BÀI GIẢNG

Lý thuyết xác suất thống kê (Probability and Statistics)

TS.Nguyễn Thị Loan

Bài toán

Anh A nghi là mình mắc một loại bệnh, anh ấy đi xét nghiệm thấy có kết quả dương tính. Anh ta rất lo lắng và tìm hiểu thì thấy tỉ lệ người dân mắc bệnh loại này là $1/1000$. Loại xét nghiệm mà anh A vừa thử, ai mắc bệnh khi xét nghiệm cũng ra phản ứng dương tính, nhưng tỉ lệ phản ứng dương tính nhầm là 5% (tức là trong những người không bị bệnh có 5% số người thử ra phản ứng dương tính). Với những thông tin này, bạn thử tính xem khả năng mắc bệnh của anh A là bao nhiêu và anh ấy nên làm thế nào?

Bài toán

Bạn đang có 100 triệu và được đề nghị ba dự án đầu tư trong cùng một thời gian với các ước tính như sau :

1. Dự án 1 : Nếu đầu tư đem về lợi nhuận 1 tỉ với xác suất 0.15 và thiệt hại 100 triệu với xác suất 0.85
2. Dự án 2 : Nếu đầu tư đem về lợi nhuận 100 triệu với xác suất 0.5 và lợi nhuận 50 triệu với xác suất 0.3 và thiệt hại 50 triệu với xác suất 0.2
3. Dự án 3 : Chắc chắn đem về lợi nhuận 40 triệu

Bạn sẽ chọn đầu tư dự án nào ?

Lịch sử ra đời

1. Lý thuyết xác suất ra đời vào nửa cuối thế kỷ 17, bắt nguồn từ những trao đổi thư từ với nhau của hai nhà toán học vĩ đại nước Pháp là Blaise Pascal và Pierre de Fermat về một số bài toán liên quan đến trò chơi may rủi.

Lịch sử ra đời

1. Lý thuyết xác suất ra đời vào nửa cuối thế kỷ 17, bắt nguồn từ những trao đổi thư từ với nhau của hai nhà toán học vĩ đại nước Pháp là Blaise Pascal và Pierre de Fermat về một số bài toán liên quan đến trò chơi may rủi.
2. Huygens, Bernoulli, De Moivre là những người có công đầu tiên sáng tạo ra cơ sở toán học của lý thuyết xác suất.

Lịch sử ra đời

1. Lý thuyết xác suất ra đời vào nửa cuối thế kỷ 17, bắt nguồn từ những trao đổi thư từ với nhau của hai nhà toán học vĩ đại nước Pháp là Blaise Pascal và Pierre de Fermat về một số bài toán liên quan đến trò chơi may rủi.
2. Huygens, Bernoulli, De Moivre là những người có công đầu tiên sáng tạo ra cơ sở toán học của lý thuyết xác suất.
3. Năm 1933, Kolmogorov đã thành công tuyệt vời trong việc tiên đề hóa lý thuyết xác suất khi ông cho ra đời cuốn sách "Foundations of the Theory of Probability" và kể từ đó thì giới toán học chính thức công nhận xác suất là một lĩnh vực toán học chặt chẽ.

Lịch sử ra đời

1. Lý thuyết xác suất ra đời vào nửa cuối thế kỷ 17, bắt nguồn từ những trao đổi thư từ với nhau của hai nhà toán học vĩ đại nước Pháp là Blaise Pascal và Pierre de Fermat về một số bài toán liên quan đến trò chơi may rủi.
2. Huygens, Bernoulli, De Moivre là những người có công đầu tiên sáng tạo ra cơ sở toán học của lý thuyết xác suất.
3. Năm 1933, Kolmogorov đã thành công tuyệt vời trong việc tiên đề hóa lý thuyết xác suất khi ông cho ra đời cuốn sách "Foundations of the Theory of Probability" và kể từ đó thì giới toán học chính thức công nhận xác suất là một lĩnh vực toán học chặt chẽ.

"Môn khoa học bắt đầu từ việc xem xét các trò chơi may rủi này sẽ hứa hẹn trở thành một đối tượng nghiên cứu quan trọng nhất của tri thức loài người." Laplace

"Môn khoa học bắt đầu từ việc xem xét các trò chơi may rủi này sẽ hứa hẹn trở thành một đối tượng nghiên cứu quan trọng nhất của tri thức loài người." Laplace
"Xác suất thống kê là một trong các quan điểm chủ chốt để xây dựng học vấn trong thời đại ngày nay." UNESCO

Đối tượng nghiên cứu

1. Hiện tượng tất yếu là hiện tượng mà nếu được thực hiện ở điều kiện giống nhau thì kết quả giống nhau.

Đối tượng nghiên cứu

1. Hiện tượng tất yếu là hiện tượng mà nếu được thực hiện ở điều kiện giống nhau thì kết quả giống nhau.
   Ví dụ : Đun nước đến $100^o C$ trong điều kiện bình thường thì nước sôi

Đối tượng nghiên cứu

1. Hiện tượng tất yếu là hiện tượng mà nếu được thực hiện ở điều kiện giống nhau thì kết quả giống nhau.
   Ví dụ : Đun nước đến 100^oC trong điều kiện bình thường thì nước sôi
2. Hiện tượng ngẫu nhiên là hiện tượng mà dù thực hiện ở điều kiện giống nhau nhưng kết quả có thể khác nhau

Đối tượng nghiên cứu

1. Hiện tượng tất yếu là hiện tượng mà nếu được thực hiện ở điều kiện giống nhau thì kết quả giống nhau.
   Ví dụ : Đun nước đến $100^o C$ trong điều kiện bình thường thì nước sôi
2. Hiện tượng ngẫu nhiên là hiện tượng mà dù thực hiện ở điều kiện giống nhau nhưng kết quả có thể khác nhau
   Ví dụ : Tung một đồng xu, ta không chắc chắn rằng mặt sấp hay mặt ngửa sẽ xuất hiện.

Đối tượng nghiên cứu

- Trong tự nhiên, xã hội : Xuất hiện nhiều hiện tượng không thể nói trước nó xảy ra hay không xảy ra khi thực hiện một lần quan sát, những hiện tượng này gọi là hiện tượng ngẫu nhiên.

Đối tượng nghiên cứu

- Trong tự nhiên, xã hội : Xuất hiện nhiều hiện tượng không thể nói trước nó xảy ra hay không xảy ra khi thực hiện một lần quan sát, những hiện tượng này gọi là hiện tượng ngẫu nhiên.
- Tuy nhiên, tiến hành quan sát nhiều lần một hiện tượng ngẫu nhiên (điều kiện như nhau) thì trong nhiều trường hợp ta có thể rút ra được những kết luận khoa học về hiện tượng này.

Đối tượng nghiên cứu

- Trong tự nhiên, xã hội : Xuất hiện nhiều hiện tượng không thể nói trước nó xảy ra hay không xảy ra khi thực hiện một lần quan sát, những hiện tượng này gọi là hiện tượng ngẫu nhiên.
- Tuy nhiên, tiến hành quan sát nhiều lần một hiện tượng ngẫu nhiên (điều kiện như nhau) thì trong nhiều trường hợp ta có thể rút ra được những kết luận khoa học về hiện tượng này.
- Xác suất là một bộ phận của toán học nghiên cứu các hiện tượng ngẫu nhiên. Lý thuyết xác suất nhằm tìm ra những quy luật trong những hiện tượng "tưởng chừng" như không có quy luật.

Mục lục

Chương 1. Xác suất/ 1.1. Biểu diễn dữ liệu

- Dữ liệu được tập hợp thông qua quan sát hoặc đo lường
- Dữ liệu có thể được biểu diễn bằng dạng số hoặc đồ thị.

Chương 1. Xác suất/ 1.1. Biểu diễn dữ liệu

- Dữ liệu được tập hợp thông qua quan sát hoặc đo lường
- Dữ liệu có thể được biểu diễn bằng dạng số hoặc đồ thị.

Ví dụ :

- bảng giá cổ phiếu ;
- tỉ giá hối đoái ;
- biểu đồ thể hiện sự phát triển kinh tế của một đất nước.

Chương 1. Xác suất/ 1.1. Biểu diễn dữ liệu

- Dữ liệu được tập hợp thông qua quan sát hoặc đo lường
- Dữ liệu có thể được biểu diễn bằng dạng số hoặc đồ thị.

Ví dụ :

- bảng giá cổ phiếu ;
- tỉ giá hối đoái ;
- biểu đồ thể hiện sự phát triển kinh tế của một đất nước.

Chúng ta nghiên cứu hai cách biểu diễn dữ liệu điển hình :

- *Stem-and-Leaf* (cây) ;
- *Histogram* (Dạng tần suất- biểu đồ).

Ví dụ

Đo lường sức căng của các tấm thép, người ta thu được dữ liệu chứa đựng 14 giá trị mẫu theo đơn vị kg/mm^2 như sau

$$89 \quad 84 \quad 87 \quad 81 \quad 89 \quad 86 \quad 91 \quad 90 \quad 78 \quad 89 \quad 87 \quad 99 \quad 83 \quad 89. \tag{1.1}$$

Lời giải Ta sắp xếp lại dữ liệu trên như sau

$$78 \quad 81 \quad 83 \quad 84 \quad 86 \quad 87 \quad 87 \quad 89 \quad 89 \quad 89 \quad 89 \quad 90 \quad 91 \quad 99. \tag{1.2}$$

Tiếp theo chúng ta chia các số này thành năm nhóm : $75-79$, $80-84$, $85-89$, $90-94$, $95-99$.

Key : $7|8 = 78$

CAF	Stem	Leaf
1	7	8
4	8	1 3 4
11	8	6 7 7 9 9 9 9
13	9	0 1
14	9	9

Figure – Biểu diễn dữ liệu dạng **Stem-and-Leaf**, CAF = Tần suất tuyệt đối tích lũy

Biểu diễn dữ liệu dạng tần suất (Histogram) : Trong trường hợp dữ liệu là một tập lớn, biểu diễn dữ liệu ở dạng tần suất sẽ hiển thị phân bố dữ liệu tốt hơn so với biểu diễn dữ liệu dạng **Stem-and-Leaf**.

Biểu diễn dữ liệu dạng tần suất (Histogram) : Trong trường hợp dữ liệu là một tập lớn, biểu diễn dữ liệu ở dạng tần suất sẽ hiển thị phân bố dữ liệu tốt hơn so với biểu diễn dữ liệu dạng **Stem-and-Leaf**.
Tần suất lớp tương đối (relative class frequency) $f_{\text{rel}}(x)$ của một lớp nhận x làm điểm chính giữa được tính như sau

$$f_{\text{rel}}(x) = \frac{\text{Số các phần tử của lớp có điểm chính giữa là } x}{\text{Tổng số phần tử của dữ liệu}}.$$

Biểu diễn dữ liệu dạng tần suất (Histogram) : Trong trường hợp dữ liệu là một tập lớn, biểu diễn dữ liệu ở dạng tần suất sẽ hiển thị phân bố dữ liệu tốt hơn so với biểu diễn dữ liệu dạng **Stem-and-Leaf**.

Tần suất lớp tương đối (relative class frequency) $f_{\text{rel}}(x)$ của một lớp nhận x làm điểm chính giữa được tính như sau

$$f_{\text{rel}}(x) = \frac{\text{Số các phần tử của lớp có điểm chính giữa là } x}{\text{Tổng số phần tử của dữ liệu}}.$$

- **Phương pháp điều tra chọn mẫu :**
  1. Chọn mẫu ngẫu nhiên (Chọn mẫu có xác suất)
  2. Chọn mẫu phi ngẫu nhiên (Chọn mẫu không xác suất)

- **Phương pháp điều tra chọn mẫu :**
  1. Chọn mẫu ngẫu nhiên (Chọn mẫu có xác suất)
  2. Chọn mẫu phi ngẫu nhiên (Chọn mẫu không xác suất)
- **Biểu diễn số liệu mẫu (bằng các loại biểu đồ) :**
  1. Số liệu thu gọn
  2. Số liệu thu gọn dạng khoảng

1.2.1. Một số khái niệm cơ bản trong lý thuyết xác xuất

- Thực nghiệm : Là một quá trình đo lường hay quan sát trong phòng thí nghiệm, nhà máy, hay bất cứ cái gì, bất cứ đâu.

1.2.1. Một số khái niệm cơ bản trong lý thuyết xác xuất

- Thực nghiệm : Là một quá trình đo lường hay quan sát trong phòng thí nghiệm, nhà máy, hay bất cứ cái gì, bất cứ đâu.
- Thực nghiệm là quá trình ngẫu nhiên và thay đổi nên ta không tiên đoán được kết quả một cách chính xác.

1.2.1. Một số khái niệm cơ bản trong lý thuyết xác xuất

- Thực nghiệm : Là một quá trình đo lường hay quan sát trong phòng thí nghiệm, nhà máy, hay bất cứ cái gì, bất cứ đâu.
- Thực nghiệm là quá trình ngẫu nhiên và thay đổi nên ta không tiên đoán được kết quả một cách chính xác.

Ví dụ 1.1. Một số các thực nghiệm ngẫu nhiên :

1. Tung một đồng xu
2. Gieo một con súc sắc
3. Đo điện thế hai đầu một mạch điện

1.2.1. Một số khái niệm cơ bản trong lý thuyết xác xuất

- Thực nghiệm : Là một quá trình đo lường hay quan sát trong phòng thí nghiệm, nhà máy, hay bất cứ cái gì, bất cứ đâu.
- Thực nghiệm là quá trình ngẫu nhiên và thay đổi nên ta không tiên đoán được kết quả một cách chính xác.

Ví dụ 1.1. Một số các thực nghiệm ngẫu nhiên :

1. Tung một đồng xu
2. Gieo một con súc sắc
3. Đo điện thế hai đầu một mạch điện

- Trong các thực nhiệm ngẫu nhiên, ta lặp đi lặp lại một nhóm các hành động để quan sát hiện tượng. Mỗi lần thực hiện một nhóm các hành động như thế gọi là một phép thử.

1.2.1. Một số khái niệm cơ bản trong lý thuyết xác xuất

- Thực nghiệm : Là một quá trình đo lường hay quan sát trong phòng thí nghiệm, nhà máy, hay bất cứ cái gì, bất cứ đâu.
- Thực nghiệm là quá trình ngẫu nhiên và thay đổi nên ta không tiên đoán được kết quả một cách chính xác.

Ví dụ 1.1. Một số các thực nghiệm ngẫu nhiên :

1. Tung một đồng xu
2. Gieo một con súc sắc
3. Đo điện thế hai đầu một mạch điện

- Trong các thực nhiệm ngẫu nhiên, ta lặp đi lặp lại một nhóm các hành động để quan sát hiện tượng. Mỗi lần thực hiện một nhóm các hành động như thế gọi là một phép thử.

Ví dụ 1.2 : Mỗi lần gieo một con xúc xắc hoặc tung một đồng xu là một phép thử.

- Kết quả của mỗi phép thử gọi là một kết cục.

- Kết quả của mỗi phép thử gọi là một kết cục.

Ví dụ 1.3. Một số phép thử :

1. Tung một đồng xu, kết cục là mặt sấp, mặt ngửa
2. Gieo một con súc sắc kết cục là có thể xuất hiện các điểm trên mặt con súc sắc 1,2,3,4,5,6

- Kết quả của mỗi phép thử gọi là một kết cục.

Ví dụ 1.3. Một số phép thử :

1. Tung một đồng xu, kết cục là mặt sấp, mặt ngửa
2. Gieo một con súc sắc kết cục là có thể xuất hiện các điểm trên mặt con súc sắc 1,2,3,4,5,6

- Tập tất cả các kết cục gọi là không gian mẫu (không gian xác suất). Kí hiệu : K

- Kết quả của mỗi phép thử gọi là một kết cục.

Ví dụ 1.3. Một số phép thử :

1. Tung một đồng xu, kết cục là mặt sấp, mặt ngửa
2. Gieo một con súc sắc kết cục là có thể xuất hiện các điểm trên mặt con súc sắc 1,2,3,4,5,6

- Tập tất cả các kết cục gọi là không gian mẫu (không gian xác suất). Kí hiệu : K
- Mỗi điểm trong không gian mẫu gọi là điểm mẫu.

- Kết quả của mỗi phép thử gọi là một kết cục.

Ví dụ 1.3. Một số phép thử :

1. Tung một đồng xu, kết cục là mặt sấp, mặt ngửa
2. Gieo một con súc sắc kết cục là có thể xuất hiện các điểm trên mặt con súc sắc 1,2,3,4,5,6

- Tập tất cả các kết cục gọi là không gian mẫu (không gian xác suất). Kí hiệu : K
- Mỗi điểm trong không gian mẫu gọi là điểm mẫu.

Ví dụ 1.4. Một số phép thử :

1. Tung một đồng xu, không gian mẫu là K={mặt sấp, mặt ngửa}
2. Gieo một con súc sắc không gian mẫu là K={1,2,3,4,5,6}

Sự kiện (biến cố)

+ Sự kiện (biến cố) : Là tập con của không gian mẫu. Nếu 1 phép thử, một kết cục a xuất hiện, $a \in A$ thì ta nói rằng sự kiện A xảy ra (xuất hiện).

Sự kiện (biến cố)

+ Sự kiện (biến cố) : Là tập con của không gian mẫu. Nếu 1 phép thử, một kết cục a xuất hiện, $a \in A$ thì ta nói rằng sự kiện A xảy ra (xuất hiện).
+ Sự kiện chắc chắn : Là sự kiện luôn luôn xảy ra, biến cố này trùng với không gian mẫu K.

Sự kiện (biến cố)

+ Sự kiện (biến cố) : Là tập con của không gian mẫu. Nếu 1 phép thử, một kết cục a xuất hiện, $a \in A$ thì ta nói rằng sự kiện A xảy ra (xuất hiện).
+ Sự kiện chắc chắn : Là sự kiện luôn luôn xảy ra, biến cố này trùng với không gian mẫu K.
+ Sự kiện không thể : Là sự kiện nhất định không xảy ra khi thực hiện phép thử. Tập $K = \emptyset$.

Ví dụ 1.5.

Ví dụ

1. Gieo một đồng tiền xu cân đối và đồng chất liên tiếp 2 lần. Đó là một phép thử với không gian mẫu :
   $K = \{SS; NN; NS; SN\};$

Ví dụ 1.5.

Ví dụ

1. Gieo một đồng tiền xu cân đối và đồng chất liên tiếp 2 lần. Đó là một phép thử với không gian mẫu :
   $K = \{SS; NN; NS; SN\}; \; |K| = 4$

Ví dụ 1.5.

Ví dụ

1. Gieo một đồng tiền xu cân đối và đồng chất liên tiếp 2 lần. Đó là một phép thử với không gian mẫu :
   $K = \{SS; NN; NS; SN\}$; $|K| = 4$
2. Gieo một con xúc xắc cân đối và đồng chất 2 lần. Đó là một phép thử với không gian mẫu :
   $K = \{11, 12, 13, 14, 15, 16; 21; \cdots; 65; 66\}$;

Ví dụ 1.5.

Ví dụ

1. Gieo một đồng tiền xu cân đối và đồng chất liên tiếp 2 lần. Đó là một phép thử với không gian mẫu :
$K = \{SS; NN; NS; SN\}$; $|K| = 4$
2. Gieo một con xúc xắc cân đối và đồng chất 2 lần. Đó là một phép thử với không gian mẫu :
$K = \{11, 12, 13, 14, 15, 16; 21; \cdots; 65; 66\}$; $|K| = 36$

Ví dụ 1.6.

Ví dụ

Bài toán gieo con súc sắc :

1. Sự kiện $A = \{1, 3, 5\}$
2. Sự kiện $B = \{2, 4, 6\}$
3. Sự kiện C : Số chấm bé hơn 5.
4. Sự kiện D : Số chấm lớn hơn 6.

Ví dụ 1.6.

Ví dụ

Bài toán gieo con súc sắc :

1. Sự kiện $A = \{1, 3, 5\}$
2. Sự kiện $B = \{2, 4, 6\}$
3. Sự kiện C : Số chấm bé hơn 5.
4. Sự kiện D : Số chấm lớn hơn 6.

Biểu diễn các sự kiện bằng sơ đồ Venn.

1.2.2. Các phép toán trên sự kiện

1. Quan hệ kéo theo

1.2.2. Các phép toán trên sự kiện

1. **Quan hệ kéo theo**
   Biến cố A kéo theo biến cố B, ký hiệu $A \subset B$, khi và chỉ khi A xảy ra thì suy ra B xảy ra.

1.2.2. Các phép toán trên sự kiện

1. **Quan hệ kéo theo**
   Biến cố A kéo theo biến cố B, ký hiệu $A \subset B$, khi và chỉ khi A xảy ra thì suy ra B xảy ra.
2. **Quan hệ tương đương**

1.2.2. Các phép toán trên sự kiện

1. **Quan hệ kéo theo**
   Biến cố A kéo theo biến cố B, ký hiệu $A \subset B$, khi và chỉ khi A xảy ra thì suy ra B xảy ra.
2. **Quan hệ tương đương**
   Hai biến cố A và B tương đương, ký hiệu $A = B$, khi và chỉ khi $A \subset B$ và $B \subset A$.

1.2.2. Các phép toán trên sự kiện

1. **Quan hệ kéo theo**
   Biến cố A kéo theo biến cố B, ký hiệu $A \subset B$, khi và chỉ khi A xảy ra thì suy ra B xảy ra.
2. **Quan hệ tương đương**
   Hai biến cố A và B tương đương, ký hiệu $A = B$, khi và chỉ khi $A \subset B$ và $B \subset A$.
3. **Tổng của hai biến cố**

1.2.2. Các phép toán trên sự kiện

1. **Quan hệ kéo theo**
   Biến cố A kéo theo biến cố B, ký hiệu $A \subset B$, khi và chỉ khi A xảy ra thì suy ra B xảy ra.
2. **Quan hệ tương đương**
   Hai biến cố A và B tương đương, ký hiệu $A = B$, khi và chỉ khi $A \subset B$ và $B \subset A$.
3. **Tổng của hai biến cố**
   Tổng của hai biến cố A và B, ký hiệu $A \cup B$, sao cho biến cố tổng $A \cup B$ xảy ra khi và chỉ khi hoặc A xảy ra hoặc B xảy ra (có ít nhất một trong hai biến cố A và B xảy ra).

Các quan hệ và các phép toán trên sự kiện (biến cố)

4. Tích của hai sự kiện (biến cố)

Các quan hệ và các phép toán trên sự kiện (biến cố)

4. Tích của hai sự kiện (biến cố)

Tích của hai biến cố A và B, ký hiệu $A \cap B$ hoặc AB, sao cho biến cố tích $A \cap B$ xảy ra khi và chỉ khi A xảy ra và B xảy ra (hai biến cố A và B cùng xảy ra).

Các quan hệ và các phép toán trên sự kiện (biến cố)

4. Tích của hai sự kiện (biến cố)
Tích của hai biến cố A và B, ký hiệu $A \cap B$ hoặc AB, sao cho biến cố tích $A \cap B$ xảy ra khi và chỉ khi A xảy ra và B xảy ra (hai biến cố A và B cùng xảy ra).

5. Hiệu của hai biến cố

Các quan hệ và các phép toán trên sự kiện (biến cố)

4. Tích của hai sự kiện (biến cố)
Tích của hai biến cố A và B, ký hiệu $A \cap B$ hoặc AB, sao cho biến cố tích $A \cap B$ xảy ra khi và chỉ khi A xảy ra và B xảy ra (hai biến cố A và B cùng xảy ra).

5. Hiệu của hai biến cố
Hiệu của biến cố A với biến cố B, ký hiệu $A \setminus B$, sao cho biến cố $A \setminus B$ xảy ra khi và chỉ khi A xảy ra, B không xảy ra.

Các quan hệ và các phép toán trên sự kiện (biến cố)

4. **Tích của hai sự kiện (biến cố)**
   Tích của hai biến cố A và B, ký hiệu $A \cap B$ hoặc AB, sao cho biến cố tích $A \cap B$ xảy ra khi và chỉ khi A xảy ra và B xảy ra (hai biến cố A và B cùng xảy ra).
5. **Hiệu của hai biến cố**
   Hiệu của biến cố A với biến cố B, ký hiệu $A \setminus B$, sao cho biến cố $A \setminus B$ xảy ra khi và chỉ khi A xảy ra, B không xảy ra.
6. **Hai biến cố xung khắc**

Các quan hệ và các phép toán trên sự kiện (biến cố)

4. **Tích của hai sự kiện (biến cố)**
   Tích của hai biến cố A và B, ký hiệu $A \cap B$ hoặc AB, sao cho biến cố tích $A \cap B$ xảy ra khi và chỉ khi A xảy ra và B xảy ra (hai biến cố A và B cùng xảy ra).
5. **Hiệu của hai biến cố**
   Hiệu của biến cố A với biến cố B, ký hiệu $A \setminus B$, sao cho biến cố $A \setminus B$ xảy ra khi và chỉ khi A xảy ra, B không xảy ra.
6. **Hai biến cố xung khắc**
   Hai biến cố A và B xung khắc nếu $AB = \emptyset$ (A và B không đồng thời xảy ra).

Các quan hệ và các phép toán trên sự kiện (biến cố)

4. **Tích của hai sự kiện (biến cố)**
   Tích của hai biến cố A và B, ký hiệu $A \cap B$ hoặc AB, sao cho biến cố tích $A \cap B$ xảy ra khi và chỉ khi A xảy ra và B xảy ra (hai biến cố A và B cùng xảy ra).
5. **Hiệu của hai biến cố**
   Hiệu của biến cố A với biến cố B, ký hiệu $A \setminus B$, sao cho biến cố $A \setminus B$ xảy ra khi và chỉ khi A xảy ra, B không xảy ra.
6. **Hai biến cố xung khắc**
   Hai biến cố A và B xung khắc nếu $AB = \emptyset$ (A và B không đồng thời xảy ra).
7. **Biến cố đối**

Các quan hệ và các phép toán trên sự kiện (biến cố)

4. **Tích của hai sự kiện (biến cố)**
   Tích của hai biến cố A và B, ký hiệu $A \cap B$ hoặc AB, sao cho biến cố tích $A \cap B$ xảy ra khi và chỉ khi A xảy ra và B xảy ra (hai biến cố A và B cùng xảy ra).
5. **Hiệu của hai biến cố**
   Hiệu của biến cố A với biến cố B, ký hiệu $A \setminus B$, sao cho biến cố $A \setminus B$ xảy ra khi và chỉ khi A xảy ra, B không xảy ra.
6. **Hai biến cố xung khắc**
   Hai biến cố A và B xung khắc nếu $AB = \emptyset$ (A và B không đồng thời xảy ra).
7. **Biến cố đối**
   Biến cố đối của biến cố A, ký hiệu $\bar{A}$, sao cho :
   $A\bar{A} = \emptyset$
   $A \cup \bar{A} = K$

Các quan hệ và các phép toán đối với các biến cố

Ví dụ

Ba xạ thủ A, B, C mỗi người bắn một viên đạn vào mục tiêu. Giả sử D là biến cố "Xạ thủ D bắn trúng" (D lần lượt là A, B, C).

1) Hãy mô tả các biến cố sau : $ABC; \bar{A}\bar{B}\bar{C}; A \cup B \cup C$
2) Biểu diễn các biến cố sau :
   a) Có ít nhất hai xạ thủ bắn trúng.
   b) Có nhiều nhất một xạ thủ bắn trúng.
   c) Chỉ có một xạ thủ bắn trúng.
   d) Chỉ có xạ thủ C bắn trúng.

1.3.1. Quy tắc cộng

Bài toán :
Để hoàn thành công việc A, có k phương án thực hiện :

- Phương án 1 : n_1 cách
- Phương án 2 : n_2 cách
- $\dots$
- Phương án k : n_k cách

Hỏi : Có bao nhiêu cách để hoàn thành công việc A ?

1.3.1. Quy tắc cộng

Bài toán :

Để hoàn thành công việc A, có k phương án thực hiện :

- Phương án 1 : n_1 cách
- Phương án 2 : n_2 cách
- $\dots$
- Phương án k : n_k cách

Hỏi : Có bao nhiêu cách để hoàn thành công việc A ?

KQ : $n_1 + n_2 + \dots + n_k$ (cách).

