MỤC LỤC

Lời nói đầu **4**

Phần I. Xác suất **6**

1 Biến cố và xác suất **7**

1.1 Bổ túc về giải tích tổ hợp 7

1.2 Phép thử ngẫu nhiên và biến cố 13

1.3 Xác suất của biến cố 18

1.4 Xác suất có điều kiện 27

1.5 Dãy phép thử Bernoulli 37

Bài tập **40**

2 Đại lượng ngẫu nhiên và phân phối xác suất **49**

2.1 Đại lượng ngẫu nhiên 49

2.2 Các số đặc trưng của đại lượng ngẫu nhiên 56

2.3 Một số phân phối xác suất quan trọng 65

2.4 Vectơ ngẫu nhiên 79

2.5 Một số định lý giới hạn 90

Bài tập **112**

Phần II. Thống kê **123**

3 Lý thuyết mẫu **124**
3.1 Khái niệm mẫu và phương pháp lấy mẫu 124
3.1.1 Khái niệm mẫu 125
3.1.2 Các phương pháp lấy mẫu 126
3.2 Cách biểu diễn mẫu 128
3.2.1 Bảng tần số và bảng tần suất 128
3.2.2 Đa giác tần số và tổ chức đồ 129
3.3 Các đặc trưng của mẫu 131
3.3.1 Hàm phân phối mẫu 132
3.3.2 Trung bình mẫu 133
3.3.3 Phương sai mẫu và phương sai hiệu chỉnh mẫu . . 134
Bài tập 137

4 Ước lượng tham số **140**
4.1 Ước lượng điểm 140
4.1.1 Định nghĩa 140
4.1.2 Ước lượng điểm cho kỳ vọng, xác suất và phương sai 142
4.2 Ước lượng khoảng 145
4.2.1 Khái niệm về khoảng tin cậy 145
4.2.2 Khoảng tin cậy cho giá trị trung bình 146
4.2.3 Khoảng tin cậy cho tỉ lệ 151
4.2.4 Độ chính xác của ước lượng 153
Bài tập 154

5 Kiểm định giả thiết **159**
5.1 Đặt vấn đề 159
5.2 Kiểm định giả thiết về giá trị trung bình và về tỉ lệ . . . 160
5.2.1 Kiểm định giả thiết về giá trị trung bình 160

5.2.2 Kiểm định giả thiết về tỉ lệ 164
5.3 So sánh các giá trị trung bình và các giá trị tỉ lệ 167
5.3.1 So sánh hai giá trị trung bình 167
5.3.2 So sánh hai tỉ lệ 171
5.4 Kiểm tra tính độc lập 174
Bài tập 177

6 Hồi quy và tương quan **182**
6.1 Hệ số tương quan mẫu 182
6.1.1 Mở đầu 182
6.1.2 Hệ số tương quan mẫu 183
6.2 Phương trình hồi quy bình phương trung bình tuyến tính 185
6.2.1 Phương trình hồi quy 185
6.2.2 Ước lượng hệ số hồi quy tuyến tính thực nghiệm . 186
Bài tập 188

Hướng dẫn giải bài tập **191**

Tài liệu tham khảo **210**

Các bảng số thông dụng **211**

LỜI NÓI ĐẦU

Giáo trình này được viết dựa trên đề cương chi tiết học phần xác suất thống kê dành cho sinh viên các ngành khối A nói chung và các ngành Công nghệ nói riêng.

Giáo trình gồm 6 chương chia thành 2 phần. Phần 1 gồm 2 chương, trình bày các kiến thức cơ sở của lý thuyết xác suất. Trong chương 1, chúng tôi trình bày về một số khái niệm và tính chất mở đầu của lý thuyết xác suất: phép thử, biến cố, xác suất của biến cố, xác suất có điều kiện, tính độc lập của các biến cố, dãy phép thử Bernoulli, ... Các khái niệm và tính chất này sẽ được dùng nhiều ở các chương sau. Chương 2 trình bày các vấn đề liên quan đến đại lượng ngẫu nhiên và vectơ ngẫu nhiên: Khái niệm đại lượng ngẫu nhiên, các loại đại lượng ngẫu nhiên, hàm phân phối, bảng phân phối và hàm mật độ xác suất, các số đặc trưng của đại lượng ngẫu nhiên, một số định lý giới hạn, ...

Phần 2 gồm 4 chương, trình bày về những vấn đề cơ bản của thống kê ứng dụng. Cụ thể là Chương 3 trình bày về lý thuyết mẫu; Chương 4 trình bày về lý thuyết ước lượng; Chương 5 trình bày về lý thuyết kiểm định và Chương 6 trình bày về lý thuyết về tương quan và hồi quy. Trong mỗi chương của phần này, chúng tôi đều chú ý đưa ra những ứng dụng của các vấn đề nêu ra vào việc giải quyết các vấn đề của thực tế.

Cuối mỗi chương đều có khá nhiều bài tập, trong đó có một số bài tương đối khó và còn ít xuất hiện ở các giáo trình khác. Đối với những bài tập này, chúng tôi có hướng dẫn giải ở cuối giáo trình.

Để đọc được giáo trình này, sinh viên đã học xong các học phần Toán A1 (Đại số Tuyến tính) và Toán A2 (Giải tích 1).

Giáo trình này được viết nhờ sự hỗ trợ tài chính của dự án Việt Nam - Hà Lan, tác giả xin cám ơn Ban quản lý dự án về sự hỗ trợ đó. Tác giả cũng xin cảm ơn các đồng nghiệp ở bộ môn Xác suất - Thống kê và Toán ứng dụng, Khoa Toán - Trường Đại học Vinh đã quan tâm động viên tác giả hoàn thành giáo trình này.

Mặc dù tác giả đã rất cố gắng nhưng chắc chắn giáo trình vẫn còn nhiều khiếm khuyết, mong được sự góp ý của bạn đọc.

Tác giả

PHẦN I. XÁC SUẤT

CHƯƠNG 1

BIẾN CỐ VÀ XÁC SUẤT

1.1 Bổ túc về giải tích tổ hợp

Trong lí thuyết xác suất, nhiều khi phải tính số phần tử của một tập hợp. Giải tích tổ hợp cho ta một phương pháp tính số phần tử đó một cách nhanh chóng và chính xác. Trong mục này chúng tôi sẽ trình bày một số vấn đề cơ bản của giải tích tổ hợp được sử dụng nhiều trong các mục sau.

Trước hết, chúng ta nghiên cứu quy tắc nhân và quy tắc cộng. Có nhiều cách trình bài các quy tắc này. Dưới đây là cách trình bày mà theo chúng tôi là tương đối đơn giản và dễ vận dụng.

1.1.1 Quy tắc nhân

Quy tắc. *Giả sử một hành động có thể được thực hiện qua k bước liên tiếp. Bước* 1 *có thể được thực hiện bằng* n_1 *cách, bước* 2 *có thể được thực hiện bằng* n_2 *cách, ..., bước k có thể được thực hiện bằng* n_k *cách. Khi đó, số cách để thực hiện hành động đó là*

$$n = n_1.n_2...n_k.$$

Như vậy, khi một hành động có thể chia nhỏ thành nhiều bước, thì số cách thực hiện hành động đó bằng tích số cách thực hiện các bước.

Ví dụ.

1. Một đội văn nghệ đã chuẩn bị được 2 vở kịch, 5 bài hát, 3 điệu múa. Hội diễn chỉ cho phép đội trình diễn một vở kịch, một bài hát, một điệu múa. Hỏi đội có bao nhiêu cách chọn chương trình biểu diễn của mình, biết rằng chất lượng các vở kịch, bài hát, điệu múa là như nhau.

Giải. Việc chọn chương trình biểu diễn có thể được thực hiện qua 3 bước:

Bước 1: Chọn 1 vở kịch trong 2 vở kịch. Có 2 cách.

Bước 2: Chọn 1 bài hát trong 5 bài hát. Có 5 cách.

Bước 3: Chọn 1 điệu múa trong 3 điệu múa. Có 3 cách.

Từ đó, áp dụng quy tắc nhân suy ra đội văn nghệ có $n = 2.3.5 = 30$ cách chọn chương trình biểu diễn.

2. Với các chữ số $1, 2, 3, 4, 5$, có thể lập được bao nhiêu số chẵn gồm 4 chữ số khác nhau?

Giải. Mỗi số cần tìm có dạng $x = \overline{a_1a_2a_3a_4}$ với

$$\{a_1, a_2, a_3, a_4\} \subset \{1, 2, 3, 4, 5\}.$$

Muốn xác định x, ta phải chọn a_1, a_2, a_3, a_4 (4 bước). Vì x là số chẵn, nên ta chọn a_4 trước, rồi chọn a_1, a_2, a_3.

Bước 1: Chọn a_4. Vì x là số chẵn nên $a_4 = 2$ hoặc $a_4 = 4$. Có 2 cách chọn a_4.

Bước 2: Chọn a_1. Vì $a_1 \neq a_4$ nên có 4 cách chọn a_1.

Bước 3: Chọn a_2. Vì $a_2 \neq a_4$, $a_2 \neq a_1$ nên còn 3 cách chọn a_2.

Bước 4: Chọn a_3. Vì $a_3 \neq a_4$, $a_3 \neq a_1$ và $a_3 \neq a_2$ nên còn 2 cách chọn a_3.

Từ đó, áp dụng quy tắc nhân suy ra số các số cần tìm là $n = 2.4.3.2 = 48$.

1.1.2 Quy tắc cộng

Quy tắc. *Giả sử các phần tử của một tập hợp có thể được chia thành k*

loại khác nhau. Loại 1 có n_1 phần tử, loại 2 có n_2 phần tử, ..., loại k có n_k phần tử. Khi đó số phần tử của tập hợp đó là

$$n = n_1 + n_2 + \cdots + n_k.$$

Dựa vào quy tắc cộng, ta có thể chuyển bài toán về tính số phần tử của một tập hợp phức tạp về các bài toán tính số phần tử của các tập hợp đơn giản hơn.

Ví dụ. Có bao nhiêu số có các chữ số khác nhau được lập từ 4 chữ số $1, 2, 3, 4$.

Giải. Tập hợp các số cần lập có thể chia làm 4 loại:

Loại 1: Các số có 1 chữ số. Có $n_1 = 4$ số: $1, 2, 3, 4$.

Loại 2: Các số có 2 chữ số khác nhau. Có $n_2 = 4.3 = 12$ số.

Loại 3: Các số có 3 chữ số khác nhau. Có $n_3 = 4.3.2 = 24$ số.

Loại 4: Các số có 4 chữ số khác nhau. Có $n_4 = 4.3.2.1 = 24$ số.

Từ đó, áp dụng quy tắc cộng, suy ra số các số cần tìm là

$$n = 4 + 12 + 24 + 24 = 64.$$

1.1.3 Hoán vị

Định nghĩa. Mỗi cách sắp xếp n phần tử cho trước theo một thứ tự nhất định gọi là một *hoán vị* của n phần tử đó.

Kí hiệu số các hoán vị của n phần tử đã cho là P_n.

Định lý. *Số hoán vị của n phần tử là*

$$P_n = n!.$$

Chứng minh. Mỗi hoán vị của n phần tử là kết quả của phép chọn gồm n bước.

Bước 1: Chọn phần tử đầu tiên cho hoán vị. Có n cách chọn.

Bước 2: Chọn phần tử thứ hai cho hoán vị. Có $(n-1)$ cách chọn.

...

Bước k: Chọn phần tử thứ k cho hoán vị. Có $(n-k+1)$ cách chọn.

...

Bước n: Chọn phần tử thứ n cho hoán vị. Có $(n-n+1)=1$ cách chọn.

Do đó, theo qui tắc nhân, số hoán vị của n phần tử là

$$P_n = n(n-1)...(n-k+1)...1 = n!.$$

Ví dụ. Giả sử một lớp có 25 sinh viên. Khi đó, số cách sắp xếp chỗ ngồi cho 25 sinh viên đó là $P_{25} = 25!$.

1.1.4 Chỉnh hợp

Định nghĩa. Mỗi bộ sắp thứ tự gồm k phần tử khác nhau, lấy từ n phần tử đã cho gọi là một *chỉnh hợp chập* k của n phần tử đó $(0 \leq k \leq n)$.

Ký hiệu số các chỉnh hợp chập k của n phần tử là A_n^k. Bằng lập luận tương tự như đối với hoán vị, ta thu được định lý sau đây:

Định lý. *Số chỉnh hợp chập k của n phần tử là*

$$A_n^k = \frac{n!}{(n-k)!} = n(n-1)(n-k+1).$$

Ví dụ. Có bao nhiêu số khác nhau gồm 3 chữ số được lập từ 4 chữ số 1, 2, 3, 4?

Giải. Mỗi số như vậy là một chỉnh hợp chập 3 của 4 phần tử. Do đó số các số như vậy là

$$A_4^3 = \frac{4!}{(4-3)!} = 4! = 24.$$

1.1.5 Chỉnh hợp lặp

Định nghĩa. Một bộ sắp thứ tự gồm k phần tử (không nhất thiết khác nhau) lấy từ n phần tử đã cho gọi là một *chỉnh hợp lặp chập* k của n phần tử đó $(k \geq 0)$.

Như vậy, khác với chỉnh hợp, trong mỗi chỉnh hợp lặp, không đòi hỏi các phần tử phải khác nhau. Chẳng hạn tất cả các chỉnh hợp lặp chập 2 của 3 phần tử của tập $M = \{1, 2, 3\}$ là

$$(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3).$$

Số các chỉnh hợp lặp chập k của n phần tử được kí hiệu là $\tilde{A}_n^k$.

Định lý. *Số các chỉnh hợp lặp chập k của n phần tử đã cho là*

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

Chứng minh. Mỗi chỉnh hợp lặp chập k của n phần tử là kết quả của hành động chọn gồm k bước. Mỗi bước đều có n cách thực hiện (vì không đòi hỏi các phần tử phải khác nhau). Từ đó, theo qui tắc nhân, ta có

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

Trong ví dụ trên, ta có $n = 3, k = 2$. Vậy số các chỉnh hợp lặp chập 2 của 3 phần tử là

$$\tilde{A}_3^2 = 3^2 = 9.$$

Ví dụ. Có bao nhiêu cách sắp xếp ngẫu nhiên 15 hành khách lên ba toa tàu?

Giải. Mỗi cách sắp xếp như vậy là một chỉnh hợp lặp chập 15 của 3 phần tử. Do đó số cách sắp xếp là

$$\tilde{A}_{15}^3 = 3^{15}.$$

1.1.6 Tổ hợp

Định nghĩa. Một tập con (không kể thứ tự) gồm k phần tử lấy từ n phần tử đã cho gọi là một *tổ hợp chập k* của n phần tử đó ($0 \le k \le n$).

Ký hiệu C_n^k là số các tổ hợp chập k của n phần tử đã cho.

Định lý. *Số tổ hợp chập k của n phần tử là*

$$C_n^k = \frac{A_n^k}{k!} = \frac{n!}{k!(n-k)!}(\textit{quy ước } 0! = 1).$$

Chứng minh. Mỗi tổ hợp chập k của n phần tử sinh ra $k!$ chỉnh hợp chập k khác nhau của n phần tử đó (các chỉnh hợp này chỉ khác nhau ở thứ tự sắp xếp của các phần tử). Do đó ta có

$$A_n^k = k! C_n^k.$$

Suy ra

$$C_n^k = \frac{A_n^k}{k!}.$$

Định lý được chứng minh.

Từ công thức nêu trong định lý trên, dễ dàng suy ra các công thức sau đây

$$C_n^k = C_n^{n-k},$$

$$C_n^k = C_{n-1}^{k-1} + C_{n-1}^k.$$

Ví dụ. Có bao nhiêu cách sắp xếp ngẫu nhiên 15 hành khách lên 3 toa tàu mà toa thứ nhất có đúng 3 hành khách?

Giải. Ta thấy có C_{15}^3 cách lấy 3 trong 15 hành khách cho vào toa I. Số cách phân ngẫu nhiên 12 người còn lại lên hai toa kia là 2^{12}. Vậy số cách phân ngẫu nhiên 15 hành khách lên 3 toa tàu mà toa I có đúng 3 hành khách là $C_{15}^3 \cdot 2^{12}$.

1.1.7 Công thức nhị thức Newton

Trên tập số thực, ta đã rất quen thuộc với các hằng đẳng thức

$$(a+b)^2 = a^2 + 2ab + b^2.$$

$$(a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3.$$

Bằng qui nạp có thể chứng minh được công thức sau đây, gọi là *công thức nhị thức Newton*

$$(a+b)^n = \Sigma_{k=0}^{n} C_n^k a^k b^{n-k} \quad (\forall a, b \in \mathbb{R}).$$

Đặc biệt, với $a = b = 1$ ta có $2^n = \Sigma_{k=0}^{n} C_n^k$.

Với $a = 1, b = -1$ ta có $0 = \Sigma_{k=0}^{n} (-1)^{n-k} C_n^k$.

Nếu thay b bởi $-b$ thì ta có công thức

$$(a-b)^n = \Sigma_{k=0}^{n} (-1)^{n-k} C_n^k a^k b^{n-k} \quad (\forall a, b \in \mathbb{R}).$$

1.2 Phép thử ngẫu nhiên và biến cố

1.2.1 Tất nhiên và ngẫu nhiên

Như ta đã biết, các hiện tượng trong tự nhiên và xã hội có thể được chia làm hai loại: tất nhiên và ngẫu nhiên.

Hiện tượng tất nhiên là hiện tượng chắc chắn xảy ra khi có một họ điều kiện nào đó được thực hiện. Chẳng hạn, với điều kiện áp suất bình thường của khí quyển và nhiệt độ 100^oC nước chắc chắn sôi; với điều kiện cho axít clohiđric (HCl) tác dụng với natri hiđrôxit (NaOH) chắc chắn xuất hiện muối ăn và nước, ...

Hiện tượng ngẫu nhiên là hiện tượng có thể xảy ra hoặc không xảy ra khi có một họ điều kiện nào đó được thực hiện. Chẳng hạn, khi ta gieo một đồng tiền cân đối đồng chất, ta không thể biết được mặt sấp hay mặt ngửa sẽ xuất hiện. Như vậy, các hiện tượng "Mặt sấp xuất hiện" và "Mặt ngửa xuất hiện" là các hiện tượng ngẫu nhiên. Kết quả của một lần kiểm tra chất lượng sản phẩm, kết quả của một lần bắn bia, ... cũng là những hiện tượng ngẫu nhiên.

