frac{n!}{(n-k)!}$$

Ví dụ 2. a) Cho ba phần tử 2,3,5. Các chỉnh hợp chập 2 của ba phần tử đó là:

$$23, 25, 32, 35, 52, 53$$

b) Mỗi lớp phải học 6 môn, mỗi ngày học 2 môn. Hỏi có bao nhiêu cách sắp xếp thời khóa biểu cho mỗi ngày.

HD: Vì mỗi cách sắp xếp thời khóa biểu trong một ngày là ghép 2 môn trong 6 môn. Các cách này do ít nhất 1 môn khác nhau hoặc chỉ do thứ tự sắp xếp trước sau giữa hai môn. Vì thế mỗi cách sắp xếp ứng với một chỉnh hợp chập 2 của 6.

$$A_6^2 = 30$$

1.1.4 Chỉnh hợp lặp

Định nghĩa 1.1.3. Chỉnh hợp lặp chập k của n phần tử là 1 nhóm thứ tự gồm k phần tử lấy từ n phần tử đã cho trong đó mỗi phần tử có thể có mặt 1,2, 3,... k lần trong nhóm tạo thành. Ký hiệu $\overline{A}_n^k = n^k$

Ví dụ 3. a) Cho ba phần tử 2,3,5. Các chỉnh hợp lặp chập 2 của ba phần tử đó là:

$$22, 23, 25, 32, 33, 35, 52, 53, 55$$

b) Để đăng ký mỗi loại máy mới người ta dùng 3 con số trong 9 con số 1,2,...9. Hỏi có thể đánh số được bao nhiêu máy.

Mỗi số của máy là chỉnh hợp lặp chập 3 của 9 số: $\overline{A}_9^3 = 9^3 = 729$

1.1.5 Tổ hợp

Định nghĩa 1.1.4. Tổ hợp chập k của n phần tử, $0 \leq k \leq n$ là một tập con của k phần tử lấy từ n phần tử đã cho và được ký hiệu là

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}$$

Ví dụ 4. Có 10 đội bóng đá thi đấu với nhau theo thể thức vòng tròn một lượt. Hỏi có bao nhiêu trận đấu?

HD: Ta thấy mỗi trận đấu giữa 2 đội đấu với nhau là 1 tổ hợp chập 2 của 10 phần tử (Vì hai đội đấu với nhau không cần xếp thứ tự) $C_{10}^2 = 45$

1.2 Phép thử ngẫu nhiên và biến cố

1.2.1 Phép thử ngẫu nhiên

Phép thử ngẫu nhiên là một hành động mà ta chưa biết trước được kết quả của nó. Tuy chưa biết trước được kết quả của phép thử nhưng biết được tập tất cả các khả

năng và ký hiệu là Ω và gọi là không gian biến cố sơ cấp.
Mỗi $\omega \in \Omega$ gọi là biến cố sơ cấp. Ta ký hiệu phép thử là G

Ví dụ 5. a) Tung đồng tiền thì $\Omega = \{S, N\}$

b) Tung con xúc xắc: $\Omega = \{1, 2, 3, 4, 5, 6\}$

1.2.2 Biến cố

Khi thực hiện một phép thử có rất nhiều câu hỏi liên quan đến kết quả của nó. Một sự kiện liên quan đến phép thử mà việc nó xảy ra hay không xảy ra phụ thuộc hoàn toàn vào phép thử gọi là một biến cố ngẫu nhiên. Ký hiệu $A, B, C, \dots$

Biến cố sơ cấp ω gọi là thuận lợi cho biến A nếu khi kết quả của phép thử là ω thì A xảy ra

Biến cố không thể là biến cố không thể xảy ra và ký hiệu là: $\emptyset$

Biến cố chắc chắn: là biến cố luôn xảy ra khi thực hiện phép thử, ký hiệu là: Ω

Ví dụ 6. Tung con xúc xắc $\Rightarrow \Omega = \{1, 2, 3, 4, 5, 6\}$

Biến cố xuất hiện mặt chấm lẻ là $A \Rightarrow A = \{1, 3, 5\}$

Biến cố xuất hiện mặt chấm nhỏ hơn 5 là $B: \Rightarrow B = \{1, 2, 3, 4\}$

1.2.3 Quan hệ và phép toán giữa các biến cố.

a. Quan hệ kéo theo

Biến cố A gọi là kéo theo biến cố B nếu A xảy ra thì B xảy ra. Ký hiệu $A \subset B$

b. Quan hệ bằng

Hai biến cố A, B gọi là bằng nhau. Ký hiệu $A=B$ nếu $A \subset B, A \supset B$

c. Giao của hai biến cố

Giao của hai biến cố là một biến cố xảy ra khi và chỉ khi A, B đồng thời xảy ra. Ký hiệu $A \cap B$ hoặc AB