1.3.2. Quy tắc nhân

Bài toán :
Để hoàn thành công việc B, chia làm k giai đoạn :

- Giai đoạn 1 : n_1 cách
- Giai đoạn 2 : n_2 cách
- $\ldots$
- Giai đoạn k : n_k cách

Hỏi : Có bao nhiêu cách để hoàn thành công việc B ?

1.3.2. Quy tắc nhân

Bài toán :
Để hoàn thành công việc B, chia làm k giai đoạn :

- Giai đoạn 1 : n_1 cách
- Giai đoạn 2 : n_2 cách
- $\cdots$
- Giai đoạn k : n_k cách

Hỏi : Có bao nhiêu cách để hoàn thành công việc B ?
KQ : $n_1 n_2 \cdots n_k$ (cách).

1.3.3. Hoán vị

Bài toán :
Cho n phần tử, có bao nhiêu cách sắp xếp có thứ tự n phần tử đó ?

1.3.3. Hoán vị

Bài toán :
Cho n phần tử, có bao nhiêu cách sắp xếp có thứ tự n phần tử đó ?
Có 2 trường hợp :

1. n phần tử đôi một khác nhau.
2. n phần tử được chia thành h lớp khác nhau (mỗi lớp chứa các vật giống nhau).

1.3.3. Hoán vị

Bài toán :
Cho n phần tử, có bao nhiêu cách sắp xếp có thứ tự n phần tử đó ?
Có 2 trường hợp :

1. n phần tử đôi một khác nhau.
2. n phần tử được chia thành h lớp khác nhau (mỗi lớp chứa các vật giống nhau).

Hoán vị : của n phần tử là cách sắp xếp có thứ tự n phần tử đó.

1.3.3. Hoán vị

Bài toán :
Cho n phần tử, có bao nhiêu cách sắp xếp có thứ tự n phần tử đó ?
Có 2 trường hợp :

1. n phần tử đôi một khác nhau.
2. n phần tử được chia thành h lớp khác nhau (mỗi lớp chứa các vật giống nhau).

Hoán vị : của n phần tử là cách sắp xếp có thứ tự n phần tử đó.
Kết quả :

1. Số hoán vị : $P_n = n! = 1 \cdot 2 \cdot 3 \cdots n.$
2. Số hoán vị : $P_n = \frac{n!}{n_1! n_2! \cdots n_h!}$ $(n_1 + n_2 + \cdots + n_h = n).$

1.3.3. Hoán vị

Bài toán :
Cho n phần tử, có bao nhiêu cách sắp xếp có thứ tự n phần tử đó ?
Có 2 trường hợp :

1. n phần tử đôi một khác nhau.
2. n phần tử được chia thành h lớp khác nhau (mỗi lớp chứa các vật giống nhau).

Hoán vị : của n phần tử là cách sắp xếp có thứ tự n phần tử đó.
Kết quả :

1. Số hoán vị : $P_n = n! = 1 \cdot 2 \cdot 3 \cdots n.$
2. Số hoán vị : $P_n = \frac{n!}{n_1!n_2!\cdots n_h!}$ $(n_1 + n_2 + \cdots + n_h = n).$

Nhận xét

1. $0! = 1$
2. $(n+1)! = n!(n+1)$
3. Công thức Stirling : $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$

Ví dụ

Có bao nhiêu cách xếp 4 quyển sách lên giá ?

Ví dụ

Có bao nhiêu cách xếp 4 quyển sách lên giá ?

KQ : $4! = 24$ cách.

Ví dụ

Từ các số $3, 5, 7$ có thể lập được bao nhiêu số có 6 chữ số, trong đó số 3 xuất hiện 1 lần, số 5 xuất hiện 2 lần, số 7 xuất hiện 3 lần ?

Ví dụ

Có bao nhiêu cách xếp 4 quyển sách lên giá ?

KQ : $4! = 24$ cách.

Ví dụ

Từ các số $3, 5, 7$ có thể lập được bao nhiêu số có 6 chữ số, trong đó số 3 xuất hiện 1 lần, số 5 xuất hiện 2 lần, số 7 xuất hiện 3 lần ?

KQ : $\frac{6!}{1!2!3!} = 60$ số.

1.3.4. Chỉnh hợp

1. **Chỉnh hợp (không lặp)** : Chỉnh hợp chập k của $n (k \leq n)$ phần tử là một nhóm có thứ tự gồm k phần tử lấy từ n phần tử đã cho. Ký hiệu A_n^k

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

1.3.4. Chỉnh hợp

1. **Chỉnh hợp (không lặp)** : Chỉnh hợp chập k của $n(k \leq n)$ phần tử là một nhóm có thứ tự gồm k phần tử lấy từ n phần tử đã cho. Ký hiệu A_n^k

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

2. **Chỉnh hợp lặp** : Chỉnh hợp lặp chập k của $n(k \leq n)$ phần tử là một nhóm có thứ tự gồm k phần tử có thể giống nhau lấy từ n phần tử đã cho. Ký hiệu $\tilde{A}_n^k$

$$\tilde{A}_n^k = n^k$$

Ví dụ

Từ 3 chữ số 2,3,4, có thể tạo ra bao nhiêu số có hai chữ số trong các trường hợp sau :

1. Mỗi chữ số chỉ được chọn một lần ;
2. mỗi chữ số có thể được chọn nhiều lần.

Ví dụ

Từ 3 chữ số 2,3,4, có thể tạo ra bao nhiêu số có hai chữ số trong các trường hợp sau :

1. Mỗi chữ số chỉ được chọn một lần ;
2. mỗi chữ số có thể được chọn nhiều lần.

KQ :

1. Số các số có 2 chữ số mà mỗi chữ số đã cho chỉ được chọn 1 lần là $A_3^2 = 3.2 = 6$. Cụ thể 23, 24, 32, 34, 42, và 43.
2. Số các số có 2 chữ số mà mỗi chữ số đã cho có thể được chọn nhiều lần là $3^2 = 9$. Cụ thể 22, 23, 24, 32, 33, 34, 42, 43, và 44.

1.3.5. Tổ hợp

Tổ hợp chập k của n phần tử là một nhóm gồm k phần tử khác nhau không phân biệt thứ tự lấy từ n phần tử đã cho, với $k \leq n$. Ký hiệu C_n^k

1.3.5. Tổ hợp

Tổ hợp chập k của n phần tử là một nhóm gồm k phần tử khác nhau không phân biệt thứ tự lấy từ n phần tử đã cho, với $k \leq n$. Ký hiệu C_n^k Tóm lại

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!} \quad = \binom{n}{k}. \tag{1.3}$$

Ví dụ

Một tổ có 7 sinh viên, có bao nhiêu cách để chọn 3 bạn tham gia công tác tình nguyện ?
Giải : Nhóm tình nguyện gồm 3 người không phân biệt thứ tự được chọn từ 7 người, vì vậy ta có $C_7^3 = 35$ cách chọn.

1.4.1.Định nghĩa cổ điển của xác suất

Định nghĩa (Xác suất cổ điển)

Không gian mẫu (phép thử) K thỏa mãn hai điều kiện :

i) Không gian mẫu có một số hữu hạn phần tử;

ii) Các kết quả là đồng khả năng xảy ra,

thì xác suất cho một sự kiện A xuất hiện ký hiệu : $\mathbb{P}(A)$

1.4.1.Định nghĩa cổ điển của xác suất

Định nghĩa (Xác suất cổ điển)

Không gian mẫu (phép thử) K thỏa mãn hai điều kiện :

i) Không gian mẫu có một số hữu hạn phần tử;

ii) Các kết quả là đồng khả năng xảy ra,

thì xác suất cho một sự kiện A xuất hiện ký hiệu : $\mathbb{P}(A)$

$$\mathbb{P}(A) = \frac{\text{Số các điểm mẫu trong A}}{\text{Số các điểm mẫu trong K}} = \frac{cardA}{cardK}, \tag{1.4}$$

ở đó $cardM$ là số các điểm mẫu trong sự kiện M.

Ví dụ

Gieo một con súc sắc đồng nhất với xác suất xuất hiện các mặt là bằng nhau. Tìm xác suất để nhận các sự kiện sau :

1. $A = \{2, 4, 6\}$ (số điểm xuất hiện trên mặt con súc sắc là chẵn).
2. $B = \{1, 2\}$.

Ví dụ

Gieo một con súc sắc đồng nhất với xác suất xuất hiện các mặt là bằng nhau. Tìm xác suất để nhận các sự kiện sau :

1. $A = \{2, 4, 6\}$ (số điểm xuất hiện trên mặt con súc sắc là chẵn).
2. $B = \{1, 2\}$.

Giải : Không gian mẫu $K = \{1, 2, 3, 4, 5, 6\}$.

1. Số điểm trong A là 3 điểm, số điểm trong K là 6 điểm, vì thế $\mathbb{P}(A) = \frac{3}{6} = \frac{1}{2}$.
2. Số điểm trong B là 2 điểm, vì thế $\mathbb{P}(B) = \frac{2}{6} = \frac{1}{3}$.

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng.

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Số cách chọn : $|K| = C_{12}^2 = 66$

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Số cách chọn : $|K| = C_{12}^2 = 66$

Gọi A là biến cố : "Hai quả cầu được chọn là hai quả cầu trắng"

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Số cách chọn : $|K| = C_{12}^2 = 66$

Gọi A là biến cố : "Hai quả cầu được chọn là hai quả cầu trắng"

$$|A| = C_5^2 = 10$$

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Số cách chọn : $|K| = C_{12}^2 = 66$

Gọi A là biến cố : "Hai quả cầu được chọn là hai quả cầu trắng"

$$|A| = C_5^2 = 10$$

Xác suất của biến cố A : $\mathbb{P}(A) = \dfrac{|A|}{|K|} = \dfrac{10}{66} = \dfrac{5}{33}$

Công thức tính xác suất thông qua tần suất

$$\mathbb{P}_{\text{rel}}(A) = \frac{\text{Số lần sự kiện } A \text{ xuất hiện}}{\text{Số lần phép thử}}. \tag{1.5}$$

Công thức tính xác suất thông qua tần suất

$$\mathbb{P}_{\text{rel}}(A) = \frac{\text{Số lần sự kiện } A \text{ xuất hiện}}{\text{Số lần phép thử}}. \tag{1.5}$$

Công thức này cho một giá trị gần đúng và độ chính xác càng cao khi số lần phép thử được thực hiện càng lớn (tiến ra vô cùng). Khi đó, xác suất được định nghĩa bằng tần suất sẽ tiến về giá trị thật sự.

1.4.2. Định nghĩa tổng quát của xác suất

Định nghĩa (Định nghĩa tổng quát về xác suất)

Cho một không gian mẫu K, với mỗi sự kiện A sẽ có một số tương ứng là $\mathbb{P}(A)$ được gọi là *xác suất* để sự kiện A xuất hiện, sao cho :

1. $\mathbb{P}(A) = 0$ nếu $A = \emptyset$;
2. $\mathbb{P}(A) = 1$ nếu $A = K$;
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ nếu $A \cap B = \emptyset$.
   Nếu K có vô hạn các phần tử, tiên đề này được thay như sau
   $$\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots, \tag{1.6}$$
   ở đây các sự kiện A_1, A_2,..., A_n,... xung khắc từng đôi một, nghĩa là $A_i \cap A_j = \emptyset$ $\forall i \neq j$.

1.4.3. Các định lý cơ bản của xác suất

1. $0 \leq \mathbb{P}(A) \leq 1$; $\mathbb{P}(\emptyset) = 0$; $\mathbb{P}(K) = 1$;

1.4.3. Các định lý cơ bản của xác suất

1. $0 \leq \mathbb{P}(A) \leq 1$; $\mathbb{P}(\emptyset) = 0$; $\mathbb{P}(K) = 1$;
2. Quy tắc phần bù
$$\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A});$$

1.4.3. Các định lý cơ bản của xác suất

1. $0 \leq \mathbb{P}(A) \leq 1$; $\mathbb{P}(\emptyset) = 0$; $\mathbb{P}(K) = 1$;
2. Quy tắc phần bù
$$\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A});$$
3. Quy tắc cộng
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB);$$
$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(AB) - \mathbb{P}(AC) - \mathbb{P}(BC) + \mathbb{P}(ABC);$$

1.4.3. Các định lý cơ bản của xác suất

1. $0 \le \mathbb{P}(A) \le 1$; $\mathbb{P}(\emptyset) = 0$; $\mathbb{P}(K) = 1$;
2. Quy tắc phần bù
$$\mathbb{P}(A) = 1 - \mathbb{P}(\bar{A});$$
3. Quy tắc cộng
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB);$$
$$\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(AB) - \mathbb{P}(AC) - \mathbb{P}(BC) + \mathbb{P}(ABC);$$
4. Quy tắc cộng cho hai sự kiện xung khắc
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B).$$

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Lấy ngẫu nhiên đồng thời từ hộp ra 4 quả cầu. Tính xác suất để lấy được ít nhất một quả cầu trắng.

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Lấy ngẫu nhiên đồng thời từ hộp ra 4 quả cầu. Tính xác suất để lấy được ít nhất một quả cầu trắng.

Giải :

Phép thử : Lấy ngẫu nhiên đồng thời từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 4 quả cầu.

$$|K| = C_{12}^4 = 495$$

Gọi B là biến cố : "Lấy được ít nhất một quả cầu trắng"
Khi đó : $\bar{B}$ là biến cố : "4 quả cầu lấy ra đều là quả cầu đỏ"

$$|\bar{B}| = C_7^4 = 35$$

$$\Rightarrow \mathbb{P}(B) = 1 - \mathbb{P}(\bar{B}) = 1 - \frac{35}{495} = \frac{92}{99}$$

Ví dụ

Trong thực nghiệm tung một con súc sắc đồng nhất, xét các sự kiện sau

$$A \equiv \text{"Số điểm xuất hiện trên mặt con súc sắc là chẵn"},$$

$$B \equiv \text{"Số điểm xuất hiện trên mặt con súc sắc nhỏ hơn 5"}.$$

Tìm xác suất để sự kiện $A \cup B$ xuất hiện.

Ví dụ

Trong thực nghiệm tung một con súc sắc đồng nhất, xét các sự kiện sau

$A \equiv$ "Số điểm xuất hiện trên mặt con súc sắc là chẵn",

$B \equiv$ "Số điểm xuất hiện trên mặt con súc sắc nhỏ hơn 5".

Tìm xác suất để sự kiện $A \cup B$ xuất hiện.
Giải : Ta có

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{3}{6} + \frac{4}{6} - \frac{2}{6} = \frac{5}{6}.$$

1.5.1. Xác suất có điều kiện

Định nghĩa

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} \text{ với } \mathbb{P}(B) \neq 0. \qquad (1.3.1)$$

Nếu phép thử là hữu hạn, đồng khả năng thì :

$$\mathbb{P}(A \mid B) = \frac{|AB|}{|B|}$$

1.5.1. Xác suất có điều kiện

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng biết rằng 2 quả được chọn là 2 quả cùng màu.

1.5.1. Xác suất có điều kiện

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng biết rằng 2 quả được chọn là 2 quả cùng màu.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

1.5.1. Xác suất có điều kiện

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng biết rằng 2 quả được chọn là 2 quả cùng màu.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Gọi A là biến cố : "Hai quả cầu được chọn là hai quả cầu trắng"

1.5.1. Xác suất có điều kiện

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng biết rằng 2 quả được chọn là 2 quả cùng màu.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Gọi A là biến cố : "Hai quả cầu được chọn là hai quả cầu trắng"

$$|A| = C_5^2 = 10$$

1.5.1. Xác suất có điều kiện

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng biết rằng 2 quả được chọn là 2 quả cùng màu.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Gọi A là biến cố : "Hai quả cầu được chọn là hai quả cầu trắng"

$$|A| = C_5^2 = 10$$

Gọi B là biến cố : "Hai quả được chọn là hai quả cùng màu"

$$|B| = C_5^2 + C_7^2 = 31$$

1.5.1. Xác suất có điều kiện

Ví dụ

Một hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ. Chọn ngẫu nhiên đồng thời từ hộp ra 2 quả cầu. Tính xác suất để hai quả cầu được chọn là hai quả cầu trắng biết rằng 2 quả được chọn là 2 quả cùng màu.

Giải :

Phép thử : Chọn ngẫu nhiên đồng thời 2 quả cầu từ hộp đựng 5 quả cầu trắng và 7 quả cầu đỏ.

Gọi A là biến cố : "Hai quả cầu được chọn là hai quả cầu trắng"

$$|A| = C_5^2 = 10$$

Gọi B là biến cố : "Hai quả được chọn là hai quả cùng màu"

$$|B| = C_5^2 + C_7^2 = 31$$

Khi đó XS cần tìm là $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(AB)}{\mathbb{P}(B)} = \dfrac{\mathbb{P}(A)}{\mathbb{P}(B)} = \dfrac{|A|}{|B|}$ (do $A \subset B$)

Vậy $\mathbb{P}(A|B) = \dfrac{10}{31}$

Tính chất của xác suất có điều kiện

1. $0 \leq \mathbb{P}(A|B) \leq 1$

   $\mathbb{P}(A|B) = 0$ nếu $AB = \emptyset; \mathbb{P}(B|B) = \mathbb{P}(K|B) = 1$

2. $\mathbb{P}(A|B) = 1 - \mathbb{P}(\bar{A}|B)$

3. $\mathbb{P}(A \cup C|B) = \mathbb{P}(A|B) + \mathbb{P}(C|B)$ nếu $AC = \emptyset$

1.5.2. Công thức nhân xác suất. Các sự kiện độc lập

Công thức

$$\begin{aligned}\mathbb{P}(AB) &= \mathbb{P}(A).\mathbb{P}(B|A) \ \ (\mathbb{P}(A) > 0)\\ &= \mathbb{P}(B).\mathbb{P}(A|B) \ \ (\mathbb{P}(B) > 0)\end{aligned}$$

1.5.2. Công thức nhân xác suất. Các sự kiện độc lập

Công thức

$$\begin{aligned}\mathbb{P}(AB) &= \mathbb{P}(A).\mathbb{P}(B|A) \ \ (\mathbb{P}(A) > 0)\\ &= \mathbb{P}(B).\mathbb{P}(A|B) \ \ (\mathbb{P}(B) > 0)\end{aligned}$$

$$\begin{aligned}\mathbb{P}(ABC) &= \mathbb{P}(A).\mathbb{P}(B|A).\mathbb{P}(C|AB) \ \ (\mathbb{P}(AB) > 0)\\ &= \mathbb{P}(B).\mathbb{P}(C|B).\mathbb{P}(A|BC) \ \ (\mathbb{P}(BC) > 0)\\ &= \cdots\end{aligned}$$

Ví dụ

Hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ. Hộp II đựng 4 quả cầu trắng và 6 quả cầu đỏ. Từ hộp I lấy ngẫu nhiên đồng thời ra 3 quả cầu, sau đó lại lấy ngẫu nhiên ra 1 quả cầu trong số các quả cầu còn lại của hộp I. Từ hộp II lấy ngẫu nhiên đồng thời ra 2 quả cầu. Tính xác suất để :

a) Lấy được 4 quả cầu đỏ trong hai lần lấy ở hộp I

b) Lấy được ít nhất một quả cầu trắng trong lần lấy đầu ở hộp I và lần lấy ở hộp II

Giải :

a) Phép thử :
HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.
HĐ 2 : Trong số các quả cầu còn lại của hộp I lấy ngẫu nhiên ra 1 quả cầu.

Giải :

a) Phép thử :
HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.
HĐ 2 : Trong số các quả cầu còn lại của hộp I lấy ngẫu nhiên ra 1 quả cầu.
Gọi A là biến cố : "Lấy được 4 quả cầu đỏ trong hai lần lấy ở hộp I"

Giải :

a) Phép thử :
HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.
HĐ 2 : Trong số các quả cầu còn lại của hộp I lấy ngẫu nhiên ra 1 quả cầu.
Gọi A là biến cố : "Lấy được 4 quả cầu đỏ trong hai lần lấy ở hộp I"
C là biến cố : "Lấy được 3 quả cầu đỏ trong lần lấy đầu ở hộp I"

Giải :

a) Phép thử :
HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.
HĐ 2 : Trong số các quả cầu còn lại của hộp I lấy ngẫu nhiên ra 1 quả cầu.
Gọi A là biến cố : "Lấy được 4 quả cầu đỏ trong hai lần lấy ở hộp I"
C là biến cố : "Lấy được 3 quả cầu đỏ trong lần lấy đầu ở hộp I"
D là biến cố : "Lấy được 1 quả cầu đỏ trong lần lấy thứ hai ở hộp I"

Giải :

a) Phép thử :
HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.
HĐ 2 : Trong số các quả cầu còn lại của hộp I lấy ngẫu nhiên ra 1 quả cầu.
Gọi A là biến cố : "Lấy được 4 quả cầu đỏ trong hai lần lấy ở hộp I"
C là biến cố : "Lấy được 3 quả cầu đỏ trong lần lấy đầu ở hộp I"
D là biến cố : "Lấy được 1 quả cầu đỏ trong lần lấy thứ hai ở hộp I"
Khi đó : $\mathbb{P}(A) = \mathbb{P}(CD) = \mathbb{P}(C).\mathbb{P}(D|C) = \dfrac{C_7^3}{C_{12}^3}.\dfrac{C_4^1}{C_9^1} = \frac{7}{99}$

b) Phép thử :
HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.
HĐ 2 : Lấy ngẫu nhiên đồng thời từ hộp II đựng 4 quả cầu trắng và 6 quả cầu đỏ ra 2 quả cầu.

b) Phép thử :
HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.
HĐ 2 : Lấy ngẫu nhiên đồng thời từ hộp II đựng 4 quả cầu trắng và 6 quả cầu đỏ ra 2 quả cầu.
Gọi B là biến cố : "Lấy được ít nhất một quả cầu trắng"

b) Phép thử :
HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.
HĐ 2 : Lấy ngẫu nhiên đồng thời từ hộp II đựng 4 quả cầu trắng và 6 quả cầu đỏ ra 2 quả cầu.
Gọi B là biến cố : "Lấy được ít nhất một quả cầu trắng"
E là biến cố : "Lấy được 2 quả cầu đỏ trong hộp II "

b) Phép thử :

HĐ 1 : Lấy ngẫu nhiên đồng thời từ hộp I đựng 5 quả cầu trắng và 7 quả cầu đỏ ra 3 quả cầu.

HĐ 2 : Lấy ngẫu nhiên đồng thời từ hộp II đựng 4 quả cầu trắng và 6 quả cầu đỏ ra 2 quả cầu.

Gọi B là biến cố : "Lấy được ít nhất một quả cầu trắng"

E là biến cố : "Lấy được 2 quả cầu đỏ trong hộp II "

Khi đó : $\mathbb{P}(B) = 1 - \mathbb{P}(\bar{B}) = 1 - \mathbb{P}(CE) = 1 - \mathbb{P}(C).\mathbb{P}(E) = 1 - \dfrac{C_7^3}{C_{12}^3} \cdot \dfrac{C_6^2}{C_{10}^2}$

Các sự kiện độc lập

Định nghĩa

Hai biến cố A và B được gọi là độc lập với nhau nếu :

$$\mathbb{P}(AB) = \mathbb{P}(A).\mathbb{P}(B)$$

Các sự kiện độc lập

Định nghĩa

Hai biến cố A và B được gọi là độc lập với nhau nếu :

$$\mathbb{P}(AB) = \mathbb{P}(A).\mathbb{P}(B)$$

Nhận xét

Hai biến cố A và B độc lập thì :

1. $\mathbb{P}(A|B) = \mathbb{P}(A); \mathbb{P}(B|A) = \mathbb{P}(B)$
2. $A; \bar{B}$ độc lập với nhau

   $\bar{A}; B$ độc lập với nhau

   $\bar{A}; \bar{B}$ độc lập với nhau

Định nghĩa

Ba biến cố $A; B; C$ được gọi là độc lập (độc lập trong toàn thể) nếu :

$$\mathbb{P}(ABC) = \mathbb{P}(A).\mathbb{P}(B).\mathbb{P}(C)$$

$$\mathbb{P}(AB) = \mathbb{P}(A).\mathbb{P}(B)$$
$$\mathbb{P}(AC) = \mathbb{P}(A).\mathbb{P}(C)$$
$$\mathbb{P}(BC) = \mathbb{P}(B).\mathbb{P}(C)$$

1.6.1. Công thức xác suất toàn phần

Định nghĩa (Hệ đầy đủ các sự kiện)

Các biến cố $\{A_1; A_2; \cdots; A_n\}$ của một không gian mẫu K được gọi là *hệ đầy đủ* nếu :

1. $A_iA_j = \emptyset\ \forall i \neq j;\ i, j = 1; 2; \cdots; n$
2. $A_1 \cup A_2 \cup \cdots \cup A_n = K$

1.6.1. Công thức xác suất toàn phần

Định nghĩa (Hệ đầy đủ các sự kiện)

Các biến cố $\{A_1; A_2; \cdots; A_n\}$ của một không gian mẫu K được gọi là *hệ đầy đủ* nếu :

1. $A_i A_j = \emptyset \; \forall i \neq j;\; i, j = 1; 2; \cdots; n$
2. $A_1 \cup A_2 \cup \cdots \cup A_n = K$

Ví dụ

1. $\{A; \bar{A}\}$
2. Tung một con xúc xắc cân đối và đồng chất 1 lần, ta có nhóm đầy đủ các biến cố : $\{A_1; A_2; A_3\}$. Trong đó $A_1 = \{1; 2\}; A_2 = \{3; 4\}; A_3 = \{5; 6\}$

1.6.2. Công thức xác suất đầy đủ

Định lý (Công thức xác suất toàn phần (đầy đủ))

Cho nhóm đầy đủ các biến cố trong không gian mẫu K :
$\{A_1; A_2; \cdots; A_n\}$; $\mathbb{P}(A_i) > 0\ \forall i = 1; 2; \cdots; n$. B là biến cố bất kỳ trong K. Ta có :

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)$$

Ví dụ

Có hai chuồng thỏ. Chuồng thứ nhất có 10 thỏ trắng và 5 thỏ đen. Chuồng thứ hai có 3 thỏ trắng và 7 thỏ đen. Từ chuồng thứ nhất bắt ngẫu nhiên đồng thời 2 con thỏ cho vào chuồng thứ hai, sau đó lại bắt ngẫu nhiên một con thỏ ở chuồng thứ hai ra. Tính xác suất để con thỏ bắt được ở chuồng thứ hai là thỏ trắng.

Giải :

Gọi $A_1; A_2; A_3$ lần lượt là các biến cố hai con thỏ bắt được từ chuồng một là hai thỏ trắng, hai thỏ đen và một thỏ trắng một thỏ đen.

Ta có $\mathbb{P}(A_1) = \dfrac{C_{10}^2}{C_{15}^2} = \dfrac{3}{7}$; $\mathbb{P}(A_2) = \dfrac{C_5^2}{C_{15}^2} = \dfrac{2}{21}$; $\mathbb{P}(A_3) = \dfrac{C_{10}^1 . C_5^1}{C_{15}^2} = \dfrac{10}{21}$

Gọi B là biến cố thỏ bắt được ở chuồng hai là thỏ trắng.

Áp dụng công thức xác suất đầy đủ :

$$\mathbb{P}(B) = \mathbb{P}(A_1).\mathbb{P}(B|A_1) + \mathbb{P}(A_2).\mathbb{P}(B|A_2) + \mathbb{P}(A_3).\mathbb{P}(B|A_3)$$

Với $\mathbb{P}(B|A_1) = \dfrac{C_5^1}{C_{12}^1} = \dfrac{5}{12}$; $\mathbb{P}(B|A_2) = \dfrac{C_3^1}{C_{12}^1} = \dfrac{1}{4}$; $\mathbb{P}(B|A_3) = \dfrac{C_4^1}{C_{12}^1} = \dfrac{1}{3}$

Thay vào công thức, ta có :

$$\mathbb{P}(B) = \frac{3}{7}.\frac{5}{12} + \frac{2}{21}.\frac{1}{4} + \frac{10}{21}.\frac{1}{3} = \frac{13}{36}$$

1.6.2. Công thức Bayes (Công thức xác suất hậu kiểm)

Định lý (Công thức Bayes)

Hệ đầy đủ các sự kiện $\{A_1; A_2; \cdots; A_n\}$; $\mathbb{P}(A_i) > 0\ \forall i = 1; 2; \cdots; n$ của không gian mẫu K. B là biến cố bất kỳ, $\mathbb{P}(B) > 0$. Ta có :

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_iB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i).\mathbb{P}(B|A_i)}{\sum_{j=1}^{n} \mathbb{P}(A_j).\mathbb{P}(B|A_j)}.$$

1.6.2. Công thức Bayes (Công thức xác suất hậu kiểm)

Định lý (Công thức Bayes)

Hệ đầy đủ các sự kiện $\{A_1; A_2; \cdots; A_n\}$; $\mathbb{P}(A_i) > 0\ \forall i = 1; 2; \cdots; n$ của không gian mẫu K. B là biến cố bất kỳ, $\mathbb{P}(B) > 0$. Ta có :

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(A_iB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A_i).\mathbb{P}(B|A_i)}{\sum_{j=1}^{n} \mathbb{P}(A_j).\mathbb{P}(B|A_j)}.$$

Chú ý

- Các xác suất $\mathbb{P}(A_1)$, $\mathbb{P}(A_2)$,..., $\mathbb{P}(A_n)$: *xác suất tiên nghiệm.*
- Các xác suất $P(A_1|B)$, $P(A_2|B)$,...,$P(A_n|B)$: *xác suất hậu kiểm.*

Do đó công thức xác suất Bayes còn gọi là công thức xác suất hậu kiểm.