Tính bất định của sự xuất hiện của các hiện tượng ngẫu nhiên làm nảy sinh nhu cầu nghiên cứu khả năng xuất hiện của chúng. Đây chính là một trong những nguyên nhân ra đời và phát triển của lý thuyết xác suất.

1.2.2 Phép thử ngẫu nhiên và không gian biến cố sơ cấp

Để nghiên cứu các hiện tượng ngẫu nhiên, người ta thường phải tiến hành các phép thử ngẫu nhiên. *Phép thử ngẫu nhiên* là một hành động mà kết quả của nó là ngẫu nhiên, không thể dự báo trước được. Khi thực hiện phép thử ngẫu nhiên thì các kết quả của nó không thể xác định trước được. Tuy nhiên, ta có thể xác định được tập hợp tất cả các kết quả có thể có của nó. Tập hợp đó được gọi là *không gian biến cố sơ cấp* và được ký hiệu bởi chữ Ω. Mỗi phần tử ω của Ω sẽ được gọi là *một biến cố sơ cấp* (*BCSC*).

Ví dụ. Khi tung một đồng tiền cân đối đồng chất ta không biết trước kết quả là xuất hiện mặt sấp (S) hay mặt ngửa (N). Tuy nhiên, có thể xác định được các kết quả có thể có là S và N. Vậy hành động tung đồng tiền là một phép thử ngẫu nhiên và không gian biến cố sơ cấp của phép thử này là $\Omega = \{S, N\}$.

Tương tự, hành động tung một con xúc xắc cân đối đồng chất, hành động kiểm tra ngẫu nhiên chất lượng sản phẩm của một nhà máy, ... cũng là những phép thử ngẫu nhiên.

1.2.3 Biến cố

Giả sử G là một phép thử ngẫu nhiên. Một sự kiện, mà việc xảy ra hay không xảy ra của nó phụ thuộc hoàn toàn vào kết quả của G, được gọi là một *biến cố* của G.

Một BCSC ω của G được gọi là thuận lợi cho biến cố A nếu khi kết quả của G là ω thì A xảy ra.

Ví dụ. Xét phép thử: Tung một con xúc xắc cân đối, đồng chất. Gọi M_i là biến cố "xuất hiện mặt i chấm", C là biến cố "xuất hiện mặt có số chấm chẵn", L là biến cố "xuất hiện mặt có số chấm lẻ". Vậy thì

Không gian các BCSC là $\Omega = \{M_1, M_2, M_3, M_4, M_5, M_6\}$.

Tập hợp các BCSC thuận lợi cho C là $\{M_2, M_4, M_6\}$.

Tập hợp các BCSC thuận lợi cho L là $\{M_1, M_3, M_5\}$.

Nhận xét rằng một biến cố được xác định hoàn toàn bởi tập hợp các BCSC thuận lợi cho nó. Vì lí do đó, trong lí thuyết xác suất, người ta đồng nhất một biến cố với tập con của Ω gồm các BCSC thuận lợi cho biến cố đó. Chẳng hạn, trong ví dụ trên

$$C = \{M_2, M_4, M_6\} \qquad L = \{M_1, M_3, M_5\}.$$

Như vậy, có thể hiểu nôm na là, các biến cố được tạo nên từ các BCSC.

Các loại biến cố

Một biến cố được gọi là *biến cố không thể có*, nếu nó không thể xảy ra khi phép thử được thực hiện. Như vậy không có BCSC nào của Ω thuận lợi cho biến cố không thể có. Do đó, biến cố không thể có được đồng nhất với tập $\emptyset$.

Một biến cố được gọi là *biến cố chắc chắn*, nếu nó chắc chắn xảy ra khi phép thử được thực hiện. Mọi BCSC của phép thử đều thuận lợi cho biến cố chắc chắn. Do đó, biến cố chắc chắn được đồng nhất với toàn bộ tập Ω.

Một biến cố được gọi là *biến cố ngẫu nhiên*, nếu nó có thể xảy ra hoặc không xảy ra khi phép thử được thực hiện. Biến cố ngẫu nhiên thường được ký hiệu bởi các chữ in $A, B, C, \ldots$ Chẳng hạn, trong ví dụ trên thì:

Biến cố "Số chấm xuất hiện > 6" là biến cố không thể có ($\emptyset$).

Biến cố "Số chấm xuất hiện $\leqslant 6$" là biến cố chắc chắn (Ω).

Các biến cố $C = \{M_2, M_4, M_6\}\}$ và $L = \{M_1, M_3, M_5\}$ là các biến cố ngẫu nhiên.

1.2.4 Quan hệ và phép toán giữa các biến cố

Quan hệ thuận lợi. Ta nói biến cố A *thuận lợi* cho biến cố B nếu khi A xảy ra thì B xảy ra. Rõ ràng, A thuận lợi cho B khi và chỉ khi tập hợp các BCSC thuận lợi cho A là tập con của tập hợp các BCSC thuận lợi cho B. Do đó, nếu A thuận lợi cho B thì ta kí hiệu

$$A \subset B.$$

Quan hệ bằng nhau. Hai biến cố A và B gọi là *bằng nhau* (*hay tương đương*) nếu A xảy ra khi và chỉ khi B xảy ra. Nói cách khác, A và B bằng nhau khi và chỉ khi $A \subset B$ và $B \subset A$. Nếu A và B bằng nhau thì ta kí hiệu

$$A = B.$$

Hợp của các biến cố. Biến cố A được gọi là *hợp* của 2 biến cố B và C nếu A xảy ra khi và chỉ khi ít nhất một trong 2 biến cố B hoặc C xảy ra. Lúc đó ta kí hiệu

$$A = B \cup C.$$

Tổng quát. Biến cố A được gọi là *hợp* của họ biến cố $A_i (i \in I)$ nếu A xảy ra khi và chỉ khi ít nhất một trong các biến cố $A_i (i \in I)$ xảy ra. Kí hiệu

$$A = \bigcup_{i \in I} A_i.$$

Giao các biến cố. Biến cố A được gọi là *giao* (*tích*) của 2 biến cố B

và C nếu A xảy ra khi và chỉ khi B và C đồng thời xảy ra. Kí hiệu

$$A = B \cap C \;\; (\;\; A = B.C).$$

Tổng quát. Biến cố A được gọi là *giao* (*tích*) của họ các biến cố A_i, $(i \in I)$ nếu A xảy ra khi và chỉ khi tất cả các A_i, $(i \in I)$ đều xảy ra. Kí hiệu

$$A = \bigcap_{i \in I} A_i \;\; (A = \prod_{i \in I} A_i).$$

Hiệu các biến cố. Biến cố A được gọi là *hiệu* của biến cố B với biến cố C nếu A xảy ra khi và chỉ khi B xảy ra và C không xảy ra. Kí hiệu

$$A = B \backslash C.$$

Các biến cố xung khắc. Hai biến cố A và B được gọi là *xung khắc* nếu chúng không thể đồng thời xảy ra. Nói cách khác, A và B được gọi là xung khắc nếu

$$A.B = \emptyset.$$

Biến cố đối lập. Biến cố $\overline{A}$ được gọi là *biến cố đối lập* của biến cố A nếu $\overline{A}$ xảy ra khi và chỉ khi A không xảy ra. Rõ ràng lúc đó ta có

$$A = \overline{A}, \quad A = \Omega \backslash \overline{A}, \quad A = \overline{\overline{A}}.$$

Họ đầy đủ các biến cố. Họ n biến cố $H_1, H_2, ..., H_n$ được gọi là *họ đầy đủ các biến cố* nếu chúng thoả mãn đồng thời hai điều kiện sau

1. Chúng xung khắc với nhau đôi một. Tức là

$$H_i H_j = \emptyset, \quad (i \neq j)$$

2. Hợp của chúng là biến cố chắc chắn. Tức là

$$H_1 \cup H_2 ..... \cup H_n = \Omega.$$

Nói cách khác, $H_1, H_2, ..., H_n$ là họ đầy đủ các biến cố nếu khi phép thử được thực hiện thì có một và chỉ một trong các biến cố đó xảy ra. Chẳng hạn nếu A là biến cố bất kì thì A và $\overline{A}$ lập thành một họ đầy đủ.

Ví dụ. Hai người cùng bắn, mỗi người bắn một viên vào bia. A_i là biến cố "Người thứ i bắn trúng" $(i = 1, 2)$. Vậy thì

A_1A_2 là biến cố "cả hai người cùng bắn trúng".

$A_1\overline{A}_2 \cup \overline{A}_1A_2$ là biến cố "có đúng một người bắn trúng".

$\overline{A}_1\overline{A}_2$ là biến cố "không ai bắn trúng".

Họ đầy đủ các biến cố là: $A_1, \overline{A}_1$ hoặc $\overline{A}_1\ \overline{A}_2$, $\overline{A_1}A_2$, $A_1\overline{A_2}$, A_1A_2.

Chú ý. Quan hệ và phép toán trên tập hợp các biến cố có tất cả các tính chất của quan hệ và phép toán trên các tập hợp. Chẳng hạn

1. $(A \cup B) \cup C = A \cup (B \cup C), \quad (AB)C = A(BC).$
2. $A(B \cup C) = AB \cup AC, \quad (A \cup B)C = AC \cup BC.$
3. $A\Omega = A, \quad A \cup \Omega = \Omega.$
4. $A \cup \emptyset = A, \quad A\emptyset = \emptyset.$
5. (Công thức De-Morgan) $\overline{\bigcup_{i\in I} A_i} = \prod_{i\in I} \overline{A_i}.$

1.3 Xác suất của biến cố

Nói chung, một biến cố có thể xảy ra hoặc không xảy ra khi phép thử được thực hiện. Do đó, nảy sinh nhu cầu đo lường khả năng xuất hiện của nó. Số biểu thị khả năng xuất hiện của biến cố A gọi là xác suất của biến cố đó và được kí hiệu là $\mathbb{P}(A)$.

Có nhiều cách định nghĩa xác suất. Dưới đây sẽ trình bày một số định nghĩa quan trọng.

1.3.1 Định nghĩa xác suất cổ điển

Định nghĩa. Giả sử không gian BCSC của phép thử có n BCSC có cùng khả năng xuất hiện; trong đó có m BCSC thuận lợi cho biến cố A. Khi đó, số

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

được gọi là *xác suất* của biến cố A.

Các biến cố có cùng khả năng xuất hiện được gọi là các biến cố đồng khả năng.

Ví dụ.

1. Xét phép thử tung một con xúc xắc cân đối đồng chất. Rõ ràng

$$\mathbb{P}(C) = \mathbb{P}(L) = \frac{1}{2}.$$

2. Một nhóm học tập có 10 học sinh trong đó có 7 học sinh yếu. Kiểm tra ngẫu nhiên 3 em. Tính xác suất để

a. Cả 3 em được kiểm tra đều là học sinh yếu.

b. Trong 3 em được kiểm tra có một học sinh yếu.

c. Có ít nhất một học sinh yếu được kiểm tra.

Giải. Gọi A, B, C là các biến cố nêu trong các câu a, b, c tương ứng. Số cách chọn 3 em trong 10 em là $n = C_{10}^3$.

a. Số cách chọn cả 3 em yếu là C_7^3. Do đó

$$\mathbb{P}(A) = \frac{C_7^3}{C_{10}^3}.$$

b. Số cách chọn 3 em, trong đó có 1 em yếu là $C_7^1 C_3^2$. Do đó

$$\mathbb{P}(B) = \frac{C_7^1 C_3^2}{C_{10}^3}.$$

c. Chỉ có một cách chọn 3 em đều không học yếu. Do đó

$$\mathbb{P}(C) = \frac{C_{10}^3 - 1}{C_{10}^3}.$$

Chú ý. Trong định nghĩa xác suất cổ điển, đòi hỏi số BCSC của phép thử phải hữu hạn và các BCSC phải có cùng khả năng xuất hiện. Đòi hỏi này thường được thoả mãn trong các trò chơi may rủi, hoặc trong các phép thử mà việc chọn lựa là vô tư, ngẫu nhiên và các dụng cụ thử là cân đối, đồng chất. Nếu số BCSC là vô hạn, hoặc hữu hạn nhưng không đồng khả năng xuất hiện, thì cách tính xác suất cổ điển không còn đúng nữa.

1.3.2 Định nghĩa xác suất bằng tần suất

Tần suất của một biến cố. Giả sử phép thử G có thể lặp đi lặp lại nhiều lần độc lập nhau; A là một biến cố của G. Thực hiện G n lần. Giả sử trong n lần đó, A xuất hiện $k_n(A)$ lần. Khi đó tỉ số

$$f_n(A) = \frac{k_n(A)}{n}$$

được gọi là *tần suất* xuất hiện của biến cố A trong n phép thử. Có thể dễ dàng nhận thấy rằng tần suất có những tính chất sau đây

1. $0 \leqslant f_n(A) \leqslant 1$.
2. $f_n(\Omega) = 1$.
3. Nếu $AB = \emptyset$ thì $f_n(A \cup B) = f_n(A) + f_n(B)$.

Định nghĩa xác suất theo tần suất. Người ta nhận thấy rằng khi số phép thử tăng lên vô hạn, tần suất của $f_n(A)$ luôn dần tới một giới hạn xác định. Giới hạn đó được định nghĩa là xác suất của biến cố A và được ký hiệu là $\mathbb{P}(A)$.

Trên thực tế $\mathbb{P}(A)$ được tính xấp xỉ bởi $f_n(A)$ với n đủ lớn. Chẳng hạn, khi tiến hành thí nghiệm tung đồng tiền nhiều lần, ta thấy tần suất xuất hiện mặt sấp (S) xấp xỉ 0,5. Do đó, ta định nghĩa $\mathbb{P}(S) = 0,5$.

Định nghĩa xác suất bằng tần suất chỉ áp dụng được cho các phép thử ngẫu nhiên có thể lặp đi lặp lại nhiều lần độc lập nhau trong những điều kiện giống hệt nhau. Ngoài ra, ta chỉ có thể xác định được tương đối chính xác giá trị của xác suất khi tiến hành một số đủ nhiều các phép thử; mà điều này đôi khi không thể làm được vì hạn chế về thời gian và kinh phí.

1.3.3 Định nghĩa xác suất hình học

Trong trường hợp không gian BCSC Ω của phép thử G có vô số BCSC có cùng khả năng xuất hiện và Ω có thể biểu diễn bởi một tập đo được thì ta có định nghĩa sau đây.

Định nghĩa. Giả sử không gian BCSC Ω của phép thử có vô số BCSC có cùng khả năng xuất hiện và Ω được biểu diễn bởi một tập đo được H_Ω. Khi đó nếu biến cố A được biểu diễn bởi một tập đo được H_A thì số

$$\mathbb{P}(A) = \frac{\text{độ đo } H_A}{\text{độ đo } H_\Omega}$$

được gọi là *xác suất* của biến cố A.

Trong đó "tập đo được" là tập trên đường thẳng (mặt phẳng, không gian) có độ dài (diện tích, thể tích) còn "độ đo" của một tập là độ dài (diện tích, thể tích) của tập đó.

Ví dụ. Hai người bạn hẹn gặp nhau trong khoảng thời gian từ 7 đến 8 giờ. Biết rằng khả năng họ đến điểm hẹn vào mọi thời điểm trong thời gian hẹn là như nhau. Ngoài ra, người đến trước đợi người đến sau đúng 20 phút, nếu không gặp sẽ bỏ đi. Tìm xác suất để 2 người gặp nhau.

Giải. Gọi thời điểm đến điểm hẹn của người thứ nhất là x, của người thứ 2 là y ($7 \leqslant x \leqslant 8, 7 \leqslant y \leqslant 8$). Vậy thì có thể biểu diễn BCSC "người thứ nhất đến vào thời điểm x, người thứ 2 đến vào thời điểm y" bởi cặp số (x, y). Khi đó, không gian BCSC Ω được biểu diễn bởi hình vuông

$$H_\Omega = \{(x, y) : 7 \leqslant x \leqslant 8, 7 \leqslant y \leqslant 8\}.$$

Biến cố $A =$ "2 người gặp nhau" được biểu diễn bởi tập

$$H_A = \{(x, y) \in H_\Omega : |x - y| \leqslant \frac{1}{3}\},$$

(20 phút $= \dfrac{1}{3}$ giờ). Vậy

$$\mathbb{P}(A) = \frac{\text{d.tích } H_A}{\text{d.tích } H_\Omega} = 1 - (\frac{2}{3})^2 = \frac{5}{9}.$$

1.3.4 Định nghĩa xác suất bằng phương pháp tiên đề

Các định nghĩa xác suất trình bày ở trên có ưu điểm là khá trực quan và rất tiện lợi trong việc giải một số lớp bài toán ứng dụng cụ thể. Tuy nhiên, chúng đều có hạn chế là không tổng quát và cũng không thật chặt chẽ. Để khắc phục hạn chế này, năm 1933, nhà toán học người Nga Kolmogorov đã xây dựng xác suất bằng một hệ tiên đề. Trước hết, chúng ta đề cập đến một số khái niệm liên quan.

σ-đại số

Giả sử Ω là một tập hợp bất kì khác rỗng. Một họ $\mathcal{F}$ gồm các tập con nào đó của Ω được gọi là một *σ-đại số*, nếu

i. $\Omega \in \mathcal{F}$,

ii. Nếu $A \in \mathcal{F}$ thì $\Omega \backslash A \in \mathcal{F}$,

iii. Nếu $A_n \in \mathcal{F}$ $(n = 1, 2, \dots)$ thì $\cup_{n=1}^{\infty} A_n \in \mathcal{F}$.

σ-đại số còn được gọi là *σ-trường*.

Không gian đo và độ đo xác suất

Giả sử Ω là một tập hợp tuỳ ý khác rỗng, $\mathcal{F}$ là một σ-đại số các tập con của Ω. Khi đó, cặp $(\Omega, \mathcal{F})$ được gọi là một *không gian đo*.

Giả sử $(\Omega, \mathcal{F})$ gọi là một không gian đo. Một ánh xạ $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ được gọi là *độ đo xác suất* trên $\mathcal{F}$ nếu

i. $\mathbb{P}(A) \geqslant 0$ với mọi $A \in \mathcal{F}$ (tính không âm),

ii. $\mathbb{P}(\Omega) = 1$ (tính chuẩn hoá), *iii*. Nếu $A_n \in \mathcal{F}$ $(n = 1, 2, 3, \dots)$, $A_i \cap A_j = A_iA_j = \emptyset$ $(i \neq j)$ thì

$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$ (tính cộng tính đếm được).