TQ: $A_1 \cap A_2 \cap \dots \cap A_n$ là một biến cố xảy ra khi và chỉ khi với mọi A_i xảy ra.

d. Hợp của hai biến cố

Hợp của hai biến cố là một biến cố xảy ra khi và chỉ khi A hoặc B xảy ra. Ký hiệu $A \cup B$

TQ: $A_1 \cup A_2 \cup \dots \cup A_n$ là một biến cố xảy ra khi và chỉ khi ít nhất một A_i xảy ra.

e. Hiệu của hai biến cố.

Hiệu của hai biến cố A và B ký hiệu là $A \setminus B$ là một biến cố xảy ra khi và chỉ khi A xảy ra và B không xảy ra.

g. Biến cố đối

Biến cố đối của biến cố A là $\overline{A}$, là biến cố xảy ra khi và chỉ khi A không xảy ra.

h. Biến cố xung khắc

Hai biến cố A và B là xung khắc nếu chúng không đồng thời xảy ra, tức là $AB = \emptyset$.

Biến cố đối thì xung khắc.

h. Nhóm đầy đủ các biến cố

Nhóm n biến cố $A_1, A_2, \dots, A_n$ gọi là nhóm đầy đủ các biến cố nếu.

- i) Chúng xung khắc với nhau đôi một $A_i A_j = \emptyset, (i \neq j)$
- ii) Hợp của chúng là biến cố chắc chắn $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$

Ví dụ 7. Hai xạ thủ mỗi người bắn một viên đạn vào đích.

Gọi A_i là biến cố người thứ i trúng đích. Hãy viết biến cố sau qua A_1, A_2

- a. Biến cố chỉ người thứ nhất trúng đích: $A_1 \overline{A_2}$
- b. Có 1 người bắn trúng đích: $A_1 \overline{A_2} \cup A_2 \overline{A_1}$
- c. Có ít nhất một người bắn trúng đích: $A_1 \cup A_2$
- d. Không có ai bắn trúng: $\overline{A_1} \overline{A_2}$

1.3 Các định nghĩa về xác suất

1.3.1 Định nghĩa xác suất theo cổ điển

Giả sử không gian biến cố sơ cấp Ω của phép thử G có n kết quả đồng khả năng và có m kết quả thuận lợi cho biến cố A . Khi đó xác suất của biến cố A được ký hiệu và được định nghĩa là

$$P(A) = \frac{m}{n}$$

Ví dụ 8. Một hộp có 16 quả cầu đen và 4 quả cầu đỏ lấy ngẫu nhiên 2 quả cầu. Hãy tính xác suất

- a) Lấy được hai quả cầu đen.
- b) Lấy được 1 quả cầu đen, một quả đỏ.

HD:

- a) Gọi A là biến cố lấy được hai quả cầu đen. Khi đó

$$n = C_{20}^2, m = C_{16}^2 \Rightarrow P(A) = \frac{C_{16}^2}{C_{20}^2}$$

- b) Gọi B là biến cố lấy được 1 quả đen, 1 quả đỏ thì

$$P(B) = \frac{C_{16}^1 C_4^1}{C_{20}^2}$$

Ví dụ 9. Một nhóm học tập có 10 hs, trong đó có 7 hs yếu. Kiểm tra ngẫu nhiên 3 em. Tính xác suất để:

- a) Ba em kiểm tra là học sinh yếu
- b) Trong 3 em được kiểm tra có 1 em yếu
- c) Có ít nhất 1 học sinh yếu được kiểm tra.