Ví dụ

Có hai chuồng thỏ. Chuồng thứ nhất có 10 thỏ trắng và 5 thỏ đen. Chuồng thứ hai có 3 thỏ trắng và 7 thỏ đen. Từ chuồng thứ nhất bắt ngẫu nhiên đồng thời 2 con thỏ cho vào chuồng thứ hai, sau đó lại bắt ngẫu nhiên một con thỏ ở chuồng thứ hai ra. Tính xác suất để hai con thỏ bắt được ở chuồng thứ nhất là hai thỏ trắng biết rằng thỏ bắt được ở chuồng thứ hai là thỏ trắng.

Ví dụ

Có hai chuồng thỏ. Chuồng thứ nhất có 10 thỏ trắng và 5 thỏ đen. Chuồng thứ hai có 3 thỏ trắng và 7 thỏ đen. Từ chuồng thứ nhất bắt ngẫu nhiên đồng thời 2 con thỏ cho vào chuồng thứ hai, sau đó lại bắt ngẫu nhiên một con thỏ ở chuồng thứ hai ra. Tính xác suất để hai con thỏ bắt được ở chuồng thứ nhất là hai thỏ trắng biết rằng thỏ bắt được ở chuồng thứ hai là thỏ trắng.

Giải :

Ta có :

$$\begin{aligned}\mathbb{P}(A_1|B) &= \frac{\mathbb{P}(A_1).\mathbb{P}(B|A_1)}{\sum_{i=1}^{3}\mathbb{P}(A_i).\mathbb{P}(B|A_i)} \\ &= \frac{\mathbb{P}(A_1).\mathbb{P}(B|A_1)}{\mathbb{P}(B)} \\ &= \frac{\frac{3}{7}\cdot\frac{5}{12}}{\frac{13}{36}} = \frac{45}{91}\end{aligned}$$

1.7. Khái niệm về biến ngẫu nhiên

Định nghĩa

Biến ngẫu nhiên là đại lượng ngẫu nhiên nhận giá trị với xác suất tương ứng nào đó.

1.7. Khái niệm về biến ngẫu nhiên

Định nghĩa

Biến ngẫu nhiên là đại lượng ngẫu nhiên nhận giá trị với xác suất tương ứng nào đó.

Ký hiệu biến ngẫu nhiên : $X, Y, Z, \cdots$
ký hiệu giá trị biến ngẫu nhiên nhận : $x, y, z, \cdots$

1.7. Khái niệm về biến ngẫu nhiên

Định nghĩa

Biến ngẫu nhiên là đại lượng ngẫu nhiên nhận giá trị với xác suất tương ứng nào đó.

Ký hiệu biến ngẫu nhiên : $X, Y, Z, \cdots$
ký hiệu giá trị biến ngẫu nhiên nhận : $x, y, z, \cdots$

Ví dụ

+ X là đại lượng mô tả số chấm khi tung một con súc sắc cân đối và đồng chất một lần.

1.7. Khái niệm về biến ngẫu nhiên

Định nghĩa

Biến ngẫu nhiên là đại lượng ngẫu nhiên nhận giá trị với xác suất tương ứng nào đó.

Ký hiệu biến ngẫu nhiên : $X, Y, Z, \cdots$
ký hiệu giá trị biến ngẫu nhiên nhận : $x, y, z, \cdots$

Ví dụ

+ X là đại lượng mô tả số chấm khi tung một con súc sắc cân đối và đồng chất một lần.
+ Y là đại lượng mô tả số lần tung một đồng xu (cân đối và đồng chất) cho tới khi mặt sấp xuất hiện thì dừng lại.

1.7. Khái niệm về biến ngẫu nhiên

Định nghĩa

Biến ngẫu nhiên là đại lượng ngẫu nhiên nhận giá trị với xác suất tương ứng nào đó.

Ký hiệu biến ngẫu nhiên : $X, Y, Z, \cdots$
ký hiệu giá trị biến ngẫu nhiên nhận : $x, y, z, \cdots$

Ví dụ

+ X là đại lượng mô tả số chấm khi tung một con súc sắc cân đối và đồng chất một lần.
+ Y là đại lượng mô tả số lần tung một đồng xu (cân đối và đồng chất) cho tới khi mặt sấp xuất hiện thì dừng lại.
+ T_1 là đại lượng mô tả thời gian đàm thoại của một thuê bao di động (đơn vị phút).

1.7. Khái niệm về biến ngẫu nhiên

Định nghĩa

Biến ngẫu nhiên là đại lượng ngẫu nhiên nhận giá trị với xác suất tương ứng nào đó.

Ký hiệu biến ngẫu nhiên : $X, Y, Z, \cdots$
ký hiệu giá trị biến ngẫu nhiên nhận : $x, y, z, \cdots$

Ví dụ

- X là đại lượng mô tả số chấm khi tung một con súc sắc cân đối và đồng chất một lần.
- Y là đại lượng mô tả số lần tung một đồng xu (cân đối và đồng chất) cho tới khi mặt sấp xuất hiện thì dừng lại.
- T_1 là đại lượng mô tả thời gian đàm thoại của một thuê bao di động (đơn vị phút).
- T_2 là đại lượng mô tả tuổi thọ của một linh kiện điện tử (đơn vị ngày).

Phân loại biến ngẫu nhiên

* Biến ngẫu nhiên rời rạc : Nhận giá trị $x_1; x_2; \cdots; x_n; \cdots$ (Tập giá trị của biến ngẫu nhiên rời rạc là một tập hữu hạn hoặc vô hạn đếm được)
  Ví dụ : Đếm số xe trên đường phố, đếm số sinh viên mặc áo xanh.

Phân loại biến ngẫu nhiên

* Biến ngẫu nhiên rời rạc : Nhận giá trị $x_1; x_2; \cdots; x_n; \cdots$ (Tập giá trị của biến ngẫu nhiên rời rạc là một tập hữu hạn hoặc vô hạn đếm được)
  Ví dụ : Đếm số xe trên đường phố, đếm số sinh viên mặc áo xanh.
* Biến ngẫu nhiên liên tục : Nhận giá trị $(a; b)$; $a < b$, hoặc $(a_1; b_1) \cup (a_2; b_2) \cup \cdots$, thậm chí $(-\infty; +\infty)$. (Tập giá trị của biến ngẫu nhiên liên tục là một tập vô hạn không đếm được)
  Ví dụ : Đo cường độ dòng điện, đo độ dài,...

1.8.1.Bảng phân phối xác suất, hàm phân phối xác suất của biến ngẫu nhiên rời rạc

Cho biến ngẫu nhiên rời rạc X nhận các giá trị $x_1; x_2; \cdots; x_n; \cdots$
Luật phân phối xác suất của X : $\mathbb{P}(X = x_i) = p_i$ với $i = 1; 2; \cdots; n; \cdots$

1.8.1.Bảng phân phối xác suất, hàm phân phối xác suất của biến ngẫu nhiên rời rạc

Cho biến ngẫu nhiên rời rạc X nhận các giá trị $x_1; x_2; \cdots; x_n; \cdots$
Luật phân phối xác suất của X : $\mathbb{P}(X = x_i) = p_i$ với $i = 1; 2; \cdots; n; \cdots$

Định nghĩa (Bảng phân phối xác suất)

X	x_1	x_2	$\cdots$	x_n	$\cdots$
$\mathbb{P}(X = x_i) = p_i$	p_1	p_2	$\cdots$	p_n	$\cdots$

1.8.1.Bảng phân phối xác suất, hàm phân phối xác suất của biến ngẫu nhiên rời rạc

Cho biến ngẫu nhiên rời rạc X nhận các giá trị $x_1; x_2; \cdots; x_n; \cdots$
Luật phân phối xác suất của X : $\mathbb{P}(X = x_i) = p_i$ với $i = 1; 2; \cdots; n; \cdots$

Định nghĩa (Bảng phân phối xác suất)

X	x_1	x_2	$\cdots$	x_n	$\cdots$
$\mathbb{P}(X = x_i) = p_i$	p_1	p_2	$\cdots$	p_n	$\cdots$

Trong đó : $p_i > 0 \ \forall i$ và $\sum_{i=1}^{\infty} p_i = 1$.

Định nghĩa (Hàm xác suất (hàm mật độ xác suất))

$$f(x) = \begin{cases} p_i, & x = x_i \\ 0, & \text{các trường hợp khác,} \end{cases} \qquad (i = 1, 2, 3, \cdots). \tag{1.7}$$

1.8.1.Bảng phân phối xác suất, hàm phân phối xác suất của biến ngẫu nhiên rời rạc

Định nghĩa (Hàm phân phối xác suất)

$$\mathbb{P}(X \leq x) \equiv F(x) = \sum_{x_i \leq x} f(x_i) = \sum_{x_i \leq x} p_i \tag{1.8}$$

1.8.1.Bảng phân phối xác suất, hàm phân phối xác suất của biến ngẫu nhiên rời rạc

Định nghĩa (Hàm phân phối xác suất)

$$\mathbb{P}(X \leq x) \equiv F(x) = \sum_{x_i \leq x} f(x_i) = \sum_{x_i \leq x} p_i \tag{1.8}$$

Định nghĩa (Xác suất tương ứng với một khoảng)

$$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \sum_{a < x_i \leq b} p_i. \tag{1.9}$$

Ví dụ

Tìm hàm xác suất và phân phối xác suất của BNN X trong đó X là số chấm xuất hiện khi thực nghiệm gieo con súc sắc đồng nhất.

Ví dụ

Tìm hàm xác suất và phân phối xác suất của BNN X trong đó X là số chấm xuất hiện khi thực nghiệm gieo con súc sắc đồng nhất.

Lời giải : Ta có

X	1	2	3	4	5	6
$f(x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
$F(x)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	1

Từ bảng trên ta có, hàm xác suất cần tìm là

$$f(x) = \begin{cases} \frac{1}{6} & \text{với } x \in \{1,2,3,4,5,6\} \\ 0 & \text{với } x \notin \{1,2,3,4,5,6\} \end{cases} \tag{1.10}$$

và hàm phân phối xác suất là

$$F(x) = \begin{cases} 0 & \text{với } x < 1, \\ \dfrac{1}{6} & \text{với } 1 \leq x < 2, \\ \dfrac{1}{3} & \text{với } 2 \leq x < 3, \\ \dfrac{1}{2} & \text{với } 3 \leq x < 4, \\ \dfrac{2}{3} & \text{với } 4 \leq x < 5, \\ \dfrac{5}{6} & \text{với } 5 \leq x < 6, \\ 1 & \text{với } x \geq 6. \end{cases}$$

Đồ thị của hàm xác suất $f(x)$ và hàm phân phối $F(x)$ được minh họa trong hình sau

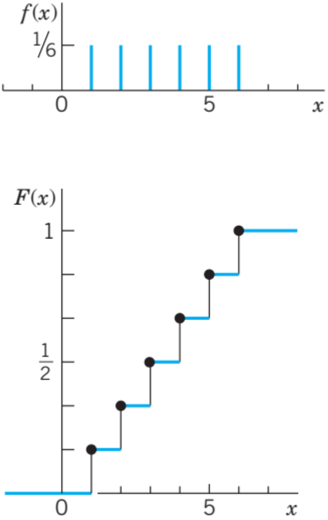

Figure – Hàm xác suất $f(x)$ và hàm phân bố $F(x)$

1.8.2. Hàm mật độ và hàm phân phối xác suất của biến ngẫu nhiên liên tục

$f(x)$ là hàm mật độ xác suất của biến ngẫu nhiên liên tục X :

i) $f(x) \geq 0$ với $\forall x \in \mathbb{R}$;

1.8.2. Hàm mật độ và hàm phân phối xác suất của biến ngẫu nhiên liên tục

$f(x)$ là hàm mật độ xác suất của biến ngẫu nhiên liên tục X :

i) $f(x) \geq 0$ với $\forall x \in \mathbb{R}$;

ii) $\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1.$

1.8.2. Hàm mật độ và hàm phân phối xác suất của biến ngẫu nhiên liên tục

$f(x)$ là hàm mật độ xác suất của biến ngẫu nhiên liên tục X :

i) $f(x) \geq 0$ với $\forall x \in \mathbb{R}$;

ii) $\int\limits_{-\infty}^{+\infty} f(x)\,dx = 1.$

Khi đó :

+ $f(X = a) = 0$ với $a \in \mathbb{R}$

+ $f(a < X < b) = f(a \leq X < b) = f(a < X \leq b) = f(a \leq X \leq b)$

$$= \int\limits_{a}^{b} f(x)\,dx.$$

1.8.2. Hàm mật độ và hàm phân phối xác suất của biến ngẫu nhiên liên tục

Định nghĩa (Hàm phân phối xác suất)

$$F(x) = f(X < x) = \int_{-\infty}^{x} f(t)\, dt, \tag{1.11}$$

trong đó $f(x)$: hàm mật độ xác suất và $f(x) = F'(x)$.

1.8.2. Hàm mật độ và hàm phân phối xác suất của biến ngẫu nhiên liên tục

Định nghĩa (Hàm phân phối xác suất)

$$F(x) = f(X < x) = \int_{-\infty}^{x} f(t)\, dt, \tag{1.11}$$

trong đó $f(x)$: hàm mật độ xác suất và $f(x) = F'(x)$.

Định nghĩa (Xác suất tương ứng với một khoảng)

$$\mathbb{P}(a < X \leq b) = F(b) - F(a) = \int_{a}^{b} f(v)dv. \tag{1.12}$$

Nhận xét

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b).$$

Nhận xét

$$\mathbb{P}(a < X < b) = \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X \leq b).$$

Ví dụ

Cho X là một BNN với hàm mật độ xác suất sau

$$f(x) = \begin{cases} 0.75(1-x^2), & -1 \leq x \leq 1 \\ 0, & \text{các trường hợp khác.} \end{cases}$$

1. Tìm hàm phân phối xác suất.
2. Tìm xác suất $\mathbb{P}(-\frac{1}{2} \leq X \leq \frac{1}{2})$.

Lời giải :

1. Hàm phân phối xác suất tương ứng với mật độ xác suất trên là

$$F(x) = \begin{cases} 0, & x < -1 \\ 0.75\displaystyle\int_{-1}^{x}(1-v^2)dv = 0.5 + 0.75x - 0.25x^3, & -1 \le x \le 1 \\ 1, & x > 1. \end{cases}$$

Lời giải :

1. Hàm phân phối xác suất tương ứng với mật độ xác suất trên là

$$F(x) = \begin{cases} 0, & x < -1 \\ 0.75\displaystyle\int_{-1}^{x}(1-v^2)dv = 0.5 + 0.75x - 0.25x^3, & -1 \le x \le 1 \\ 1, & x > 1. \end{cases}$$

2. Xác suất $\left(-\frac{1}{2} \le X \le \frac{1}{2}\right)$ được tính như sau

$$\mathbb{P}(-\frac{1}{2} \le X \le \frac{1}{2}) = F(\frac{1}{2}) - F(-\frac{1}{2}) = 68.75\%$$

Kỳ vọng (trung bình)

Định nghĩa

Kỳ vọng của biến ngẫu nhiên X, ký hiệu $\mu(X)$, là một số thực, được xác định như sau :

$$\mu(X) = \begin{cases} \sum_{i=1}^{+\infty} x_i.p_i & \text{nếu } X \text{ có } \mathbb{P}(X = x_i) = p_i;\ i = 1; 2; \cdots; \\ \int_{-\infty}^{+\infty} x.f(x)\,dx & \text{nếu } X \text{ có hàm mật độ xác suất } f(x) \end{cases}$$

Kỳ vọng (trung bình)

Định nghĩa

Kỳ vọng của biến ngẫu nhiên X, ký hiệu $\mu(X)$, là một số thực, được xác định như sau :

$$\mu(X) = \begin{cases} \sum\limits_{i=1}^{+\infty} x_i.p_i & \text{nếu } X \text{ có } \mathbb{P}(X = x_i) = p_i;\ i = 1; 2; \cdots; \\ \int\limits_{-\infty}^{+\infty} x.f(x)\,dx & \text{nếu } X \text{ có hàm mật độ xác suất } f(x) \end{cases}$$

Chú ý : Trong bài giảng này $\mu(X)$ luôn xác định, tức chuỗi số hoặc tích phân suy rộng trên là hữu hạn.

Kỳ vọng

Ý nghĩa : Kỳ vọng của biến ngẫu nhiên X :

i) Là giá trị trung bình mà biến ngẫu nhiên X nhận ;

ii) Là trọng tâm của phân phối có khối lượng 1.

Kỳ vọng

Ý nghĩa : Kỳ vọng của biến ngẫu nhiên X :

i) Là giá trị trung bình mà biến ngẫu nhiên X nhận ;

ii) Là trọng tâm của phân phối có khối lượng 1.

Một số tính chất :

i) $\mu(C) = C$ với C là hằng số;

ii) $\mu(CX) = C.\mu(X)$;

iii) $\mu(X+Y) = \mu(X) + \mu(Y)$;

iv) $\mu(X.Y) = \mu(X).\mu(Y)$ nếu X và Y là hai biến ngẫu nhiên độc lập.

v)

$$\mu(X^2) = \begin{cases} \sum_{i=1}^{\infty} x_i^2 p_i & \text{nếu } X \text{ có } \mathbb{P}(X = x_i) = p_i;\ i = 1; 2; \cdots; \\ \int_{-\infty}^{+\infty} x^2 f(x)\, dx & \text{nếu } X \text{ có hàm mật độ xác suất } f(x). \end{cases}$$

Phương sai

Định nghĩa

Phương sai của biến ngẫu nhiên X, ký hiệu $\sigma^2(X)$, là một số không âm, được xác định bởi công thức :

$$\sigma^2(X) = \mu(X - \mu(X))^2 = \mu(X^2) - (\mu(X))^2 .$$

Ý nghĩa : Phương sai là số đo mức độ phân tán giữa các giá trị của biến ngẫu nhiên X với giá trị trung bình $\mu(X)$.

Độ lệch chuẩn

Độ lệch tiêu chuẩn : $\sigma = \sqrt{\sigma^2(X)}$ là sự chênh lệch trung bình giữa các giá trị của biến ngẫu nhiên X so với giá trị trung bình $\mu(X)$.

Một số tính chất :

i) $\sigma^2(C) = 0$ với C là hằng số ;
ii) $\sigma^2(C.X) = C^2.\sigma^2(X)$;
iii) $\sigma^2(X \pm Y) = \sigma^2(X) + \sigma^2(Y)$ nếu X và Y là hai biến ngẫu nhiên độc lập.

Ví dụ

Cho thực nghiệm tung một đồng xu đồng nhất. Gọi biến ngẫu nhiên X là số mặt sấp (S) xuất hiện. Tính kỳ vọng, phương sai của biến ngẫu nhiên này.

Ví dụ

Cho thực nghiệm tung một đồng xu đồng nhất. Gọi biến ngẫu nhiên X là số mặt sấp (S) xuất hiện. Tính kỳ vọng, phương sai của biến ngẫu nhiên này.

Lời giải : Các giá trị có thể của X là 0 (nếu mặt ngửa (N) xuất hiện) và 1 (nếu mặt sấp xuất hiện).

Ví dụ

Cho thực nghiệm tung một đồng xu đồng nhất. Gọi biến ngẫu nhiên X là số mặt sấp (S) xuất hiện. Tính kỳ vọng, phương sai của biến ngẫu nhiên này.

Lời giải : Các giá trị có thể của X là 0 (nếu mặt ngửa (N) xuất hiện) và 1 (nếu mặt sấp xuất hiện).
Do đồng xu đồng nhất nên ta có các xác suất

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

Ví dụ

Cho thực nghiệm tung một đồng xu đồng nhất. Gọi biến ngẫu nhiên X là số mặt sấp (S) xuất hiện. Tính kỳ vọng, phương sai của biến ngẫu nhiên này.

Lời giải : Các giá trị có thể của X là 0 (nếu mặt ngửa (N) xuất hiện) và 1 (nếu mặt sấp xuất hiện).
Do đồng xu đồng nhất nên ta có các xác suất

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

Bằng cách áp dụng các công thức trên cho phân bố rời rạc, chúng ta có thể tính kỳ vọng và phương sai như sau

$$\mu = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2},$$

Ví dụ

Cho thực nghiệm tung một đồng xu đồng nhất. Gọi biến ngẫu nhiên X là số mặt sấp (S) xuất hiện. Tính kỳ vọng, phương sai của biến ngẫu nhiên này.

Lời giải : Các giá trị có thể của X là 0 (nếu mặt ngửa (N) xuất hiện) và 1 (nếu mặt sấp xuất hiện).
Do đồng xu đồng nhất nên ta có các xác suất

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{2}.$$

Bằng cách áp dụng các công thức trên cho phân bố rời rạc, chúng ta có thể tính kỳ vọng và phương sai như sau

$$\mu = 0 \times \frac{1}{2} + 1 \times \frac{1}{2} = \frac{1}{2},$$

$$\sigma^2 = (0 - \frac{1}{2})^2 \times \frac{1}{2} + (1 - \frac{1}{2})^2 \times \frac{1}{2} = \frac{1}{4}.$$

Ví dụ

Cho hàm mật độ xác suất của một biến ngẫu nhiên X như sau

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{khác.} \end{cases}$$

Tính kỳ vọng và phương sai của phân bố này.

Ví dụ

Cho hàm mật độ xác suất của một biến ngẫu nhiên X như sau

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{khác.} \end{cases}$$

Tính kỳ vọng và phương sai của phân bố này.

Lời giải : Bằng cách áp dụng các công thức trên cho phân phối liên tục, chúng ta có thể tính kỳ vọng và phương sai như sau

$$\mu = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2},$$

Ví dụ

Cho hàm mật độ xác suất của một biến ngẫu nhiên X như sau

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{khác.} \end{cases}$$

Tính kỳ vọng và phương sai của phân bố này.

Lời giải : Bằng cách áp dụng các công thức trên cho phân phối liên tục, chúng ta có thể tính kỳ vọng và phương sai như sau

$$\mu = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2},$$

$$\sigma^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} dx = \frac{(b-a)^2}{12}.$$

Định lý (Kỳ vọng của phân phối đối xứng)

Nếu một phân bố X đối xứng tương ứng với $x = c$, có nghĩa là $f(x-c) = f(x+c)$ với mọi x, thì $\mu = c$.

Định lý (Kỳ vọng của phân phối đối xứng)

Nếu một phân bố X đối xứng tương ứng với $x = c$, có nghĩa là $f(x-c) = f(x+c)$ với mọi x, thì $\mu = c$.

Định lý (Phép biến đổi của kỳ vọng và phương sai)

Cho biến ngẫu nhiên X có kỳ vọng μ và phương sai σ^2. Khi đó

1. Biến ngẫu nhiên $X^* = a_1 + a_2 X$ $(a_2 > 0)$ có kỳ vọng μ^* và phương sai σ^{*2} được tính bởi
$$\mu^* = a_1 + a_2\mu \quad \sigma^{*2} = a_2^2\sigma^2.$$

Định lý (Kỳ vọng của phân phối đối xứng)

Nếu một phân bố X đối xứng tương ứng với $x = c$, có nghĩa là $f(x - c) = f(x + c)$ với mọi x, thì $\mu = c$.

Định lý (Phép biến đổi của kỳ vọng và phương sai)

Cho biến ngẫu nhiên X có kỳ vọng μ và phương sai σ^2. Khi đó

1. Biến ngẫu nhiên $X^* = a_1 + a_2 X$ $(a_2 > 0)$ có kỳ vọng μ^* và phương sai σ^{*2} được tính bởi
$$\mu^* = a_1 + a_2\mu \quad \sigma^{*2} = a_2^2\sigma^2.$$
2. Đặc biệt, BNN chuẩn hóa (standardized random variable) Z tương ứng với X, có nghĩa,
$$Z = \frac{X - \mu}{\sigma}$$
có kỳ vọng $\mu_Z = 0$ và phương sai $\sigma_Z^2 = 1$.

1.10.1. Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n, p)$

Định nghĩa (Phép thử Bernoulli)

Phép thử Bernoulli là phép thử ngẫu nhiên mà nó có thể nhận một trong hai kết quả thành công hay thất bại, trong đó xác suất thành công giống nhau mỗi khi phép thử này được tiến hành.

1.10.1.Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n, p)$

Định nghĩa (Phép thử Bernoulli)

Phép thử Bernoulli là phép thử ngẫu nhiên mà nó có thể nhận một trong hai kết quả thành công hay thất bại, trong đó xác suất thành công giống nhau mỗi khi phép thử này được tiến hành.

Xét dãy n phép thử Bernoulli, A là biến cố liên quan đến phép thử, xác suất để biến cố A xuất hiện tại mỗi phép thử là p. Gọi X là số lần biến cố A xuất hiện trong n phép thử Bernoulli. Khi đó : $X \simeq B(n, p)$.

1.10.1.Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n,p)$

Định nghĩa (Phép thử Bernoulli)

Phép thử Bernoulli là phép thử ngẫu nhiên mà nó có thể nhận một trong hai kết quả thành công hay thất bại, trong đó xác suất thành công giống nhau mỗi khi phép thử này được tiến hành.

Xét dãy n phép thử Bernoulli, A là biến cố liên quan đến phép thử, xác suất để biến cố A xuất hiện tại mỗi phép thử là p. Gọi X là số lần biến cố A xuất hiện trong n phép thử Bernoulli. Khi đó : $X \simeq B(n,p)$.
Luật phân phối xác suất của X :

$$\mathbb{P}(X=m) = C_n^m p^m (1-p)^{n-m} \text{ với } m = 0; 1; 2; \cdots, n.$$

1.10.1.Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n,p)$

Định nghĩa (Phép thử Bernoulli)

Phép thử Bernoulli là phép thử ngẫu nhiên mà nó có thể nhận một trong hai kết quả thành công hay thất bại, trong đó xác suất thành công giống nhau mỗi khi phép thử này được tiến hành.

Xét dãy n phép thử Bernoulli, A là biến cố liên quan đến phép thử, xác suất để biến cố A xuất hiện tại mỗi phép thử là p. Gọi X là số lần biến cố A xuất hiện trong n phép thử Bernoulli. Khi đó : $X \simeq B(n,p)$.
Luật phân phối xác suất của X :

$$\mathbb{P}(X=m) = C_n^m p^m (1-p)^{n-m} \text{ với } m = 0;1;2;\cdots,n.$$

Bảng phân phối xác suất của X

X	0	1	$\cdots$	n
$\mathbb{P}(X=x_i)=p_i$	$C_n^0 p^0 (1-p)^n$	$C_n^1 p^1 (1-p)^{n-1}$	$\cdots$	$C_n^n p^n (1-p)^0$

1.10.1.Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n, p)$

- Hàm xác suất tương ứng được cho như sau

$$f(x) = P(X = x) = C_n^x p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x} \quad (x = 0, 1, 2, \cdots, n). \tag{1.13}$$

1.10.1. Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n,p)$

- Hàm xác suất tương ứng được cho như sau

$$f(x) = P(X = x) = C_n^x p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x} \quad (x = 0, 1, 2, \cdots, n). \tag{1.13}$$

- Phân phối xác suất tương ứng với hàm xác suất này được gọi là *phân phối nhị thức* (binomial distribution) hay *phân phối Bernoulli* (Bernoulli distribution).