Bây giờ, giả sử Ω là tập bất kỳ khác rỗng, $\mathcal{F}$ là một σ- đại số các tập con của Ω, $\mathbb{P}$ là độ đo xác suất trên $\mathcal{F}$. Khi đó, bộ ba $(\Omega, \mathcal{F}, \mathbb{P})$ được gọi là *không gian xác suất*.

Tập Ω được gọi là *không gian biến cố sơ cấp*.

σ- đại số $\mathcal{F}$ được gọi là *σ- đại số các biến cố*.

Mỗi $A \in \mathcal{F}$ được gọi là một *biến cố*.

Nếu $A \cap B = AB = \emptyset$ thì A, B được gọi là các *biến cố xung khắc*.

Biến cố $\Omega \in \mathcal{F}$ gọi là *biến cố chắc chắn*.

Biến cố $\emptyset \in \mathcal{F}$ gọi là *biến cố không thể có*.

Biến cố $\overline{A} = \Omega \backslash A$ được gọi là *biến cố đối lập* của biến cố A.

Không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$ gọi là *không gian xác suất đầy đủ* nếu mọi tập con của biến cố có xác suất không đều là biến cố. Để đơn giản, từ nay về sau, khi nói đến không gian xác suất $(\Omega, \mathcal{F}, \mathbb{P})$, ta luôn xem đó là không gian xác suất đầy đủ.

Chú ý. Điều kiện (ii) trong định nghĩa trên đảm bảo rằng biến cố chắc chắn có xác suất bằng 1. Tuy nhiên, sau này ta sẽ gặp những biến cố có xác suất bằng 1 nhưng chưa chắc đã là biến cố chắc chắn. Những biến cố như vậy gọi là *biến cố hầu chắc chắn*.

Ví dụ.

1. Giả sử G là một phép thử mà không gian BCSC Ω của nó gồm n BCSC có cùng khả năng xuất hiện. Gọi $\mathcal{F}$ là họ tất cả các tập con của Ω. Với mỗi $A \in \mathcal{F}$ ta đặt

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

trong đó $|A|$ là số phần tử của A.

Rõ ràng $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ thoả mãn các tiên đề xác suất. Đây chính là định nghĩa xác suất cổ điển. Như vậy, trong mô hình cổ điển, mọi tập $A \subset \Omega$ đều là biến cố.

2. Giả sử G là một phép thử mà không gian các BCSC của nó có vô số BCSC có cùng khả năng xuất hiện và được biểu diễn bởi một tập đo được H_Ω. Gọi $\mathcal{F}$ là họ tất cả các tập con đo được của H_Ω. Mỗi $H \in \mathcal{F}$ đặt

$$\mathbb{P}(H) = \frac{\text{độ đo } (H)}{\text{độ đo } (H_\Omega)}.$$

Dễ thấy $(\Omega, \mathcal{F}, \mathbb{P})$ lập thành một không gian xác suất.

Bây giờ, giả sử A là một biến cố của phép thử. Nếu A được biểu diễn bởi tập đo được H_A thì ta đặt

$$\mathbb{P}(A) = \mathbb{P}(H_A).$$

Đây chính là định nghĩa xác suất hình học.

3. Giả sử Ω là một tập vô hạn đếm được $\Omega = (\omega_1, \omega_2, \ldots, \omega_n, \ldots)$ còn (p_n) là một dãy số không âm thoả mãn điều kiện $\sum_{n=1}^{\infty} p_n = 1$. Gọi $\mathcal{F}$ là σ- đại số tất cả các tập con của Ω. Với mỗi $A \in \mathcal{F}$ đặt

$$\mathbb{P}(A) = \sum_{\omega_i \in A} p_i.$$

Dễ thấy ánh xạ $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ xác định như trên là xác suất và $(\Omega, \mathcal{F}, \mathbb{P})$ trở thành một không gian xác suất. Mô hình này được gọi là mô hình xác suất rời rạc.

Như vậy, có nhiều cách định nghĩa xác suất thoả mãn tiên đề Kolmogorov. Vấn đề là ở chỗ cần định nghĩa xác suất sao cho phù hợp với thực tiễn khách quan.

1.3.5 Các tính chất của xác suất

Giả sử $A, B, C, \ldots$ là những biến cố. Khi đó, xác suất của chúng có các tính chất sau

1. $\mathbb{P}(\emptyset) = 0$.

Thật vậy, đặt $A_1 = \Omega;\ A_n = \emptyset (\forall n > 1)$. Khi đó, sử dụng (iii) ta được

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n) = \mathbb{P}(\Omega) + \sum_{n=2}^{\infty} \mathbb{P}(A_n).$$

Suy ra

$$\sum_{n=2}^{\infty} \mathbb{P}(A_n) = 0.$$

Điều này, cùng với (i) cho ta

$$\mathbb{P}(\emptyset) = \mathbb{P}(A_n) = 0; (\forall n > 1).$$

2. *Nếu* $AB = \emptyset$ *thì* $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Tính chất này là hệ quả trực tiếp của tiên đề về tính cộng tính đếm được của độ đo xác suất và tính chất 1.

3. $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$.

Thật vậy, ta có

$$A \cup \overline{A} = \Omega, \quad A\overline{A} = \emptyset.$$

Suy ra

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup \overline{A}) = \mathbb{P}(A) + \mathbb{P}(\overline{A}).$$

Nên

$$\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A).$$

4. *Nếu* $A \subset B$ *thì* $\mathbb{P}(B \backslash A) = \mathbb{P}(B) - \mathbb{P}(A)$ *và do đó* $\mathbb{P}(A) \leqslant \mathbb{P}(B)$.

Thật vậy

$$B = A \cup (B \backslash A), \quad A(B \backslash A) = \emptyset.$$

Nên

$$\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \backslash A) \geqslant \mathbb{P}(A).$$

Suy ra

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

5. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$.

Thật vậy, ta có

$$A \cup B = A \cup (B\backslash AB); \quad A(B\backslash AB) = \emptyset; \quad AB \subset B.$$

Suy ra

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B\backslash AB) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB).$$

6. $\mathbb{P}(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} \mathbb{P}(A_k) - \sum_{1 \leqslant k \leqslant i \leqslant n} \mathbb{P}(A_k A_i) + \sum_{1 \leqslant k \leqslant l \leqslant m \leqslant n} \mathbb{P}(A_k A_l A_m) - \cdots + (-1)^{n-1} \mathbb{P}(A_1 A_2 \ldots A_n)$.

Thật vậy, với $n = 2$, công thức trên chính là tính chất (5). Bằng phương pháp quy nạp toán học, sẽ chứng minh được kết quả cho trường hợp $n > 2$.

7. $\mathbb{P}(\cup_{n=1}^{\infty} A_n) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(A_n)$.

Thật vậy, đặt

$$B_1 = A_1, B_2 = A_2\backslash B_1, \ldots, B_n = A_n\backslash \cup_{k=1}^{n-1} B_k.$$

Lúc đó

$$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \mathbb{P}(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) \leqslant \sum_{n=1}^{\infty} \mathbb{P}(A_n).$$

8. (*Tính liên tục của xác suất*)

i. Nếu (A_n) *là dãy đơn điệu tăng, nghĩa là* $A_1 \subset A_2 \subset \cdots \subset A_n \subset \ldots$, *thì tồn tại*

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cup_{n=1}^{\infty} A_n).$$

ii. Nếu (A_n) *là dãy đơn điệu giảm, nghĩa là* $A_1 \supset A_2 \supset \cdots \supset A_n \supset \ldots$, *thì tồn tại*

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cap_{n=1}^{\infty} A_n).$$

Thật vậy, giả sử (A_n) là dãy tăng. Đặt

$$B_1 = A_1, B_2 = A_2 \backslash B_1, \ldots, B_n = A_n \backslash A_{n-1} = A_n \backslash \cup_{k=1}^{n-1} B_k.$$

Lúc đó

$$\cup_{k=1}^{n} B_k = \cup_{k=1}^{n} A_k = A_n, \cup_{n=1}^{\infty} B_n = \cup_{n=1}^{\infty} A_n, B_i \cap B_j = \emptyset.$$

Suy ra
$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \mathbb{P}(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mathbb{P}(B_n) = \lim_{n\to\infty} \mathbb{P}(\cup_{k=1}^{n} B_k) = \lim_{n\to\infty} \mathbb{P}(A_n)$.
Bây giờ giả sử (A_n) là dãy đơn điệu giảm. Vậy thì $(\overline{A_n})$ là dãy đơn điệu tăng, nên theo (i) ta có

$$\lim_{n\to\infty} \mathbb{P}(\overline{A_n}) = \mathbb{P}(\cup_{n=1}^{\infty} \overline{A_n}) = \mathbb{P}(\overline{\cap_{n=1}^{\infty} A_n}) = 1 - \mathbb{P}(\cap_{n=1}^{\infty} A_n).$$

Suy ra
$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty}(1 - \mathbb{P}(\overline{A_n})) = 1 - \lim_{n\to\infty} \mathbb{P}(\overline{A_n}) = 1 - (1 - \mathbb{P}(\cap_{n=1}^{\infty} A_n)) = \mathbb{P}(\cap_{n=1}^{\infty} A_n)$.
Đó là điều cần chứng minh.

1.4 Xác suất có điều kiện

1.4.1 Định nghĩa và ví dụ

Trong thực tế, nhiều khi ta phải tính xác suất của một biến cố khi biết một biến cố khác đã xảy ra. Xác suất đó gọi là xác suất có điều kiện và được định nghĩa như sau.

Định nghĩa 1. Giả sử A và B là các biến cố. Xác suất của B được tính với giả thiết A đã xảy ra được gọi là *xác suất của B với điều kiện A* và được kí hiệu là $\mathbb{P}(B/A)$.

Xác suất của B với điều kiện A còn được gọi là *xác suất có điều kiện của A đối với B.*

Ví dụ. Một lớp học có n học sinh, trong đó có m học sinh nữ ($m \leqslant n$). Trong m học sinh nữ đó có k em học yếu ($k \leqslant m$). Kiểm tra ngẫu nhiên một học sinh. Gọi A là biến cố "học sinh đó là nữ", B là biến cố "học sinh đó học yếu". Vậy thì rõ ràng $\mathbb{P}(B/A)$ là tỉ lệ học sinh nữ học yếu. Do đó

$$\mathbb{P}(B/A) = \frac{k}{m}.$$

Mặt khác, dễ dàng nhận thấy rằng

$$\mathbb{P}(A) = \frac{m}{n}, \quad \mathbb{P}(B) = \frac{k}{n}.$$

Suy ra

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

Qua kiểm nghiệm thực tế, người ta nhận thấy công thức trên luôn luôn đúng. Điều này gợi ý cho ta đi đến định nghĩa sau:

Định nghĩa 2. Giả sử $(\Omega, \mathcal{F}, \mathbb{P})$ là không gian xác suất. $A, B \in \mathcal{F}$, $\mathbb{P}(A) > 0$. Khi đó số

$$\mathbb{P}(B/A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} \tag{1}$$

được gọi là *xác suất của B với điều kiện A.*

Chú ý.

1. Nếu $\mathbb{P}(A) = 0$ thì vẫn có $\mathbb{P}(B/A)$ nhưng không thể áp dụng công thức (1).

2. Tuỳ theo tình huống cụ thể mà có thể tính $\mathbb{P}(B/A)$ theo một trong hai định nghĩa trên.

1.4.2 Tính chất.

1. $\mathbb{P}(B/A) \geqslant 0$. (Hiển nhiên).

2. *Nếu* $B \supset A$ *thì* $\mathbb{P}(B/A) = 1$. *Đặc biệt* $\mathbb{P}(\Omega/A) = 1$.

Chứng minh.

$$B \supset A \Rightarrow AB = A \Rightarrow \mathbb{P}(B/A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A)}{\mathbb{P}(A)} = 1.$$

3. *Nếu* (B_n) *là dãy các biến cố đôi một xung khắc thì*

$$\mathbb{P}(\cup_{n=1}^{\infty} B_n/A) = \sum_{n=1}^{\infty} \mathbb{P}(B_n/A).$$

Chứng minh. Trước hết nhận xét rằng nếu (B_n) là dãy biến cố đôi một xung khắc thì (AB_n) cũng là dãy biến cố đôi một xung khắc. Do đó

$$\mathbb{P}(\cup_{n=1}^{\infty} B_n/A) = \frac{\mathbb{P}((\cup_{n=1}^{\infty} B_n)A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(\cup_{n=1}^{\infty}(AB_n))}{\mathbb{P}(A)}$$

$$= \frac{\sum_{n=1}^{\infty} \mathbb{P}(AB_n)}{\mathbb{P}(A)} = \sum_{n=1}^{\infty} \frac{\mathbb{P}(AB_n)}{\mathbb{P}(A)} = \sum_{n=1}^{\infty} \mathbb{P}(B_n/A).$$

Từ (1) - (3) suy ra rằng nếu A là một biến cố, $\mathbb{P}(A) > 0$ thì ánh xạ $\mathbb{P}_A : \mathcal{F} \to \mathbb{R}$ xác định bởi công thức

$$\mathbb{P}_A(B) = \mathbb{P}(B/A) \quad (\forall B \in \mathcal{F})$$

cũng là xác suất trên $\mathcal{F}$.

1.4.3 Quy tắc nhân

Định lý. *Giả sử* $A_1, A_2, ..., A_n$ $(n \geqslant 2)$, *là* n *biến cố bất kì sao cho* $\mathbb{P}(A_1A_2...A_{n-1}) > 0$. *Khi đó*

$$\mathbb{P}(A_1A_2...A_n) = \mathbb{P}(A_1).\mathbb{P}(A_2/A_1)...\mathbb{P}(A_n/A_1...A_{n-1}). \qquad (2)$$

Chứng minh. Ta chứng minh bằng phương pháp quy nạp theo n. Ta có, với mọi $A_1, A_2 \in \mathcal{F}$
$$\mathbb{P}(A_2/A_1) = \frac{\mathbb{P}(A_1A_2)}{\mathbb{P}(A_1)}.$$
Do đó
$$\mathbb{P}(A_1A_2) = \mathbb{P}(A_1).\mathbb{P}(A_2).$$
Vậy (2) đúng với $n = 2$.
Giả sử (2) đúng đến $n = k$, ta sẽ chứng minh nó đúng với $n = k + 1$. Thật vậy, giả sử $A_i \in \mathcal{F}$ với $i = 1, 2, \dots, k, k + 1$. Lúc đó
$$\mathbb{P}(A_1A_2 \dots A_kA_{k+1}) = \mathbb{P}[(A_1A_2 \dots A_k)A_{k+1}]$$
$$= \mathbb{P}(A_1A_2 \dots A_k).\mathbb{P}(A_{k+1}/A_1A_2 \dots A_k).$$

Mặt khác, do giả thiết qui nạp
$$\mathbb{P}(A_1A_2 \dots A_k) = \mathbb{P}(A_1).\mathbb{P}(A_2/A_1) \dots \mathbb{P}(A_k/A_1A_2 \dots A_{k-1}).$$
Suy ra
$$\mathbb{P}(A_1A_2 \dots A_kA_{k+1}) = \mathbb{P}(A_1).\mathbb{P}(A_2/A_1) \dots \mathbb{P}(A_{k+1}/A_1A_2 \dots A_k).$$

Ví dụ. Một thủ kho có một chùm chìa khoá gồm 9 chiếc bề ngoài giống hệt nhau nhưng chỉ có 2 chìa mở được kho. Anh ta thử ngẫu nhiên từng chìa (chìa nào không mở được thì bỏ ra). Tìm xác suất để anh ta mở được cửa ở lần mở thứ 3?
Giải. Gọi A_i là biến cố "thử đúng chìa ở lần thứ i". Ta phải tính $\mathbb{P}(\overline{A}_1\overline{A}_2A_3)$. áp dụng qui tắc nhân, ta có
$$\mathbb{P}(\overline{A}_1\overline{A}_2A_3) = \mathbb{P}(\overline{A_1})\mathbb{P}(\overline{A_2}/\overline{A_1})\mathbb{P}(A_3/\overline{A}_1\overline{A}_2).$$
Dễ thấy
$$\mathbb{P}(\overline{A}_1) = \frac{7}{9}, \quad \mathbb{P}(\overline{A}_2/\overline{A}_1) = \frac{6}{8}, \quad \mathbb{P}(A_3/\overline{A}_1\overline{A}_2) = \frac{2}{7}.$$

Do đó
$$\mathbb{P}(\overline{A}_2\overline{A}_1A_3) = \frac{7}{9}\frac{6}{8}\frac{2}{7} = \frac{1}{6}.$$

1.4.4 Công thức xác suất đầy đủ và công thức Bayes

Định lý. *Giả sử* $H_1, H_2, \ldots, H_n$ *là họ đầy đủ các biến cố và* $\mathbb{P}(H_i) > 0$, $(\forall i = 1, 2, \ldots, n)$. *Khi đó, với biến cố* A *bất kì, ta có*

i) $\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A/H_i)\mathbb{P}(H_i)$.

ii) Nếu $\mathbb{P}(A) > 0$ *thì*
$$\mathbb{P}(H_k/A) = \frac{\mathbb{P}(A/H_k)\mathbb{P}(H_k)}{\sum_{i=1}^{n} \mathbb{P}(A/H_i)\mathbb{P}(H_i)} \quad (k = 1, 2, \ldots, n).$$

Chứng minh.

$i)$ Ta có
$$A = A\Omega = A(\cup_{i=1}^{n} H_i) = \cup_{i=1}^{n} AH_i,$$
$$(AH_i)(AH_j) = A(H_iH_j) = A\emptyset = \emptyset \quad (i \neq j).$$

Do đó
$$\mathbb{P}(A) = \mathbb{P}(A(\cup_{i=1}^{n} H_i)) = \sum_{i=1}^{n} \mathbb{P}(AH_i) = \sum_{i=1}^{n} \mathbb{P}(A/H_i)\mathbb{P}(H_i).$$

$ii)$ $\mathbb{P}(H_k/A) = \dfrac{\mathbb{P}(AH_k)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(A/H_k)\mathbb{P}(H_k)}{\sum_{i=1}^{n} \mathbb{P}(A/H_i)\mathbb{P}(H_i)}.$

Công thức (i) gọi là công thức xác suất đầy đủ, còn công thức (ii) gọi là công thức Bayes.