1.3.2 Định nghĩa xác suất theo tần suất và thống kê

Một phép thử được thực hiện n lần mà có m biến cố A xuất hiện thì tỷ số m/n gọi là tần suất của biến cố A

Khi n thay đổi, tần suất m/n cũng thay đổi nhưng nó luôn dao động quanh một số cố định nào đó, n càng lớn thì m/n càng gần số cố định đó. Số cố định này được gọi là

xác suất của biến cố A theo nghĩa thống kê. Trên thực tế khi n đủ lớn ta xấp xỉ $P(A)$ bởi m/n tức là

$$P(A) = \frac{m}{n}$$

1.3.3 Tính chất của xác suất

- 1) $0 \leq P(A) \leq 1, P(\emptyset) = 0, P(\Omega) = 1$
- 2) $P(\bar{A}) = 1 - P(A)$.
- 3) Nếu $A \subset B$ thì $P(A) \leq P(B)$ với $P(B/A) = P(B) - P(A)$.
- 4) $P(A \cup B) = P(A) + P(B) - P(AB)$

Ví dụ 10. Một hộp chứa 5 cầu trắng, 3 cầu xanh và 4 cầu đen cùng kích thước, chọn ngẫu nhiên cùng lúc 3 cầu. Tìm xác suất để.

- a) Cả 3 cầu cùng màu (A)
- b) Có đúng 2 cầu cùng màu(B)
- c) Có ít nhất hai cầu cùng màu(C)
- d) Cả 3 cầu khác màu nhau(D)

HD:

a) Gọi $A_1 = \{ 3 \text{ quả cầu rút ra đều màu trắng} \}$

$A_2 = \{ 3 \text{ quả cầu rút ra cùng màu đen} \}$

$A_3 = \{ 3 \text{ quả cầu rút ra đều màu xanh} \}$

Khi đó: $A = A_1 + A_2 + A_3 \implies P(A) = P(A_1) + P(A_2) + P(A_3)$

$$\implies P(A) = \frac{C_5^3 + C_3^3 + C_4^3}{C_{12}^3} = \frac{3}{44}$$

b) Gọi $B_1 = \{ 2 \text{ quả cầu rút ra cùng màu trắng} \}$

$B_2 = \{ 2 \text{ quả cầu rút ra cùng màu đen} \}$

$B_3 = \{ 2 \text{ quả cầu rút ra cùng màu xanh} \}$

$$\implies P(B) = P(B_1) + P(B_2) + P(B_3) = \frac{C_5^2 C_7^1 + C_4^2 C_8^1 + C_3^2 C_9^1}{C_{12}^3} = \frac{29}{44}$$

c) $P(C) = P(A) + P(B) = \frac{32}{44}$

d) Cách1: $P(D) = 1 - P(C)$

Cách2: làm trực tiếp

$$P(D) = \frac{C_5^1 C_3^1 C_4^1}{C_{12}^3} = \frac{3}{11}$$

1.4 Dãy phép thử Bernoulli và công thức nhị thức

Hai biến cố A và B gọi là độc lập với nhau nếu việc xảy ra biến cố này không ảnh hưởng đến khả năng xảy ra của biến cố khác

Hai phép thử gọi là độc lập với nhau nếu việc thực hiện và kết quả của phép thử này không ảnh hưởng và không phụ thuộc vào phép thử kia

Định nghĩa 1.4.1. Dãy n phép thử gọi là dãy n phép thử Bernoulli đối với biến cố A nếu thoả mãn các điều kiện sau:

- Chúng là n phép thử lặp.
- Các phép thử đó là độc lập.
- Mỗi phép thử biến cố A xuất hiện với xác suất đều bằng p .

Công thức nhị thức

Xác suất để trong n phép thử Bernoulli biến cố A xuất hiện đúng k lần là:

$$P_n(k) = C_n^k p^k (1-p)^{n-k} = C_n^k p^k (q)^{n-k} = P_n(k, p)$$

Công thức trên gọi là công thức xác suất nhị thức.

Số khả năng nhất

Giả sử $G_1, G_2, \dots, G_n$ là n phép thử Bernoulli, xác suất xuất hiện A k lần là

$$P_k = C_n^k p^k q^{n-k}, \quad (0 \leq k \leq n)$$

Khi đó số k_0 , ($0 \leq k_0 \leq n$) được gọi là số có khả năng nhất nếu

$$P_{k_0} = \max_{0 \leq k \leq n} P_k$$

trong đó k_0 được tính theo công thức sau:

$$k_0 = \begin{cases} np - q & \text{và } np - q + 1 \text{ nếu } np - q \text{ nguyên} \\ [p(n+1)] & \text{nếu } np - q \text{ không nguyên} \end{cases}$$

Ví dụ 11. Tung đồng tiền 5 lần. Tính xác suất để mặt sấp xuất hiện k lần.