1.10.1.Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n,p)$

- Hàm xác suất tương ứng được cho như sau

$$f(x) = P(X = x) = C_n^x p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x} \quad (x = 0, 1, 2, \cdots, n). \tag{1.13}$$

- Phân phối xác suất tương ứng với hàm xác suất này được gọi là *phân phối nhị thức* (binomial distribution) hay *phân phối Bernoulli* (Bernoulli distribution).
- Các tham số đặc trưng của phân phối : $\mu(X) = np$ và $\sigma^2(X) = np(1-p)$.

1.10.1.Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n,p)$

- Hàm xác suất tương ứng được cho như sau

$$f(x) = P(X = x) = C_n^x p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x} \quad (x = 0, 1, 2, \cdots, n). \tag{1.13}$$

- Phân phối xác suất tương ứng với hàm xác suất này được gọi là *phân phối nhị thức* (binomial distribution) hay *phân phối Bernoulli* (Bernoulli distribution).
- Các tham số đặc trưng của phân phối : $\mu(X) = np$ và $\sigma^2(X) = np(1-p)$.
- Các tham số đặc trưng của phân phối nhị thức :
  $\mu(X) = np$ và $\sigma^2(X) = np(1-p)$.

1.10.1.Dãy phép thử độc lập Bernoulli và Phân phối nhị thức $X \sim B(n, p)$

- Hàm xác suất tương ứng được cho như sau

$$f(x) = P(X = x) = C_n^x p^x (1-p)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x} \quad (x = 0, 1, 2, \cdots, n). \tag{1.13}$$

- Phân phối xác suất tương ứng với hàm xác suất này được gọi là *phân phối nhị thức* (binomial distribution) hay *phân phối Bernoulli* (Bernoulli distribution).
- Các tham số đặc trưng của phân phối : $\mu(X) = np$ và $\sigma^2(X) = np(1-p)$.
- Các tham số đặc trưng của phân phối nhị thức : $\mu(X) = np$ và $\sigma^2(X) = np(1-p)$.
- **Đặc biệt** : Trong trường hợp *đối xứng* của khả năng xuất hiện và không xuất hiện của sự kiện A ($p = \frac{1}{2}$), trung bình $\mu = \frac{n}{2}$, phương sai $\sigma^2 = \frac{n}{4}$ và hàm xác suất là

$$f(x) = P(X = x) = C_n^x \left(\frac{1}{2}\right)^n \quad (x = 0, 1, \ldots, n).$$

Ví dụ

Tính xác suất để nhận được ít nhất hai mặt sáu điểm (mặt sáu điểm xuất hiện ít nhất hai lần) trong thực nghiệm tung con súc sắc đồng nhất bốn lần.

Ví dụ

Tính xác suất để nhận được ít nhất hai mặt sáu điểm (mặt sáu điểm xuất hiện ít nhất hai lần) trong thực nghiệm tung con súc sắc đồng nhất bốn lần.

Lời giải : Ta có hàm xác suất tương ứng với thực nghiệm này như sau

$$f(x) = C_4^x \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{4-x} \quad (x = 0, 1, 2, 3, 4).$$

Ví dụ

Tính xác suất để nhận được ít nhất hai mặt sáu điểm (mặt sáu điểm xuất hiện ít nhất hai lần) trong thực nghiệm tung con súc sắc đồng nhất bốn lần.

Lời giải : Ta có hàm xác suất tương ứng với thực nghiệm này như sau

$$f(x) = C_4^x \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{4-x} \quad (x = 0, 1, 2, 3, 4).$$

Sử dụng hàm xác suất này, chúng ta có thể tính được xác suất nhận được ít nhất hai mặt sáu điểm như sau

$$\mathbb{P}[X \geq 2] = f(2) + f(3) + f(4) = \frac{171}{1296} = 13.2\%.$$

1.10.3.Phân phối chuẩn, phân phối chuẩn tắc

Định nghĩa

Phân phối chuẩn (normal distribution) hay *phân phối Gauss* (Gauss distribution) là phân phối liên tục.
Ký hiệu : $X \simeq N(\mu; \sigma^2);\ \sigma > 0$

1.10.3.Phân phối chuẩn, phân phối chuẩn tắc

Định nghĩa

Phân phối chuẩn (normal distribution) hay *phân phối Gauss* (Gauss distribution) là phân phối liên tục.
Ký hiệu : $X \simeq N(\mu;\sigma^2);\ \sigma > 0$

Nhận xét

Đây là phân phối liên tục quan trọng nhất vì trong nhiều ứng dụng, các biến ngẫu nhiên hoặc có phân phối chuẩn hoặc được xấp xỉ bởi phân phối chuẩn hoặc có thể biến đổi về phân phối chuẩn.

1.10.3.Phân phối chuẩn, phân phối chuẩn tắc

Định nghĩa

Phân phối chuẩn (normal distribution) hay *phân phối Gauss* (Gauss distribution) là phân phối liên tục.
Ký hiệu : $X \simeq N(\mu; \sigma^2)$; $\sigma > 0$

Nhận xét

Đây là phân phối liên tục quan trọng nhất vì trong nhiều ứng dụng, các biến ngẫu nhiên hoặc có phân phối chuẩn hoặc được xấp xỉ bởi phân phối chuẩn hoặc có thể biến đổi về phân phối chuẩn.

Hàm mật độ xác suất của X :

$$\mathbb{P}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \qquad \forall x \in \mathbb{R}$$

1.10.3.Phân phối chuẩn, phân phối chuẩn tắc

Định nghĩa

Phân phối chuẩn (normal distribution) hay *phân phối Gauss* (Gauss distribution) là phân phối liên tục.
Ký hiệu : $X \simeq N(\mu; \sigma^2)$; $\sigma > 0$

Nhận xét

Đây là phân phối liên tục quan trọng nhất vì trong nhiều ứng dụng, các biến ngẫu nhiên hoặc có phân phối chuẩn hoặc được xấp xỉ bởi phân phối chuẩn hoặc có thể biến đổi về phân phối chuẩn.

Hàm mật độ xác suất của X :

$$\mathbb{P}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-\mu)^2}{2\sigma^2}} \qquad \forall x \in \mathbb{R}$$

Các tham số đặc trưng :

$$E(X) = \mu \text{ và } D(X) = \sigma^2.$$

Hàm phân bố tương ứng được cho như sau

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2} dv. \tag{1.14}$$

Hàm phân bố tương ứng được cho như sau

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int\limits_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2} dv. \tag{1.14}$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right). \tag{1.15}$$

Hàm phân bố tương ứng được cho như sau

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\left(\frac{v-\mu}{\sigma}\right)^2} dv. \quad (1.14)$$

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right). \quad (1.15)$$

Trong trường hợp trung bình $\mu = 0$ và độ lệch chuẩn $\sigma = 1$, phân bố chuẩn được gọi là *phân bố chuẩn tắc*, được ký hiệu là $\Phi(z)$ và được cho như sau

$$F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du. \quad (1.16)$$

Tính chất quan trọng của hàm Φ :

Tính chất quan trọng của hàm Φ :

1. $\Phi(+\infty) = 1$ hay $\displaystyle\int_{-\infty}^{\infty} e^{-u^2/2} du = \sqrt{2\pi}$.

Tính chất quan trọng của hàm Φ :

1. $\Phi(+\infty) = 1$ hay $\int_{-\infty}^{\infty} e^{-u^2/2} du = \sqrt{2\pi}$.
2. Với mọi x,
$$\Phi(-x) = 1 - \Phi(x).$$

Tính chất quan trọng của hàm Φ :

1. $\Phi(+\infty) = 1$ hay $\int_{-\infty}^{\infty} e^{-u^2/2} du = \sqrt{2\pi}$.
2. Với mọi x,
$$\Phi(-x) = 1 - \Phi(x).$$
3. $$P(a < X \leq b) = F(b) - F(a) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right). \tag{1.17}$$

Ví dụ

Cho BNN X có phân bố chuẩn với trung bình $\mu = 5$ và phương sai $\sigma^2 = 0.04$. Tìm c và k sao cho

1. $\mathbb{P}(X \leq c) = 95\%$.
2. $\mathbb{P}(5 - k \leq X \leq 5 + k) = 90\%$.
3. $\mathbb{P}(X \geq c) = 1\%$.

Ví dụ

Cho BNN X có phân bố chuẩn với trung bình $\mu = 5$ và phương sai $\sigma^2 = 0.04$. Tìm c và k sao cho

1. $\mathbb{P}(X \leq c) = 95\%$.
2. $\mathbb{P}(5 - k \leq X \leq 5 + k) = 90\%$.
3. $\mathbb{P}(X \geq c) = 1\%$.

Lời giải. (a) Ta có $\sigma = 0.2$ và

$$\mathbb{P}(X \leq c) = 95\% \Leftrightarrow \Phi\left(\frac{c-5}{0.2}\right) = 95\% \Leftrightarrow \frac{c-5}{0.2} = 1.654 \Leftrightarrow c = 5.329 \quad (A8).$$

Ví dụ

Cho BNN X có phân bố chuẩn với trung bình $\mu = 5$ và phương sai $\sigma^2 = 0.04$. Tìm c và k sao cho

1. $\mathbb{P}(X \leq c) = 95\%$.
2. $\mathbb{P}(5 - k \leq X \leq 5 + k) = 90\%$.
3. $\mathbb{P}(X \geq c) = 1\%$.

Lời giải. (a) Ta có $\sigma = 0.2$ và

$$\mathbb{P}(X \leq c) = 95\% \Leftrightarrow \Phi\left(\frac{c-5}{0.2}\right) = 95\% \Leftrightarrow \frac{c-5}{0.2} = 1.654 \Leftrightarrow c = 5.329 \quad (A8).$$

(b) Ta có

$$\begin{aligned}\mathbb{P}(5 - k \leq X \leq 5 + k) &= \Phi\left(\frac{(5+k)-5}{0.2}\right) - \Phi\left(\frac{(5-k)-5}{0.2}\right)\\ &= \Phi(k/0.2) - \Phi(-k/0.2) = 2\Phi(k/0.2) - 1.\end{aligned}$$

Ta có $\Phi(k/0.2) = 95\% \Leftrightarrow k = 1.645 \times 0.2 = 0.329 \quad (A8)$.

(c) Ta có

$$\mathbb{P}(X \leq c) = 1 - \mathbb{P}(X > c) = 1 - P(X \geq c) = 99\%.$$

Tương tự câu (a),

$$(c-5)/0.2 = 2.326 \Leftrightarrow c = 5.465 \quad (A8).$$

Chương 2. Ước lượng tham số

Từ chương này chúng ta bắt đầu nghiên cứu thống kê (toán học) : Rút ra những kết luận quan trọng từ bảng dữ liệu thô.

Chương 2. Ước lượng tham số

Từ chương này chúng ta bắt đầu nghiên cứu thống kê (toán học) : Rút ra những kết luận quan trọng từ bảng dữ liệu thô.
Mục đích chính của thống kê :

Chương 2. Ước lượng tham số

Từ chương này chúng ta bắt đầu nghiên cứu thống kê (toán học) : Rút ra những kết luận quan trọng từ bảng dữ liệu thô.
Mục đích chính của thống kê :

- mô tả số liệu ;
- ước lượng và dự đoán các đại lượng ;
- tìm ra các mối quan hệ giữa các đại lượng ;
- kiểm định các giả thuyết.

Chương 2. Ước lượng tham số

Từ chương này chúng ta bắt đầu nghiên cứu thống kê (toán học) : Rút ra những kết luận quan trọng từ bảng dữ liệu thô.
Mục đích chính của thống kê :

- mô tả số liệu ;
- ước lượng và dự đoán các đại lượng ;
- tìm ra các mối quan hệ giữa các đại lượng ;
- kiểm định các giả thuyết.

Các phương pháp quan trọng của thống kê :

- ước lượng điểm của các tham số đặc trưng θ (kỳ vọng, phương sai, xác suất, ...) ;
- xác định khoảng tin cậy và kiểm định các giả thuyết.

Mẫu ngẫu nhiên

Khái niệm 1

Ta gọi dãy các biến ngẫu nhiên $X_1; X_2; \cdots; X_n$ là mẫu ngẫu nhiên cỡ n rút ra từ biến ngẫu nhiên X nếu chúng độc lập và có cùng phân phối với X

x_i : Giá trị thực nghiệm (giá trị quan sát cụ thể) của X_i
Dãy $x_1; x_2; \cdots; x_n$: Dãy n giá trị quan sát cụ thể (dãy số liệu mẫu) thu được từ mẫu ngẫu nhiên $(X_1; X_2; \cdots; X_n)$ rút ra từ X.

Mẫu ngẫu nhiên

Số liệu thu gọn

$x_1; x_2; \cdots; x_n$: Dãy n giá trị quan sát cụ thể

$$\begin{array}{cccc} x_{(1)} < & x_{(2)} < & \cdots < & x_{(k)} \\ n_1 & n_2 & \cdots & n_k \end{array}$$

$$\sum_{i=1}^{k} n_i = n$$

n_i : Tần số (số lần số liệu mẫu nhận giá trị x_i)
k : Số các giá trị mẫu khác nhau (cỡ mẫu thu gọn)

Mẫu ngẫu nhiên

Số liệu thu gọn dạng khoảng
Chia miền giá trị số liệu mẫu thành k khoảng con rời nhau $[a_i; a_{i+1})$ bởi các điểm chia $a_i \quad (i = 1; 2; \cdots; k)$

$$\begin{array}{cccc} [a_1; a_2) & [a_2; a_3) & \cdots & [a_k; a_{k+1}) \\ n_1 & n_2 & \cdots & n_k \end{array}$$

n_i : Số giá trị quan sát rơi vào khoảng con $[a_i; a_{i+1})$

$$\sum_{i=1}^{k} n_i = n$$

Các đặc trưng mẫu

Số liệu thu gọn :
$(X_1; X_2; \cdots; X_n)$ là mẫu ngẫu nhiên cỡ n rút ra từ biến ngẫu nhiên X.

1. **Kỳ vọng mẫu (trung bình mẫu).**
   Ký hiệu : $\bar{x}$, xác định bởi :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$\bar{x}$ là biến ngẫu nhiên có : $\mathbb{E}(\bar{x}) = \mathbb{E}(x); \mathbb{D}(\bar{x}) = \dfrac{\mathbb{D}(x)}{n}$

Các đặc trưng mẫu

2. **Phương sai mẫu (độ lệch chuẩn).**
Ký hiệu : s^2, xác định bởi :

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{n-1}\left[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\cdots+(x_n-\bar{x})^2\right] \\ &= \frac{n}{n-1}\left(\overline{x^2}-\bar{x}^2\right). \end{aligned} \quad (2.18)$$

Chú ý 1

Nếu số liệu cho dưới dạng khoảng, ta chọn mỗi khoảng một điểm đại diện x_i, thông thường là điểm giữa khoảng, lúc đó ta lại có số liệu thu gọn.

Ví dụ

Tính trung bình mẫu và phương sai mẫu của một mẫu gồm chiều cao của 400 cây con được cho bởi trong bảng phân bố ghép lớp sau

Chiều cao (cm)	Tần số r_i	Chiều cao (cm)	Tần số r_i
$[4.5, 9.5]$	18	$(16.5, 19.5]$	57
$(9.5, 11.5]$	58	$(19.5, 22.5]$	42
$(11.5, 13.5]$	62	$(22.5, 26.5]$	36
$(13.5, 16.5]$	72	$(26.5, 29.5]$	55

Giải. Kích thước mẫu $n = 400, k = 8$.

Ví dụ

Tính trung bình mẫu và phương sai mẫu của một mẫu gồm chiều cao của 400 cây con được cho bởi trong bảng phân bố ghép lớp sau

Chiều cao (cm)	Tần số r_i	Chiều cao (cm)	Tần số r_i
$[4.5, 9.5]$	18	$(16.5, 19.5]$	57
$(9.5, 11.5]$	58	$(19.5, 22.5]$	42
$(11.5, 13.5]$	62	$(22.5, 26.5]$	36
$(13.5, 16.5]$	72	$(26.5, 29.5]$	55

Giải. Kích thước mẫu $n = 400, k = 8$.
Các điểm đại diện của các lớp :

$$\begin{aligned} &x_1 = 7, \quad x_2 = 10.5, \quad x_3 = 12.5, \quad x_4 = 15, \\ &x_5 = 18, \quad x_6 = 21, \quad x_7 = 24.5, \quad x_8 = 28\,(cm); \end{aligned}$$

Ví dụ

Tính trung bình mẫu và phương sai mẫu của một mẫu gồm chiều cao của 400 cây con được cho bởi trong bảng phân bố ghép lớp sau

Chiều cao (cm)	Tần số r_i	Chiều cao (cm)	Tần số r_i
$[4.5, 9.5]$	18	$(16.5, 19.5]$	57
$(9.5, 11.5]$	58	$(19.5, 22.5]$	42
$(11.5, 13.5]$	62	$(22.5, 26.5]$	36
$(13.5, 16.5]$	72	$(26.5, 29.5]$	55

Giải. Kích thước mẫu $n = 400, k = 8$.
Các điểm đại diện của các lớp :

$$x_1 = 7, \quad x_2 = 10.5, \quad x_3 = 12.5, \quad x_4 = 15,$$
$$x_5 = 18, \quad x_6 = 21, \quad x_7 = 24.5, \quad x_8 = 28\,(cm);$$

tần số xuất hiện của các lớp là

$$r_1 = 18, \quad r_2 = 58, \quad r_3 = 62, \quad r_4 = 72,$$
$$r_5 = 57, \quad r_6 = 42, \quad r_7 = 36, \quad r_8 = 55.$$

Ta có $\sum_{i=1}^{k} r_i = 400 = n$.

Ta có $\sum_{i=1}^{k} r_i = 400 = n$.
Trung bình của mẫu là

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i r_i = 17.3\,(cm).$$

Phương sai của mẫu là

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k} r_i(x_i - \bar{x})^2 = 38.01\,(cm^2).$$

2.2. Ước lượng điểm

X là biến ngẫu nhiên có hàm phân phối xác suất $F(x, \theta)$ phụ thuộc vào tham số đặc trưng θ chưa biết nào đó (θ : trung bình, phương sai, xác suất, ...).

Để xác định hoàn toàn phân phối của X ta phải ước lượng được giá trị của tham số θ mà phân phối nhận, đó là số $\hat{\theta}$ được tính từ một tập mẫu $(X_1, X_2, \cdots, X_n)$ đã cho và được hiểu như là một xấp xỉ của một giá trị chính xác chưa biết của tham số θ.

2.2. Ước lượng điểm

X là biến ngẫu nhiên có hàm phân phối xác suất $F(x,\theta)$ phụ thuộc vào tham số đặc trưng θ chưa biết nào đó (θ : trung bình, phương sai, xác suất, ...).
Để xác định hoàn toàn phân phối của X ta phải ước lượng được giá trị của tham số θ mà phân phối nhận, đó là số $\hat{\theta}$ được tính từ một tập mẫu $(X_1, X_2, \cdots, X_n)$ đã cho và được hiểu như là một xấp xỉ của một giá trị chính xác chưa biết của tham số θ.

2.2. Ước lượng điểm

X là biến ngẫu nhiên có hàm phân phối xác suất $F(x, \theta)$ phụ thuộc vào tham số đặc trưng θ chưa biết nào đó (θ : trung bình, phương sai, xác suất, ...).
Để xác định hoàn toàn phân phối của X ta phải ước lượng được giá trị của tham số θ mà phân phối nhận, đó là số $\hat{\theta}$ được tính từ một tập mẫu $(X_1, X_2, \cdots, X_n)$ đã cho và được hiểu như là một xấp xỉ của một giá trị chính xác chưa biết của tham số θ.

2.2. Ước lượng điểm của : Kỳ vọng, phương sai, xác suất

Định nghĩa 1

Hàm số $\hat{\theta} = \hat{\theta}(X_1; X_2; \cdots; X_n)$ được gọi là ước lượng điểm cho θ

Chú ý 2

i) Hàm số $\hat{\theta}$ được gọi là thống kê

ii) Hàm số $\hat{\theta}$ là một biến ngẫu nhiên và nhận giá trị cụ thể khi mẫu $(X_1; X_2; \cdots; X_n)$ nhận các giá trị $(x_1; x_2; \cdots; x_n)$

Các tính chất của ước lượng điểm

1. Tính vững.
2. Tính không chệch.
3. Tính hiệu quả

Ước lượng điểm cho các tham số

1. Ước lượng điểm cho kỳ vọng của biến ngẫu nhiên X : $\bar{x}$

$$\hat{\mu} = \bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n).$$

2. Ước lượng điểm cho phương sai của biến ngẫu nhiên X : s_x^2

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}\left(\overline{x^2} - \bar{x}^2\right).$$

3. Ước lượng điểm cho tỷ lệ :

$$p^* = \frac{m}{n}$$

Với n : Cỡ mẫu ban đầu ;
m : Số lần xuất hiện của biến cố cần ước lượng xác suất.

Ước lượng điểm cho tham số

(4) Ước lượng giá trị của xác suất Cho biến ngẫu nhiên X :

$$X = \begin{cases} 0 & \text{nếu không xảy ra sự kiện } A \text{ trong phép thử} \\ 1 & \text{nếu xảy ra sự kiện } A \text{ trong phép thử.} \end{cases}$$

p là xác suất xuất hiện sự kiện A.

$$\mathbb{P}(X = 0) = 1 - p \quad \text{và} \quad \mathbb{P}(X = 1) = p.$$

Xét một mẫu $(X_1, X_2, ..., X_n)$ có kích thước n.
Tần suất xuất hiện sự kiện A trong mẫu là

$$f = \frac{(x_1 + x_2 + \cdots + x_n)}{n} = \bar{x}.$$

Khi đó $f = \bar{x}$ là một ước lượng của xác suất tổng thể p.

Ước lượng điểm cho tham số

(4) Ước lượng giá trị của xác suất Cho biến ngẫu nhiên X :

$$X = \begin{cases} 0 & \text{nếu không xảy ra sự kiện } A \text{ trong phép thử} \\ 1 & \text{nếu xảy ra sự kiện } A \text{ trong phép thử.} \end{cases}$$

p là xác suất xuất hiện sự kiện A.

$$\mathbb{P}(X = 0) = 1 - p \quad \text{và} \quad \mathbb{P}(X = 1) = p.$$

Xét một mẫu $(X_1, X_2, ..., X_n)$ có kích thước n.
Tần suất xuất hiện sự kiện A trong mẫu là

$$f = \frac{(x_1 + x_2 + \cdots + x_n)}{n} = \bar{x}.$$

Khi đó $f = \bar{x}$ là một ước lượng của xác suất tổng thể p.

Ước lượng điểm cho tham số

(4) Ước lượng giá trị của xác suất Cho biến ngẫu nhiên X :

$$X = \begin{cases} 0 & \text{nếu không xảy ra sự kiện } A \text{ trong phép thử} \\ 1 & \text{nếu xảy ra sự kiện } A \text{ trong phép thử.} \end{cases}$$

p là xác suất xuất hiện sự kiện A.

$$\mathbb{P}(X = 0) = 1 - p \quad \text{và} \quad \mathbb{P}(X = 1) = p.$$

Xét một mẫu $(X_1, X_2, ..., X_n)$ có kích thước n.
Tần suất xuất hiện sự kiện A trong mẫu là

$$f = \frac{(x_1 + x_2 + \cdots + x_n)}{n} = \bar{x}.$$

Khi đó $f = \bar{x}$ là một ước lượng của xác suất tổng thể p.

Ước lượng điểm cho tham số

(4) Ước lượng giá trị của xác suất Cho biến ngẫu nhiên X :

$$X = \begin{cases} 0 & \text{nếu không xảy ra sự kiện } A \text{ trong phép thử} \\ 1 & \text{nếu xảy ra sự kiện } A \text{ trong phép thử.} \end{cases}$$

p là xác suất xuất hiện sự kiện A.

$$\mathbb{P}(X = 0) = 1 - p \quad \text{và} \quad \mathbb{P}(X = 1) = p.$$

Xét một mẫu $(X_1, X_2, ..., X_n)$ có kích thước n.
Tần suất xuất hiện sự kiện A trong mẫu là

$$f = \frac{(x_1 + x_2 + \cdots + x_n)}{n} = \bar{x}.$$

Khi đó $f = \bar{x}$ là một ước lượng của xác suất tổng thể p.

Phương pháp hợp lý cực đại (Maximum likelihood method)

Xét BNN (rời rạc hoặc liên tục) X với hàm mật độ $f(x, \theta)$. Lấy một mẫu có n giá trị độc lập

$$x_1, x_2, ..., x_n$$

tương ứng với các BNN độc lập

$$X_1, X_2, ..., X_n.$$

Phương pháp hợp lý cực đại (Maximum likelihood method)

Xét BNN (rời rạc hoặc liên tục) X với hàm mật độ $f(x,\theta)$. Lấy một mẫu có n giá trị độc lập

$$x_1, x_2, ..., x_n$$

tương ứng với các BNN độc lập

$$X_1, X_2, ..., X_n.$$

Hàm xác suất đồng thời của các BNN này có dạng

$$l(x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n, \theta). \tag{2.19}$$

Hàm l trong công thức (2.19) gọi là hàm hợp lý cực đại của tham số θ.

Phương pháp hợp lý cực đại (Maximum likelihood method)

Xét BNN (rời rạc hoặc liên tục) X với hàm mật độ $f(x,\theta)$. Lấy một mẫu có n giá trị độc lập

$$x_1, x_2, ..., x_n$$

tương ứng với các BNN độc lập

$$X_1, X_2, ..., X_n.$$

Hàm xác suất đồng thời của các BNN này có dạng

$$l(x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n, \theta). \tag{2.19}$$

Hàm l trong công thức (2.19) gọi là hàm hợp lý cực đại của tham số θ.

Ý tưởng : Tìm ước lượng $\hat{\theta}$ của tham số θ sao cho xác suất l của mẫu quan sát lớn nhất khi $\hat{\theta} = \theta$.

Phương pháp hợp lý cực đại (Maximum likelihood method)

Xét BNN (rời rạc hoặc liên tục) X với hàm mật độ $f(x, \theta)$. Lấy một mẫu có n giá trị độc lập

$$x_1, x_2, ..., x_n$$

tương ứng với các BNN độc lập

$$X_1, X_2, ..., X_n.$$

Hàm xác suất đồng thời của các BNN này có dạng

$$l(x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n, \theta). \tag{2.19}$$

Hàm l trong công thức (2.19) gọi là hàm hợp lý cực đại của tham số θ.

Ý tưởng : Tìm ước lượng $\hat{\theta}$ của tham số θ sao cho xác suất l của mẫu quan sát lớn nhất khi $\hat{\theta} = \theta$.Điều kiện cần là

$$\frac{\partial l}{\partial \theta}|_{\theta=\hat{\theta}} = 0 \Leftrightarrow \frac{\partial \ln l}{\partial \theta}|_{\theta=\hat{\theta}} = 0 \tag{2.20}$$

Phương pháp hợp lý cực đại (Maximum likelihood method)

Xét BNN (rời rạc hoặc liên tục) X với hàm mật độ $f(x, \theta)$. Lấy một mẫu có n giá trị độc lập

$$x_1, x_2, ..., x_n$$

tương ứng với các BNN độc lập

$$X_1, X_2, ..., X_n.$$

Hàm xác suất đồng thời của các BNN này có dạng

$$l(x_1, x_2, ..., x_n; \theta) = f(x_1; \theta) f(x_2; \theta) \cdots f(x_n, \theta). \tag{2.19}$$

Hàm l trong công thức (2.19) gọi là hàm hợp lý cực đại của tham số θ.

Ý tưởng : Tìm ước lượng $\hat{\theta}$ của tham số θ sao cho xác suất l của mẫu quan sát lớn nhất khi $\hat{\theta} = \theta$.Điều kiện cần là

$$\frac{\partial l}{\partial \theta}\Big|_{\theta=\hat{\theta}} = 0 \Leftrightarrow \frac{\partial \ln l}{\partial \theta}\Big|_{\theta=\hat{\theta}} = 0 \tag{2.20}$$

Nghiệm $\hat{\theta}$ của phương trình (2.20) được gọi là ước lượng hợp lý cực đại của tham số θ.