Ví dụ. Trong một trường PTTH số học sinh các khối 10, 11, 12 tương ứng chiếm 40%, 35% và 25% tổng số học sinh của trường. Tỷ lệ học sinh yếu của khối lớp 10 là 25%, khối lớp 11 là 20% và của khối lớp 12 là 10%. Kiểm tra ngẫu nhiên một học sinh của trường.

a. Tính xác suất để em đó là học sinh yếu?

b. Giả sử biết rằng em đó là học sinh yếu. Tính xác suất để em đó là học sinh khối 12?

Giải. Kí hiệu A, H_1, H_2, H_3 là các biến cố sau:

A "học sinh được kiểm tra là học sinh yếu".

H_1 "học sinh được kiểm tra là học sinh yếu khối 10".

H_2 "học sinh được kiểm tra là học sinh yếu khối 11".

H_3 "học sinh được kiểm tra là học sinh yếu khối 12".

Khi đó H_1, H_2, H_3 lập thành họ đầy đủ và $\mathbb{P}(H_1) = 0,4$; $\mathbb{P}(H_2) = 0,35$; $\mathbb{P}(H_3) = 0,25$; $\mathbb{P}(A/H_1) = 0,25$; $\mathbb{P}(A/H_2) = 0,2$; $\mathbb{P}(A/H_3) = 0,1$. Ta cần tính $\mathbb{P}(A)$ và $\mathbb{P}(H_3/A)$.

áp dụng công thức xác suất đầy đủ, ta có

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(A/H_1).\mathbb{P}(H_1) + \mathbb{P}(A/H_2).\mathbb{P}(H_2) + \mathbb{P}(A/H_3).\mathbb{P}(H_3)\\ &= 0,25.0,4 + 0,2.0,35 + 0,1.0,25 = 0,195 = 19,5\%.\end{aligned}$$

(Đây chính là tỉ lệ học sinh yếu của trường!).

áp dụng công thức Bayes, ta có

$$\mathbb{P}(H_3/A) = \frac{\mathbb{P}(A/H_3)\mathbb{P}(H_3)}{\mathbb{P}(H_3)} = \frac{0,1.0,25}{0,195} \approx 0,128 = 12,8\%.$$

(Điều này có nghĩa là số học sinh yếu của lớp 12 chiếm $12,8\%$ số học sinh yếu của trường!).

1.4.5 Tính độc lập của các biến cố

Định nghĩa 1. Hai biến cố A và B được gọi là độc lập nếu

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

Tính chất

1. *A, B độc lập khi và chỉ khi* $\mathbb{P}(A/B) = \mathbb{P}(A)$ *hoặc* $\mathbb{P}(B/A) = \mathbb{P}(B)$.

Chứng minh. Giả sử A, B độc lập. Khi đó

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

Mặt khác theo công thức nhân

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

Suy ra

$$\mathbb{P}(A/B) = \mathbb{P}(A); \qquad \mathbb{P}(B/A) = \mathbb{P}(B).$$

Ngược lại, giả sử chẳng hạn $\mathbb{P}(A/B) = \mathbb{P}(A)$, khi đó

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

Nhận xét. Qua chứng minh trên, ta thấy rằng hai đẳng thức $\mathbb{P}(A/B) = \mathbb{P}(A)$ và $\mathbb{P}(B/A) = \mathbb{P}(B)$ đều tương đương với định nghĩa độc lập và do đó tương đương với nhau.

2. *Hai biến cố A và B độc lập khi và chỉ khi một trong các điều kiện sau thoả mãn*

i) $\overline{A}$, B *độc lập;*

ii) A, $\overline{B}$ *độc lập;*

iii) $\overline{A}$, $\overline{B}$ *độc lập.*

Chứng minh. Ta sẽ chứng minh trường hợp i) các trường hợp còn lại chứng minh tương tự. Giả sử A,B độc lập. Vậy thì

$$\mathbb{P}(\overline{A}B) = \mathbb{P}(B \backslash AB) = \mathbb{P}(B) - \mathbb{P}(AB)$$

$$= \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = P(B)(1 - \mathbb{P}(A)) = \mathbb{P}(\overline{A})\mathbb{P}(B).$$

Do đó $\overline{A}$, B độc lập. Ngược lại nếu $\overline{A}$, B độc lập thì

$$\mathbb{P}(AB) = \mathbb{P}(B \backslash \overline{A}B) = \mathbb{P}(B) - \mathbb{P}(\overline{A}B)$$

$$= \mathbb{P}(B) - \mathbb{P}(\overline{A})\mathbb{P}(B) = \mathbb{P}(B)(1 - \mathbb{P}(\overline{A})) = \mathbb{P}(A)\mathbb{P}(B).$$

Do đó A, B độc lập.

Ví dụ.

1. Tung đồng thời hai đồng tiền cân đối đồng chất. Gọi A là biến cố "Đồng tiền thứ nhất sấp", B là biến cố "Đồng tiền thứ hai ngửa". Ta sẽ chứng minh rằng A, B độc lập.

Thật vậy, ta có

$$\Omega = \{(S,S)(S,N)(N,S)(N,N)\},$$

$$A = \{(S,S)(S,N)\} \quad B = \{(S,N)(N,N)\} \quad AB = \{(S,N)\}.$$

Do đó

$$\mathbb{P}(AB) = \frac{1}{4} = \frac{1}{2}\frac{1}{2} = \mathbb{P}(A).\mathbb{P}(B).$$

Suy ra A, B độc lập.

Rõ ràng kết quả này phù hợp với nhận thức trực quan của chúng ta.

2. Hai người độc lập nhau cùng bắn vào một máy bay. Xác suất bắn trúng của người thứ nhất là 0,3; của người thứ hai là 0,2. Biết rằng máy bay sẽ rơi khi có ít nhất một người bắn trúng. Tìm xác suất để máy bay rơi?

Giải. Gọi A là biến cố "người thứ nhất bắn trúng",

B là biến cố "người thứ hai bắn trúng",

C là biến cố "máy bay rơi".

Vậy thì A, B độc lập và $C = A \cup B$. Do đó

$$\mathbb{P}(C) = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB)$$

$$= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) = 0,3 + 0,2 - 0,2.0,3 = 0,44.$$

Dưới đây sẽ trình bày khái niệm độc lập của một họ biến cố.

Định nghĩa 2. Họ các biến cố $(A_i)_{i\in I}$ được gọi là *độc lập đôi một* nếu hai biến cố bất kỳ của họ đều độc lập.

Họ các biến cố $(A_i)_{i\in I}$ được gọi là *độc lập toàn cục* (*gọi vắn tắt là* *độc lập*), nếu đối với mọi họ hữu hạn các biến cố $A_{i_1}, A_{i_2}, \ldots, A_{i_n}$ của họ đó, ta đều có

$$\mathbb{P}(A_{i_1}A_{i_2}\ldots A_{i_n}) = P(A_{i_1})\mathbb{P}(A_{i_2})...\mathbb{P}(A_{i_n}).$$

Một họ độc lập thì độc lập đôi một. Tuy nhiên điều ngược lại nói chung không đúng. Ví dụ sau đây của Bernstein chứng tỏ điều đó.

Ví dụ. Cho một khối tứ diện đều đồng chất có ba mặt sơn ba màu trắng, xanh, vàng, còn mặt thứ tư sơn cả ba màu đó. Tung khối tứ diện đó lên mặt phẳng và chú ý đến mặt tiếp xúc với mặt phẳng. Gọi T là biến cố "xuất hiện màu trắng"; V là biến cố "xuất hiện màu vàng"; X là biến cố "xuất hiện màu xanh". Khi đó

$$\mathbb{P}(T) = \mathbb{P}(X) = \mathbb{P}(V) = \frac{2}{4} = \frac{1}{2}, \mathbb{P}(TV)$$

$$= \mathbb{P}(VX) = \mathbb{P}(XT) = \frac{1}{4} = \frac{1}{2}.\frac{1}{2}.$$

Do đó họ các biến cố T, V, X độc lập đôi một. Tuy nhiên

$$\mathbb{P}(TVX) = \frac{1}{4} \neq \mathbb{P}(T)\mathbb{P}(V)\mathbb{P}(X) = 1/8.$$

Cho nên họ trên không độc lập.

Đối với dãy độc lập các biến cố, ta có tính chất quan trọng sau đây, gọi là Luật $0-1$ Borel-Cantelli.

Định lý. (Luật $0-1$ Borel-Cantelli) *Giả sử* $(A_n, n \geq 1)$ *là dãy các biến cố. Khi đó*

i) Nếu $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$ *thì* $\mathbb{P}(\limsup A_n) = 0$.

ii) Nếu $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$ *và* (A_n) *độc lập thì* $\mathbb{P}(\limsup A_n) = 1$, *trong đó*

$$\limsup A_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k.$$

Chứng minh. $i)$ Giả sử $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$. Khi đó

$$\mathbb{P}(\limsup A_n) = \mathbb{P}(\cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k) = \lim_{n\to\infty} \mathbb{P}(\cup_{k=n}^{\infty} A_k)$$

$$\leqslant \lim_{n\to\infty} \sum_{k=n}^{\infty} \mathbb{P}(A_k) = 0.$$

$ii)$ Trước hết, nhận xét rằng nếu $0 \leqslant x < 1$ thì

$$1 - x \leqslant e^{-x} \quad .$$

Giả sử $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$. Khi đó, vì dãy (A_n) độc lập nên dãy $(\overline{A}_n)$ cũng độc lập. Do đó, với mọi $n = 1, 2, \ldots$ và mọi $m > n$, ta có

$$1 - \mathbb{P}(\cup_{k=n}^{m} A_k) = \mathbb{P}(\overline{\cup_{=n}^{m} A_k}) = \mathbb{P}(\cap_{k=n}^{m} \overline{A_k})$$

$$= \prod_{k=n}^{m} (1 - \mathbb{P}(A_k)) \leqslant e^{-\sum_{k=n}^{m} \mathbb{P}(A_k)}.$$

Suy ra

$$0 \leqslant 1 - \mathbb{P}(\cup_{k=n}^{\infty} A_k) = \lim_{m\to\infty} (1 - \mathbb{P}(\cup_{k=n}^{m} A_k))$$

$$\leqslant \lim_{m\to\infty} e^{-\sum_{k=n}^{m} \mathbb{P}(A_k)} = 0.$$

Do đó $\mathbb{P}(\cup_{k=n}^{\infty} A_k) = 1$ với mọi $n = 1, 2, \ldots$. Điều này kéo theo

$$\mathbb{P}(\limsup A_n) = \lim_{n\to\infty} \mathbb{P}(\cup_{k=n}^{\infty} A_k) = 1.$$

Đó là điều phải chứng minh.

Như vậy, nếu (A_n) là dãy biến cố độc lập thì $\mathbb{P}(\limsup A_n)$ chỉ có thể nhận một trong hai giá trị 0 hoặc 1 tùy theo chuỗi $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ hội tụ hay phân kỳ.

1.5 Dãy phép thử Bernoulli

1.5.1 Định nghĩa và ví dụ

Định nghĩa. Dãy n phép thử $G_1, G_2, ..., G_n$ được gọi là dãy *n phép thử Bernoulli đối với biến cố A*, nếu

i) Các kết quả của chúng độc lập với nhau.

ii) Xác suất của biến cố A là $\mathbb{P}(A) = p$ như nhau đối với mỗi phép thử trong n phép thử này.

Ví dụ.

1. Tung đồng tiền 10 lần và quan sát sự xuất hiện của mặt sấp. Đó là 10 phép thử Bernoulli đối với biến cố A "xuất hiện mặt sấp".

2. Một người có 5 viên đạn, bắn từng viên một vào mục tiêu, đó là 5 phép thử Bernoulli đối với biến cố A "viên đạn trúng mục tiêu".

(Tuy nhiên nếu có 5 người, mỗi người bắn một viên vào mục tiêu thì nói chung đây không phải là 5 phép thử Bernoulli !).

1.5.2 Định lý Bernoulli. Khi thực hiện n phép thử Bernoulli. Biến cố A có thể xuất hiện 0 lần, 1 lần, ..., n lần. Vậy xác suất để A xuất hiện k lần $(0 \leqslant k \leqslant n)$ là bao nhiêu? Bernoulli đã trả lời được câu hỏi đó trong định lý sau đây.

Định lý. *Giả sử xác suất xuất hiện biến cố A trong mỗi phép thử là $\mathbb{P}(A) = p$. Khi đó xác suất để A xuất hiện k lần trong n phép thử Bernoulli là*

$$p_n(k; p) = C_n^k p^k (1-p)^{n-k} \qquad (0 \leqslant k \leqslant n).$$

Chứng minh. Ký hiệu A_k là biến cố "A xuất hiện đúng k lần khi thực hiện n phép thử Bernoulli". Vậy thì A_k là hợp của các biến cố dạng $(\omega_1, \omega_2, \ldots, \omega_n)$, trong đó ω_i có thể là A hoặc $\overline{A}$ và A có mặt ở k vị trí,

$\overline{A}$ có mặt ở $(n-k)$ vị trí. Mỗi biến cố như vậy được xác định hoàn toàn bởi k vị trí của A trong n vị trí. Do đó có C_n^k biến cố thuận lợi cho A_k. Do tính độc lập của các phép thử nên xác suất của mỗi biến cố là

$$\mathbb{P}(\omega_1)\mathbb{P}(\omega_2)\ldots\mathbb{P}(\omega_n) = (\mathbb{P}(A)\mathbb{P}(A)\ldots\mathbb{P}(A))(\mathbb{P}(\overline{A})\mathbb{P}(\overline{A})\ldots\mathbb{P}(\overline{A}))$$

$$= p^k(1-p)^{n-k}.$$

Vậy

$$\mathbb{P}(A_k) = C_n^k p^k (1-p)^{n-k}.$$

Chú ý. Trong nhiều trường hợp, ngoài việc tính xác suất để A xuất hiện k lần, ta còn phải tính xác suất để số lần xuất hiện A nằm giữa hai số k_1 và k_2 $(0 \leqslant k_1 \leqslant k_2)$. Tức là tính xác suất của biến cố B: "Số lần xuất hiện A lớn hơn hoặc bằng k_1 và nhỏ hơn hoặc bằng k_2". Xác suất này thường được kí hiệu là $p_n(k_1, k_2; p)$.

Rõ ràng

$$B = \cup_{i=k_1}^{k_2} A_i, \quad A_iA_j = \emptyset \quad (i \neq j).$$

Suy ra

$$p_n(k_1, k_2; p) = \mathbb{P}(B) = \sum_{k=k_1}^{k_2} A_k = \sum_{k=k_1}^{k_2} C_n^k p^k (1-p)^{n-k}.$$

Ví dụ. Tung con xúc xắc 10 lần. Tìm xác suất để

1. Mặt 6 chấm xuất hiện 2 lần.

2. Mặt 6 chấm xuất hiện không quá 2 lần.

Giải.

1. $p = \dfrac{1}{6}, \quad n = 10, \quad k = 2$ nên ta có $p_{10}(2; \dfrac{1}{6}) = C_{10}^2 (\dfrac{1}{6})^2 (\dfrac{5}{8})^8.$

2. $p_{10}(0, 2; \dfrac{1}{6}) = \sum_{k=0}^2 C_{10}^k (\dfrac{1}{6})^k (\dfrac{5}{6})^{10-k}$

$= C_{10}^0 (\dfrac{5}{6})^{10} + C_{10}^1 (\dfrac{1}{6})(\dfrac{5}{6})^9 + C_{10}^2 (\dfrac{1}{6})^2 (\dfrac{5}{8})^8.$

Chú ý. Khi n, k và $(n-k)$ đều khá lớn thì việc tính $p_n(k;p)$ và $p_n(k_1,k_2;p)$ theo các công thức trên là rất nặng nề. Vì thế, trên thực tế, người ta thường dùng các công thức gần đúng sau đây:

$$p_n(k;p) \approx \frac{1}{\sqrt{npq}}\varphi(\frac{k-np}{\sqrt{npq}}),$$

$$p_n(k_1,k_2;p) \approx \Phi(\frac{k_2-np}{\sqrt{npq}}) - \Phi(\frac{k_1-np}{\sqrt{npq}})$$

với

$$q = 1-p, \quad \varphi(x) = \frac{1}{\sqrt{2n}}e^{\frac{-x^2}{2}}, \quad \Phi(x) = \int_{-\infty}^{x}\varphi(t)dt = \frac{1}{\sqrt{2n}}\int_{-\infty}^{x}e^{\frac{-t^2}{2}}dt.$$

Các công thức trên được rút ra từ các định lý giới hạn. Các định lý này sẽ được nghiên cứu trong Chương 3.

1.5.3 Số có khả năng nhất

Định nghĩa. Số k_0 thoả mãn điều kiện:

$$p_n(k_0;p) = \max_{1\leqslant k\leqslant n} p_n(k;p)$$

được gọi là *số có khả năng nhất.*

Như vậy k_0 là số có khả năng nhất nếu trong các biến cố A_k "A xảy ra k lần" $(k = 0,1,2,\ldots,n)$ thì biến cố A_{k_0} "A xảy ra k_0 lần" là biến cố có khả năng xảy ra nhất.

Định lý. Nếu $np-q$ nguyên thì $k_0 = np-q$ và $k_0 = np-q+1$. Nếu $np-q$ không nguyên thì $k_0 = [p(n+1)]$.

Chứng minh. Ta có

$$\frac{p_n(k+1;p)}{p_n(k;p)} = \frac{C_n^{k+1}p^{k+1}q^{n-k-1}}{C_n^k p^k q^{n-k}}$$

$$= \frac{n!}{(k+1)!(n-k-1)!}.\frac{k!(n-k)!}{n!}.\frac{p}{q} = \frac{n-k}{k+1}.\frac{p}{q}.$$

Suy ra

$$p_n(k+1;p) > p_n(k;p) \Leftrightarrow (n-k)p > (k+1)q \Leftrightarrow k < np - q,$$

$$p_n(k+1;p) = p_n(k;p) \Leftrightarrow k = np - q,$$

$$p_n(k+1;p) < p_n(k;p) \Leftrightarrow k < np - q.$$

Từ đây suy ra rằng:

Nếu $np - q$ là số nguyên thì $p_n(k;p)$ đạt cực đại tại hai giá trị của k là $k_0 = np - q$ và $k_0 = np - q + 1 = np + p = (n+1)p$.

Nếu $np-q$ không nguyên thì $p_n(k;p)$ đạt cực đại tại giá trị của k_0 là số nguyên nằm giữa $np-q$ và $np-q+1$. Tức là $k_0 = [np-q+1] = [(n+1)p]$.

Ví dụ . Trong một cuộc thi bắn quốc tế, mỗi xạ thủ bắn 60 viên vào bia. Xạ thủ Việt Nam bắn trúng tâm với xác suất 0,92. Tìm số viên trúng tâm có khả năng nhất và xác suất tương ứng?