HD: Đây là 5 phép thử Bernoulli đối với biến cố A xuất hiện mặt sấp với $p = \frac{1}{2}$.

- Xác suất biến cố A xuất hiện 0 lần là:

$$P_5(0) = C_5^0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5 = \frac{1}{32}$$

- Xác suất để A xuất hiện 1, 2, 3, 4, 5 lần

$$P_5(1) = C_5^1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^4 = \frac{5}{32}; \quad P_5(2) = C_5^2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^3 = \frac{10}{32}$$

$$P_5(3) = C_5^3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^2 = \frac{10}{32}; \quad P_5(4) = C_5^4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^1 = \frac{5}{32}$$

$$P_5(5) = C_5^5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^0 = \frac{1}{32}$$

Ta thấy $k = 2$ hoặc $k = 3$ thì $P_5(k)$ lớn nhất và ta nói 2, 3 là số có khả năng nhất.

Ví dụ 12. Kết quả điều tra về bệnh lao, tỷ lệ người bị lao ở vùng nọ là 0,001. Tìm xác suất để khi khám cho 10 người.

a. Không có ai bị lao.

b. 5 người bị lao.

c. Ít nhất một người bị lao.

d. Số người không bị lao có khả năng nhất.

HD: Ta có 10 phép thử Bernoulli, với biến cố A là " người được khám bị lao" suy ra $P(A) = 0.001$

a.

$$\begin{aligned} P_{10}(n, p) &= P_{10}(0, 0.001) \\ &= C_{10}^0 (0.001)^0 (1 - 0.001)^{10} = (0.999)^{10} \end{aligned}$$

b.

$$P_{10}(5, 0.001) = C_{10}^5 (0.001)^5 (0.999)^5$$

c.

$$\begin{aligned} P_{10}(k \geq 1, 0.001) &= \sum_{k=1}^{10} C_{10}^k (0.001)^k (0.999)^{10-k} \\ &= 1 - P_{10}(0, 0.001) = 1 - (0.999)^{10} \end{aligned}$$

d. Ta có $q = 1 - p = 1 - 0.001 = 0.999$ mà $q(1 + n) = 11.0, 999 = 10, 989$ không phải là số nguyên do đó số người không bị bệnh lao có khả năng cao nhất là 10.

1.5 Biến ngẫu nhiên và hàm phân phối

1.5.1 Khái niệm biến ngẫu nhiên và hàm phân phối

Khái niệm biến ngẫu nhiên

Một đại lượng mà giá trị của nó là ngẫu nhiên, không dự đoán trước được gọi là một đại lượng ngẫu nhiên (biến ngẫu nhiên) và ký hiệu bằng chữ $X, Y, Z, \dots$

Hoặc một đại lượng ngẫu nhiên nhận các giá trị của nó với xác suất tương ứng nào đó gọi là đại lượng ngẫu nhiên hay là biến ngẫu nhiên

Có hai loại biến ngẫu nhiên chính đó là: Biến ngẫu nhiên rời rạc và biến ngẫu nhiên liên tục

Hàm phân phối xác suất

Hàm phân phối xác suất của biến ngẫu nhiên X được ký hiệu và xác định như sau:

$$F_X(x) = P[X < x] ; x \in R$$

• Nếu biến ngẫu nhiên rời rạc thì

$$F_X(x) = P(X < x) = \sum_{x_i < x} p_i$$

- Nếu biến ngẫu nhiên liên tục thì

$$F_X(x) = \int_{-\infty}^x p(t)dt$$

1.5.2 Biến ngẫu nhiên rời rạc

Định nghĩa 1.5.1. Biến ngẫu nhiên X được gọi là rời rạc nếu tập giá trị của nó là hữu hạn hoặc đếm được.

Bảng phân phối xác suất

Giả sử X là biến ngẫu nhiên rời rạc nhận các giá trị $x_1, \dots, x_n, \dots$. Khi đó $p_i = P(X = x_i)$ thì bảng sau gọi là bảng phân phối xác suất của X .