Ví dụ 2.2.1(Phương pháp hợp lý cực đại)

Tìm ước lượng hợp lý cực đại của xác suất (tần suất) tổng thể p, biết hàm xác suất

$$f(x,p) = p^x(1-p)^{1-x}, x \in \{0,1\}.$$

Ví dụ 2.2.1(Phương pháp hợp lý cực đại)

Tìm ước lượng hợp lý cực đại của xác suất (tần suất) tổng thể p, biết hàm xác suất

$$f(x,p) = p^x(1-p)^{1-x}, x \in \{0,1\}.$$

Lời giải. Hàm hợp lý cực đại của xác suất p có dạng

$$\begin{aligned} l(x_1, x_2, ..., x_n; \theta) &= f_1(x_1;\theta) f_2(x_2;\theta) \cdots f_n(x_n,\theta) \\ &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \end{aligned}$$

Ví dụ 2.2.1(Phương pháp hợp lý cực đại)

Tìm ước lượng hợp lý cực đại của xác suất (tần suất) tổng thể p, biết hàm xác suất

$$f(x,p) = p^x(1-p)^{1-x}, x \in \{0,1\}.$$

Lời giải. Hàm hợp lý cực đại của xác suất p có dạng

$$\begin{aligned} l(x_1, x_2, ..., x_n; \theta) &= f_1(x_1;\theta) f_2(x_2;\theta) \cdots f_n(x_n, \theta) \\ &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \end{aligned}$$

$$\Leftrightarrow \ln l(p) = \sum_{i=1}^{n} \left(x_i l_p + (1-x_i) ln(1-p)\right).$$

Ví dụ 2.2.1(Phương pháp hợp lý cực đại)

Tìm ước lượng hợp lý cực đại của xác suất (tần suất) tổng thể p, biết hàm xác suất

$$f(x,p) = p^x(1-p)^{1-x}, x \in \{0,1\}.$$

Lời giải. Hàm hợp lý cực đại của xác suất p có dạng

$$\begin{aligned} l(x_1, x_2, ..., x_n; \theta) &= f_1(x_1;\theta) f_2(x_2;\theta) \cdots f_n(x_n, \theta) \\ &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \end{aligned}$$

$$\Leftrightarrow \ln l(p) = \sum_{i=1}^{n} (x_i l_p + (1-x_i) ln(1-p)).$$

$$\Leftrightarrow \frac{\partial \ln l(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n}(1-x_i).$$

Ví dụ 2.2.1(Phương pháp hợp lý cực đại)

Tìm ước lượng hợp lý cực đại của xác suất (tần suất) tổng thể p, biết hàm xác suất

$$f(x,p) = p^x(1-p)^{1-x}, x \in \{0,1\}.$$

Lời giải. Hàm hợp lý cực đại của xác suất p có dạng

$$\begin{aligned} l(x_1, x_2, ..., x_n; \theta) &= f_1(x_1; \theta) f_2(x_2; \theta) \cdots f_n(x_n, \theta) \\ &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \end{aligned}$$

$$\Leftrightarrow \ln l(p) = \sum_{i=1}^{n} \left(x_i l_p + (1-x_i) ln(1-p)\right).$$

$$\Leftrightarrow \frac{\partial \ln l(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n}(1-x_i).$$

Giải phương trình

$$\frac{\partial \ln l(p)}{\partial p} = 0 \Leftrightarrow p = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \bar{x}.$$

Ví dụ 2.2.1(Phương pháp hợp lý cực đại)

Tìm ước lượng hợp lý cực đại của xác suất (tần suất) tổng thể p, biết hàm xác suất

$$f(x,p) = p^x(1-p)^{1-x}, x \in \{0,1\}.$$

Lời giải. Hàm hợp lý cực đại của xác suất p có dạng

$$\begin{aligned} l(x_1, x_2, ..., x_n; \theta) &= f_1(x_1;\theta) f_2(x_2;\theta) \cdots f_n(x_n,\theta) \\ &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \end{aligned}$$

$$\Leftrightarrow \ln l(p) = \sum_{i=1}^{n} \left(x_i l_p + (1-x_i) ln(1-p)\right).$$

$$\Leftrightarrow \frac{\partial \ln l(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}\sum_{i=1}^{n}(1-x_i).$$

Giải phương trình

$$\frac{\partial \ln l(p)}{\partial p} = 0 \Leftrightarrow p = \frac{1}{n}(x_1 + x_2 + \cdots + x_n) = \bar{x}.$$

Vậy, ước lượng hợp lý cực đại của p là $\hat{p} = \bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$.

Phương pháp hợp lý cực đại

Ví dụ 2.2.2.

Tìm ước lượng hợp lý cực đại của kỳ vọng, độ lệch chuẩn của phân bố chuẩn từ một tập mẫu có kích thước n, trung bình mẫu $\bar{x}$, phương sai mẫu s^2.

Phương pháp hợp lý cực đại

Ví dụ 2.2.2.

Tìm ước lượng hợp lý cực đại của kỳ vọng, độ lệch chuẩn của phân bố chuẩn từ một tập mẫu có kích thước n, trung bình mẫu $\bar{x}$, phương sai mẫu s^2.

Lời giải. Ta có hàm hợp lý cực đại l trong trường hợp này là

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-h},\ h = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 .$$

Ta có

$$\ln l = -n\ln\sqrt{2\pi} - n\ln\sigma - h.$$

Suy ra

$$\frac{\partial \ln l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)$$

Áp dụng phương pháp hợp lý cực đại, giải phương trình

$$\frac{\partial \ln l}{\partial \mu} = 0 \Leftrightarrow \mu = \bar{x}.$$

Phương pháp hợp lý cực đại

Ví dụ 2.2.2.

Tìm ước lượng hợp lý cực đại của kỳ vọng, độ lệch chuẩn của phân bố chuẩn từ một tập mẫu có kích thước n, trung bình mẫu $\bar{x}$, phương sai mẫu s^2.

Phương pháp hợp lý cực đại

Ví dụ 2.2.2.

Tìm ước lượng hợp lý cực đại của kỳ vọng, độ lệch chuẩn của phân bố chuẩn từ một tập mẫu có kích thước n, trung bình mẫu $\bar{x}$, phương sai mẫu s^2.

Lời giải. Ta có hàm hợp lý cực đại l trong trường hợp này là

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-h},\ h = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Ta có

$$\ln l = -n\ln\sqrt{2\pi} - n\ln\sigma - h.$$

Áp dụng phương pháp hợp lý cực đại, giải phương trình

$$\frac{\partial \ln l}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(x_i-\bar{x})^2 = 0 \Leftrightarrow \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} = s\sqrt{\frac{n-1}{n}}.$$

Phương pháp hợp lý cực đại

Ví dụ 2.2.3.

Giả sử điểm thi cuối kỳ môn Lý thuyết xác suất thống kê của sinh viên một trường đại học là biến ngẫu nhiên có phân phối chuẩn. Khảo sát điểm thi cuối kỳ môn này một của số bạn sinh viên người ta thu được mẫu sau :

Điểm	0-2	2-4	4-6	6-8	8-10
Số SV	5	10	15	18	8

Hãy sử dụng phương pháp hợp lý cực đại để tính ước lượng điểm cho điểm thi trung bình môn Lý thuyết xác suất thống kê của sinh viên trường đại học nói trên.

Phương pháp hợp lý cực đại

Lời giải. $n = 56$. Ta có hàm mật độ xác suất cực đại l trong trường hợp này là

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-h},\ h = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 .$$

Ta có

$$\ln l = -n\ln\sqrt{2\pi} - n\ln\sigma - h.$$

Do đó

$$\frac{\partial \ln l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) .$$

Áp dụng phương pháp hợp lý cực đại, xét phương trình

$$\frac{\partial \ln l}{\partial \mu} = 0$$

Phương pháp hợp lý cực đại

Lời giải. $n = 56$. Ta có hàm mật độ xác suất cực đại l trong trường hợp này là

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-h},\ h = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2 .$$

Ta có

$$\ln l = -n \ln\sqrt{2\pi} - n\ln\sigma - h.$$

Do đó

$$\frac{\partial \ln l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) .$$

Áp dụng phương pháp hợp lý cực đại, xét phương trình

$$\frac{\partial \ln l}{\partial \mu} = 0 \Leftrightarrow \mu = \bar{x}$$

Phương pháp hợp lý cực đại

Lời giải. $n = 56$. Ta có hàm mật độ xác suất cực đại l trong trường hợp này là

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-h},\ h = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 .$$

Ta có

$$\ln l = -n\ln\sqrt{2\pi} - n\ln\sigma - h.$$

Do đó

$$\frac{\partial \ln l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) .$$

Áp dụng phương pháp hợp lý cực đại, xét phương trình

$$\frac{\partial \ln l}{\partial \mu} = 0 \Leftrightarrow \mu = \bar{x}$$

với

$$\bar{x} = \frac{1\cdot 5 + 3\cdot 10 + 5\cdot 15 + 7\cdot 18 + 9\cdot 8}{56} = 5,5.$$

Vậy

$$\mu = 5,5.$$

Phương pháp hợp lý cực đại

Ví dụ 2.2.4.

Cho biến ngẫu nhiên X có hàm mật độ xác suất

$$f(x) = \begin{cases} \alpha e^{-\alpha(x-2)} & \text{với } x \geq 2 \\ 0 & \text{với } x < 2 \end{cases} \tag{2.21}$$

Hãy sử dụng phương pháp hợp lý cực đại để tìm α thông qua mẫu dữ liệu sau của $X : 2; 3; 4; 5; 6.$

Phương pháp hợp lý cực đại

Ví dụ 2.2.4.

Cho biến ngẫu nhiên X có hàm mật độ xác suất

$$f(x) = \begin{cases} \alpha e^{-\alpha(x-2)} & \text{với } x \geq 2 \\ 0 & \text{với } x < 2 \end{cases} \tag{2.21}$$

Hãy sử dụng phương pháp hợp lý cực đại để tìm α thông qua mẫu dữ liệu sau của $X : 2; 3; 4; 5; 6$.

Lời giải. $n = 6$. Ta có hàm mật độ xác suất cực đại l trong trường hợp này là

$$L = f(2) \cdot f(3) \cdot f(4) \cdot f(5) \cdot f(6) = \alpha^5 e^{-10\alpha}.$$

Ta có

$$\ln l = 5 \ln \alpha - 10\alpha \Leftrightarrow \frac{\partial \ln l}{\partial \alpha} = \frac{5}{\alpha} - 10.$$

Áp dụng phương pháp hợp lý cực đại, xét phương trình

$$\frac{\partial \ln l}{\partial \alpha} = 0$$

Phương pháp hợp lý cực đại

Ví dụ 2.2.4.

Cho biến ngẫu nhiên X có hàm mật độ xác suất

$$f(x) = \begin{cases} \alpha e^{-\alpha(x-2)} & \text{với } x \geq 2 \\ 0 & \text{với } x < 2 \end{cases} \tag{2.21}$$

Hãy sử dụng phương pháp hợp lý cực đại để tìm α thông qua mẫu dữ liệu sau của $X: 2; 3; 4; 5; 6$.

Lời giải. $n = 6$. Ta có hàm mật độ xác suất cực đại l trong trường hợp này là

$$L = f(2) \cdot f(3) \cdot f(4) \cdot f(5) \cdot f(6) = \alpha^5 e^{-10\alpha}.$$

Ta có

$$\ln l = 5 \ln \alpha - 10\alpha \Leftrightarrow \frac{\partial \ln l}{\partial \alpha} = \frac{5}{\alpha} - 10.$$

Áp dụng phương pháp hợp lý cực đại, xét phương trình

$$\frac{\partial \ln l}{\partial \alpha} = 0 \Leftrightarrow \alpha = \frac{1}{2}$$

Phương pháp hợp lý cực đại

Ví dụ 2.2.5.

Sử dụng phương pháp hợp lý cực đại để tìm ước lượng điểm cho trung bình của tập nền từ mẫu dữ liệu sau với giả sử rằng chỉ số IQ của học sinh lớp 9 tuân theo phân bố chuẩn

Chỉ số IQ	80-86	86-92	92-98	98-104	104-110
Số HS	4	8	12	10	17

Phương pháp hợp lý cực đại

Ví dụ 2.2.5.

Sử dụng phương pháp hợp lý cực đại để tìm ước lượng điểm cho trung bình của tập nền từ mẫu dữ liệu sau với giả sử rằng chỉ số IQ của học sinh lớp 9 tuân theo phân bố chuẩn

Chỉ số IQ	80-86	86-92	92-98	98-104	104-110
Số HS	4	8	12	10	17

Lời giải. $n = 51$. Ta có hàm mật độ xác suất cực đại l trong trường hợp này là

$$l = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{-h},\ h = \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

Ta có

$$\ln l = -n\ln\sqrt{2\pi} - n\ln\sigma - h.$$

Do đó

$$\frac{\partial \ln l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu).$$

Phương pháp hợp lý cực đại

Ví dụ 2.2.5.

Sử dụng phương pháp hợp lý cực đại để tìm ước lượng điểm cho trung bình của tập nền từ mẫu dữ liệu sau với giả sử rằng chỉ số IQ của học sinh lớp 9 tuân theo phân bố chuẩn

Chỉ số IQ	80-86	86-92	92-98	98-104	104-110
Số HS	4	8	12	10	17

Phương pháp hợp lý cực đại

Ví dụ 2.2.5.

Sử dụng phương pháp hợp lý cực đại để tìm ước lượng điểm cho trung bình của tập nền từ mẫu dữ liệu sau với giả sử rằng chỉ số IQ của học sinh lớp 9 tuân theo phân bố chuẩn

Chỉ số IQ	80-86	86-92	92-98	98-104	104-110
Số HS	4	8	12	10	17

Lời giải. Áp dụng phương pháp hợp lý cực đại, xét phương trình

$$\frac{\partial \ln l}{\partial \mu} = 0$$

Phương pháp hợp lý cực đại

Ví dụ 2.2.5.

Sử dụng phương pháp hợp lý cực đại để tìm ước lượng điểm cho trung bình của tập nền từ mẫu dữ liệu sau với giả sử rằng chỉ số IQ của học sinh lớp 9 tuân theo phân bố chuẩn

Chỉ số IQ	80-86	86-92	92-98	98-104	104-110
Số HS	4	8	12	10	17

Lời giải. Áp dụng phương pháp hợp lý cực đại, xét phương trình

$$\frac{\partial \ln l}{\partial \mu} = 0 \Leftrightarrow \mu = \bar{x}$$

Phương pháp hợp lý cực đại

Ví dụ 2.2.5.

Sử dụng phương pháp hợp lý cực đại để tìm ước lượng điểm cho trung bình của tập nền từ mẫu dữ liệu sau với giả sử rằng chỉ số IQ của học sinh lớp 9 tuân theo phân bố chuẩn

Chỉ số IQ	80-86	86-92	92-98	98-104	104-110
Số HS	4	8	12	10	17

Lời giải. Áp dụng phương pháp hợp lý cực đại, xét phương trình

$$\frac{\partial \ln l}{\partial \mu} = 0 \Leftrightarrow \mu = \bar{x}$$

với

$$\bar{x} = \frac{83 \cdot 4 + 89 \cdot 8 + 95 \cdot 12 + 101 \cdot 10 + 107 \cdot 17}{51} = 96,171.$$

Vậy

$$\mu = 96,171.$$

2.3.1. Lý thuyết chung về ước lượng khoảng

- Hạn chế của ước lượng điểm :
    - Khi mẫu có kích thước nhỏ thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng ;
    - không thể biết được độ chính xác của các giá trị ước lượng $\hat{\theta}$ so với giá trị của đại lượng cần ước lượng θ.

2.3.1. Lý thuyết chung về ước lượng khoảng

- Hạn chế của ước lượng điểm :
  - Khi mẫu có kích thước nhỏ thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng ;
  - không thể biết được độ chính xác của các giá trị ước lượng $\hat{\theta}$ so với giá trị của đại lượng cần ước lượng θ.
- Khắc phục : Ước lượng bằng một khoảng giá trị (Ước lượng khoảng).

2.3.1. Lý thuyết chung về ước lượng khoảng

- Hạn chế của ước lượng điểm :
  - Khi mẫu có kích thước nhỏ thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng ;
  - không thể biết được độ chính xác của các giá trị ước lượng $\hat{\theta}$ so với giá trị của đại lượng cần ước lượng θ.
- Khắc phục : Ước lượng bằng một khoảng giá trị (Ước lượng khoảng).
- Ưu điểm của ước lượng khoảng : Xác suất sai lầm có thể biết và trong chừng mực nào đó có thể hy vọng kiểm soát được.

2.3.1. Lý thuyết chung về ước lượng khoảng

- Hạn chế của ước lượng điểm :
  - Khi mẫu có kích thước nhỏ thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng ;
  - không thể biết được độ chính xác của các giá trị ước lượng $\hat{\theta}$ so với giá trị của đại lượng cần ước lượng θ.
- Khắc phục : Ước lượng bằng một khoảng giá trị (Ước lượng khoảng).
- Ưu điểm của ước lượng khoảng : Xác suất sai lầm có thể biết và trong chừng mực nào đó có thể hy vọng kiểm soát được.

Khoảng tin cậy cho tham số θ : một khoảng $\theta_1 \leq \theta \leq \theta_2$ với một *xác suất cao (độ tin cậy)* γ.

2.3.1. Lý thuyết chung về ước lượng khoảng

- Hạn chế của ước lượng điểm :
  - Khi mẫu có kích thước nhỏ thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng ;
  - không thể biết được độ chính xác của các giá trị ước lượng $\hat{\theta}$ so với giá trị của đại lượng cần ước lượng θ.
- Khắc phục : Ước lượng bằng một khoảng giá trị (Ước lượng khoảng).
- Ưu điểm của ước lượng khoảng : Xác suất sai lầm có thể biết và trong chừng mực nào đó có thể hy vọng kiểm soát được.

Khoảng tin cậy cho tham số θ : một khoảng $\theta_1 \leq \theta \leq \theta_2$ với một *xác suất cao (độ tin cậy)* γ.
θ_1, θ_2 : độ tin cậy dưới và trên.

2.3.1. Lý thuyết chung về ước lượng khoảng

- Hạn chế của ước lượng điểm :
  - Khi mẫu có kích thước nhỏ thì ước lượng điểm có thể sai lệch rất nhiều so với giá trị của tham số cần ước lượng ;
  - không thể biết được độ chính xác của các giá trị ước lượng $\hat{\theta}$ so với giá trị của đại lượng cần ước lượng θ.
- Khắc phục : Ước lượng bằng một khoảng giá trị (Ước lượng khoảng).
- Ưu điểm của ước lượng khoảng : Xác suất sai lầm có thể biết và trong chừng mực nào đó có thể hy vọng kiểm soát được.

Khoảng tin cậy cho tham số θ : một khoảng $\theta_1 \leq \theta \leq \theta_2$ với một *xác suất cao (độ tin cậy)* γ.
θ_1, θ_2 : độ tin cậy dưới và trên.
Bài toán : Xác định θ_1 và θ_2 sao cho xác suất để tìm thấy θ trong miền $[\theta_1, \theta_2]$ bằng γ, nghĩa là

$$P(\theta_1 \leq \theta \leq \theta_2) = \gamma. \tag{2.22}$$

2.3.2. Phân phối student ; ước lượng khoảng cho kỳ vọng của phân phối chuẩn

1. Khoảng tin cậy cho μ của phân phối chuẩn khi biết phương sai σ^2

Định lý (Tổng của các BNN độc lập có phân phối chuẩn)

$X_1, \cdots, X_n$: các BNN có phân phối chuẩn, có chung kỳ vọng μ và phương sai σ^2. Khi đó, các khẳng định sau là đúng :

1. Tổng $X_1 + \cdots + X_n$ là BNN có phân phối chuẩn với kỳ vọng $n\mu$, phương sai $n\sigma^2$.
2. Biến ngẫu nhiên
$$\overline{X} := \frac{1}{n}(X_1 + \cdots + X_n),$$
có phân phối chuẩn với kỳ vọng μ, phương sai $\frac{\sigma^2}{n}$.
3. Biến ngẫu nhiên
$$Z := \frac{\overline{X} - \mu}{\sigma/\sqrt{n}},$$
có phân phối chuẩn với kỳ vọng 0 và phương sai 1.

1. Khoảng tin cậy cho μ của phân phối chuẩn khi biết phương sai σ^2

Định nghĩa

BNN $X \sim N(0,1)$ có phân bố chuẩn tắc, $\alpha \in (0,1)$. U_α được gọi là *phân vị mức* α (*giá trị tới hạn mức* α) của X nếu

$$P(X \leq U_\alpha) = 1 - \alpha \tag{2.23}$$

hay

$$\Phi(U_\alpha) = 1 - \alpha.$$

ở đó $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du$: hàm Laplace.

1. Khoảng tin cậy cho μ của phân phối chuẩn khi biết phương sai σ^2

Định nghĩa

BNN $X \sim N(0,1)$ có phân bố chuẩn tắc, $\alpha \in (0,1)$. U_α được gọi là *phân vị mức* α (*giá trị tới hạn mức* α) của X nếu

$$P(X \leq U_\alpha) = 1 - \alpha \tag{2.23}$$

hay

$$\Phi(U_\alpha) = 1 - \alpha.$$

ở đó $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du$: hàm Laplace.

Giá trị của U_α được cho bởi bảng A8.

1. Khoảng tin cậy cho μ của phân phối chuẩn khi biết phương sai σ^2

Định nghĩa

BNN $X \sim N(0,1)$ có phân bố chuẩn tắc, $\alpha \in (0,1)$. U_α được gọi là *phân vị mức* α (*giá trị tới hạn mức* α) của X nếu

$$P(X \leq U_\alpha) = 1 - \alpha \tag{2.23}$$

hay

$$\Phi(U_\alpha) = 1 - \alpha.$$

ở đó $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du$: hàm Laplace.

Giá trị của U_α được cho bởi bảng A8.

Ví dụ

Với $\alpha = 5\%$, thì $\Phi(U_\alpha) = 95\%$ và $U_\alpha = 1.645$.

Các bước xác định khoảng tin cậy cho μ khi biết phương sai σ^2

1. Chọn độ tin cậy γ (95%, 99% hoặc khác).

Các bước xác định khoảng tin cậy cho μ khi biết phương sai σ^2

1. Chọn độ tin cậy γ (95%, 99% hoặc khác).
2. Xác định giá trị tới hạn $U_{\alpha/2}$ tương ứng với mức $\frac{\alpha}{2} = \frac{1-\gamma}{2}$ từ Bảng A8, thông qua hàm $\Phi(U_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$.

Các bước xác định khoảng tin cậy cho μ khi biết phương sai σ^2

1. Chọn độ tin cậy γ (95%, 99% hoặc khác).
2. Xác định giá trị tới hạn $U_{\alpha/2}$ tương ứng với mức $\frac{\alpha}{2} = \frac{1-\gamma}{2}$ từ Bảng A8, thông qua hàm $\Phi(U_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$.
3. Tính trung bình mẫu
$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$
từ tập mẫu $x_1, x_2, \cdots, x_n$ đã cho.

Các bước xác định khoảng tin cậy cho μ khi biết phương sai σ^2

1. Chọn độ tin cậy γ (95%, 99% hoặc khác).
2. Xác định giá trị tới hạn $U_{\alpha/2}$ tương ứng với mức $\frac{\alpha}{2} = \frac{1-\gamma}{2}$ từ Bảng A8, thông qua hàm $\Phi(U_{\frac{\alpha}{2}}) = 1 - \frac{\alpha}{2}$.
3. Tính trung bình mẫu
$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$
từ tập mẫu x_1, x_2, $\cdots$, x_n đã cho.
4. Tính $k = \frac{U_{\alpha/2}\sigma}{\sqrt{n}}$. Khoảng tin cậy cho μ với độ tin cậy γ là
$$\bar{x} - k \le \mu \le \bar{x} + k. \tag{2.24}$$

Ví dụ

Tìm khoảng ước lượng với độ tin cậy 95% cho kỳ vọng μ của phân bố chuẩn với phương sai $\sigma^2 = 9$ sử dụng tập mẫu có kích thước $n = 100$ và trung bình mẫu $\bar{x} = 5$.

Ví dụ

Tìm khoảng ước lượng với độ tin cậy 95% cho kỳ vọng μ của phân bố chuẩn với phương sai $\sigma^2 = 9$ sử dụng tập mẫu có kích thước $n = 100$ và trung bình mẫu $\bar{x} = 5$.

Giải : Độ tin cậy $\gamma = 95\%$, vì thế $\alpha/2 = (1 - 0.95)/2 = 2.5\%$ và $U_{\alpha/2} = 1.960$.

Ví dụ

Tìm khoảng ước lượng với độ tin cậy 95% cho kỳ vọng μ của phân bố chuẩn với phương sai $\sigma^2 = 9$ sử dụng tập mẫu có kích thước $n = 100$ và trung bình mẫu $\bar{x} = 5$.

Giải : Độ tin cậy $\gamma = 95\%$, vì thế $\alpha/2 = (1 - 0.95)/2 = 2.5\%$ và $U_{\alpha/2} = 1.960$.
Trung bình mẫu $\bar{x} = 5$.

Ví dụ

Tìm khoảng ước lượng với độ tin cậy 95% cho kỳ vọng μ của phân bố chuẩn với phương sai $\sigma^2 = 9$ sử dụng tập mẫu có kích thước $n = 100$ và trung bình mẫu $\bar{x} = 5$.

Giải : Độ tin cậy $\gamma = 95\%$, vì thế $\alpha/2 = (1 - 0.95)/2 = 2.5\%$ và $U_{\alpha/2} = 1.960$.
Trung bình mẫu $\bar{x} = 5$.
Độ lệch chuẩn $\sigma = \sqrt{9} = 3$.

Ví dụ

Tìm khoảng ước lượng với độ tin cậy 95% cho kỳ vọng μ của phân bố chuẩn với phương sai $\sigma^2 = 9$ sử dụng tập mẫu có kích thước $n = 100$ và trung bình mẫu $\bar{x} = 5$.

Giải : Độ tin cậy $\gamma = 95\%$, vì thế $\alpha/2 = (1 - 0.95)/2 = 2.5\%$ và $U_{\alpha/2} = 1.960$.
Trung bình mẫu $\bar{x} = 5$.
Độ lệch chuẩn $\sigma = \sqrt{9} = 3$.Suy ra,

$$k = \frac{U_0\sigma}{\sqrt{n}} = \frac{1.960 \times 3}{\sqrt{100}} = 0.588.$$

Khoảng ước lượng cho μ với độ tin cậy 95% là

$$5 - 0.588 \leq \mu \leq 5 + 0.588 \quad \text{hay} \quad 4.412 \leq \mu \leq 5.588.$$

Ví dụ

Tìm kích thước mẫu n nhỏ nhất trong ví dụ trên sao cho khoảng tin cậy có độ dài $L = 0.4$ với độ tin cậy $\gamma = 0.95$?

Ví dụ

Tìm kích thước mẫu n nhỏ nhất trong ví dụ trên sao cho khoảng tin cậy có độ dài $L = 0.4$ với độ tin cậy $\gamma = 0.95$?

Giải. Độ dài khoảng tin cậy là

$$L = 2k = 2U_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

Vì $\gamma = 0.95$ nên $U_{\alpha/2} = 1.960$. Do đó

$$n = \left(\frac{2U_{\alpha/2}\sigma}{L}\right)^2 = \left(\frac{2 \times 1.960 \times 3}{0.4}\right)^2 \approx 865.$$

2. Khoảng tin cậy cho μ của phân bố chuẩn khi chưa biết σ^2

Định lý (Phân phối Student)

$X_1, \cdots, X_n$: các BNN độc lập có phân bố chuẩn, có cùng kỳ vọng μ, phương sai σ^2. BNN được định nghĩa :

$$T := \frac{\overline{X} - \mu}{S/\sqrt{n}}, \tag{2.25}$$

có phân bố Student $(n-1)$ bậc tự do và với hàm phân bố xác suất

$$F(z) = K_{n-1} \int\limits_{-\infty}^{z} \left(1 + \frac{u^2}{n-1}\right)^{-\frac{n}{2}} du, K_m = \frac{\Gamma(m/2 + 1/2)}{\sqrt{m\pi}\Gamma(m/2)} \tag{2.26}$$

với

$$\Gamma(\alpha) = \int\limits_{0}^{\infty} e^{-t} t^{\alpha-1} dt, \overline{X} = \frac{1}{n}(X_1 + \cdots + X_n), \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Nhận xét

1. Đồ thị hàm mật độ của phân bố Student $(n-1)$ bậc tự do cũng có dạng "hình chuông" tương tự phân bố chuẩn tắc (xem Hình 3.)