Giải. Ta có

$$n = 60, p = 0,92, np - q = 60.0,92 - 0,08 = 55,12.$$

Suy ra

$$k_0 = [np - q + 1] = [56,12] = 56.$$

$$p_{60}(56;0,92) = C_{60}^{56}(0,92)^{56}(0,08)^4.$$

BÀI TẬP

1. Gieo đồng thời hai con xúc xắc. Tìm xác suất để

a) Tổng số nốt xuất hiện là 7.

b) Tổng số nốt xuất hiện là 8.

c) Số nốt xuất hiện hơn kém nhau 2.

2. Một khách sạn có 6 phòng đơn. Có 10 khách đến thuê phòng, trong đó có 6 nam và 4 nữ. Người quản lý chọn ngẫu nhiên 6 người. Tính xác suất để trong đó

a) Cả 6 người đều là nam.

b) Có 4 nam và 2 nữ.

c) Có ít nhất 2 nữ.

3. Một hòm có 9 tấm thẻ đánh số từ 1 đến 9. Chọn ngẫu nhiên ra hai tấm thẻ. Tính xác suất để tích của hai số trên hai tấm thẻ là một số chẵn.

4. Gieo n con xúc xắc. Tìm xác suất để

a) Số nốt xuất hiện trên chúng là như nhau.

b) Tổng số nốt xuất hiện bằng $n+1$.

c) Tổng số nốt xuất hiện bằng $n+2$.

5. Chứng minh rằng, nếu A, B, C là các biến cố thì

a) $\mathbb{P}(A\triangle B) = \mathbb{P}(A) + \mathbb{P}(B) - 2\mathbb{P}(AB)$,

b) $\mathbb{P}(AB) + \mathbb{P}(BC) + \mathbb{P}(CA) \geqslant \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - 1$,

c) $\mathbb{P}(AB) + \mathbb{P}(AC) - \mathbb{P}(BC) \leqslant \mathbb{P}(A)$,

d) $\mathbb{P}(A\triangle B) \leqslant \mathbb{P}(A\triangle C) + \mathbb{P}(C\triangle B)$,

ở đây $A\triangle B = (A \setminus B) \cup (B \setminus A)$.

6. Giả sử A, B là các biến cố. Chứng minh rằng

a) $\mathbb{P}(A\cup B)\mathbb{P}(AB) \leqslant \mathbb{P}(A)\mathbb{P}(B)$.

b) $|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| \leqslant \dfrac{1}{4}$.

7. Giả sử $A_1, \ldots, A_n, C$ là các biến cố. Chứng minh rằng

$$\mathbb{P}(A_1 \ldots A_n | C) = \mathbb{P}(A_1|C)\mathbb{P}(A_2|A_1C) \ldots \mathbb{P}(A_n|A_1 \ldots A_{n-1}C)$$

nếu vế phải xác định.

8. Giả sử $B_1, \ldots, B_n$ là họ đầy đủ các biến cố, C là biến cố thoả mãn điều kiện $\mathbb{P}(B_iC) > 0$ $(i = 1, \ldots, n)$, A là biến cố bất kỳ. Chứng minh rằng

$$\mathbb{P}(A|C) = \sum_{k=1}^{n} \mathbb{P}(A|B_kC)\mathbb{P}(B_k|C).$$

9. Giả sử A và B là các biến cố độc lập và $\mathbb{P}(A \cup B) = 1$. Chứng minh rằng hoặc A hoặc B có xác suất bằng 1.

10. Giả sử A và B là các biến cố độc lập. Chứng minh rằng nếu $A \cap B$ và $A \cup B$ độc lập thì hoặc $\mathbb{P}(A) = 0$, hoặc $\mathbb{P}(B) = 0$, hoặc $\mathbb{P}(A) = 1$, hoặc $\mathbb{P}(B) = 1$.

11. Giả sử A và B là các biến cố độc lập, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B), \mathbb{P}(A \triangle B) = p$ và $\mathbb{P}(A \setminus B) < p$. Tính $\mathbb{P}(A), \mathbb{P}(B), \mathbb{P}(A \setminus B)$.

12. Các biến cố A, B và C độc lập và đều có xác suất khác 0 và khác 1. Chứng minh rằng các biến cố AB, BC và AC không độc lập đôi một, do đó không độc lập.

13. Giả sử A, B, C là các biến cố thoả mãn: A độc lập với BC và với $B \cup C$; B độc lập với AC ; C độc lập với AB; A, B, C đều có xác suất dương. Chứng minh rằng A, B, C độc lập.

14. Giả sử $\varepsilon > 0$, hai biến cố A, B được gọi là hai biến cố ε-độc lập nếu

$$|\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| \leqslant \varepsilon.$$

Chứng minh rằng

a) Nếu A, B là ε - độc lập thì A, $\overline{B}$ và $\overline{A}$, B cũng ε - độc lập.

b) Nếu A là biến cố thoả mãn $\mathbb{P}(A) \leqslant \varepsilon$ hoặc $\mathbb{P}(A) \geqslant 1 - \varepsilon$ thì A là biến cố ε - độc lập với mọi biến cố B.

c) Nếu biến cố A là biến cố ε - độc lập với chính nó thì hoặc $\mathbb{P}(A) \leqslant 2\varepsilon$ hoặc $\mathbb{P}(A) \geqslant 1 - 2\varepsilon$.

15. Ta có 10 hộp bi trong đó có 4 hộp loại I, mỗi hộp có 3 bi trắng và 5 bi đỏ; 3 hộp loại II, mỗi hộp có 4 bi trắng, 6 bi đỏ; 3 hộp loại III, mỗi hộp 2 bi trắng, 5 bi đỏ.

a) Rút hú hoạ một hộp và từ đó lấy ngẫu nhiên ra một viên bi. Tìm xác suất để được viên bi đỏ.

b) Rút hú hoạ một hộp và từ đó lấy ngẫu nhiên một viên bi thì được bi trắng. Tìm xác suất để viên bi đó rút ra từ hộp loại II.

16. Một nhà máy sản xuất bóng đèn có tỉ lệ bóng đèn đạt tiêu chuẩn là 90%. Trước khi xuất ra thị trường, mỗi bóng đèn đều được qua kiểm tra chất lượng. Vì sự kiểm tra không thể tuyệt đối hoàn hảo nên một bóng đèn tốt có xác suất $0,8$ được công nhận là tốt và một bóng đèn hỏng có xác suất $0,95$ bị loại bỏ. Hãy tính tỉ lệ bóng đạt tiêu chuẩn trong số bóng được công nhận là tốt.

17. Có hai lô sản phẩm. Lô I có 10 sản phẩm loại I và 2 sản phẩm loại II. Lô 2 có 16 sản phẩm loại I và 4 sản phẩm loại II. Từ mỗi lô ta lấy ngẫu nhiên ra một sản phẩm. Sau đó trong 2 sản phẩm thu được ta lại lấy hú hoạ ra một sản phẩm. Tìm xác suất để sản phẩm lấy ra sau cùng là sản phẩm loại I.

18. Biết rằng tỷ lệ người mắc bệnh nào đó ở một địa phương là 2%. Người ta sử dụng một phản ứng mà nếu người bị bệnh thì phản ứng luôn luôn dương tính, nếu người không bị bệnh thì phản ứng có thể dương tính với xác suất $0,30$.

a) Tìm xác suất phản ứng dương tính.

b) Tìm xác suất bị bệnh, không bị bệnh trong nhóm người có phản ứng dương tính.

19. Có 3 hộp phấn, trong đó hộp 1 chứa 20 viên tốt và 5 viên xấu, hộp 2 chứa 10 viên tốt và 5 viên xấu, hộp 3 chứa 5 viên tốt và 15 viên xấu. Ta gieo một con xúc xắc cân đối: Nếu xuất hiện mặt 1 chấm thì chọn hộp 1, nếu xuất hiện mặt 2 hoặc 3 chấm thì chọn hộp 2, nếu xuất hiện các mặt còn lại thì chọn hộp 3. Từ hộp được chọn đó lấy ngẫu nhiên ra 1 viên phấn. Tìm xác suất đó là viên phấn tốt.

20. Một lô hàng gồm a sản phẩm loại I và b sản phẩm loại II được đóng gói để gửi cho khách hàng. Nơi nhận kiểm tra thấy thất lạc một sản phẩm. Chọn ngẫu nhiên ra một sản phẩm thì thấy nó là sản phẩm loại I; tính xác suất để sản phẩm thất lạc cũng là sản phẩm loại I.

21. Theo kết quả điều tra về bệnh lao, tỷ lệ người bị lao ở vùng nọ là $0,001$. Tìm xác suất để khi khám cho 10 người thì:

a) Không ai bị lao.

b) 5 người bị lao.

c) ít nhất 1 người bị lao.

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

22. Trong một cuộc thi bắn quốc tế, mỗi xạ thủ bắn 60 viên vào bia. Xạ thủ của Việt Nam bắn trúng tâm với xác suất $0,95$. Tìm xác suất để a) Xạ thủ này bắn trúng tâm cả 60 viên.

b) Xạ thủ này bị trượt ngoài tâm 2 viên.

c) Xạ thủ này bị trượt ít nhất 1 viên.

d) Tìm số viên trúng tâm có khả năng nhất. Tinh xác suất tương ứng.

23.Trong một thành phố nào đó 65% dân cư thích xem đá bóng. Chọn ngẫu nhiên 12 người, hãy tính xác suất để trong đó có đúng 5 người thích xem đá bóng.

24. Một bài thi trắc nghiệm (multiple - choice test) gồm 12 câu hỏi, mỗi

câu hỏi có 5 câu trả lời, trong đó chỉ có một câu trả lời đúng. Giả sử mỗi câu trả lời đúng được 4 điểm và mỗi câu trả lời sai bị trừ 1 điểm. Một học sinh học kém làm bài bằng cách chọn hú hoạ một câu trả lời cho mỗi câu hỏi. Tìm xác suất để

a) Anh ta được 13 điểm.

b) Anh ta được điểm âm.

25. Hai đấu thủ chơi cờ ngang tài ngang sức thi đấu với nhau. Hỏi rằng khả năng nào cao hơn giữa hai khả năng:

a) Thắng 2 ván trong 4 ván.

b) Thắng 3 ván trong 6 ván.

26. Giả sử $A_1, A_2, \dots$ và $B_1, B_2, \dots$ là hai dãy biến cố và $\mathbb{P}(B_n) \to 1$ khi $n \to \infty$. Chứng minh rằng

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \lim_{n\to\infty} \mathbb{P}(A_n B_n)$$

với điều kiện là một trong hai giới hạn đó tồn tại.

27. Giả sử $A_1, A_2, \dots$ là dãy các biến cố. Khi đó, giới hạn trên và giới hạn dưới của dãy đó được định nghĩa như sau

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \text{ và } \liminf A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

Chứng minh rằng

a. $\liminf A_n \subset \limsup A_n$.

b. $\overline{\limsup A_n} = \liminf \overline{A_n}, \quad \overline{\liminf A_n} = \limsup \overline{A_n}$.

c. Giả sử A, B là hai biến cố. Đặt

$$A_n = \begin{cases} A & \text{nếu } n \text{ chẵn} \\ B & \text{nếu } n \text{ lẻ.} \end{cases}$$

Tìm $\limsup A_n$ và $\liminf A_n$.

28. Chứng minh các hệ thức sau đây

$$\limsup(A_n \cup B_n) = \limsup A_n \cup \limsup B_n,$$

$$\liminf(A_n \cap B_n) = \liminf A_n \cap \liminf B_n,$$

$$\limsup A_n \cap \liminf B_n \subset \limsup(A_n \cap B_n) \subset \limsup A_n \cap \limsup B_n.$$

29. Giả sử $A_1, A_2, \ldots$ là dãy các biến cố. Chứng minh rằng

$$\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \mathbb{P}(A_1) + \mathbb{P}(\bar{A}_1 A_2) + \mathbb{P}(\bar{A}_1 \bar{A}_2 A_3) + \cdots$$

30. Chứng minh rằng, đối với mọi dãy biến cố $A_1, A_2, \ldots$, ta có

a. $\mathbb{P}(\limsup A_n) = \lim\limits_{n\to\infty} \mathbb{P}(\bigcup_{i=n}^{\infty} A_i)$,

b. $\mathbb{P}(\liminf A_n) = \lim\limits_{n\to\infty} \mathbb{P}(\bigcap_{i=n}^{\infty} A_i)$,

c. $\mathbb{P}(\liminf A_n) \leqslant \liminf \mathbb{P}(A_n) \leqslant \limsup \mathbb{P}(A_n) \leqslant \mathbb{P}(\limsup A_n)$.

31. Chứng minh rằng nếu họ các biến cố $(A_i)_{i\in I}$ độc lập thì họ các biến cố $(\overline{A_i})_{i\in I}$ cũng độc lập.

32. Giả sử $A_1, A_2, \ldots, A_n$ là các biến cố độc lập. Chứng minh rằng

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = 1 - \prod_{k=1}^{n} \mathbb{P}(\bar{A}_k).$$

33. Giả sử $A_1, A_2, \ldots$ là dãy các biến cố độc lập. Chứng minh rằng

$$\mathbb{P}(\bigcap_{k=1}^{\infty} A_k) = \prod_{k=1}^{\infty} \mathbb{P}(A_k).$$

34. Giả sử $A_1, A_2, \ldots$ là dãy các biến cố độc lập. Chứng minh rằng chuỗi

$$\sum_{n=1}^{\infty}(1 - \mathbb{P}(A_n))$$

hội tụ khi và chỉ khi

$$\mathbb{P}(\bigcap_{n=1}^{\infty} A_n) > 0.$$

35. Giả sử $(\Omega, \mathcal{F}, \mathbb{P})$ là không gian xác suất sao cho đối với mọi biến cố A, tập hợp các biến cố độc lập với A lập thành một đại số. Chứng minh rằng khi đó mọi họ các biến cố độc lập đôi một sẽ độc lập.

36. Giả sử $(\Omega, \mathcal{F}, \mathbb{P})$ là không gian xác suất. Chứng minh rằng cần và đủ để tất cảc các biến cố của nó độc lập đôi một là các biến cố đó có xác suất bằng 0 hoặc bằng 1.

37. Quan hệ độc lập giữa các biến cố trong một không gian xác suất bất kỳ có tính bắc cầu hay không ?

38. Quan hệ phụ thuộc (không độc lập) giữa các biến cố trong một không gian xác suất bất kỳ có tính bắc cầu hay không ?

39. Chứng minh rằng, điều kiện cần và đủ để quan hệ độc lập giữa các biến cố trong một không gian xác suất có tính bắc cầu là các biến cố của không gian đó có xác suất bằng 0 hoặc bằng 1.

40. Giả sử $A, B, A_n, B_n (n \geqslant 1)$ là các biến cố, Chứng minh rằng:

a) $A \triangle B = \bar{A} \triangle B$,

b) $\bigcup_{n=1}^{\infty} A_n \triangle \bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty} (A_n \triangle B_n)$,

c) $\bigcap_{n=1}^{\infty} A_n \triangle \bigcap_{n=1}^{\infty} B_n \subset \bigcap_{n=1}^{\infty} (A_n \triangle B_n)$.

41. Giả sử $(\Omega, \mathcal{F}, P)$ là không gian xác suất, σ-đại số $\mathcal{F}$ sinh bởi đại số $\mathcal{A}$. Chứng minh rằng, với mọi $\varepsilon > 0$ và mọi $A \in \mathcal{F}$, tồn tại $B \in \mathcal{A}$ sao cho $\mathbb{P}(A \triangle B) \leqslant \varepsilon$.

42. Chứng minh rằng nếu hai σ-đại số độc lập và đều chứa các biến cố khác 0 và khác 1 thì hợp của chúng không thể là một đại số.

43. Giả sử $A_1, A_2, \ldots$ và $B_1, B_2, \ldots$ là hai dãy các biến cố và $\mathbb{P}(B_n) \to 1$ khi $n \to \infty$. Chứng minh rằng nếu

$$\liminf \mathbb{P}(A_n) \geqslant a > 0$$

thì

$$\lim_{n\to\infty} \frac{\mathbb{P}(A_n)}{\mathbb{P}(A_n B_n)} = 1.$$

44. Giả sử $A_1, A_2, \ldots$ là dãy các biến cố thoả mãn

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n \bar{A}_{n+1}) < \infty \text{ và } \mathbb{P}(A_n) \to 0 \;\; khi \;\; n \to \infty.$$

Chứng minh rằng $\mathbb{P}(\limsup A_n) = 0$.

CHƯƠNG 2

ĐẠI LƯỢNG NGẪU NHIÊN VÀ PHÂN PHỐI XÁC SUẤT

2.1 Đại lượng ngẫu nhiên

2.1.1 Đại lượng ngẫu nhiên rời rạc

Định nghĩa. Một đại lượng nhận các giá trị bằng số, phụ thuộc vào kết quả của phép thử được gọi là *đại lượng ngẫu nhiên*.

Một đại lượng ngẫu nhiên gọi là *đại lượng ngẫu nhiên rời rạc* nếu nó chỉ nhận một số hữu hạn hoặc đếm được giá trị.

Đối với đại lượng ngẫu nhiên rời rạc, tập hợp tất cả các giá trị có thể có của nó có thể được liệt kê bằng một dãy hữu hạn hay vô hạn $x_1, x_2, x_3, \ldots, x_n, \ldots$.

Tập hợp các giá trị có thể có của đại lượng ngẫu nhiên X được kí hiệu là $X(\Omega)$.

Ví dụ.

1. Tung hai đồng tiền cân đối đồng chất. Gọi X là số mặt sấp xuất hiện. Khi đó X là đại lượng ngẫu nhiên rời rạc và $X(\Omega) = \{0, 1, 2\}$.

2. Kiểm tra ngẫu nhiên 3 học sinh trong một nhóm gồm 4 em học khá và 5 em học trung bình. Gọi X là số em khá được kiểm tra. Khi đó X là đại lượng ngẫu nhiên rời rạc và $X(\Omega) = \{0, 1, 2, 3\}$.

Như vậy, mỗi đại lượng nhẫu nhiên là một đặc trưng về lượng nào đó của các biến cố.

Bảng phân phối. Khi nghiên cứu về đại lượng ngẫu nhiên rời rạc X, ta cần biết tất cả các giá trị của nó cùng với các xác suất tương ứng. Các thông tin này được xác định tiện lợi trong một bảng gọi là bảng phân phối.