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

Chú ý: $\sum_i p_i = 1$

Hàm phân phối xác suất

$$F_X(x) = P(X \leq x) = \sum_{i: x_i \leq x} p_i = \begin{cases} 0 & \text{nếu } x \leq x_1 \\ p_1 & \text{nếu } x_1 < x \leq x_2 \\ p_1 + p_2 & \text{nếu } x_2 < x \leq x_3 \\ \dots & \dots \\ p_1 + p_2 + \dots + p_k & \text{nếu } x_k < x \leq x_{k+1} \\ \dots & \dots \\ 1 & \text{nếu } x > x_n \end{cases}$$

Tính chất của hàm phân phối

Giả sử $F(x)$ là hàm phân phối của biến ngẫu nhiên X

- $F(x)$ là hàm không giảm tức nếu $x_1 < x_2$ thì $F(x_1) \leq F(x_2)$
- $P(a \leq X < b) = F(b) - F(a)$
- $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$
- $F(x)$ liên tục trái $\lim_{x \rightarrow x_0^-} F(x) = F(x_0)$
- $P(X \leq x) = F(x+0) \Rightarrow P(X = x) = F(x+0) - F(x)$

Ví dụ 13. Một xí nghiệp có hai ô tô vận tải. Xác suất bị hỏng trong thời gian t của hai ô tô tương ứng là: 0, 1; 0, 2.

Gọi X là số ô tô bị hỏng trong thời gian t . Hãy lập bảng phân phối xác suất và tính hàm phân phối xác suất của X .

HD:

Gọi A_i là biến cố ô tô thứ i hỏng ($i=1, 2$). Khi đó ta có X nhận 3 giá trị: 0, 1, 2.

$$P(X = 0) = P(\overline{A_1} \overline{A_2}) = P(\overline{A_1})P(\overline{A_2}) = 0,9.0,8 = 0,72$$

$$\begin{aligned} P(X = 1) &= P(\overline{A_1}A_2 + A_1\overline{A_2}) \\ &= P(\overline{A_1}A_2) + P(A_1\overline{A_2}) \\ &= P(\overline{A_1})P(A_2) + P(A_1)P(\overline{A_2}) \\ &= 0,1.0,8 + 0,9.0,2 = 0,26 \end{aligned}$$

$$P(X = 2) = P(A_1A_2) = 0,1.0,2 = 0,02$$

Ta có bảng phân phối xác suất của X là

X	0	1	2
P	0,72	0,26	0,02

Hàm phân phối là:

$$F(x) = P(X < x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ 0,72 & \text{nếu } 0 < x \leq 1 \\ 0,98 & \text{nếu } 1 < x \leq 2 \\ 1 & \text{nếu } x > 2 \end{cases}$$

Ví dụ 14. Chọn ngẫu nhiên 3 đứa trẻ từ 1 nhóm gồm 6 bé trai và 4 bé gái. Gọi X là số bé gái trong nhóm. Lập bảng phân phối xác suất của X .

HD: X nhận các giá trị 0, 1, 2, 3. Khi đó ta có

$$\begin{aligned} P\{X = 0\} &= \frac{C_6^3}{C_{10}^3} = \frac{5}{30} \\ P\{X = 1\} &= \frac{C_4^1 C_6^2}{C_{10}^3} = \frac{15}{30} \\ P\{X = 2\} &= \frac{C_4^2 C_6^1}{C_{10}^3} = \frac{9}{30} \\ P\{X = 3\} &= \frac{C_4^3}{C_{10}^3} = \frac{1}{30} \end{aligned}$$

Từ đó suy ra bảng phân phối hàm phân phối.

1.5.3 Biến ngẫu nhiên liên tục

Định nghĩa 1.5.2. Biến ngẫu nhiên X được gọi là biến ngẫu nhiên liên tục nếu:

i) $F_X(x)$ là hàm liên tục.

ii) Tồn tại hàm số $p(x) \geq 0, \forall x \in R$ sao cho

$$F_X(x) = \int_{-\infty}^x p(t)dt \quad x \in R$$

$p(x)$ được gọi là hàm mật độ xác suất.