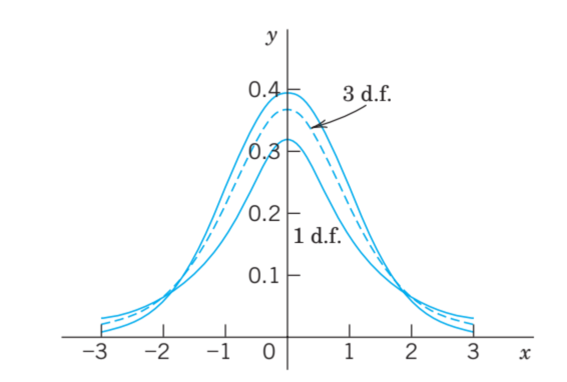

Figure – Hàm mật độ của phân bố Student với 1 và 3 bậc tự do và hàm mật độ của phân bố chuẩn tắc

Nhận xét

(2) Hàm phân bố xác suất $F(z)$ được cho bởi công thức (2.26) có các tính chất sau :

- $F(+\infty) = 1$;
- $F(-z) = 1 - F(z)$ với mọi $z \in \mathbb{R}$.

Nhận xét

(2) Hàm phân bố xác suất $F(z)$ được cho bởi công thức (2.26) có các tính chất sau :

- $F(+\infty) = 1$;
- $F(-z) = 1 - F(z)$ với mọi $z \in \mathbb{R}$.

(3) Giá trị t_α được gọi là *giá trị tới hạn mức* α của phân bố Student $(n-1)$ bậc tự do nếu

$$F(t_\alpha) = 1 - \alpha \tag{2.27}$$

với F là hàm được cho bởi công thức (2.26). Giá trị của t_α được cho bởi bảng $A9$.

Nhận xét

(2) Hàm phân bố xác suất $F(z)$ được cho bởi công thức (2.26) có các tính chất sau :

- $F(+\infty) = 1$;
- $F(-z) = 1 - F(z)$ với mọi $z \in \mathbb{R}$.

(3) Giá trị t_α được gọi là *giá trị tới hạn mức* α của phân bố Student $(n-1)$ bậc tự do nếu

$$F(t_\alpha) = 1 - \alpha \tag{2.27}$$

với F là hàm được cho bởi công thức (2.26). Giá trị của t_α được cho bởi bảng $A9$. Ví dụ : xét $\alpha = 0.005$ và $n - 1 = 4$, ta có $F(t_\alpha) = 0.995$ suy ra $t_\alpha = 4.60$.

Các bước xác định khoảng tin cậy cho μ khi chưa biết phương sai :

1. Chọn một độ tin cậy γ.

Các bước xác định khoảng tin cậy cho μ khi chưa biết phương sai :

1. Chọn một độ tin cậy γ.
2. Xác định $t_{\alpha/2}$ ứng với mức $\frac{\alpha}{2} = \frac{1-\gamma}{2}$ từ phương trình

$$F(t_{\alpha/2}) = 1 - \frac{\alpha}{2} = \frac{1}{2}(1+\gamma), \tag{2.28}$$

nhờ bảng $A9$, bậc $(n-1)$.

Các bước xác định khoảng tin cậy cho μ khi chưa biết phương sai :

1. Chọn một độ tin cậy γ.
2. Xác định $t_{\alpha/2}$ ứng với mức $\frac{\alpha}{2} = \frac{1-\gamma}{2}$ từ phương trình
$$F(t_{\alpha/2}) = 1 - \frac{\alpha}{2} = \frac{1}{2}(1+\gamma), \tag{2.28}$$
nhờ bảng $A9$, bậc $(n-1)$.
3. Tính
$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$
và phương sai mẫu
$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}\left(\overline{x^2} - \bar{x}^2\right)$$
từ tập mẫu kích thước n đã cho.

Các bước xác định khoảng tin cậy cho μ khi chưa biết phương sai :

1. Chọn một độ tin cậy γ.
2. Xác định $t_{\alpha/2}$ ứng với mức $\frac{\alpha}{2} = \frac{1-\gamma}{2}$ từ phương trình
$$F(t_{\alpha/2}) = 1 - \frac{\alpha}{2} = \frac{1}{2}(1+\gamma), \tag{2.28}$$
nhờ bảng $A9$, bậc $(n-1)$.
3. Tính
$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \cdots + x_n)$$
và phương sai mẫu
$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}\left(\overline{x^2} - \bar{x}^2\right)$$
từ tập mẫu kích thước n đã cho.
4. Tính $k = \frac{t_{\alpha/2}s}{\sqrt{n}}$. Khoảng tin cậy cho μ với kỳ vọng γ được xác định như sau
$$\bar{x} - k \le \mu \le \bar{x} + k. \tag{2.29}$$

Ví dụ

Nhiệt độ (^oF) ở điểm đốt cháy của dầu Diesel được đo năm lần với các giá trị lần lượt :

$$144 \quad 147 \quad 146 \quad 142 \quad 144.$$

Giả sử tập nền có phân bố chuẩn, hãy tìm khoảng ước lượng cho kỳ vọng với độ tin cậy 99%.

Ví dụ

Nhiệt độ (^oF) ở điểm đốt cháy của dầu Diesel được đo năm lần với các giá trị lần lượt :

$$144 \quad 147 \quad 146 \quad 142 \quad 144.$$

Giả sử tập nền có phân bố chuẩn, hãy tìm khoảng ước lượng cho kỳ vọng với độ tin cậy 99%.

Giải : Độ tin cậy $\gamma = 0.99$, suy ra $\alpha = 0.01$. Ta có

$$F(t_{\alpha/2}) = 1 - \alpha/2 = 0.995$$

, Do đó, $t_{\alpha/2} = 4.60$ (bảng $A9$).
Từ tập mẫu ta tính được $\bar{x} = 144.6$ và $s^2 = 3.8$. Do đó

$$k = \frac{4.60\sqrt{3.8}}{\sqrt{5}} = 4.01.$$

Vậy khoảng ước lượng cho μ với cấp độ tin cậy 99% là

$$144.6 - 4.01 \leq \mu \leq 144.6 + 4.01 \quad \text{hay} \quad 140.59 \leq \mu \leq 148.61.$$

2.3.3.Ước lượng khoảng cho xác suất (tỷ lệ)

$X_1, X_2, \ldots, X_n$- các BNN có cùng trung bình μ và kỳ vọng σ^2. Khi đó $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ có tính chất sau :

2.3.3.Ước lượng khoảng cho xác suất (tỷ lệ)

$X_1, X_2, \ldots, X_n$- các BNN có cùng trung bình μ và kỳ vọng σ^2. Khi đó $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ có tính chất sau :

1. $\overline{X}$ có trung bình là μ và kỳ vọng là $\frac{\sigma^2}{n}$;
2. Nếu $X_1, X_2, \ldots, X_n$ là các BNN có phân bố chuẩn, thì $\overline{X}$ cũng có phân bố chuẩn.

Định lý (Định lý Giới hạn trung tâm (Central Limit Theorem))

$X_1, X_2, \ldots, X_n$ là các BNN độc lập có cùng trung bình μ và kỳ vọng σ^2. Đặt $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Khi đó

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{2.30}$$

được xấp xỉ tiệm cận bởi một phân bố chuẩn có trung bình 0 và kỳ vọng 1 (phân bố chuẩn tắc), có nghĩa là

$$\lim_{n \to +\infty} F_n(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-u^2/2} du \tag{2.31}$$

với $F_n(x)$ là hàm phân bố của Z_n.

Áp dụng Định lý Giới hạn trung tâm để xây dựng khoảng tin cậy cho xác suất tổng thể p.

p : xác suất tổng thể của một tập nền để xuất hiện sự kiện A.

$$X := \begin{cases} 0 & \text{nếu không xảy ra sự kiện } A \text{ trong phép thử} \\ 1 & \text{nếu xảy ra sự kiện } A \text{ trong phép thử.} \end{cases}$$

Ta có

$$X \sim A(p) = B(1,p), \mu = EX = p, \sigma^2 = DX = p(1-p).$$

Xét một mẫu $(x_1, x_2, ..., x_n)$ có kích thước n.

Áp dụng Định lý Giới hạn trung tâm để xây dựng khoảng tin cậy cho xác suất tổng thể p.

p : xác suất tổng thể của một tập nền để xuất hiện sự kiện A.

$$X := \begin{cases} 0 & \text{nếu không xảy ra sự kiện } A \text{ trong phép thử} \\ 1 & \text{nếu xảy ra sự kiện } A \text{ trong phép thử.} \end{cases}$$

Ta có

$$X \sim A(p) = B(1,p), \mu = EX = p, \sigma^2 = DX = p(1-p).$$

Xét một mẫu $(x_1, x_2, ..., x_n)$ có kích thước n.
Khi đó $(x_1 + x_2 + \cdots + x_n)$ bằng số lần xuất hiện sự kiện A.

Áp dụng Định lý Giới hạn trung tâm để xây dựng khoảng tin cậy cho xác suất tổng thể p.

p : xác suất tổng thể của một tập nền để xuất hiện sự kiện A.

$$X := \begin{cases} 0 & \text{nếu không xảy ra sự kiện } A \text{ trong phép thử} \\ 1 & \text{nếu xảy ra sự kiện } A \text{ trong phép thử.} \end{cases}$$

Ta có

$$X \sim A(p) = B(1,p), \mu = EX = p, \sigma^2 = DX = p(1-p).$$

Xét một mẫu $(x_1, x_2, ..., x_n)$ có kích thước n.
Khi đó $(x_1 + x_2 + \cdots + x_n)$ bằng số lần xuất hiện sự kiện A.
Tần suất xuất hiện sự kiện A trong mẫu là

$$f = \frac{(x_1 + x_2 + \cdots + x_n)}{n} = \bar{x}.$$

Theo Định lý Giới hạn trung tâm, ta có thể xấp xỉ

$$Z_n = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

bởi phân bố chuẩn tắc khi n đủ lớn.

Do đó, theo công thức tính khoảng tin cậy cho kỳ vọng $\mu = p$ với độ tin cậy γ là

$$\bar{x} - \frac{U_{\alpha/2}\sigma}{\sqrt{n}} \leq \mu = p \leq \bar{x} + \frac{U_{\alpha/2}\sigma}{\sqrt{n}}, \tag{2.32}$$

ở đây $U_{\alpha/2}$ ứng với mức $\alpha/2$ của phân bố chuẩn tắc.

Do đó, theo công thức tính khoảng tin cậy cho kỳ vọng $\mu = p$ với độ tin cậy γ là

$$\bar{x} - \frac{U_{\alpha/2}\sigma}{\sqrt{n}} \leq \mu = p \leq \bar{x} + \frac{U_{\alpha/2}\sigma}{\sqrt{n}}, \tag{2.32}$$

ở đây $U_{\alpha/2}$ ứng với mức $\alpha/2$ của phân bố chuẩn tắc.
Tuy nhiên vì p chưa biết, nên thay p bởi f, thay σ bởi $\sqrt{f(1-f)}$ vào công thức trên, ta được khoảng tin cậy cho xác suất p với độ tin cậy γ (với n đủ lớn) là

$$f - U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} \leq p \leq f + U_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}, \tag{2.33}$$

Ví dụ

Trong một cuộc bầu cử tổng thống ở nước Mỹ, người ta phỏng vấn ngẫu nhiên 1600 cử tri và được biết có 960 người bầu cử cho ứng viên A. Với độ tin cậy $\gamma = 95\%$, tối thiểu ứng cử viên A sẽ chiếm được bao nhiêu phần trăm số phiếu bầu ?

Ví dụ

Trong một cuộc bầu cử tổng thống ở nước Mỹ, người ta phỏng vấn ngẫu nhiên 1600 cử tri và được biết có 960 người bầu cử cho ứng viên A. Với độ tin cậy $\gamma = 95\%$, tối thiểu ứng cử viên A sẽ chiếm được bao nhiêu phần trăm số phiếu bầu ?

Giải. Gọi p : tỷ lệ phiếu bầu cho ứng cử viên A. Gọi X là BNN :

$$X = \begin{cases} 0 & \text{nếu ứng cử viên A không được bầu} \\ 1 & \text{nếu ứng cử viên A được bầu.} \end{cases}$$

Khi đó X là BNN có phân bố không-một $A(p) \sim B(1, p)$. Ta có, tần suất mẫu là

$$f = \frac{960}{1600} = 0.6.$$

Vì $\gamma = 0.95$ nên $\alpha/2 = 2.5\%$ và $U_{\alpha/2} = 1.960$ và đó đó khoảng tin cậy của p là

$$0.6 - 1.960 \times \sqrt{\frac{0.6(1-0.6)}{1600}} \leq p \leq 0.6 + 1.960 \times \sqrt{\frac{0.6(1-0.6)}{1600}}$$

hay

$$0.576 \leq p \leq 0.624.$$

Vậy với độ tin cậy 95%, thì tối thiểu có 57.6% và tối đa 62.4% cử tri sẽ bỏ phiếu cho ứng viên A.

Chương 3. Kiểm định giả thuyết thống kê

- Thống kê hiện đại :
  1. Khoảng tin cậy
  2. Kiểm định giả thiết thống kê

Chương 3. Kiểm định giả thuyết thống kê

- Thống kê hiện đại :
  1. Khoảng tin cậy
  2. Kiểm định giả thiết thống kê
- Kiểm định giả thiết thống kê :
  1. Đưa ra suy luận từ mẫu đến tổng thể (tập nền) thông qua việc kiểm tra một giả thiết

Chương 3. Kiểm định giả thuyết thống kê

- Thống kê hiện đại :
  1. Khoảng tin cậy
  2. Kiểm định giả thiết thống kê
- Kiểm định giả thiết thống kê :
  1. Đưa ra suy luận từ mẫu đến tổng thể (tập nền) thông qua việc kiểm tra một giả thiết
  2. dựa trên kinh nghiệm hoặc quan sát, dựa trên lý thuyết hoặc yêu cầu chất lượng,...

Ví dụ

Ta sẽ mua 100 cuộn dây kim loại nào đó nếu có thể xác minh được khẳng định của nhà sản xuất rằng dây có giới hạn đứt $\mu = \mu_0 = 90kg$.

Ví dụ

Ta sẽ mua 100 cuộn dây kim loại nào đó nếu có thể xác minh được khẳng định của nhà sản xuất rằng dây có giới hạn đứt $\mu = \mu_0 = 90kg$.

Đây là một bài kiểm định của *giả thuyết (hypothesis)*

$$H_0 : \mu = \mu_0 = 90kg$$

Ví dụ

Ta sẽ mua 100 cuộn dây kim loại nào đó nếu có thể xác minh được khẳng định của nhà sản xuất rằng dây có giới hạn đứt $\mu = \mu_0 = 90kg$.

Đây là một bài kiểm định của *giả thuyết (hypothesis)*

$$H_0 : \mu = \mu_0 = 90kg$$

Không đạt yêu cầu nếu kiểm định (thống kê) cho thấy rằng giới hạn đứt của cuộn dây thực sự là

$$H_1 : \mu = \mu_1 < \mu_0.$$

Ví dụ

Ta sẽ mua 100 cuộn dây kim loại nào đó nếu có thể xác minh được khẳng định của nhà sản xuất rằng dây có giới hạn đứt $\mu = \mu_0 = 90kg$.

Đây là một bài kiểm định của *giả thuyết (hypothesis)*

$$H_0 : \mu = \mu_0 = 90kg$$

Không đạt yêu cầu nếu kiểm định (thống kê) cho thấy rằng giới hạn đứt của cuộn dây thực sự là

$$H_1 : \mu = \mu_1 < \mu_0.$$

Khi đó, H_1 : gọi là *đối thiết* (hay *giả thuyết thay thế*) của phép thử.

Ví dụ

Ta sẽ mua 100 cuộn dây kim loại nào đó nếu có thể xác minh được khẳng định của nhà sản xuất rằng dây có giới hạn đứt $\mu = \mu_0 = 90kg$.

Đây là một bài kiểm định của *giả thuyết (hypothesis)*

$$H_0 : \mu = \mu_0 = 90kg$$

Không đạt yêu cầu nếu kiểm định (thống kê) cho thấy rằng giới hạn đứt của cuộn dây thực sự là

$$H_1 : \mu = \mu_1 < \mu_0.$$

Khi đó, H_1 : gọi là *đối thiết* (hay *giả thuyết thay thế*) của phép thử.
Giả thuyết sẽ được *chấp nhận* nếu thử nghiệm cho thấy giả thuyết đó đúng, ngoại trừ một xác suất lỗi nhỏ α–được gọi là *mức ý nghĩa* của kiểm định.

Ví dụ

Ta sẽ mua 100 cuộn dây kim loại nào đó nếu có thể xác minh được khẳng định của nhà sản xuất rằng dây có giới hạn đứt $\mu = \mu_0 = 90kg$.

Đây là một bài kiểm định của *giả thuyết (hypothesis)*

$$H_0 : \mu = \mu_0 = 90kg$$

Không đạt yêu cầu nếu kiểm định (thống kê) cho thấy rằng giới hạn đứt của cuộn dây thực sự là

$$H_1 : \mu = \mu_1 < \mu_0.$$

Khi đó, H_1 : gọi là *đối thiết* (hay *giả thuyết thay thế*) của phép thử.
Giả thuyết sẽ được *chấp nhận* nếu thử nghiệm cho thấy giả thuyết đó đúng, ngoại trừ một xác suất lỗi nhỏ α–được gọi là *mức ý nghĩa* của kiểm định.
Thông thường ta chọn mức ý nghĩa $\alpha = 5\%$ hay $\alpha = 1\%$.

3.1.1. Thủ tục kiểm định giả thuyết thống kê

B1 :

- Phát biểu giả thiết
$$H_0 : \theta = \theta_0$$
(trong Ví dụ ta có $H_0 : \mu = \mu_0 = 90kg, \theta = \mu, \theta_0 = \mu_0$.)
- Phát biểu đối thiết
$$H_1 : \theta < (\neq, >)\theta_0$$
(trong Ví dụ ta có $H_1 : \mu < \mu_0 = 90kg, \theta = \mu, \theta_0 = \mu_0$.)
- Chọn mức ý nghĩa α (thường chọn $\alpha = 5\%,\ 1\%,\ 0.01\%$.)

3.1.1. Thủ tục kiểm định giả thuyết thống kê

B1 :
- Phát biểu giả thiết
$$H_0 : \theta = \theta_0$$
(trong Ví dụ ta có $H_0 : \mu = \mu_0 = 90kg, \theta = \mu, \theta_0 = \mu_0$.)
- Phát biểu đối thiết
$$H_1 : \theta < (\neq, >)\theta_0$$
(trong Ví dụ ta có $H_1 : \mu < \mu_0 = 90kg, \theta = \mu, \theta_0 = \mu_0$.)
- Chọn mức ý nghĩa α (thường chọn $\alpha = 5\%,\ 1\%,\ 0.01\%$.)

B2 : Chọn *tiêu chuẩn kiểm định*
$$T = g(X_1, X_2, ..., X_n; \theta_0)$$
và xác định quy luật phân bố của T với điều kiện giả thiết H_0 đúng.
($(X_1, X_2, ..., X_n)$ là biến ngẫu nhiên được lập từ một mẫu có kích thước n).

3.1.1. Các bước kiểm định giả thuyết thống kê

B3 : Tìm *miền bác bỏ* W_α sao cho

$$\mathbb{P}(T \in W_\alpha | H_0) = \alpha, \tag{3.34}$$

(*trong trường hợp giả thiết* H_0 *đúng thì xác suất xảy ra lỗi* $T \in W_\alpha$ *bằng* α.

3.1.1. Các bước kiểm định giả thuyết thống kê

B3 : Tìm *miền bác bỏ* W_α sao cho

$$\mathbb{P}(T \in W_\alpha | H_0) = \alpha, \tag{3.34}$$

(*trong trường hợp giả thiết* H_0 *đúng thì xác suất xảy ra lỗi* $T \in W_\alpha$ *bằng* α.

B4 : Từ mẫu cụ thể $(x_1, x_2, ..., x_n)$ tính *giá trị quan sát* của tiêu chuẩn kiểm định

$$T_{qs} = g(x_1, x_2, ..., x_n; \theta_0).$$

3.1.1. Các bước kiểm định giả thuyết thống kê

B3 : Tìm *miền bác bỏ* W_α sao cho

$$\mathbb{P}(T \in W_\alpha | H_0) = \alpha, \tag{3.34}$$

(*trong trường hợp giả thiết* H_0 *đúng thì xác suất xảy ra lỗi* $T \in W_\alpha$ *bằng* α.

B4 : Từ mẫu cụ thể $(x_1, x_2, ..., x_n)$ tính *giá trị quan sát* của tiêu chuẩn kiểm định

$$T_{qs} = g(x_1, x_2, ..., x_n; \theta_0).$$

B5 : Chấp nhận hoặc bác bỏ giả thiết H_0 dựa trên mối quan hệ giữa giá trị quan sát T_{qs} và miền bác bỏ W_α :

- Nếu $T_{qs} \in W_\alpha$, theo nguyên tắc kiểm định thì H_0 sai, do đó **bác bỏ** H_0 và thừa nhận H_1.
- Nếu $T_{qs} \notin W_\alpha$, thì **chưa có đủ cơ sở để bác bỏ** H_0 (trên thực tế ta thừa nhận H_0).

3.1.2. Sai lầm của kiểm định thống kê

Kiểm định (thống kê) có thể mắc hai loại sai lầm sau :

1. Bác bỏ giả thiết H_0 trong khi nó đúng.
   (Sai lầm loại này sinh ra khi kích thước mẫu quá nhỏ, do phương pháp lấy mẫu. Xác suất mắc sai lầm loại 1 đúng bằng mức ý nghĩa α).
2. Thừa nhận H_0 trong khi nó sai.
   (Sai lầm loại này xảy ra khi giá trị quan sát $T_{qs} \notin W_\alpha$ trong khi H_1 đúng).

3.1.2. Sai lầm của kiểm định thống kê

Kiểm định (thống kê) có thể mắc hai loại sai lầm sau :

1. Bác bỏ giả thiết H_0 trong khi nó đúng.
   (Sai lầm loại này sinh ra khi kích thước mẫu quá nhỏ, do phương pháp lấy mẫu. Xác suất mắc sai lầm loại 1 đúng bằng mức ý nghĩa α).
2. Thừa nhận H_0 trong khi nó sai.
   (Sai lầm loại này xảy ra khi giá trị quan sát $T_{qs} \notin W_\alpha$ trong khi H_1 đúng).
   Xác suất xảy ra sai lầm loại 2 được xác định như sau

$$\mathbb{P}(T \notin W_\alpha | H_1) = \beta.$$

Đại lượng $\eta = 1 - \beta = \mathbb{P}(T \in W_\alpha | H_1)$: *lực lượng kiểm định* (power of the test).

3.1.2. Sai lầm của kiểm định thống kê

Kiểm định (thống kê) có thể mắc hai loại sai lầm sau :

1. Bác bỏ giả thiết H_0 trong khi nó đúng.
   (Sai lầm loại này sinh ra khi kích thước mẫu quá nhỏ, do phương pháp lấy mẫu. Xác suất mắc sai lầm loại 1 đúng bằng mức ý nghĩa α).
2. Thừa nhận H_0 trong khi nó sai.
   (Sai lầm loại này xảy ra khi giá trị quan sát $T_{qs} \notin W_\alpha$ trong khi H_1 đúng).
   Xác suất xảy ra sai lầm loại 2 được xác định như sau

$$\mathbb{P}(T \notin W_\alpha | H_1) = \beta.$$

Đại lượng $\eta = 1 - \beta = \mathbb{P}(T \in W_\alpha | H_1)$: *lực lượng kiểm định* (power of the test).

3.1.3. Hai loại bài toán kiểm định

1. *Kiểm định tham số*
   - Kiểm định các tham số đặc trưng (như kỳ vọng, phương sai, tần suất,...) của biến ngẫu nhiên gốc trong tổng thể.
   - Kiểm định về sự bằng nhau của hai kỳ vọng, của hai phương sai, hay của hai tần suất của tổng thể.
2. *Kiểm định phi tham số*

   Bao gồm các bài toán về quy luật phân bố xác suất của biến ngẫu nhiên gốc của tổng thể, về tính độc lập của các dấu hiệu nghiên cứu định tính...

3.2. Kiểm định kỳ vọng của phân phối chuẩn

BNN X có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$. Ta có giả thiết thống kê

$$H_0 : \mu = \mu_0. \tag{3.35}$$

3.2. Kiểm định kỳ vọng của phân phối chuẩn

BNN X có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$. Ta có giả thiết thống kê

$$H_0 : \mu = \mu_0. \tag{3.35}$$

Bài toán : Kiểm định giả thiết H_0.

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai σ^2

Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên

$$W = (X_1, ..., X_n),$$

có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai σ^2

Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên

$$W = (X_1, ..., X_n),$$

có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.
Theo Định lý tổng của BNN có phân bố chuẩn, kiểm định thống kê

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

là BNN có phân bố chuẩn với kỳ vọng 0 và phương sai 1 (phân bố chuẩn tắc) nếu giả thiết H_0 đúng.

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai σ^2

Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên

$$W = (X_1, ..., X_n),$$

có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.
Theo Định lý tổng của BNN có phân bố chuẩn, kiểm định thống kê

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

là BNN có phân bố chuẩn với kỳ vọng 0 và phương sai 1 (phân bố chuẩn tắc) nếu giả thiết H_0 đúng.
Ta xét các bài toán kiểm định sau.

Bài toán 1 (Kiểm định hai phía)

1.
$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0, \\ \text{Đối thiết } H_1 : \mu \neq \mu_0 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.36}$$

2. Miền bác bỏ
$$W_\alpha = \left\{T \,\middle|\, |T| > U_{\alpha/2}\right\} \tag{3.37}$$
ở đó giá trị tới hạn $U_{\alpha/2}$ được xác định bởi (bảng A8)
$$\Phi(U_{\alpha/2}) = 1 - \frac{\alpha}{2}.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 1 (Kiểm định hai phía)

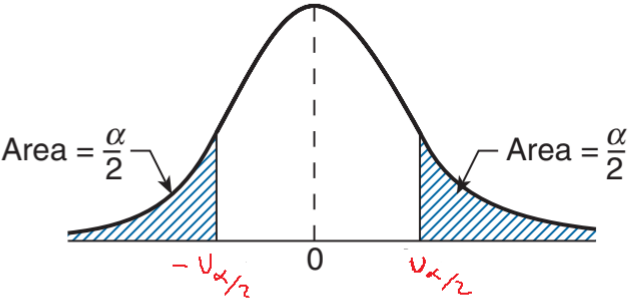

Figure – Miền bác bỏ W_α - phần gạch sọc

Bài toán 2 (Kiểm định một phía bên phải) (Kiểm định phải)

1.
$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0, \\ \text{Đối thiết } H_1 : \mu > \mu_0 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.38}$$

2. Miền bác bỏ
$$W_\alpha = \{T \,|\, T > U_\alpha\} \tag{3.39}$$
ở đó giá trị tới hạn U_α được xác định bởi (bảng A8)
$$\Phi(U_\alpha) = 1 - \alpha.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 2 (Kiểm định một phía bên phải) (Kiểm định phải)

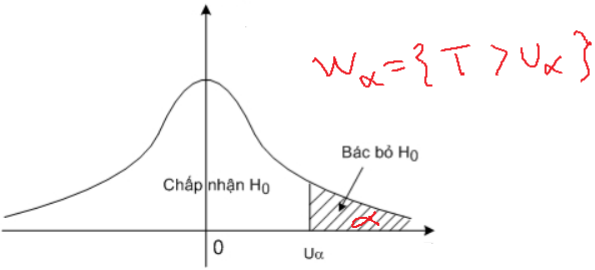

Figure – Miền bác bỏ W_α - phần gạch sọc

Bài toán 3 (Kiểm định một phía bên trái) (Kiểm định trái)

1. $$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0, \\ \text{Đối thiết } H_1 : \mu < \mu_0 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.40}$$

2. Miền bác bỏ
$$W_\alpha = \{T \,|\, T < -U_\alpha\} \tag{3.41}$$
ở đó giá trị tới hạn U_α được xác định bởi (bảng A8)
$$\Phi(U_\alpha) = 1 - \alpha.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 3 (Kiểm định một phía bên trái) (Kiểm định trái)

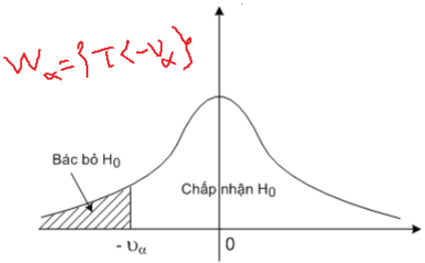

Figure – Miền bác bỏ W_α - phần gạch sọc

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai

Ví dụ

X là BNN có phân bố chuẩn với phương sai $\sigma^2 = 9$. Sử dụng một mẫu kích cỡ $n = 20$ với trung bình $\bar{x} = 23$, kiểm định giả thiết $\mu = \mu_0 = 24$ với các đối thiết sau

1. $\mu \neq \mu_0$,
2. $\mu > \mu_0$,
3. $\mu < \mu_0$.