Giả sử đại lượng ngẫu nhiên rời rạc X nhận các giá trị $x_1, x_2, \ldots, x_n, \ldots$ với các xác suất tương ứng là $\mathbb{P}(X = x_i) = p_i$ $(i = 1, 2, 3, \ldots, n \ldots)$. Khi đó bảng có dạng

X	x_1	x_2	$\ldots$	x_n	$\ldots$
$\mathbb{P}$	p_1	p_2	$\ldots$	p_n	$\ldots$

được gọi là *bảng phân phối* của X (chú ý rằng $\sum_i p_i = 1$).

Ví dụ. Tung hai đồng tiền cân đối đồng chất. Gọi X là số mặt sấp xuất hiện. Vậy thì

$$\Omega = \{(S, N), (S, S), (N, S), (N, N)\},$$

$$X(\Omega) = \{0, 1, 2\},$$

$\mathbb{P}(X = 0) = \mathbb{P}(N, N) = \dfrac{1}{4},$

$\mathbb{P}(X = 1) = \mathbb{P}((S, N)(N, S)) = \dfrac{2}{4} = \dfrac{1}{2},$

$\mathbb{P}(X = 2) = \mathbb{P}(S, S) = \dfrac{1}{4}.$

Do đó, bảng phân phối của X là

X	0	1	2
$\mathbb{P}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$

Chú ý. Trong các bài toán thực tế, các giá trị của đại lượng ngẫu nhiên thường gắn với các biến cố của phép thử. Khi đó, xác suất của các biến cố đó chính là xác suất của các giá trị tương ứng.

2.1.2 Hàm phân phối

Định nghĩa. *Hàm phân phối* của đại lượng ngẫu nhiên X là hàm số $F(x) = F_X(x)$ được xác định bởi công thức

$$F_X(x) = \mathbb{P}(X < x).$$

Ví dụ. Tung hai đồng tiền cân đối và đồng chất. Gọi X là số mặt sấp xuất hiện. Khi đó, bảng phân phối của X là

X	0	1	2
$\mathbb{P}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$

Nếu $x \leqslant 0$ thì không có giá trị nào của X nhỏ hơn x nên:

$$F(x) = \mathbb{P}(X < x) = \mathbb{P}(\emptyset) = 0.$$

Nếu $0 < x \leqslant 1$ thì

$$F(x) = \mathbb{P}(X = 0) = \frac{1}{4}.$$

Nếu $1 < x \leqslant 2$ thì

$$F(x) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}.$$

Nếu $x > 2$ thì

$$F(x) = \mathbb{P}(X = 0) + \mathbb{P}(X = 1) + \mathbb{P}(X = 2) = 1.$$

Vậy:

$$F(x) = \begin{cases} 0 & \text{nếu } x \leqslant 0, \\ \dfrac{1}{4} & \text{nếu } 0 < x \leqslant 1, \\ \dfrac{3}{4} & \text{nếu } 1 < x \leqslant 2, \\ 1 & \text{nếu } x > 2. \end{cases}$$

Tính chất.

1.

$$0 \leqslant F(x) \leqslant 1.$$

Tính chất này suy ra từ định nghĩa và tính chất tương ứng của xác suất.

2. *Nếu $a < b$ thì $F(b) - F(a) = \mathbb{P}(a \leqslant X < b)$; do đó $F(x)$ là hàm không giảm.*

Thật vậy, giả sử $a < b$, ta có

$$F(b) = \mathbb{P}(X < b) = \mathbb{P}(X < a) + \mathbb{P}(a \leqslant X < b) \geqslant \mathbb{P}(X < a) = F(a).$$

3. $\lim_{x\to+\infty} F(x) = 1$ *và* $\qquad \lim_{x\to-\infty} F(x) = 0.$

Thật vậy

$$\lim_{x\to+\infty} F(x) = \lim_{x\to+\infty} \mathbb{P}(X < x) = \mathbb{P}(X < +\infty) = 1;$$

$$\lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} \mathbb{P}(X < x) = \mathbb{P}(X < -\infty) = 0.$$

Để thuận tiện, người ta thường dùng ký hiệu

$$F(+\infty) = \lim_{x\to+\infty} F(x), \qquad F(-\infty) = \lim_{x\to-\infty} F(x).$$

Lúc đó Tính chất 3 có thể viết

$$F(+\infty) = 1, \qquad F(-\infty) = 0.$$

2.2.2. Đại lượng ngẫu nhiên liên tục

Định nghĩa. Đại lượng ngẫu nhiên X được gọi là *đại lượng ngẫu nhiên liên tục* nếu hàm phân phối $F(x)$ của nó là hàm liên tục và tồn tại hàm số $p(x)$ sao cho

1. $p(x) \geqslant 0, \qquad -\infty < x < +\infty,$
2. $F(x) = \int_{-\infty}^{x} p(t)dt, \qquad -\infty < x < +\infty.$

Hàm số $p(x)$ nêu trên được gọi là *hàm mật độ xác suất* của X.

Ví dụ.

1. Giả sử đại lượng ngẫu nhiên X có hàm phân phối là

$$F(x) = \begin{cases} 0 & \text{nếu } x \leqslant 0, \\ x & \text{nếu } 0 < x \leqslant 1, \\ 1 & \text{nếu } x \geqslant 1. \end{cases}$$

Khi đó, dễ dàng kiểm tra trực tiếp được rằng X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất là

$$p(x) = \begin{cases} 0 & \text{nếu } x \notin (0;1), \\ 1 & \text{nếu } x \in (0;1). \end{cases}$$

2. Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ

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

Tìm hàm phân phối của X .

Giải. Nếu $x < 0$ thì $p(t) = 0, \forall t \leqslant x$. Khi đó

$$F(x) = \int_{-\infty}^{x} 0dt = 0.$$

Nếu $0 \leqslant x \leqslant \dfrac{\pi}{2}$ thì

$$F(x) = \int_{-\infty}^{0} 0dt + \int_{0}^{x} \sin 2tdt = \frac{-\cos 2t}{2}\bigg|_{0}^{x} = \frac{1-\cos 2x}{2}.$$

Nếu $x > \dfrac{\pi}{2}$ thì

$$F(x) = \int_{-\infty}^{0} 0dt + \int_{0}^{\frac{\pi}{2}} \sin 2tdt + \int_{\frac{\pi}{2}}^{x} 0dt = \frac{-\cos 2t}{2}\bigg|_{0}^{\frac{\pi}{2}} = 1.$$

Vậy

$$F(x) = \begin{cases} 0 & \text{nếu } x < 0, \\ \dfrac{1-\cos 2x}{2} & \text{nếu } 0 \leqslant x \leqslant \dfrac{\pi}{2}, \\ 1 & \text{nếu } x > \dfrac{\pi}{2}. \end{cases}$$

Tính chất. Từ định nghĩa, suy ra

1. *Với mọi $a, b : -\infty \leqslant a < b \leqslant +\infty$ ta có*

$$\mathbb{P}(a < X < b) = \int_{a}^{b} p(x)dx.$$

2. $\int_{-\infty}^{+\infty} p(x)dx = 1.$

3. $p(x) = F^{'}(x)$ *tại mọi điểm x mà $p(x)$ liên tục.*

Chứng minh

1. $\mathbb{P}(a < X < b) = \mathbb{P}(X < b) - \mathbb{P}(X < a) = \int_{-\infty}^{b} p(t)dt - \int_{-\infty}^{a} p(t)dt = \int_{a}^{b} p(t)dt.$

2. $\int_{-\infty}^{+\infty} p(x)dx = \mathbb{P}(X < +\infty) = 1.$

3.

$$\begin{aligned} F^{'}_{+}(x) &= \lim_{\Delta x \to 0\oplus} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0\oplus} \frac{\mathbb{P}(x < X < x+\Delta x)}{\Delta x} \\ &= \lim_{\Delta x \to 0^{+}} \frac{\int_{x}^{x+\Delta x} p(t)dt}{\Delta x} = p(x) \end{aligned}$$

(theo định lý giá trị trung bình của tích phân).

Tương tự

$$F_-^{'}(x) = p(x).$$

Suy ra

$$F^{'}(x) = p(x).$$

2.2.3. Các đại lượng ngẫu nhiên độc lập

Định nghĩa. Hai đại lượng ngẫu nhiên X và Y gọi là *độc lập* nếu với mọi $x, y \in \mathbb{R}$, ta có

$$\mathbb{P}(X < x, Y < y) = \mathbb{P}(X < x)\mathbb{P}(Y < y).$$

Họ các đại lượng ngẫu nhiên $(X_i)_{i\in I}$ được gọi là *độc lập đôi một* nếu với mọi $i, j \in I,\ i \neq j$ thì X_i và X_j độc lập.

Từ định nghĩa suy ra rằng nếu X, Y là các đại lượng ngẫu nhiên rời rạc:

$$X(\Omega) = \{x_1, x_2, \ldots, x_n, \ldots\},\ \ Y(\Omega) = \{y_1, y_2, \ldots, y_m, \ldots\},$$

thì X và Y độc lập khi và chỉ khi

$$\mathbb{P}(X = x_i; Y = y_j) = \mathbb{P}(X = x_i) \cdot \mathbb{P}(Y = y_j)$$

$$(i = 1, 2, \ldots, n, \ldots; j = 1, 2, \ldots, m, \ldots).$$

Ví dụ.

1. Tung hai con xúc xắc. Gọi X là số chấm xuất hiện trên con xúc xắc thứ nhất; Y là số chấm xuất hiện trên con xúc xắc thứ 2. Khi đó X và Y độc lập.

2. Hai người độc lập nhau bắn vào một cái bia. Gọi X và Y lần lượt là số viên trúng của người thứ nhất và người thứ 2. Vậy thì X và Y cũng độc lập.

2.2 Các số đặc trưng của đại lượng ngẫu nhiên

2.3.1 Kỳ vọng

Định nghĩa. *Kỳ vọng* của đại lượng ngẫu nhiên X là số $\mathbb{E}X$ xác định bởi công thức

$$\mathbb{E}X = \begin{cases} \sum_i x_i p_i \ \text{ nếu } X \text{ rời rạc và } \mathbb{P}(X = x_i) = p_i, \\ \int_{-\infty}^{+\infty} xp(x)dx \ \text{ nếu } X \text{ liên tục có hàm mật độ } p(x). \end{cases}$$

Ví dụ.

1. Cho X là đại lượng ngẫu nhiên rời rạc nhận n giá trị $x_1, x_2, \dots, x_n$ với xác suất như nhau

$$p_i = \mathbb{P}(X = x_i) = \frac{1}{n}.$$

Khi đó

$$\mathbb{E}X = x_1p_1 + x_2p_2 + \cdots + x_np_n = \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

2. Cho X là đại lượng ngẫu nhiên liên tục có hàm mật độ là

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

Khi đó

$$\mathbb{E}X = \int_{-\infty}^{+\infty} xp(x)dx = \int_{-\infty}^{a} xp(x)dx + \int_{a}^{b} xp(x)dx$$

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

3. Qua hai ví dụ trên, ta có suy nghĩ rằng $\mathbb{E}X$ là giá trị trung bình (trung bình cộng) của đại lượng ngẫu nhiên. Song điều này không hoàn toàn chính xác. Chẳng hạn ta xét hai đại lượng ngẫu nhiên

X	0	1
$\mathbb{P}$	$\frac{1}{4}$	$\frac{3}{4}$

Y	0	1
$\mathbb{P}$	$\frac{3}{4}$	$\frac{1}{4}$

Khi đó $\mathbb{E}X = \frac{3}{4}$, $\mathbb{E}Y = \frac{1}{4}$.

ý nghĩa: Qua các ví dụ trên, ta thấy rằng kỳ vọng của đại lượng ngẫu nhiên X là giá trị trung bình theo xác suất của đại lượng ngẫu nhiên đó. Trong trường hợp X nhận các giá trị với xác suất như nhau thì kỳ vọng chính là trung bình cộng của nó.

Tính chất. Kỳ vọng có các tính chất sau đây

1. *Nếu $X \geqslant 0$ thì $\mathbb{E}X \geqslant 0$.*
2. *Nếu $X = C$ thì $\mathbb{E}X = C$.*
3. *Nếu tồn tại $\mathbb{E}X$ thì với mọi $C \in \mathbb{R}$, ta có $\mathbb{E}(CX) = C\mathbb{E}X$.*
4. *Nếu tồn tại $\mathbb{E}X$ và $\mathbb{E}Y$ thì $\mathbb{E}(X \pm Y) = \mathbb{E}X \pm \mathbb{E}Y$.*
5. *Nếu X và Y là hai đại lượng ngẫu nhiên độc lập thì*

$$\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y.$$

6. *Nếu $f : \mathbb{R} \to \mathbb{R}$ là hàm liên tục và $Y = f(X)$ thì*

$$\mathbb{E}Y = \begin{cases} \sum_i f(x_i)p_i \;\; \textit{nếu } X \textit{ rời rạc nhận các giá trị } x_1, x_2, \dots \\ \;\textit{với } \mathbb{P}(X = x_i) = p_i. \\ \int_{-\infty}^{+\infty} f(x)p(x)dx \;\; \textit{nếu } X \textit{ liên tục có hàm mật độ } p(x). \end{cases}$$

7. (*Bất đẳng thức Markov*) *Giả sử X là đại lượng ngẫu nhiên không âm. Khi đó nếu tồn tại $\mathbb{E}X$ thì với mọi $\varepsilon > 0$, ta có*

$$\mathbb{P}(X \geq \varepsilon) \leqslant \frac{\mathbb{E}X}{\varepsilon}\ .$$

Chứng minh. Ta chứng minh các tính chất trên cho các đại lượng ngẫu nhiên rời rạc.

1. Hiển nhiên.

2. Vì $\mathbb{P}(X = C) = 1$ nên $\mathbb{E}X = C \cdot 1 = C$.

3. Ta có $\mathbb{P}(CX = Cx_i) = \mathbb{P}(CX = Cx_i)$ nên $\mathbb{E}(CX) = \sum_i Cx_ip_i \cdot 1 = C\sum_i x_ip_i = C\mathbb{E}X$.

4. Giả sử

$$X(\Omega) = \{x_1, x_2, \dots, x_m \cdots\}; \qquad Y(\Omega) = \{y_1, y_2, \dots, y_n \cdots\}.$$

Khi đó $Z = X + Y$ là đại lượng ngẫu nhiên nhận các giá trị $z_{ij} = x_i + y_j$ với các xác suất tương ứng $p_{ij} = \mathbb{P}(X = x_i; Y = y_j)$, $(i = 1, 2, \dots;$ $j = 1, 2, \dots)$. Do đó

$$\mathbb{E}Z = \sum_i \sum_j (x_i + y_j)p_{ij} = \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij}.$$

Mặt khác

$$\mathbb{P}(X = x_i) = \mathbb{P}(X = x_i, Y = y_1) + \mathbb{P}(X = x_i, Y = y_2) + \cdots + \mathbb{P}(X = x_i, Y = y_n) + \cdots$$

Tương tự

$$\mathbb{P}(Y = y_j) = \sum_i p_{ij}.$$

Suy ra

$$\mathbb{E}Z = \sum_i x_i \mathbb{P}(X = x_i) + \sum_j y_j \mathbb{P}(Y = y_j) = \mathbb{E}X + \mathbb{E}Y.$$

5. Giả sử

$$X(\Omega) = \{x_1, x_2, \ldots, x_m, \ldots\}; \qquad Y(\Omega) = \{y_1, y_2, \ldots, y_n, \ldots\}.$$

Khi đó $Z = X.Y$ nhận các giá trị $z_{ij} = x_i y_j$ với các xác suất tương ứng $p_{ij} = \mathbb{P}(X = x_i; Y = y_j)$, $(i = 1, 2, \ldots; j = 1, 2, \ldots)$. Do X, Y độc lập nên

$$p_{ij} = \mathbb{P}(X = x_i; Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j) = p_i q_j.$$

Suy ra

$$\begin{aligned}\mathbb{E}Z &= \sum_i \sum_j x_i y_j p_{ij} = \sum_i \sum_j x_i y_j p_i q_j \\ &= \sum_i x_i p_i \sum_j y_j q_j = \mathbb{E}X\mathbb{E}Y.\end{aligned}$$

6. Nếu X nhận giá trị x_i thì $Y = f(X)$ nhận giá trị $f(x_i)$ nên $\mathbb{E}Y = \sum_i f(x_i) p_i$.

7. Giả sử G là tập giá trị của X, ta đặt

$$G_1 = [c \in G : \ 0 \leq c < \varepsilon]; \ \ G_2 = [c \in G : \ c \geqslant \varepsilon].$$

Khi đó

$$\begin{aligned}\mathbb{E}X &= \sum_{c_i \in G} c_i \, \mathbb{P}[X = c_i] = \sum_{c_i \in G_1} c_i \, \mathbb{P}[X = c_i] + \sum_{c_i \in G_2} c_i \, \mathbb{P}[X = c_i] \\ &\geqslant \sum_{c_i \in G_2} c_i \, \mathbb{P}[X = c_i] \geqslant \varepsilon \sum_{c_i \in G_2} \mathbb{P}[X = c_i] = \varepsilon \, \mathbb{P}[X \geq \varepsilon].\end{aligned}$$

Từ đó suy ra điều phải chứng minh.

Ví dụ. Cho hai đại lượng ngẫu nhiên X, Y độc lập với nhau có bảng phân phối

X	0	1	2
$\mathbb{P}$	$0,3$	$0,4$	$0,3$

Y	-1	1
$\mathbb{P}$	$0,4$	$0,6$

Tính $\mathbb{E}X^2$; $\mathbb{E}(X+Y)$ và $\mathbb{E}(XY)$.

Giải. Ta có

$\mathbb{E}X = 0.0,3 + 1.0,4 + 2.0,3 = 1;$

$\mathbb{E}Y = -1.0,4 + 1.0,6 = 0,2;$

$\mathbb{E}X^2 = 0^2.0,3 + 1^2.0,4 + 2^2.0,3 = 1,6;$

$\mathbb{E}(X+Y) = \mathbb{E}X + \mathbb{E}Y = 1 + 0,2 = 1,2;$

$\mathbb{E}(X.Y) = \mathbb{E}X.\mathbb{E}Y = 1.0,2 = 0,2.$

ý nghĩa: Qua các ví dụ trên, ta thấy rằng kỳ vọng của đại lượng ngẫu nhiên X là giá trị trung bình theo xác suất của đại lượng ngẫu nhiên đó. Trong trường hợp X nhận các giá trị với xác suất như nhau thì kỳ vọng chính là trung bình cộng của nó.