Tính chất hàm mật độ

- i) $\int_{-\infty}^{+\infty} p(x)dx = 1$
 - ii) $P(a \leq X < b) = \int_a^b p(x)dx$ nếu X liên tục có hàm mật độ $p(x)$.
 - iii) $p(x) \geq 0, \forall x \in R$
 - iv) $p(x) = F'(x)$ tại mọi điểm x mà $p(x)$ liên tục
- Chú ý: Nếu X liên tục thì

$$P(a \leq X < b) = P(a < X < b) = P(a < X \leq b) = F(b) - F(a) = P(a \leq X \leq b)$$

vì $P(X = a) = P(X = b) = 0$

Ví dụ 15. Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$p(x) = \begin{cases} 0 & \text{nếu } x \notin [0, \pi/2] \\ a \sin x & \text{nếu } x \in [0, \pi/2] \end{cases}$$

Tìm hàm phân phối của X

HD

Tìm a ta có $1 = \int_{-\infty}^{+\infty} p(x)dx = a \int_0^{\pi/2} \sin x dx = -a \cos x \Big|_0^{\pi/2} = a$ suy ra $a = 1$

Ta có $F(x) = \int_{-\infty}^x p(t)dt$

+) Nếu $x \leq 0$ thì $F(x) = 0$

+) Nếu $0 < x \leq \pi/2 \rightarrow F(x) = \int_{-\infty}^x p(t)dt = \int_0^x \sin t dt = 1 - \cos x$

+) Nếu $x > \pi/2 \rightarrow F(x) = \int_{-\infty}^0 p(x)dx + \int_0^{\pi/2} p(x)dx + \int_{\pi/2}^{+\infty} p(x)dx = 1$

Vậy

$$F(x) = \begin{cases} 0 & \text{nếu } x \leq 0 \\ 1 - \cos x & \text{nếu } 0 < x \leq \pi/2 \\ 1 & \text{nếu } x > \pi/2 \end{cases}$$

Ví dụ 16. cho biến ngẫu nhiên có hàm phân phối $F(x) = 1/2 + 1/\pi \arctg x$

a) Tính xác suất của biến cố $\{0 < x < 1\}$

b) Tìm hàm mật độ của X.

HD:

a) Ta có:

$$\begin{aligned} P(0 < x < 1) &= P(0 \leq x < 1) - P(x = 0) \\ &= P(0 \leq x < 1) \\ &= F(1) - F(0) \\ &= 1/\pi (\arctg 1 - \arctg 0) = 1/\pi (\pi/4 - 0) = 1/4 \end{aligned}$$

b) Ta có hàm mật độ: $p(x) = F'(x) = \frac{1}{\pi(x^2+1)}$

Ví dụ 17. Cho X là biến ngẫu nhiên có hàm mật độ $p(x)$ như sau:

$$p(x) = \begin{cases} 1+x & \text{nếu } x \in [-1, 0) \\ 1-x & \text{nếu } x \in [0, 1) \\ 0 & \text{nếu } |x| > 1 \end{cases}$$

Tính $P\{-\frac{1}{2} < X < 1\}$, tìm hàm phân phối $F(x)$

1.6 Các đặc trưng của biến ngẫu nhiên

1.6.1 Kỳ vọng

Định nghĩa 1.6.1. Kỳ vọng của biến ngẫu nhiên X được ký hiệu EX và xác định bởi:

- Nếu X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

thì $EX = \sum_i x_i p_i$

- Nếu X liên tục với hàm mật độ xác suất $p(x)$ thì

$$EX = \int_{-\infty}^{+\infty} xp(x)dx$$

Tính chất của kỳ vọng.

- $EC = C$; ($c = \text{const}$)
- $E(X \pm Y) = EX \pm EY$, (nếu hai vế có nghĩa)
- $E(aX) = aEX$; a là hằng số.
- $EXY = EXEY$; nếu X, Y độc lập.

Ý nghĩa của kỳ vọng

Kỳ vọng của một biến ngẫu nhiên là giá trị trung bình theo xác suất, nếu đối với hệ cơ học thì kỳ vọng là trọng tâm của hệ, nếu nó nhận xác suất như nhau thì kỳ vọng chính là trung bình số học.