Biết mức ý nghĩa $\alpha = 5\%$.

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai

Lời giải. (i) Giá trị tới hạn $U_{\alpha/2}$ thỏa mãn

$$\Phi(U_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975.$$

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai

Lời giải. (i) Giá trị tới hạn $U_{\alpha/2}$ thỏa mãn

$$\Phi(U_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975.$$

Ta có $U_{\alpha/2} = 1.960(A8)$,

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai

Lời giải. (i) Giá trị tới hạn $U_{\alpha/2}$ thỏa mãn

$$\Phi(U_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975.$$

Ta có $U_{\alpha/2} = 1.960(A8)$,suy ra miền bác bỏ

$$W_\alpha = \{T \,|\, |T| > 1.960\}.$$

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai

Lời giải. (i) Giá trị tới hạn $U_{\alpha/2}$ thỏa mãn

$$\Phi(U_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975.$$

Ta có $U_{\alpha/2} = 1.960(A8)$,suy ra miền bác bỏ

$$W_\alpha = \{T \mid |T| > 1.960\}.$$

Giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{-1}{3/\sqrt{10}} = -1.054.$$

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai

Lời giải. (i) Giá trị tới hạn $U_{\alpha/2}$ thỏa mãn

$$\Phi(U_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975.$$

Ta có $U_{\alpha/2} = 1.960(A8)$,suy ra miền bác bỏ

$$W_\alpha = \{T \mid |T| > 1.960\}.$$

Giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{-1}{3/\sqrt{10}} = -1.054.$$

Vậy ta có $T_{qs} \notin W_\alpha$ suy ra chấp nhận giả thiết H_0.

3.2.1.Kiểm định kỳ vọng khi đã biết phương sai

Lời giải. (i) Giá trị tới hạn $U_{\alpha/2}$ thỏa mãn

$$\Phi(U_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975.$$

Ta có $U_{\alpha/2} = 1.960(A8)$,suy ra miền bác bỏ

$$W_\alpha = \{T \,|\, |T| > 1.960\}.$$

Giá trị quan sát của tiêu chuẩn kiểm định

$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{-1}{3/\sqrt{10}} = -1.054.$$

Vậy ta có $T_{qs} \notin W_\alpha$ suy ra chấp nhận giả thiết H_0.
Tương tự với (ii), (iii)

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai σ^2

Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên

$$W = (X_1, ..., X_n),$$

có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai σ^2

Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên

$$W = (X_1, ..., X_n),$$

có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.Đặt

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n),$$

$$S^2 = \frac{1}{n-1}\left[(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right].$$

Theo Định lý về phân phối Student, BNN

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

có phân bố Student $(n-1)$ bậc tự do nếu giả thiết H_0 đúng.

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai σ^2

Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên

$$W = (X_1, ..., X_n),$$

có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.Đặt

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n),$$

$$S^2 = \frac{1}{n-1}\left[(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right].$$

Theo Định lý về phân phối Student, BNN

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

có phân bố Student $(n-1)$ bậc tự do nếu giả thiết H_0 đúng.
Ta xét các bài toán kiểm định sau.

Bài toán 1 (Kiểm định hai phía)

1. $$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0, \\ \text{Đối thiết } H_1 : \mu \neq \mu_0 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.42}$$

2. Miền bác bỏ
$$W_\alpha = \left\{T \,\middle|\, |T| > t_{\alpha/2}\right\} \tag{3.43}$$
ở đó giá trị tới hạn $t_{\alpha/2}$ được xác định bởi (bảng A9) bậc $(n-1)$
$$F(t_{\alpha/2}) = 1 - \frac{\alpha}{2}.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 1 (Kiểm định hai phía)

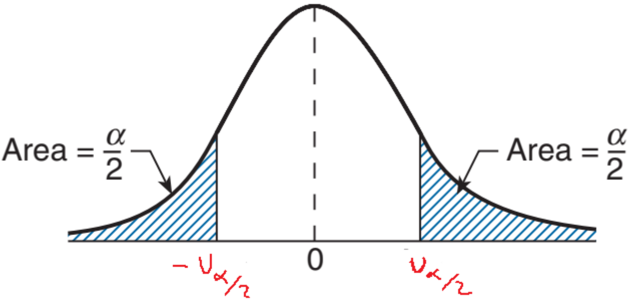

Figure – Miền bác bỏ W_α - phần gạch sọc

Bài toán 2 (Kiểm định một phía bên phải)

1.
$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0, \\ \text{Đối thiết } H_1 : \mu > \mu_0 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.44}$$

2. Miền bác bỏ
$$W_\alpha = \{T \,|\, T > t_\alpha\} \tag{3.45}$$
ở đó giá trị tới hạn t_α được xác định bởi (bảng A9, bậc $(n-1)$)
$$F(t_\alpha) = 1 - \alpha.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

4. Xét giá trị T_{qs} :
    - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
    - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 3 (Kiểm định một phía bên trái)

1.
$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0, \\ \text{Đối thiết } H_1 : \mu < \mu_0 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.46}$$

2. Miền bác bỏ
$$W_\alpha = \{T \mid T < -t_\alpha\} \tag{3.47}$$
ở đó giá trị tới hạn t_α được xác định bởi (bảng A9 bậc $(n-1)$)
$$F(t_\alpha) = 1 - \alpha.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai

Ví dụ

Thử nghiệm sức căng của một mẫu gồm $n = 16$ sợi dây thừng (đường kính 3 inh), thu được trung bình mẫu $\bar{x} = 4482$ kg, độ lệch chuẩn mẫu $s = 115$ kg. Biết rằng sức căng của dây thừng là một biến ngẫu nhiên có phân bố chuẩn. Với mức ý nghĩa 5%, hãy kiểm định giả thiết sức căng trung bình của dây thừng là $\mu = \mu_0 = 4500$ kg so với đối thiết $\mu = \mu_1 = 4400$ kg.

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0 = 4500, \\ \text{Đối thiết } H_1 : \mu = \mu_1 < \mu_0, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0 = 4500, \\ \text{Đối thiết } H_1 : \mu = \mu_1 < \mu_0, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Giá trị tới hạn t_α thỏa mãn

$$F(t_\alpha) = 1 - \alpha = 0.95$$

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0 = 4500, \\ \text{Đối thiết } H_1 : \mu = \mu_1 < \mu_0, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Giá trị tới hạn t_α thỏa mãn

$$F(t_\alpha) = 1 - \alpha = 0.95$$

Tra bảng A9 với $(n-1) = 15$ bậc tự do ta có $t_\alpha = 1.75$.

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0 = 4500, \\ \text{Đối thiết } H_1 : \mu = \mu_1 < \mu_0, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Giá trị tới hạn t_α thỏa mãn

$$F(t_\alpha) = 1 - \alpha = 0.95$$

Tra bảng A9 với $(n-1) = 15$ bậc tự do ta có $t_\alpha = 1.75$.
Miền bác bỏ

$$W_\alpha = \{T \,|\, T < -1.75\}\,. \tag{3.48}$$

3.2.2.Kiểm định kỳ vọng khi chưa biết phương sai

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu = \mu_0 = 4500, \\ \text{Đối thiết } H_1 : \mu = \mu_1 < \mu_0, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Giá trị tới hạn t_α thỏa mãn

$$F(t_\alpha) = 1 - \alpha = 0.95$$

Tra bảng A9 với $(n-1) = 15$ bậc tự do ta có $t_\alpha = 1.75$.
Miền bác bỏ

$$W_\alpha = \{T \mid T < -1.75\}. \tag{3.48}$$

Giá trị quan sát

$$T_{qs} = \frac{4482 - 4500}{115/\sqrt{16}} = -0.626.$$

Vì $T_{qs} > -1.75$ nên $T_{qs} \notin W_\alpha$ và do đó chấp nhận H_0.

3.3.1.Phân bố χ^2 (khi-bình phương)

Định nghĩa (Phân bố khi-bình phương)

BNN X được gọi là có phân bố *khi-bình phương* với m bậc tự do và ký hiệu $X \sim \chi^2(m)$ nếu hàm phân bố $F(x)$ của nó có dạng

$$F(x) = \begin{cases} 0 & \text{nếu } x < 0, \\ C_m \int_0^x e^{-u/2} u^{(m-2)/2} du & \text{nếu } x \geq 0 \end{cases}, \quad (3.49)$$

$C_m = \frac{1}{2^{m/2}\Gamma(m/2)}$, $\Gamma(t)$ là hàm Gamma.

3.3.1.Phân bố χ^2 (khi-bình phương)

Hàm mật độ của phân bố khi-bình phương ứng với 2,3, và 5 bậc tự do

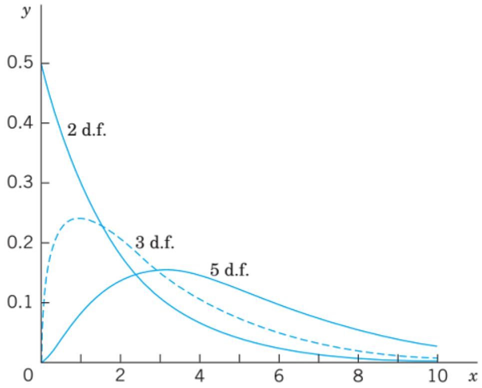

Figure – Hàm mật độ của phân bố χ^2 với 2,3, và 5 bậc tự do

3.3.1.Phân bố χ^2 (khi-bình phương)

Định lý (Phân bố khi-bình phương)

$X_1, \cdots, X_n$: BNN độc lập có phân bố chuẩn, có cùng kỳ vọng μ và phương sai σ^2. Khi đó, BNN

$$Y = (n-1)\frac{S^2}{\sigma^2}, \tag{3.50}$$

có phân bố khi-bình phương với $(n-1)$ bậc tự do, với

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

BNN X có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$. Ta có giả thiết thống kê

$$H_0 : \sigma^2 = \sigma_0^2. \tag{3.51}$$

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

BNN X có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$. Ta có giả thiết thống kê

$$H_0 : \sigma^2 = \sigma_0^2. \tag{3.51}$$

Bài toán : Kiểm định giả thiết H_0.

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

BNN X có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$. Ta có giả thiết thống kê

$$H_0 : \sigma^2 = \sigma_0^2. \tag{3.51}$$

Bài toán : Kiểm định giả thiết H_0.
Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên $W = (X_1, ..., X_n)$, ở đó $X_1, ..., X_n$ là các biến ngẫu nhiên có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

BNN X có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$. Ta có giả thiết thống kê

$$H_0 : \sigma^2 = \sigma_0^2. \tag{3.51}$$

Bài toán : Kiểm định giả thiết H_0.
Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên $W = (X_1, ..., X_n)$, ở đó $X_1, ..., X_n$ là các biến ngẫu nhiên có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.
Đặt

$$S^2 = \frac{1}{n-1}\left[(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right].$$

Theo Định lý Phân bố khi- bình phương, kiểm định thống kê

$$T = (n-1)\frac{S^2}{\sigma_0^2}$$

là biến ngẫu nhiên có phân bố khi-bình phương $(n-1)$ bậc tự do nếu giả thiết H_0 đúng.

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

BNN X có phân bố chuẩn $\mathcal{N}(\mu, \sigma^2)$. Ta có giả thiết thống kê

$$H_0 : \sigma^2 = \sigma_0^2. \tag{3.51}$$

Bài toán : Kiểm định giả thiết H_0.
Xét mẫu ngẫu nhiên $(x_1, ..., x_n)$ tương ứng với biến ngẫu nhiên $W = (X_1, ..., X_n)$, ở đó $X_1, ..., X_n$ là các biến ngẫu nhiên có phân bố chuẩn và có cùng kỳ vọng và phương sai với X.
Đặt

$$S^2 = \frac{1}{n-1}\left[(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right].$$

Theo Định lý Phân bố khi- bình phương, kiểm định thống kê

$$T = (n-1)\frac{S^2}{\sigma_0^2}$$

là biến ngẫu nhiên có phân bố khi-bình phương $(n-1)$ bậc tự do nếu giả thiết H_0 đúng.
Ta xét các bài toán kiểm định sau.

Bài toán 1 (Kiểm định hai phía)

1. $$\begin{cases} \text{Giả thiết } H_0 : \sigma^2 = \sigma_0^2, \\ \text{Đối thiết } H_1 : \sigma^2 \neq \sigma_0^2 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.52}$$

2. Miền bác bỏ
$$W_\alpha = \{T \mid T < \chi^2_{1-\alpha/2} \text{ hoặc } T > \chi^2_{\alpha/2}\} \tag{3.53}$$
ở đó giá trị tới hạn χ^2_z được xác định bởi (bảng A10 $(n-1)bc$)
$$F(\chi^2_z) = 1 - z.$$

3. Tính giá trị quan sát
$$T_{qs} = (n-1)\frac{s^2}{\sigma_0^2}.$$

4. Xét giá trị T_{qs} :
    - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
    - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 1 (Kiểm định hai phía)

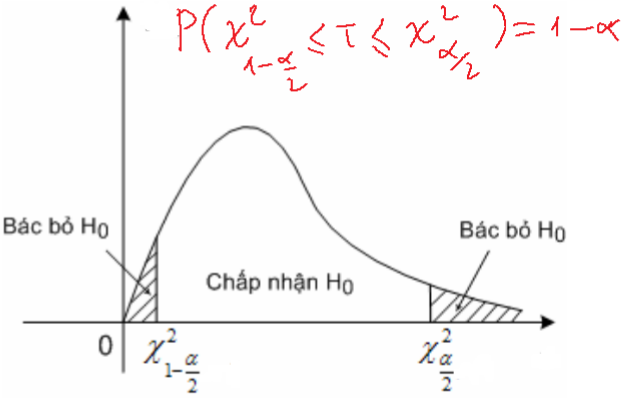

Figure – Miền bác bỏ W_α - phần gạch sọc

Bài toán 2 (Kiểm định phải)

1.
$$\begin{cases} \text{Giả thiết } H_0 : \sigma^2 = \sigma_0^2, \\ \text{Đối thiết } H_1 : \sigma^2 > \sigma_0^2 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.54}$$

2. Miền bác bỏ
$$W_\alpha = \{T \mid T > \chi_\alpha^2\} \tag{3.55}$$
ở đó giá trị tới hạn χ_α^2 được xác định bởi (bảng A10 $(n-1)$ bậc)
$$F(\chi_z^2) = 1 - z..$$

3. Tính giá trị quan sát
$$T_{qs} = (n-1)\frac{S^2}{\sigma_0^2}.$$

4. Xét giá trị T_{qs} :
    - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
    - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 2 (Kiểm định phải)

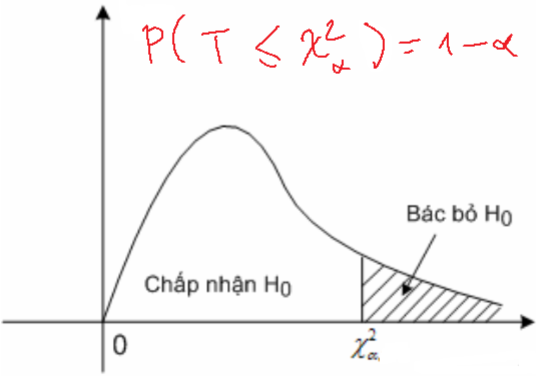

Figure – Miền bác bỏ W_α - phần gạch sọc

Bài toán 3 (Kiểm định trái)

1. $$\begin{cases} \text{Giả thiết } H_0 : \sigma^2 = \sigma_0^2, \\ \text{Đối thiết } H_1 : \sigma^2 \le \sigma_0^2 \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.56}$$

2. Miền bác bỏ
$$W_\alpha = \{T \,|\, T < \chi^2_{1-\alpha}\} \tag{3.57}$$
ở đó giá trị tới hạn χ^2_α được xác định bởi (bảng A10 $(n-1)$ bậc)
$$F(\chi^2_z) = 1 - z..$$

3. Tính giá trị quan sát
$$T_{qs} = (n-1)\frac{S^2}{\sigma_0^2}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 3 (Kiểm định trái)

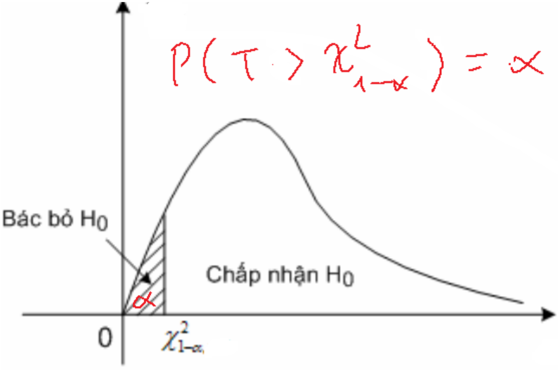

Figure – Miền bác bỏ W_α - phần gạch sọc

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

Ví dụ

Sử dụng mẫu ngẫu nhiên kích thước $n = 15$ và phương sai mẫu $s^2 = 13$ từ một tập nền có phân bố chuẩn, hãy kiểm định giả thiết $\sigma^2 = \sigma_0^2 = 10$ và đối thiết $\sigma^2 = \sigma_1^2 = 20$, với mức ý nghĩa $\alpha = 5\%$.

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \sigma^2 = \sigma_0^2 = 10, \\ \text{Đối thiết } H_1 : \sigma^2 = \sigma_1^2 > \sigma_0^2, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \sigma^2 = \sigma_0^2 = 10, \\ \text{Đối thiết } H_1 : \sigma^2 = \sigma_1^2 > \sigma_0^2, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Giá trị tới hạn χ^2_α thỏa mãn

$$F(\chi^2_\alpha) = 1 - \alpha = 0.95.$$

Tra bảng A10 với $(n-1) = 14$ bậc tự do, ta có $\chi^2_\alpha = 23.68$

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \sigma^2 = \sigma_0^2 = 10, \\ \text{Đối thiết } H_1 : \sigma^2 = \sigma_1^2 > \sigma_0^2, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Giá trị tới hạn χ^2_α thỏa mãn

$$F(\chi^2_\alpha) = 1 - \alpha = 0.95.$$

Tra bảng A10 với $(n-1) = 14$ bậc tự do, ta có $\chi^2_\alpha = 23.68$
Miền bác bỏ

$$W_\alpha = \{T \,|\, T > 23.68\}. \tag{3.58}$$

3.3.2.Kiểm định cho phương sai của phân bố chuẩn

Lời giải. Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \sigma^2 = \sigma_0^2 = 10, \\ \text{Đối thiết } H_1 : \sigma^2 = \sigma_1^2 > \sigma_0^2, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Giá trị tới hạn χ^2_α thỏa mãn

$$F(\chi^2_\alpha) = 1 - \alpha = 0.95.$$

Tra bảng A10 với $(n-1) = 14$ bậc tự do, ta có $\chi^2_\alpha = 23.68$
Miền bác bỏ

$$W_\alpha = \{T \,|\, T > 23.68\}. \tag{3.58}$$

Giá trị quan sát

$$T_{qs} = (n-1)\frac{S^2}{\sigma_0^2} = (15-1) \times \frac{13}{10} = 18.2 \leq 23.68.$$

Do đó $T_{qs} \notin W_\alpha$ và vì vậy ta chấp nhận H_0.

Xét cùng lúc hai tổng thể.

1. Tổng thể có dấu hiệu nghiên cứu là BNN X với phân bố chuẩn $\mathcal{N}(\mu_X, \sigma_X^2)$.
2. Tổng thể có dấu hiện nghiên cứu là BNN Y với phân bố chuẩn $\mathcal{N}(\mu_Y, \sigma_Y^2)$.

Xét cùng lúc hai tổng thể.

1. Tổng thể có dấu hiệu nghiên cứu là BNN X với phân bố chuẩn $\mathcal{N}(\mu_X, \sigma_X^2)$.
2. Tổng thể có dấu hiện nghiên cứu là BNN Y với phân bố chuẩn $\mathcal{N}(\mu_Y, \sigma_Y^2)$.

Giả sử $\sigma_X = \sigma_Y$.

Bài toán : Kiểm định giả thiết thống kê

$$H_0 : \mu_X = \mu_Y.$$

Xét cùng lúc hai tổng thể.

1. Tổng thể có dấu hiệu nghiên cứu là BNN X với phân bố chuẩn $\mathcal{N}(\mu_X, \sigma_X^2)$.
2. Tổng thể có dấu hiện nghiên cứu là BNN Y với phân bố chuẩn $\mathcal{N}(\mu_Y, \sigma_Y^2)$.

Giả sử $\sigma_X = \sigma_Y$.
Bài toán : Kiểm định giả thiết thống kê

$$H_0 : \mu_X = \mu_Y.$$

Phương pháp : Sử dụng hai mẫu tương ứng với hai tổng thể :

1. $(x_1, ..., x_{n_1})$, có kích thước n_1 ứng với biến ngẫu nhiên
$$(X_1, ..., X_{n_1})$$
2. $(y_1, ..., y_{n_2})$ có kích thước n_2 ứng với biến ngẫu nhiên
$$(Y_1, ..., Y_{n_2}).$$

Xét cùng lúc hai tổng thể.

1. Tổng thể có dấu hiệu nghiên cứu là BNN X với phân bố chuẩn $\mathcal{N}(\mu_X, \sigma_X^2)$.
2. Tổng thể có dấu hiện nghiên cứu là BNN Y với phân bố chuẩn $\mathcal{N}(\mu_Y, \sigma_Y^2)$.

Giả sử $\sigma_X = \sigma_Y$.
Bài toán : Kiểm định giả thiết thống kê

$$H_0 : \mu_X = \mu_Y.$$

Phương pháp : Sử dụng hai mẫu tương ứng với hai tổng thể :

1. $(x_1, ..., x_{n_1})$, có kích thước n_1 ứng với biến ngẫu nhiên
$$(X_1, ..., X_{n_1})$$
2. $(y_1, ..., y_{n_2})$ có kích thước n_2 ứng với biến ngẫu nhiên
$$(Y_1, ..., Y_{n_2}).$$

Đặt

$$\bar{X} = \frac{1}{n_1}(X_1 + \cdots + X_{n_1}), \quad S_X^2 = \frac{1}{n_1 - 1}\left[(X_1 - \bar{X})^2 + \cdots + (X_{n_1} - \bar{X})^2\right]$$
$$\bar{Y} = \frac{1}{n_2}(Y_1 + \cdots + Y_{n_2}), \quad S_Y^2 = \frac{1}{n_2 - 1}\left[(Y_1 - \bar{Y})^2 + \cdots + (Y_{n_2} - \bar{Y})^2\right]$$

Đặt

$$S = \sqrt{\frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}}.$$

Đặt

$$S = \sqrt{\frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}}.$$

Xét tiêu chuẩn kiểm định

$$T = \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Khi đó, nếu giả thiết H_0 đúng thì T là biến ngẫu nhiên có phân bố Student $(n_1 + n_2 - 2)$ bậc tự do.

Đặt

$$S = \sqrt{\frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}}.$$

Xét tiêu chuẩn kiểm định

$$T = \frac{\bar{X} - \bar{Y}}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

Khi đó, nếu giả thiết H_0 đúng thì T là biến ngẫu nhiên có phân bố Student $(n_1 + n_2 - 2)$ bậc tự do.

Đặc biệt : Khi $n_1 = n_2 = n$, ta có

$$T = \sqrt{n}\frac{\bar{X} - \bar{Y}}{\sqrt{S_X^2 + S_Y^2}}$$

và T là biến ngẫu nhiên có phân bố Student $(2n - 2)$ bậc tự do khi giả thiết H_0 đúng.
Ta xét các bài toán kiểm định sau.

Bài toán 1 (Kiểm định hai phía)

1.
$$\begin{cases} \text{Giả thiết } H_0 : \mu_X = \mu_Y, \\ \text{Đối thiết } H_1 : \mu_X \neq \mu_Y \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.59}$$

2. Miền bác bỏ
$$W_\alpha = \left\{T \,\middle|\, |T| > t_{\alpha/2}\right\} \tag{3.60}$$
ở đó giá trị tới hạn $t_{\alpha/2}$ được xác định bởi (bảng A9) bậc $(n_1 + n_2 - 2)$
$$F(t_{\alpha/2}) = 1 - \frac{\alpha}{2}.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 2 (Kiểm định bên phải)

1. $$\begin{cases} \text{Giả thiết } H_0 : \mu_X = \mu_Y, \\ \text{Đối thiết } H_1 : \mu_X > \mu_Y \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.61}$$

2. Miền bác bỏ
$$W_\alpha = \{T \,|\, T > t_\alpha\} \tag{3.62}$$
ở đó giá trị tới hạn t_α được xác định bởi (bảng A9, bậc $(n_1 + n_2 - 2)$)
$$F(t_\alpha) = 1 - \alpha.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

Bài toán 3 (Kiểm định bên trái)

1. $$\begin{cases} \text{Giả thiết } H_0 : \mu_X = \mu_Y, \\ \text{Đối thiết } H_1 : \mu_X < \mu_Y \\ \text{Mức ý nghĩa } : \alpha. \end{cases} \tag{3.63}$$

2. Miền bác bỏ
$$W_\alpha = \{T \,|\, T < -t_\alpha\} \tag{3.64}$$
ở đó giá trị tới hạn t_α được xác định bởi (bảng A9 bậc $(n_1 + n_2 - 2)$)
$$F(t_\alpha) = 1 - \alpha.$$

3. Tính giá trị quan sát
$$T_{qs} = \frac{\bar{x} - \bar{y}}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

4. Xét giá trị T_{qs} :
   - $T_{qs} \in W_\alpha$,**bác bỏ** H_0, thừa nhận H_1.
   - $T_{qs} \notin W_\alpha$, thì **thừa nhận** H_0.

3.4.So sánh kỳ vọng của hai phân phối chuẩn

Ví dụ

Xét hai mẫu ngẫu nhiên thể hiện sản lượng của các công nhân sản xuất tấm thiếc trong hai môi trường làm việc khác nhau như sau

X	105	108	86	103	103	107	124	105
Y	89	92	84	97	103	107	111	97

Giả thiết rằng hai tổng thể tương ứng có phân bố chuẩn có cùng phương sai. Với mức ý nghĩa 5%, có thể nói : Sản lượng trung bình của các công nhân trong hai môi trường làm việc trên là như nhau hay không ?

Lời giải.

Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu_X = \mu_Y, \\ \text{Đối thiết } H_1 : \mu_X \neq \mu_Y, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Lời giải.

Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu_X = \mu_Y, \\ \text{Đối thiết } H_1 : \mu_X \neq \mu_Y, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Dựa vào giá trị của hai mẫu trên ta có

$$n_1 = n_2 = 8, \quad \bar{x} = 105.125, \quad \bar{y} = 97.500, \quad s_x^2 = 106.125, \quad s_x^2 = 84.000.$$

Với mức ý nghĩa $\alpha = 5\%$, từ bảng A9 ứng với $(n_1 + n_2 - 2) = 14$ bậc tự do và công thức

$$F(t_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975,$$

ta thu được $t_{\alpha/2} = 2.14$. Do đó miền bác bỏ

$$W_\alpha = \{T \mid |T| > 2.14\}$$

với $n = 8$.

Lời giải.

Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu_X = \mu_Y, \\ \text{Đối thiết } H_1 : \mu_X \neq \mu_Y, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Dựa vào giá trị của hai mẫu trên ta có

$$n_1 = n_2 = 8, \quad \bar{x} = 105.125, \quad \bar{y} = 97.500, \quad s_x^2 = 106.125, \quad s_x^2 = 84.000.$$

Với mức ý nghĩa $\alpha = 5\%$, từ bảng A9 ứng với $(n_1 + n_2 - 2) = 14$ bậc tự do và công thức

$$F(t_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975,$$

ta thu được $t_{\alpha/2} = 2.14$. Do đó miền bác bỏ

$$W_\alpha = \{T \mid |T| > 2.14\}$$

với $n = 8$.Giá trị quan sát của tiêu chuẩn kiểm định T là

$$T_{qs} = \sqrt{n}\frac{\bar{x} - \bar{y}}{\sqrt{s_X^2 + s_Y^2}}\sqrt{8} = \frac{105.125 - 97.500}{\sqrt{106.125 + 84.000}} = 1.56.$$

Lời giải.

Ta có

$$\begin{cases} \text{Giả thiết } H_0 : \mu_X = \mu_Y, \\ \text{Đối thiết } H_1 : \mu_X \neq \mu_Y, \\ \text{Mức ý nghĩa } : \alpha = 0.05. \end{cases}$$

Dựa vào giá trị của hai mẫu trên ta có

$$n_1 = n_2 = 8, \quad \bar{x} = 105.125, \quad \bar{y} = 97.500, \quad s_x^2 = 106.125, \quad s_x^2 = 84.000.$$

Với mức ý nghĩa $\alpha = 5\%$, từ bảng A9 ứng với $(n_1 + n_2 - 2) = 14$ bậc tự do và công thức

$$F(t_{\alpha/2}) = 1 - \frac{\alpha}{2} = 0.975,$$

ta thu được $t_{\alpha/2} = 2.14$. Do đó miền bác bỏ

$$W_\alpha = \{T \mid |T| > 2.14\}$$

với $n = 8$.Giá trị quan sát của tiêu chuẩn kiểm định T là

$$T_{qs} = \sqrt{n}\frac{\bar{x} - \bar{y}}{\sqrt{s_X^2 + s_Y^2}}\sqrt{8} = \frac{105.125 - 97.500}{\sqrt{106.125 + 84.000}} = 1.56.$$

Vì $|T_{qs}| \leq 2.14$ nên $T_{qs} \notin W_\alpha$ và vì vậy ta chấp nhận H_0.

Kiểm định mức độ phù hợp của mô hình : Thông qua một mẫu ngẫu nhiên

$$(x_1, ..., x_n)$$

ta kiểm định một hàm $F(x)$ có phải là hàm phân bố của biến ngẫu nhiên gốc X trong tổng thể hay không ?

Kiểm định mức độ phù hợp của mô hình : Thông qua một mẫu ngẫu nhiên

$$(x_1, ..., x_n)$$

ta kiểm định một hàm $F(x)$ có phải là hàm phân bố của biến ngẫu nhiên gốc X trong tổng thể hay không ?
Ta có giả thiết thống kê

$$H_0 : F(x) \text{ là hàm phân bố của } X$$

và đối thiết

$$H_1 : F(x) \text{ không là hàm phân bố của } X.$$

Kiểm định mức độ phù hợp của mô hình : Thông qua một mẫu ngẫu nhiên

$$(x_1, ..., x_n)$$

ta kiểm định một hàm $F(x)$ có phải là hàm phân bố của biến ngẫu nhiên gốc X trong tổng thể hay không ?
Ta có giả thiết thống kê

$$H_0 : F(x) \text{ là hàm phân bố của } X$$

và đối thiết

$$H_1 : F(x) \text{ không là hàm phân bố của } X.$$

Định nghĩa (hàm phân bố mẫu)

$\tilde{F}(x)$ được gọi là hàm phân bố mẫu nếu

$$\tilde{F}(x) = \text{Tổng của tần suất tương đối của tất cả các giá trị mẫu } x_j \text{ với } x_j \leq x.$$

Kiểm định mức độ phù hợp của mô hình : Thông qua một mẫu ngẫu nhiên

$$(x_1, ..., x_n)$$

ta kiểm định một hàm $F(x)$ có phải là hàm phân bố của biến ngẫu nhiên gốc X trong tổng thể hay không ?
Ta có giả thiết thống kê

$$H_0 : F(x) \text{ là hàm phân bố của } X$$

và đối thiết

$$H_1 : F(x) \text{ không là hàm phân bố của } X.$$

Định nghĩa (hàm phân bố mẫu)

$\tilde{F}(x)$ được gọi là hàm phân bố mẫu nếu

$$\tilde{F}(x) = \text{Tổng của tần suất tương đối của tất cả các giá trị mẫu } x_j \text{ với } x_j \leq x.$$

Nếu $\tilde{F}(x)$ "đủ gần" so với $F(x)$, thì chấp nhận giả thiết rằng $F(x)$ là hàm phân bố của tổng thể, ngược lại ta bác bỏ giả thiết đó.

Lược đồ kiểm định giả thiết

B1 : Chia trục số thực thành K khoảng : $I_1, I_2, ..., I_K$ sao cho mỗi khoảng chứa ít nhất 5 giá trị của mẫu $(x_1, ..., x_n)$.

Lược đồ kiểm định giả thiết

B1 : Chia trục số thực thành K khoảng : $I_1, I_2, ..., I_K$ sao cho mỗi khoảng chứa ít nhất 5 giá trị của mẫu $(x_1, ..., x_n)$.
Xác định *tần số thực nghiệm*

$$b_j = \text{Số các giá trị mẫu } x_1, x_2, ..., x_n \text{ nằm trong khoảng } I_j, \quad 1 \leq j \leq K.$$

Lược đồ kiểm định giả thiết

B1 : Chia trục số thực thành K khoảng : $I_1, I_2, ..., I_K$ sao cho mỗi khoảng chứa ít nhất 5 giá trị của mẫu $(x_1, ..., x_n)$.

Xác định *tần số thực nghiệm*

$$b_j = \text{Số các giá trị mẫu } x_1, x_2, ..., x_n \text{ nằm trong khoảng } I_j, \quad 1 \leq j \leq K.$$

Nếu một giá trị mẫu nào đó nằm trên phần giao của hai khoảng, thì cộng 0.5 cho hai số b_j tương ứng với hai khoảng đó.

Lược đồ kiểm định giả thiết

B1 : Chia trục số thực thành K khoảng : $I_1, I_2, ..., I_K$ sao cho mỗi khoảng chứa ít nhất 5 giá trị của mẫu $(x_1, ..., x_n)$.
Xác định *tần số thực nghiệm*

$$b_j = \text{Số các giá trị mẫu } x_1, x_2, ..., x_n \text{ nằm trong khoảng } I_j, \quad 1 \le j \le K.$$

Nếu một giá trị mẫu nào đó nằm trên phần giao của hai khoảng, thì cộng 0.5 cho hai số b_j tương ứng với hai khoảng đó.

B2 : Giả sử H_0 đúng,

$$p_j = P(X \in I_j), \quad 1 \le j \le K.$$

Lược đồ kiểm định giả thiết

B1 : Chia trục số thực thành K khoảng : $I_1, I_2, ..., I_K$ sao cho mỗi khoảng chứa ít nhất 5 giá trị của mẫu $(x_1, ..., x_n)$.
Xác định *tần số thực nghiệm*

$$b_j = \text{Số các giá trị mẫu } x_1, x_2, ..., x_n \text{ nằm trong khoảng } I_j, \quad 1 \leq j \leq K.$$

Nếu một giá trị mẫu nào đó nằm trên phần giao của hai khoảng, thì cộng 0.5 cho hai số b_j tương ứng với hai khoảng đó.

B2 : Giả sử H_0 đúng,

$$p_j = P(X \in I_j), \quad 1 \leq j \leq K.$$

Tần số lý thuyết

$$e_j = np_j, \quad 1 \leq j \leq K.$$

Lược đồ kiểm định giả thiết

B1 : Chia trục số thực thành K khoảng : $I_1, I_2, ..., I_K$ sao cho mỗi khoảng chứa ít nhất 5 giá trị của mẫu $(x_1, ..., x_n)$.
Xác định *tần số thực nghiệm*

$$b_j = \text{Số các giá trị mẫu } x_1, x_2, ..., x_n \text{ nằm trong khoảng } I_j, \quad 1 \leq j \leq K.$$

Nếu một giá trị mẫu nào đó nằm trên phần giao của hai khoảng, thì cộng 0.5 cho hai số b_j tương ứng với hai khoảng đó.

B2 : Giả sử H_0 đúng,

$$p_j = P(X \in I_j), \quad 1 \leq j \leq K.$$

Tần số lý thuyết

$$e_j = np_j, \quad 1 \leq j \leq K.$$

B3 : Tính *độ lệch*

$$\chi_0^2 = \sum_{j=1}^{K} \frac{(b_j - e_j)^2}{e_j}. \tag{3.65}$$

Lược đồ kiểm định giả thiết

B1 : Chia trục số thực thành K khoảng : $I_1, I_2, ..., I_K$ sao cho mỗi khoảng chứa ít nhất 5 giá trị của mẫu $(x_1, ..., x_n)$.
Xác định *tần số thực nghiệm*

$$b_j = \text{Số các giá trị mẫu } x_1, x_2, ..., x_n \text{ nằm trong khoảng } I_j, \quad 1 \leq j \leq K.$$

Nếu một giá trị mẫu nào đó nằm trên phần giao của hai khoảng, thì cộng 0.5 cho hai số b_j tương ứng với hai khoảng đó.

B2 : Giả sử H_0 đúng,

$$p_j = P(X \in I_j), \quad 1 \leq j \leq K.$$

Tần số lý thuyết

$$e_j = np_j, \quad 1 \leq j \leq K.$$

B3 : Tính *độ lệch*

$$\chi_0^2 = \sum_{j=1}^{K} \frac{(b_j - e_j)^2}{e_j}. \tag{3.65}$$

B4 : Chọn mức ý nghĩa α (5%, 1%, hoặc giá trị khác).

Lược đồ kiểm định giả thiết

B5 : Xác định giá trị tới hạn c từ phương trình

$$\hat{F}(c) = 1 - \alpha$$

và từ bảng A10 với $K - 1$ bậc tự do.

Lược đồ kiểm định giả thiết

B5 : Xác định giá trị tới hạn c từ phương trình

$$\hat{F}(c) = 1 - \alpha$$

và từ bảng A10 với $K - 1$ bậc tự do.

Nếu r tham số của $F(x)$ chưa biết và phương pháp ước lượng hợp lý cực đại được sử dụng, thì $(K - r - 1)$ bậc tự do sẽ thay thế cho $(K - 1)$ bậc tự do.

Lược đồ kiểm định giả thiết

B5 : Xác định giá trị tới hạn c từ phương trình

$$\hat{F}(c) = 1 - \alpha$$

và từ bảng A10 với $K - 1$ bậc tự do.

Nếu r tham số của $F(x)$ chưa biết và phương pháp ước lượng hợp lý cực đại được sử dụng, thì $(K - r - 1)$ bậc tự do sẽ thay thế cho $(K - 1)$ bậc tự do.

Khi đó

- Nếu $\chi_0^2 \leq c$, chấp nhận H_0.
- Nếu $\chi_0^2 > c$, bác bỏ H_0.

Chương 4. Tương quan và hồi quy

Nghiên cứu đồng thời hai đại lượng X và Y. Tức là mẫu ngẫu nhiên được cho bởi các cặp

$$\{(x_1, y_1), (x_2, y_2), \dots (x_{,n}, y_n)\}.$$

Chương 4. Tương quan và hồi quy

Nghiên cứu đồng thời hai đại lượng X và Y. Tức là mẫu ngẫu nhiên được cho bởi các cặp

$$\{(x_1, y_1), (x_2, y_2), \ldots (x_{,n}\ , y_n)\}.$$

Nội dung chính :

1. **Phân tích hồi quy** :
   - X có thể được đo lường với độ chính xác cao (*biến cơ sở, biến độc lập, biến điều khiển*).

Chương 4. Tương quan và hồi quy

Nghiên cứu đồng thời hai đại lượng X và Y. Tức là mẫu ngẫu nhiên được cho bởi các cặp

$$\{(x_1, y_1), (x_2, y_2), \dots (x_{,n}, y_n)\}.$$

Nội dung chính :

1. **Phân tích hồi quy** :
    - X có thể được đo lường với độ chính xác cao (*biến cơ sở, biến độc lập, biến điều khiển*).
    - Y là một BNN và ta quan tâm tới sự phụ thuộc của Y vào X.

Chương 4. Tương quan và hồi quy

Nghiên cứu đồng thời hai đại lượng X và Y. Tức là mẫu ngẫu nhiên được cho bởi các cặp

$$\{(x_1, y_1), (x_2, y_2), \ldots (x_{,n}, y_n)\}.$$

Nội dung chính :

1. **Phân tích hồi quy** :
    - X có thể được đo lường với độ chính xác cao (*biến cơ sở, biến độc lập, biến điều khiển*).
    - Y là một BNN và ta quan tâm tới sự phụ thuộc của Y vào X.
      Ví dụ :
        - Sự phụ thuộc của huyết áp Y của một người vào độ tuổi X của người đó.
        - Mối liên hệ giữa lượng cân tăng Y của một loài động vật nào đó phụ thuộc vào chế độ dinh dưỡng hàng ngày X.

Chương 4. Tương quan và hồi quy

2 **Phân tích tương quan** : Mối quan hệ (mối tương quan) giữa hai đại lượng ngẫu nhiên X và Y.

Chương 4. Tương quan và hồi quy

2 **Phân tích tương quan** : Mối quan hệ (mối tương quan) giữa hai đại lượng ngẫu nhiên X và Y.

Ví dụ :

- Nghiên cứu sự tương quan độ mòn X của lốp trước bên trái của một xe ô tô so với độ mòn Y của lốp trước bên phải.
- Quan hệ giữa điểm số môn toán X và điểm số môn vật lý Y của một học sinh lớp 12.

Sự phụ thuộc của đại lượng Y vào đại lượng $X = x$ giống như sự phụ thuộc của kỳ vọng μ vào x, tức là $\mu = \mu(x)$.

Sự phụ thuộc của đại lượng Y vào đại lượng $X = x$ giống như sự phụ thuộc của kỳ vọng μ vào x, tức là $\mu = \mu(x)$.
Đồ thị của hàm $\mu(x)$ gọi là *đường hồi quy* của Y.

Sự phụ thuộc của đại lượng Y vào đại lượng $X = x$ giống như sự phụ thuộc của kỳ vọng μ vào x, tức là $\mu = \mu(x)$.
Đồ thị của hàm $\mu(x)$ gọi là *đường hồi quy* của Y.
Khi

$$\mu(x) = \kappa_0 + \kappa_1 x, \tag{4.66}$$

thì đường hồi quy của Y là đường thẳng và gọi là *đường hồi quy tuyến tính*.

Sự phụ thuộc của đại lượng Y vào đại lượng $X = x$ giống như sự phụ thuộc của kỳ vọng μ vào x, tức là $\mu = \mu(x)$.
Đồ thị của hàm $\mu(x)$ gọi là *đường hồi quy* của Y.
Khi

$$\mu(x) = \kappa_0 + \kappa_1 x, \tag{4.66}$$

thì đường hồi quy của Y là đường thẳng và gọi là *đường hồi quy tuyến tính*.
Cho mẫu ngẫu nhiên gồm n cặp giá trị :

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

tương ứng với một đám mây điểm trên mặt phẳng Oxy.

Sự phụ thuộc của đại lượng Y vào đại lượng $X = x$ giống như sự phụ thuộc của kỳ vọng μ vào x, tức là $\mu = \mu(x)$.
Đồ thị của hàm $\mu(x)$ gọi là *đường hồi quy* của Y.
Khi

$$\mu(x) = \kappa_0 + \kappa_1 x, \tag{4.66}$$

thì đường hồi quy của Y là đường thẳng và gọi là *đường hồi quy tuyến tính*.
Cho mẫu ngẫu nhiên gồm n cặp giá trị :

$$\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$$

tương ứng với một đám mây điểm trên mặt phẳng Oxy.
Mục tiêu : Xây dựng một đường thẳng "phù hợp" đi xuyên qua đám mây điểm trên để từ đó có thể xấp xỉ giá trị của $\mu(x)$ tại một giá trị cụ thể của x.

Đồ thị của đường hồi quy tuyến tính và đám mây điểm được thể hiện trong Hình

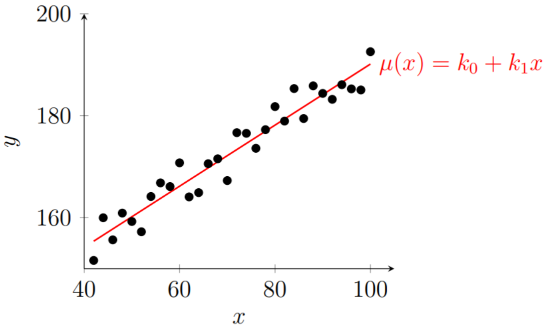

Figure – Đường hồi quy tuyến tính tương ứng với đám mây điểm

Phương pháp tìm phương trình đường hồi quy tuyến tính (Nguyên lý bình phương tối thiểu)

- *Đường hồi quy tuyến tính đi xuyên qua đám mây điểm sao cho tổng của bình phương khoảng cách từ các điểm tới đường hồi quy là nhỏ nhất.*
- Điều kiện để đường hồi quy tuyến tính duy nhất :

Phương pháp tìm phương trình đường hồi quy tuyến tính (Nguyên lý bình phương tối thiểu)

- *Đường hồi quy tuyến tính đi xuyên qua đám mây điểm sao cho tổng của bình phương khoảng cách từ các điểm tới đường hồi quy là nhỏ nhất.*
- Điều kiện để đường hồi quy tuyến tính duy nhất :

Có ít nhất hai số có giá trị khác nhau trong các số $x_1, x_2, \ldots, x_n$**. (A1)**

G/s có đám mây điểm

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

thỏa mãn giả thiết (A1). Gọi đường thẳng

$$y = k_0 + k_1 x \tag{4.67}$$

là **đường hồi quy (tuyến tính) mẫu**.

G/s có đám mây điểm

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

thỏa mãn giả thiết (A1). Gọi đường thẳng

$$y = k_0 + k_1 x \tag{4.67}$$

là **đường hồi quy (tuyến tính) mẫu**. Khoảng cách đứng từ điểm (x_i, y_i) đến đường hồi quy mẫu là

$$\epsilon_i = |y_i - k_0 - k_1 x_i|.$$

G/s có đám mây điểm

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

thỏa mãn giả thiết (A1). Gọi đường thẳng

$$y = k_0 + k_1 x \tag{4.67}$$

là **đường hồi quy (tuyến tính) mẫu**. Khoảng cách đứng từ điểm (x_i, y_i) đến đường hồi quy mẫu là

$$\epsilon_i = |y_i - k_0 - k_1 x_i|.$$

Tổng bình phương các khoảng cách đó là

$$q = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - k_0 - k_1 x_i)^2.$$

G/s có đám mây điểm

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

thỏa mãn giả thiết (A1). Gọi đường thẳng

$$y = k_0 + k_1 x \tag{4.67}$$

là **đường hồi quy (tuyến tính) mẫu**. Khoảng cách đứng từ điểm (x_i, y_i) đến đường hồi quy mẫu là

$$\epsilon_i = |y_i - k_0 - k_1 x_i|.$$

Tổng bình phương các khoảng cách đó là

$$q = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - k_0 - k_1 x_i)^2.$$

Để q bé nhất, áp dụng phương pháp cực trị của hàm hai biến số k_0 và k_1, nghĩa là k_0, k_1 thỏa mãn hệ phương trình

$$\begin{cases} \frac{\partial q}{\partial k_0} = 0 \\ \frac{\partial q}{\partial k_1} = 0 \end{cases}$$

G/s có đám mây điểm

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$$

thỏa mãn giả thiết (A1). Gọi đường thẳng

$$y = k_0 + k_1 x \tag{4.67}$$

là **đường hồi quy (tuyến tính) mẫu**. Khoảng cách đứng từ điểm (x_i, y_i) đến đường hồi quy mẫu là

$$\epsilon_i = |y_i - k_0 - k_1 x_i|.$$

Tổng bình phương các khoảng cách đó là

$$q = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} (y_i - k_0 - k_1 x_i)^2.$$

Để q bé nhất, áp dụng phương pháp cực trị của hàm hai biến số k_0 và k_1, nghĩa là k_0, k_1 thỏa mãn hệ phương trình

$$\begin{cases} \frac{\partial q}{\partial k_0} = 0 \\ \frac{\partial q}{\partial k_1} = 0 \end{cases}$$

hay

$$\begin{cases} n k_0 + k_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ k_0 \sum_{i=1}^{n} x_i + k_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases} \tag{4.68}$$

Phương trình (4.68) được gọi là *hệ phương trình chuẩn* với hai ẩn số k_0, k_1. Định thức của ma trận các hệ số là

$$\begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix} = n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2$$
$$= n(n-1)s_x^2 = n\sum_{i=1}^n (x_i - \bar{x})^2 > 0$$

với

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n), \quad s_x^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2.$$

Phương trình (4.68) được gọi là *hệ phương trình chuẩn* với hai ẩn số k_0, k_1. Định thức của ma trận các hệ số là

$$\begin{vmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{vmatrix} = n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2$$
$$= n(n-1)s_x^2 = n\sum_{i=1}^n (x_i - \bar{x})^2 > 0$$

với

$$\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n), \quad s_x^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})^2.$$

Theo Định lý Cramer, ta có

$$k_1 = \frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n(n-1)s_x^2} = \frac{n(\overline{xy} - \bar{x}\bar{y})}{(n-1)s_x^2} \tag{4.69}$$

với

$$\bar{y} = \frac{1}{n}(y_1 + \cdots + y_n), \quad \overline{xy} = \frac{1}{n}(x_1 y_1 + \cdots + x_n y_n).$$

Đại lượng s_{xy} được gọi là **hiệp phương sai mẫu** :

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) \qquad (4.70)$$

Đại lượng s_{xy} được gọi là **hiệp phương sai mẫu** :

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) \qquad (4.70)$$

hay

$$s_{xy} = \frac{n}{n-1}[\overline{xy} - \bar{x}\bar{y}]. \qquad (4.71)$$

Đại lượng s_{xy} được gọi là **hiệp phương sai mẫu** :

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) \tag{4.70}$$

hay

$$s_{xy} = \frac{n}{n-1}[\overline{xy} - \bar{x}\bar{y}]. \tag{4.71}$$

Từ (4.69), ta suy ra

$$k_1 = \frac{s_{xy}}{s_x^2}. \tag{4.72}$$

Đại lượng s_{xy} được gọi là **hiệp phương sai mẫu** :

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) \tag{4.70}$$

hay

$$s_{xy} = \frac{n}{n-1}[\overline{xy} - \bar{x}\bar{y}]. \tag{4.71}$$

Từ (4.69), ta suy ra

$$k_1 = \frac{s_{xy}}{s_x^2}. \tag{4.72}$$

Từ phương trình thứ nhất của (4.68), hệ số k_0 được tính như sau

$$k_0 = \bar{y} - k_1\bar{x}.$$

Đại lượng s_{xy} được gọi là **hiệp phương sai mẫu** :

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) \tag{4.70}$$

hay

$$s_{xy} = \frac{n}{n-1}[\overline{xy} - \bar{x}\bar{y}]. \tag{4.71}$$

Từ (4.69), ta suy ra

$$k_1 = \frac{s_{xy}}{s_x^2}. \tag{4.72}$$

Từ phương trình thứ nhất của (4.68), hệ số k_0 được tính như sau

$$k_0 = \bar{y} - k_1\bar{x}.$$

Thay vào phương trình (4.67), ta có

$$y - \bar{y} = k_1(x - \bar{x}).$$

Đại lượng s_{xy} được gọi là **hiệp phương sai mẫu** :

$$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right) \tag{4.70}$$

hay

$$s_{xy} = \frac{n}{n-1}[\overline{xy} - \bar{x}\bar{y}]. \tag{4.71}$$

Từ (4.69), ta suy ra

$$k_1 = \frac{s_{xy}}{s_x^2}. \tag{4.72}$$

Từ phương trình thứ nhất của (4.68), hệ số k_0 được tính như sau

$$k_0 = \bar{y} - k_1\bar{x}.$$

Thay vào phương trình (4.67), ta có

$$y - \bar{y} = k_1(x - \bar{x}).$$

Do đó hệ số k_1 được cho bởi công thức (4.72) còn được gọi là **hệ số hồi quy** hay **hệ số xác định**.

Ví dụ

Đo sự giảm thể tích $y[\%]$ của một tấm da dưới áp lực x và nhận được giá trị sau

x_i	4000	6000	8000	10000
y_i	2.3	4.1	5.7	6.9

Tìm đường hồi quy tuyến tính của y theo x.

Ví dụ

Đo sự giảm thể tích $y[\%]$ của một tấm da dưới áp lực x và nhận được giá trị sau

x_i	4000	6000	8000	10000
y_i	2.3	4.1	5.7	6.9

Tìm đường hồi quy tuyến tính của y theo x.

Giải. Ta có $n = 4$, $\bar{x} = 7000$, $\bar{y} = 4.75$ và

$$s_x^2 = \frac{20000000}{3}, \quad s_{xy} = \frac{15400}{3}.$$

Ví dụ

Đo sự giảm thể tích $y[\%]$ của một tấm da dưới áp lực x và nhận được giá trị sau

x_i	4000	6000	8000	10000
y_i	2.3	4.1	5.7	6.9

Tìm đường hồi quy tuyến tính của y theo x.

Giải. Ta có $n = 4$, $\bar{x} = 7000$, $\bar{y} = 4.75$ và

$$s_x^2 = \frac{20000000}{3}, \quad s_{xy} = \frac{15400}{3}.$$

Do đó hệ số xác định $k_1 = \frac{s_{xy}}{s_x^2} = 0.00077$ và đường hồi quy cần tìm là

$$y - 4.74 = 0.00077(x - 7000) \quad \text{hay} \quad y = 0.00077x - 0.64.$$

Cho mẫu với n cặp giá trị

$$(x_1, y_1), \ldots, (x_n, y_n).$$

Mối quan hệ giữa các giá trị x_i và y_i $(1 \leq i \leq n)$ được thể hiện thông qua hiệp phương sai mẫu s_{xy} hay thông qua **hệ số tương quan mẫu**

$$r = \frac{s_{xy}}{s_x s_y} \tag{4.73}$$

Cho mẫu với n cặp giá trị
$$(x_1, y_1), \ldots, (x_n, y_n).$$
Mối quan hệ giữa các giá trị x_i và y_i $(1 \leq i \leq n)$ được thể hiện thông qua hiệp phương sai mẫu s_{xy} hay thông qua **hệ số tương quan mẫu**
$$r = \frac{s_{xy}}{s_x s_y} \tag{4.73}$$
với s_{xy} được cho bởi (4.70) và
$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}[\overline{x^2} - (\bar{x})^2],$$
$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{n}{n-1}[\overline{y^2} - (\bar{y})^2].$$

Định lý (Hệ số tương quan mẫu)

Hệ số tương quan mẫu r có các tính chất :

1. $-1 \leq r \leq 1$;
2. $r = \pm 1 \Leftrightarrow$ tất cả các điểm của mẫu nằm trên một đường thẳng.

Một số giá trị của hệ số tương quan mẫu r được minh họa

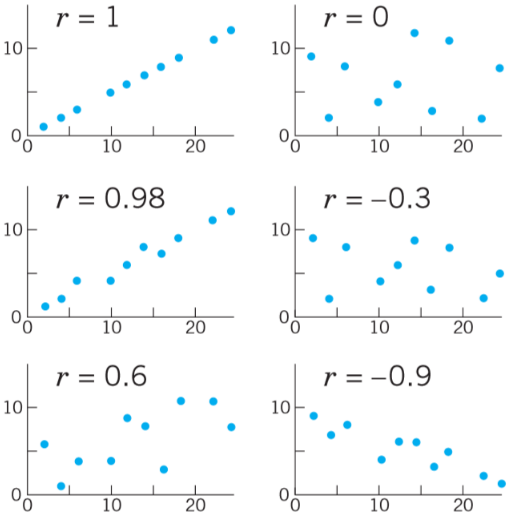

Figure – Mẫu với các giá trị của hệ số tương quan mẫu r