2.3.2 Phương sai

Định nghĩa. Giả sử X là đại lượng ngẫu nhiên. Khi đó, số $DX = \mathbb{E}(X - \mathbb{E}X)^2$ (nếu tồn tại) được gọi là *phương sai* của X.

Chú ý. Từ định nghĩa trên và từ tính chất của kỳ vọng, suy ra rằng phương sai DX của đại lượng ngẫu nhiên X có thể tồn tại hoặc không tồn tại và nếu tồn tại thì có thể được tính theo công thức

$$DX = \begin{cases} \sum (x_i - \mathbb{E}X)^2 p_i & \text{nếu } X \text{ rời rạc và } \mathbb{P}(X = x_i) = p_i. \\ \int_{-\infty}^{+\infty} (x - \mathbb{E}X)^2 p(x)dx & \text{nếu } X \text{ liên tục có hàm mật độ là } p(x). \end{cases}$$

Ví dụ. 1. Giả sử X là đại lượng ngẫu nhiên rời rạc có bảng phân phối

X	0	1
$\mathbb{P}$	p	q

Khi đó

$$\mathbb{E}X = 0.q + 1.p = p.$$

Suy ra

$$DX = (0-p)^2.q + (1-p)^2 = p^2q + q^2p = pq.$$

2. Giả sử X là đại lượng ngẫu nhiên liên tục có hàm mật độ

$$p(x) = \begin{cases} 0 & \text{nếu } x \notin [0,2] \\ \dfrac{1}{2} & \text{nếu } x \in [0,2]. \end{cases}$$

Khi đó

$$\mathbb{E}X = 1$$

Suy ra

$$DX = \int_{-\infty}^{+\infty} (x-1)^2 p(x)dx$$

$$= \int_0^2 \frac{(x-1)^2}{2}dx = \frac{(x-1)^3}{6}\bigg|_0^2 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

ý nghĩa: $|X - \mathbb{E}X|$ là độ lệch của các giá trị của đại lượng ngẫu nhiên X khỏi $\mathbb{E}X$. Do đó phương sai $DX = \mathbb{E}(X - EX)^2$ chính là trung bình của bình phương độ lệch của X quanh $\mathbb{E}X$. Phương sai cho biết mức độ phân tán của các giá trị của đại lượng ngẫu nhiên X quanh kỳ vọng của nó. Tuy nhiên, phương sai DX có hạn chế là không cùng thứ nguyên với X. Do đó, cùng với DX, người ta còn dùng $\sqrt{DX}$ để nghiên cứu mức độ phân tán của đại lượng ngẫu nhiên X quanh $\mathbb{E}X$. $\sqrt{DX} = \sigma_X$ gọi là *độ lệch chuẩn của* X. σ_X có cùng thứ nguyên với X.

Tính chất. *Phương sai có những tính chất cơ bản sau đây:*

1. $DX = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

2. $DX \geqslant 0$.

3. $DX = 0$ *khi và chỉ khi* $X = \mathbb{E}X =$ *hằng số h.c.c.*

4. $D(CX) = C^2 DX$.

5. Nếu X, Y *độc lập thì* $D(X \pm Y) = DX + DY$.

Tổng quát: Nếu $(X_i)_{i=\overline{1,n}}$ *là họ đôi một độc lập thì*

$$D(X_1 + \cdots + X_n) = DX_1 + \cdots + DX_n.$$

Chứng minh.

1. $DX = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}(X^2 - 2X.\mathbb{E}X + (\mathbb{E}X)^2) = \mathbb{E}X^2 - 2\mathbb{E}X.\mathbb{E}X + (\mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$.

2. $(X - \mathbb{E}X)^2 \geqslant 0 \Rightarrow DX = \mathbb{E}(X - \mathbb{E}X)^2 \geqslant 0$.

3. $DX = 0 \iff \mathbb{E}(X - \mathbb{E}X)^2 = 0 \iff (X - \mathbb{E}X)^2 = 0 \iff P(X = \mathbb{E}X) = 1 \iff X = \mathbb{E}X$ *h.c.c.*

4. $D(CX) = \mathbb{E}(CX - \mathbb{E}CX) = \mathbb{E}(CX - C\mathbb{E}X)^2 = \mathbb{E}(C^2(X - \mathbb{E}X)^2) = C^2\mathbb{E}(X - \mathbb{E}X)^2 = C^2 DX$.

5. $D(X + Y) = \mathbb{E}(X + Y)^2 - (\mathbb{E}(X + Y))^2 = \mathbb{E}(X^2 + 2XY + Y^2) - (\mathbb{E}X + \mathbb{E}Y)^2 = (\mathbb{E}X^2 - (\mathbb{E}X)^2) + (\mathbb{E}Y^2 - (\mathbb{E}Y)^2) + 2(\mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y) = DX + DY + 2(\mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y) = DX + DY$ (vì $\mathbb{E}(XY) = \mathbb{E}X.\mathbb{E}Y$ (Do X và Y độc lập)).

Kết hợp với tính chất 4 ta được

$$D(X - Y) = D(X + (-Y)) = DX + D(-Y) = DX + DY.$$

Bây giờ giả sử $(X_i)_{i=\overline{1,n}}$ là họ độc lập đôi một. Vậy thì

$$D(\sum_{i=1}^{n} X_i) = \mathbb{E}(\sum_{i=1}^{n} X_i)^2 - (\mathbb{E}(\sum_{i=1}^{n} X_i))^2 = \mathbb{E}(\sum_{i=1}^{n} X_i^2$$

$$+2\sum_{1\leqslant i<j\leqslant n} X_iX_j) - \sum_{i=1}^{n}(\mathbb{E}X_i)^2 - 2\sum_{1\leqslant i<j\leqslant n}\mathbb{E}X_i\mathbb{E}X_j$$

$$= \sum_{i=1}^{n}(\mathbb{E}X_i^2 - (\mathbb{E}Xi)^2) + 2\sum_{1\leqslant i<j\leqslant n}(\mathbb{E}(X_iX_j) - \mathbb{E}X_i\mathbb{E}X_j) = \sum_{i=1}^{n} DX_i.$$

(Vì X_i, X_j độc lập nên $\mathbb{E}(X_iX_j) - \mathbb{E}X_i.\mathbb{E}X_j = 0$).

2.3.3 Mode

Khái niệm mode được định nghĩa riêng rẽ cho hai trường hợp đại lượng ngẫu nhiên rời rạc và đại lượng ngẫu nhiên liên tục.

Nếu X là đại lượng ngẫu nhiên rời rạc thì giá trị x_0 được gọi là *mode* của X, nếu X có xác suất lớn nhất tại x_0.

Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ $p(x)$ thì giá trị x_0 được gọi là *mode* của X, nếu $p(x)$ đạt giá trị lớn nhất tại x_0.

Nếu x_0 là mode của X thì ta viết $x_0 = \mod X$.

2.3.4 Phân vị cấp p

Số x_p $(0 < p < 1)$ được gọi là *phân vị cấp p* của hàm phân phối $F(x)$ của đại lượng ngẫu nhiên X nếu

$$F(x_p) \leqslant p \quad \text{và} \quad F(x_p + 0) \geqslant p \qquad (F(x_p+0) = \lim_{x\to x_p} F(x)).$$

Rõ ràng, nếu $F(x)$ là hàm liên tục thì $F(x_p) = p$.
Nếu $p = \dfrac{1}{2}$ thì $x_p = x_{\frac{1}{2}}$ được gọi là *trung vị* hay *median* của X và được ký hiệu là $m(X)$.

2.3.5 Moment, hệ số bất đối xứng và hệ số nhọn

Giả sử X là đại lượng ngẫu nhiên, khi đó số

$$m_k = \mathbb{E}X^k$$

(nếu tồn tại) được gọi là *moment cấp k* của X, còn số

$$\alpha_k = \mathbb{E}(X - \mathbb{E}X)^k$$

(nếu tồn tại) được gọi là *moment trung tâm* cấp k của X (moment cấp 1 chính là kỳ vọng, còn moment trung tâm cấp 2 chính là phương sai).

Số

$$S = \frac{\alpha_3}{\alpha_2^{\frac{3}{2}}}$$

được gọi là *hệ số bất đối xứng* của X.

Số

$$E = \frac{\alpha_4}{\alpha_2^2} - 3$$

được gọi là *hệ số nhọn* của X.

Ví dụ. Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ

$$p(x) = \begin{cases} 0 & \text{nếu } x < 0, \\ \dfrac{3x^2}{2} & \text{nếu } 0 \leqslant x < 1, \\ \dfrac{3}{2}(2-x)^2 & \text{nếu } 1 \leqslant x < 2, \\ 0 & \text{nếu } x \geqslant 2. \end{cases}$$

Hãy tính các moment và moment trung tâm cấp 1, 2, 3, 4. Tính hệ số bất đối xứng và hệ số nhọn.

Giải.

$m_1 = 1$; $m_2 = 1,1$;

$m_3 = \dfrac{3}{2}\int_0^1 x^5 dx + \dfrac{3}{2}\int_1^2 x^3(2-x)^2 dx = 1,3$;

$m_4 = \dfrac{3}{2}\int_0^1 x^6 dx + \dfrac{3}{2}\int_1^2 x^4(2-x)^2 dx = 1\dfrac{22}{55}$.

Từ đó

$\alpha_1 = 0$, $\alpha_2 = m_2 - m_1^2 = 0,1$;

$\alpha_3 = m_3 - 3m_1m_2 + 2m_1^3 = 1,3 - 3(1,1) + 2 = 0;$

$\alpha_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4 = -0,2;$

$S = \dfrac{\alpha_3}{\alpha_2^{\frac{3}{2}}} = 0.$

$E = \dfrac{\alpha_4}{\alpha_2^2} - 3 = -\dfrac{1}{7}.$

2.3 Một số phân phối xác suất quan trọng

2.4.1 Phân phối nhị thức

Định nghĩa. Đại lượng ngẫu nhiên X được gọi là có *phân phối nhị thức* với các tham số n, p và được ký hiệu là $X \sim B(n;p)$ ($n \geq 1$, $0 < p < 1$), nếu

$$X(\Omega) = \{0, 1, 2, \ldots, n\}$$

và

$$p_k = \mathbb{P}(X = k) = C_n^k p^k q^{n-k},$$

trong đó $q = 1 - p$.

Ví dụ. Trong một thành phố nào đó 75% gia đình có tivi màu. Chọn ngẫu nhiên 12 gia đình và gọi X là số gia đình có tivi màu.

a. Gọi tên phân phối xác suất của X.

b. Tính xác suất để có đúng 5 gia đình có tivi màu.

c. Tính xác suất để trong mẫu có ít nhất 2 gia đình có tivi màu.

Giải.

a. X có phân phối nhị thức với tham số $n = 12$, $p = 0,75$.

b. $\mathbb{P}(X = 5) = C_{12}^5(0,75)^5(0,75)^7 = 0,0591.$

c. $\mathbb{P}(X \geqslant 2) = 1 - P(X = 0) - P(X = 1) = 1 - (0,25)^{12} - 12(0,75)(0,25)^{11}.$

Tìm kỳ vọng, phương sai và mode của phân bố nhị thức.

Định lý. *Nếu* $X \sim B(n;p)$ *thì*

i) $\mathbb{E}X = np$,

ii) $DX = npq$,

iii)

$$\text{mod } X = \begin{cases} [np - q + 1] \quad \textit{nếu } np - q \textit{ không nguyên}, \\ np - q \textit{ và } np - q + 1 \textit{ nếu } np - q \textit{ nguyên}. \end{cases}$$

Chứng minh.

i) Ta có

$$\mathbb{E}X = \sum_{k=0}^{n} k C_n^k p^k q^{n-k} = \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

$$= n \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^k q^{n-k} = np \sum_{k=1}^{n} C_{n-1}^{k-1} p^k q^{n-k}$$

$$= np \sum_{i=0}^{n-1} C_{n-1}^{i-1} p^i q^{n-1-i} = np(p+q)^{n-1} = np.$$

ii)

$$\begin{aligned} \mathbb{E}X^2 &= \sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k} \\ &= \sum_{k=1}^{n} \frac{k^2 n!}{k!(n-k)!} p^k q^{n-k} \\ &= \sum_{k=1}^{n} \frac{k(k-1)n!}{k!(n-k)!} p^k q^{n-k} + \sum_{k=1}^{n} \frac{kn!}{k!(n-k)!} p^k q^{n-k} \\ &= S_1 + S_2. \end{aligned}$$

Ta có

$S_1 = \sum_{k=2}^{n} \dfrac{n!}{(k-2)!(n-k)!} p^k q^{n-k} = n(n-1)p^2$,

$S_2 = np$ (đã tính ở trên).

Thành thử

$$DX = \mathbb{E}X^2 - (\mathbb{E}X)^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq.$$

Để chứng minh *iii*) ta chỉ cần nhận xét rằng $\mod X$ chính là số có khả năng nhất.

Ví dụ. Tỉ lệ cử tri ủng hộ ứng cử viên A trong một cuộc bầu cử tổng thống là 60%. Người ta hỏi ý kiến 20 cử tri được chọn một cách ngẫu nhiên. Gọi X là số người bỏ phiếu cho ứng cử viên A trong 20 người đó.

a. Tìm giá trị trung bình, độ lệch tiêu chuẩn của X và $\mod X$.

b. Tính $\mathbb{P}(X \leqslant 10)$.

Giải. a. X có phân phối nhị thức $B(20; 0,6)$. Vậy $\mathbb{E}X = 20.(0,6) = 12$; $DX = 20.(0,6).(0,4) = 4,8$; $\sigma_X = \sqrt{4,8} \approx 2,2$; $\mod X = [21.0,6] = [12,6] = 12$.

b. $\mathbb{P}(X \leqslant 10) = \sum_{k=0}^{10} C_{20}^k (0,6)^k (0,4)^{20-k}$.

2.4.2 Phân phối Poisson

Định nghĩa. Ta nói rằng đại lượng ngẫu nhiên X có phân phối Poisson với tham số λ và kí hiệu là $X \sim P(\lambda)$ nếu

$$X(\Omega) = \{0, 1, 2, \dots\}$$

và

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

trong đó λ là một số dương cho trước.

Người ta đã lập bảng để tính sẵn $\mathbb{P}(X \leqslant k) = \sum_{i=1}^{k} e^{-\lambda}\dfrac{\lambda^i}{i!}$ với các giá trị λ khác nhau.

Ví dụ. Một gara cho thuê ôtô thấy rằng số người đến thuê ôtô vào ngày thứ Bảy cuối tuần là một đại lượng ngẫu nhiên X có phân phối Poisson với tham số $\lambda = 2$. Giả sử gara có 4 chiếc ôtô. Hãy tìm xác suất để

a. Không phải tất cả 4 chiếc đều được thuê.

b. Tất cả 4 ôtô đều được thuê.

c. Gara không đáp ứng được yêu cầu.

d. Trung bình có bao nhiêu ôtô được thuê.

e. Gara cần có ít nhất bao nhiêu ôtô để xác suất không đáp ứng nhu cầu thuê bé hơn 2%.

Giải.

a. $\mathbb{P}(X \leqslant 3) = 0,857$.

b. $\mathbb{P}(X \geqslant 4) = 1 - \mathbb{P}(X \leqslant 3) = 1 - 0,857 = 0,143$.

c. $\mathbb{P}(X > 4) = 1 - \mathbb{P}(X \leq 4) = 1 - 0,947 = 0,053$.

d. Gọi Y là số ôtô được thuê. Khi đó

$\mathbb{P}(Y = 0) = \mathbb{P}(X = 0) = 0,135$;

$\mathbb{P}(Y = 1) = \mathbb{P}(X \leqslant 1) - \mathbb{P}(X = 0) = 0,406 - 0,135 = 0,271$;

$\mathbb{P}(Y = 2) = \mathbb{P}(X = 2) = \mathbb{P}(X \leqslant 2) - \mathbb{P}(X \leqslant 1) = 0,677 - 0,406 = 0,271$;

$\mathbb{P}(Y = 3) = \mathbb{P}(X = 3) = \mathbb{P}(X \leqslant 3) - \mathbb{P}(X \leqslant 2) = 0,857 - 0,677 = 0,18$;

$\mathbb{P}(Y = 4) = \mathbb{P}(X \geqslant 4) = 1 - \mathbb{P}(X \leqslant 3) = 1 - 0,857 = 0,143$.

Từ đó $\mathbb{E}Y = 1,925$.

e. Ta cần xác định n nhỏ nhất để

$$\mathbb{P}(X > n) < 0,02 \Leftrightarrow \mathbb{P}(X \leqslant n) > 0,98.$$

Vì $\mathbb{P}(X \leqslant 4) = 0,947$; $\mathbb{P}(X \leqslant 5) = 0,983$.

Do đó $n = 5$.

Ta hãy tìm kỳ vọng, phương sai và mode của phân phối Poisson.

Định lý. *Giả sử $X \sim P(\lambda)$. Khi đó $\mathbb{E}X = \lambda$, $DX = \lambda$, $\mod X = [\lambda]$.*

Chứng minh. Ta có

$$\mathbb{E}X = \sum_{k=0}^{\infty} k e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda}\lambda e^{\lambda} = \lambda$$

$$\mathbb{E}X^2 = \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=2}^{\infty} \frac{k(k-1)\lambda^k}{k!} + e^{-\lambda} \sum_{k=0}^{\infty} k \frac{\lambda^k}{k!}$$

$$= e^{-\lambda}\lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda = \lambda^2 + \lambda.$$

Vậy $DX = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \lambda$.

Để tìm mod$X$ ta xét

$$\frac{\mathbb{P}(X = k)}{\mathbb{P}(X = k-1)} = \frac{\lambda}{k} > 1 \Leftrightarrow \lambda > k.$$

Vậy $\mathbb{P}(X = k)$ lớn nhất khi k là số nguyên lớn nhất bé hơn λ. Nói cách khác mod$X = [\lambda]$.

Ví dụ. ở một tổng đài bưu điện, các cú điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau và tốc độ trung bình 2 cuộc gọi trong một phút. Tìm xác suất để:

a. Có đúng 5 cú điện thoại trong 2 phút.

b. Không có cú điện thoại nào trong khoảng thời gian 30 giây.

c. Có ít nhất 1 cú điện thoại trong khoảng thời gian 10 giây.