Ví dụ 18. Cho X có bảng phân phối xác suất

X	x_1	x_2	$\dots$	x_n	$\dots$
P	$\frac{1}{n}$	$\frac{1}{n}$	$\dots$	$\frac{1}{n}$	$\dots$

Khi đó

$$EX = \frac{x_1 + x_2 + \dots + x_n}{n}$$

Ví dụ 19. Cho X liên tục có hàm mật độ

$$p(x) = \begin{cases} 0 & \text{nếu } x \notin [a, b] \\ \frac{1}{b-a} & \text{nếu } x \in [a, b] \end{cases}$$

Khi đó

$$EX = \int_{-\infty}^{+\infty} xp(x)dx = EX = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{a+b}{2}$$

1.6.2 Phương sai

Định nghĩa 1.6.2. Phương sai của biến ngẫu nhiên X ký hiệu là DX được xác định

$$DX = E(X - EX)^2$$

$$DX = \begin{cases} \sum (x_i - EX)^2 p_i & \text{nếu X rời rạc và có bảng phân phối} \\ \int_{-\infty}^{+\infty} (x - EX)^2 f(x) dx & \text{nếu X liên tục có hàm mật độ f(x)} \end{cases}$$

Tính chất của phương sai

- $DX = EX^2 - (EX)^2$
- $D(aX) = a^2 DX$, ($a = \text{const}$)
- $D(X + a) = DX$
- $D(X \pm Y) = DX + DY$, nếu X, Y độc lập.
- $DX \geq 0$

Ý nghĩa của phương sai

Phương sai của biến ngẫu nhiên là độ lệch trung bình của X xung quanh giá trị kỳ vọng EX nếu DX bé thì giá trị của X tập trung xung quanh kỳ vọng, ngược lại DX lớn thì giá trị của X phân tán xung quanh kỳ vọng

1.6.3 Mod

Mốt là giá trị của biến ngẫu nhiên X, ký hiệu là x_{mod} mà tại đó hàm mật độ $f(x)$ đạt cực đại, trường hợp X là biến ngẫu nhiên rời rạc, x_{mod} là giá trị, mà xác suất để $X = x_{mod}$.

1.6.4 Median

Trung vị (Median) là giá trị của biến ngẫu nhiên X, ký hiệu x_{Me} hoặc $m(X)$, mà tại đó

- Nếu X là rời rạc thì

$$F(x_i) \leq 1/2 \leq F(x_{i+1}) \Rightarrow m(X) = x_i$$

- Nếu X là liên tục thì $F(x_{Me}) = \frac{1}{2}$ hoặc $F[m(X)] = \frac{1}{2}$

1.7 Một số phân phối thường gặp

1.7.1 Phân phối nhị thức

Định nghĩa 1.7.1. Biến ngẫu nhiên X gọi là phân phối nhị thức với tham số n, p ký hiệu $X \sim B(n, p)$ nếu X nhận các giá trị $0, 1, \dots, n$ với xác suất

$$P_k = P(X = k) = C_n^k p^k q^{n-k}$$

Các số đặc trưng.

Nếu $X \sim B(n, p)$ thì

$$EX = np, DX = npq, (n+1)p - 1 \leq \text{Mod} X \leq (n+1)p$$

Ví dụ 20. Tỷ lệ phế phẩm của một loại sản phẩm là 1% , người ta lấy ngẫu nhiên có hoàn lại 100 sản phẩm để kiểm tra.

1. Tính xác suất có 2 sản phẩm
2. Hỏi trung bình có bao nhiêu sản phẩm
3. Khả năng có bao nhiêu sản phẩm (Mod)

Có 2-5 sản phẩm

HD

1. Gọi X là số sản phẩm suy ra $X \sim (100; 0,01)$

$$P[X = 2] = C_{100}^2 (0,01)^2 (0,99)^{98}$$

$$2. EX = np = 0,01 \times 100$$

$$3. np - q \leq \text{Mod} X \leq np + q$$

$$\Rightarrow 0,01 \times 100 - 0,99 \leq \text{Mod} X \leq 0,01 \times 100 + 0,99 \Rightarrow \text{Mod} X = \dots (\in Z)$$

$$4. P[2 \leq X \leq 5] = P[X = 2] + P[X = 3] + P[X = 4] + P[X = 5]$$

1.7.2 Phân phối Poisson

Định nghĩa 1.7.2. Biến ngẫu nhiên X gọi là có phân phối theo quy luật Poisson với tham số $\lambda > 0$. Ký hiệu $X \sim f(\lambda)$, nếu X nhận các giá trị $0, 1, \dots$ với xác suất tương ứng

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

Tức là ta có bảng:

X	0	...	k	...	n	
P	$e^{-\lambda}$	...	$e^{-\lambda} \frac{\lambda^k}{k!}$		$e^{-\lambda} \frac{\lambda^n}{n!}$	...

Chú ý: Ta có bảng tính sẵn $P(X \leq k) = \sum_{i=0}^k e^{-\lambda} \frac{\lambda^i}{i!}$

Các đặc trưng

Nếu $X \sim f(\lambda)$ thì $EX = DX = \lambda; ModX = [\lambda]$

Ví dụ 21. Một ga ra cho thuê ô tô, thấy rằng số người đến thuê ô tô vào ngày thứ 7 là một biến ngẫu nhiên có phân phối Poisson với tham số $\lambda = 2$. Giả sử ga ra có 4 chiếc ô tô. Hãy tính xác suất.

- Không phải cả 4 chiếc đều được thuê.
- Tất cả 4 ô tô đều được thuê
- Ga ra không đáp ứng được nhu cầu
- Trung bình có bao nhiêu ô tô được thuê?

1.7.3 Phân phối chuẩn

a) Phân phối chuẩn tắc

Định nghĩa 1.7.3. Biến ngẫu nhiên X gọi là có phân phối chuẩn tắc, ký hiệu $X \sim N(0, 1)$ nếu X có hàm mật độ

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad (x \in R)$$

Đặc trưng

Nếu $X \sim N(0, 1)$ thì $EX = 0, DX = 1$

Hàm phân phối chuẩn tắc của X được ký hiệu là:

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

Chú ý

- $\phi(-x) = 1 - \phi(x)$
- $\phi(x > 3, 9) = 1$

b) Phân phối chuẩn

Định nghĩa 1.7.4. Biến ngẫu nhiên gọi là có phân phối chuẩn với tham số (μ, σ^2) , ký hiệu là $X \sim N(\mu, \sigma^2)$, nếu $\frac{X-\mu}{\sigma}$ có phân phối chuẩn tắc.

Đặc trưng:

Nếu $X \sim N(\mu, \sigma^2)$ thì $EX = \mu, DX = \sigma^2$

Xác suất để biến ngẫu nhiên có phân phối chuẩn nhận giá trị trong một đoạn.

Nếu $X \sim N(\mu, \sigma^2)$. Khi đó

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{b - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

Hàm mật độ của phân phối chuẩn $N(\mu, \sigma^2)$ là:

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

1.7.4 Tính gần đúng phân phối nhị thức

Khi n lớn và p không quá bé (thường xét $n > 30; np > 5$)
Ta có thể tính xấp xỉ phân phối nhị thức $B(n, p)$ bằng phân phối chuẩn $N(\mu; \sigma^2)$ với $\mu = np, \sigma^2 = npq$. Cụ thể là:

- $p_k = p(X = k) = C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} \varphi(z)$ với $z = \frac{(x-np)}{\sqrt{npq}}$
- $p(a \leq X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$

Ví dụ 22. Xác suất của biến cố A là 0,8. Tìm xác suất khi thực hiện 100 phép thử thì:

- Số lần xảy ra $A > 75$
- Số lần xảy ra A không quá 74
- số lần xảy ra A trong khoảng 75-90.

HD

Gọi X là số lần xảy ra A : $X \sim B(100, 0,8)$

a) Ta có $\mu = 0,8 \cdot 100 = 80; \sigma^2 = 0,2 \cdot 0,8 \cdot 100 = 16 \Rightarrow \sigma = 4$. Khi đó

$$P[X > 75] = P[75 < X < 100] = \Phi(5) - \Phi(-1,25) = \Phi(5) - 1 + \Phi(1,25) = 0,8944$$

b)

$$P[0 \leq X \leq 74] = \Phi\left[\frac{74 - 80}{4}\right] - \Phi\left[\frac{0 - 80}{4}\right]$$

c)

$$P[75 \leq X \leq 90] = \Phi\left[\frac{90 - 80}{4}\right] - \Phi\left[\frac{75 - 80}{4}\right]$$

Ví dụ 23. Chiều cao của một loại cây lấy gỗ là biến ngẫu nhiên có phân phối chuẩn với chiều cao trung bình là 20m, độ lệch chuẩn là 2,5m. Cây đạt tiêu chuẩn khai thác là cây có chiều cao tối thiểu 15m.

Hãy tính tỷ lệ cây đạt tiêu chuẩn khai thác, nếu cây đạt tiêu chuẩn sẽ lãi 10000 đồng, ngược lại cây không đạt tiêu chuẩn lỗ 50.000 đồng. Người ta khai thác một lô 100 cây, tính tiền lãi trung bình cho lô cây đó.