Giải. a. Số cú điện thoại xuất hiện trong khoảng thời gian 2 phút là đại lượng ngẫu nhiên $X \sim P(4)$. Vậy

$$\mathbb{P}(X = 5) = e^{-4}\frac{4^5}{5!}$$

$$= \mathbb{P}(X \leqslant 5) - \mathbb{P}(X \leqslant 4) = 0,785 - 0,629 = 0,156.$$

b. Số cú điện thoại xuất hiện trong khoảng thời gian 30 giây là đại lượng ngẫu nhiên $X \sim \mathbb{P}(1)$. Vậy

$$\mathbb{P}(X = 0) = e^{-1} = 0,3679.$$

c. Số cú điện thoại xuất hiện trong khoảng thời gian 10 giây là đại lượng ngẫu nhiên $X \sim P(\dfrac{1}{3})$. Vậy

$$\mathbb{P}(X \geqslant 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-\dfrac{1}{3}} = 0,2835.$$

Định lý. *Nếu* $X \sim P(\mu)$, $Y \sim P(\lambda)$ *và* X, Y *độc lập thì* $Z = (X+Y) \sim P(\lambda + \mu)$.

Chứng minh. Ta có

$$\begin{aligned}
\mathbb{P}(Z = k) = \mathbb{P}(X + Y = k) &= \sum_{i=0}^{k} \mathbb{P}(X = i, Y = k - i) \\
&= \sum_{i=0}^{k} \mathbb{P}(X = i)\mathbb{P}(Y = k - i) = \sum_{i=0}^{k} e^{-\mu}\frac{\mu^i}{i!}\frac{e^{-\lambda}\lambda^{k-1}}{(k-1)!} \\
&= \frac{e^{-(\lambda+\mu)}}{k!}\sum_{i=0}^{k} C_k^i \mu^i \lambda^{k-1} = \frac{e^{-(\lambda+\mu)}(\lambda+\mu)^k}{k!}.
\end{aligned}$$

Ví dụ. Một cửa hàng bán đồ điện tử gồm hai mặt hàng: tivi và radio. Số tivi và radio bán trong một ngày đều có phân phối Poisson và chúng độc lập với nhau. Trung bình mỗi ngày bán được 1 tivi và 2 radio. Tìm xác suất để một ngày cửa hàng bán được ít nhất 4 chiếc (radio và tivi).

Giải. Gọi X và Y tương ứng là số tivi và số radio bán được trong ngày. Ta có $X \sim P(1)$ và $Y \sim P(2)$, X và Y độc lập. Theo định lý trên thì $(X + Y) \sim P(3)$. Vậy

$$\mathbb{P}(X + Y > 3) = 1 - \mathbb{P}(X + Y \leqslant 3) = 1 - 0,647 = 0,353.$$

2.4.3 Phân phối chuẩn

Phân phối chuẩn tắc. Đại lượng ngẫu nhiên liên tục Z được gọi là đại lượng ngẫu nhiên có *phân phối chuẩn tắc* nếu hàm mật độ của nó là

$$p(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}}e^{\frac{-\pi^2}{2}}, \qquad \forall x \in \mathbb{R}.$$

Đồ thị hàm mật độ của Z có dạng

Đó là một đường cong đối xứng qua trục tung, có điểm cực đại tại $x = 0$. Các điểm uốn của đồ thị là $x = \pm 1$. Hàm phân phối của Z, kí hiệu bởi $\Phi(x)$, là

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

Rất tiếc rằng $\Phi(x)$ không biểu diễn được qua các hàm sơ cấp đã biết. Người ta đã lập bảng tính sẵn các giá trị của $\Phi(x)$.
Chú ý rằng nhiều khi người ta chỉ cho các giá trị của $\Phi(x)$ với $x > 0$. Với $x < 0$ ta sử dụng công thức: $\Phi(-x) = 1 - \Phi(x)$.
Chẳng hạn $\Phi(-0,62) = 1 - \Phi(0,62) = 1 - 0,7324 = 0,2676$.
Như vậy ta có

$$\mathbb{P}(Z < a) = \Phi(a), \quad \mathbb{P}(Z > a) = 1 - \Phi(a),$$

$$\mathbb{P}(a < Z < b) = \Phi(b) - \Phi(a).$$

Dễ thấy rằng $\mathbb{E}Z = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} xe^{-\frac{x^2}{2}} dx = 0$ (Vì hàm dưới dấu tích phân là hàm lẻ). Sử dụng công thức tích phân từng phần ta có

$$DZ = \mathbb{E}Z^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{2}} dx$$

$$= -\frac{1}{\sqrt{2\pi}} xe^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}} dx = 1.$$

Ta cũng dễ thấy mode và median của Z là 0.
Như vậy

$$\mathbb{E}Z = 0; \qquad DZ = 1; \qquad \mathrm{mod}\ Z = 0; \qquad m(Z) = 0.$$

Phân phối chuẩn. Đại lượng ngẫu nhiên liên tục X được gọi là có *phân phối chuẩn với tham số μ và σ^2* (trong đó $\sigma > 0$) nếu đại lượng ngẫu nhiên $Z = \dfrac{X-\mu}{\sigma}$ có phân phối chuẩn tắc. Khi đó ta kí hiệu $X \sim N(\mu, \sigma^2)$.

Ta hãy tìm hàm mật độ của đại lượng ngẫu nhiên X có phân phối chuẩn.

Giả sử $X \sim N(\mu, \sigma^2)$. Khi đó

$$X = \sigma Z + \mu,$$

trong đó Z có phân phối chuẩn tắc. Ta có

$$\mathbb{P}(X < x) = \mathbb{P}(\sigma Z + \mu < x) = \mathbb{P}(Z < \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma}).$$

Vậy

$$p_X(x) = \Phi'(\frac{x-\mu}{\sigma})\frac{1}{\sigma} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma}}.$$

Tiếp theo ta tính các tham số đặc trưng của X.

Ta có

$$\mathbb{E}X = \mathbb{E}(\sigma Z + \mu) = \sigma \mathbb{E}Z + \mu = \mu.$$

$$DX = \mathbb{E}(X-\mu)^2 = \mathbb{E}Z(\sigma^2 Z^2) = \sigma^2 \mathbb{E}Z^2 = \sigma^2.$$

Dễ thấy mode và median của X đều bằng μ.

Như vậy tham số μ là kỳ vọng, đồng thời cũng là mode và median của phân phối chuẩn $N(\mu, \sigma^2)$. Tham số σ là độ lệch chuẩn của nó.

Ta có thể tính được các xác suất liên quan tới X bằng cách biến đổi nó về một biến cố có liên quan tới Z rồi tra bảng, chẳng hạn

$\mathbb{P}(X < \alpha) = \mathbb{P}(Z < \dfrac{\alpha-\mu}{\sigma}) = \Phi(\dfrac{\alpha-\mu}{\sigma})$.

$\mathbb{P}(X > \alpha) = 1 - \mathbb{P}(X < \alpha) = 1 - \Phi(\dfrac{\alpha-\mu}{\sigma})$.

$\mathbb{P}(\alpha < X < \beta) = \mathbb{P}(\dfrac{\alpha-\mu}{\sigma} < Z < \dfrac{\beta-\mu}{\sigma}) = \Phi(\dfrac{\beta-\mu}{\sigma}) - \Phi(\dfrac{\alpha-\mu}{\sigma})$.

Ví dụ.

1. Giả sử X là đại lượng ngẫu nhiên có phân phối chuẩn với kỳ vọng $\mu = 2100$ và độ lệch chuẩn $\sigma = 200$. Hãy tìm

a. $\mathbb{P}(X > 2400)$.

b. $\mathbb{P}(1700 < X < 2200)$.

c. Xác định a để $\mathbb{P}(X > a) = 0,03$.

Giải.

a. $\mathbb{P}(X > 2400) = 1 - \Phi(\frac{2400 - 2100}{200}) = 1 - \Phi(1,5)$

$= 1 - 0,9332 = 0,0668.$

b. $\mathbb{P}(1700 < X < 2200) = \Phi(\frac{2200 - 2100}{200}) - \Phi(\frac{1700 - 2100}{200})$

$= \Phi(0,5) - \Phi(-2) = 0,6915 - (1 - 0,9772) = 0,6687.$

c. $\mathbb{P}(X > a) = 1 - \Phi(\frac{a - 2100}{200}) = 0,03$

$\Rightarrow \Phi(\frac{a - 2100}{200}) = 0,97.$

Tra ngược bảng $\Phi(x)$ ta tìm được $\Phi(1,881) = 0,97$.

Vậy $\frac{a - 2100}{200} = 1,881$. Từ đó $a = 2476,2$.

2. Trọng lượng của một gói đường (đóng bằng máy tự động) có phân phối chuẩn. Trong 1000 gói đường có 70 gói có trọng lượng lớn hơn 1015 g. Hãy ước lượng xem có bao nhiêu gói đường có trọng lượng ít hơn 1008g, biết rằng trọng lượng trung bình của 1000 gói đường là 1012g.

Giải. Theo giả thiết trọng lượng của một gói đường X có phân phối chuẩn với $\mu = 1012g$ và độ lệch chuẩn σ. Ta cần ước lượng σ.

Xác suất để trọng lượng gói đường lớn hơn $1015g$ là $\frac{70}{1000} = 0,07$. Vậy ta có

$$\mathbb{P}(X > 1015) = 0,07 \Leftrightarrow 1 - \Phi(\frac{1015 - 1012}{\sigma}) = 0,07.$$

Suy ra

$$\Phi(\frac{3}{\sigma}) = \Phi(1,476).$$

Từ đó $\sigma = \frac{3}{1,470} = 2,0325.$

Vậy

$$\mathbb{P}(X < 1008) = \Phi(\frac{1008 - 1012}{2,0325})$$
$$= \Phi(-1,968) = 1 - \Phi(1,968) = 0,0245.$$

Do đó trong 1000 gói đường sẽ có khoảng $1000.0,0245 = 24,5$ gói có trọng lượng ít hơn $1008g$.

2.4.4 Phân phối mũ

Định nghĩa. Đại lượng ngẫu nhiên liên tục X được gọi là có *phân phối mũ* với tham số $\lambda > 0$ và kí hiệu $X \sim E(\lambda)$, nếu nó có hàm mật độ

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & \text{nếu } x \geqslant 0, \\ 0 & \text{nếu } x < 0. \end{cases}$$

Đồ thị hàm mật độ của phân phối mũ có dạng sau **Hàm phân phối và các số đặc trưng**

Giả sử X có phân phối mũ với tham số λ. Hàm phân phối của X là

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad \text{với } x > 0$$

và

$$F(x) = 0 \quad \text{với } x \leqslant 0.$$

Kỳ vọng của X là

$$\mathbb{E}X = \lambda \int_0^\infty x e^{-\lambda x} dx = (-x e^{-\lambda x})\Big|_0^\infty + \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda}.$$

(Sử dụng tích phân từng phần).

Phương sai của X là $DX = \frac{1}{\lambda^2}$. Thật vậy, bằng cách tích phân từng phần ta được

$$\int_0^\infty x^2 \lambda e^{-\lambda x} dx = (-x^2 \lambda e^{-\lambda x})\Big|_0^\infty + 2\int_0^\infty x e^{-\lambda x} dx = \frac{2}{\lambda^2}.$$

Thành thử $DX = \dfrac{1}{\lambda^2}$ và do đó độ lệch chuẩn của X là $\sigma_X = \sqrt{DX} = \dfrac{1}{\lambda}$.
Ví dụ. Giả sử tuổi thọ (tính bằng năm) của một mạch điện tử trong máy tính là một đại lượng ngẫu nhiên có phân phối mũ với kỳ vọng là $6,25$. Thời gian bảo hành của mạch điện tử này là 5 năm. Hỏi có bao nhiêu phần trăm mạch điện tử bán ra phải thay thế trong thời gian bảo hành.
Giải. Ta có $\lambda = \dfrac{1}{\mathbb{E}X} = \dfrac{1}{6,25}$.

$$\mathbb{P}(X \leqslant 5) = 1 - e^{-5\lambda} = 1 - e^{-\frac{5}{6,25}} = 1 - e^{-0,8} \approx 1 - 0,449 \approx 0,5506.$$

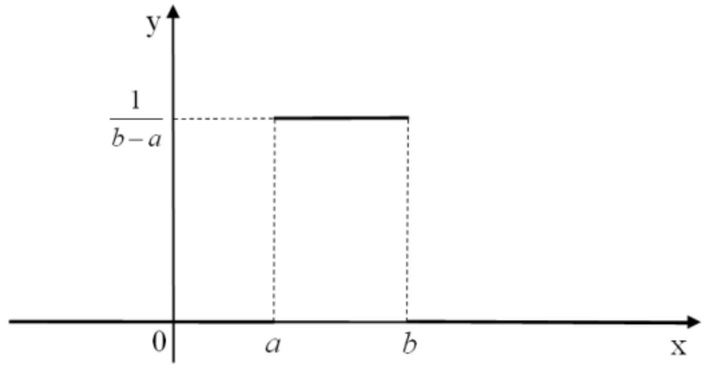
Vậy có khoảng 55% số mạch điện tử bán ra phải thay thế trong thời gian bảo hành.

2.4.5 Phân phối đều

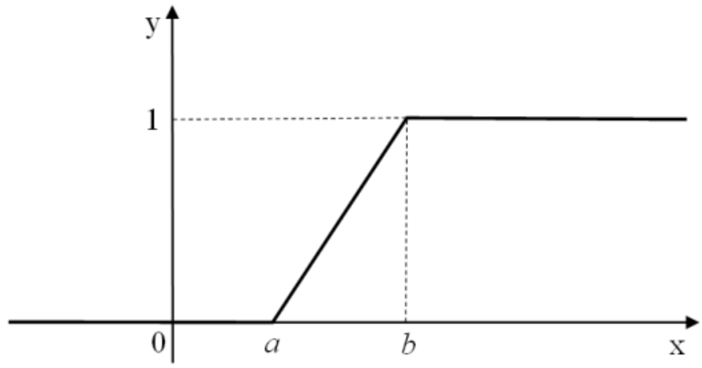
Định nghĩa. Đại lượng ngẫu nhiên liên tục X được gọi là có *phân phối đều* trên đoạn $[a, b]$ và kí hiệu là $X \sim U_{[a,b]}$, nếu hàm mật độ của nó cho bởi công thức

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

Hàm phân phối và các số đặc trưng.
Hàm phân phối $F(x)$ của đại lượng ngẫu nhiên X có phân phối đều trên đoạn $[a, b]$ được xác định như sau
$F(x) = 0$ nếu $x < a$;
$F(x) = \int_{-\infty}^{x} p(x)dx = \int_{a}^{x} \dfrac{dx}{b-a} = \dfrac{x-a}{b-a}$ nếu $a \leqslant x \leqslant b$
$F(x) = 1$ nếu $x > b$.
Đồ thị của hàm mật độ và hàm phân phối của X



Đồ thị của $p(x)$



Đồ thị của $F(x)$

Giả sử $(\alpha, \beta) \subset [a, b]$. Xác suất để X rơi vào (α, β) là

$$\mathbb{P}(\alpha < X < \beta) = \int_{\alpha}^{\beta} p(x)dx = \frac{\beta - \alpha}{b - a}.$$

Như vậy xác suất để X rơi vào một khoảng (α, β) chỉ phụ thuộc vào độ dài khoảng đó và tỉ lệ thuận với độ dài của khoảng.
Ta hãy tính một số tham số đặc trưng của phân phối đều.
Ta có

$$\mathbb{E}X = \int_{a}^{b} \frac{xdx}{b-a} = \frac{1}{b-a}(\frac{x^2}{2})\Big|_a^b = \frac{1}{b-a}\frac{(b^2-a^2)}{2} = \frac{a+b}{2}.$$

Như vậy kỳ vọng là trung điểm của đoạn $[a, b]$.

$$\begin{aligned} DX &= \int_a^b \frac{x^2dx}{b-a} - (\mathbb{E}X)^2 = \frac{1}{b-a}(\frac{x^3}{3})\Big|_a^b - (\frac{a+b}{2})^2 \\ &= \frac{b^2+ab+a^2}{3} - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}. \end{aligned}$$

Dễ thấy median của X là $m(X) = \dfrac{a+b}{2}$.

Mode của phân phối đều là bất cứ điểm nào của $[a, b]$.

Ví dụ. Đại lượng ngẫu nhiên X có phân phối đều trên đoạn $[-1, 2]$. Tìm hàm phân phối và hàm mật độ của $Y = X^2$.

Giải. Rõ ràng với $x \leqslant 0$ thì

$$F_Y(x) = \mathbb{P}(Y < x) = \mathbb{P}(X^2 < 0) = 0.$$

Với $x > 4$ thì

$$F_Y(x) = \mathbb{P}(X^2 < x) = 1.$$

Với $1 \leqslant x \leqslant 4$ thì

$$F_Y(x) = \mathbb{P}(X^2 < x) = \mathbb{P}(-\sqrt{x} < X < \sqrt{x}) = \frac{\sqrt{x}+1}{3} \qquad (\text{ vì } -\sqrt{x} < -1).$$

Với $0 \leqslant x \leqslant 1$ thì

$$F_Y(x) = \mathbb{P}(-\sqrt{x} < X < \sqrt{x}) = \frac{2\sqrt{x}}{3} \;\; (\text{ vì } \; -1 < -\sqrt{x}).$$

Tóm lại

$$F_Y(x) = \begin{cases} 0 & \text{nếu } x \leqslant 0, \\ \dfrac{2\sqrt{x}}{3} & \text{nếu } 0 \leqslant x \leqslant 1, \\ \dfrac{\sqrt{x}+1}{3} & \text{nếu } 1 \leqslant x \leqslant 4, \\ 1 & \text{nếu } x > 4. \end{cases}$$

Từ đó hàm mật độ của Y là

$$p_Y(x) = \begin{cases} \dfrac{1}{3\sqrt{x}} & \text{nếu } 0 \leqslant x \leqslant 1, \\ \dfrac{1}{6\sqrt{x}} & \text{nếu } 1 \leqslant x \leqslant 4, \\ 0 & \text{nếu } x \leqslant 0 \text{ hoặc } x \geqslant 4. \end{cases}$$

2.4.6 Phân phối χ^2 và phân phối Student

1. *Phân phối* χ^2 (*Khi bình phương*)

Đại lượng ngẫu nhiên X được gọi là có phân phối χ^2 (khi bình phương) với n bậc tự do, nếu hàm mật độ của nó có dạng

$$p(x) = \begin{cases} 0 & \text{nếu } x < 0, \\ Cx^{\frac{n-1}{2}} e^{-\frac{x}{2}} & \text{nếu } x \geqslant 0, \end{cases}$$

trong đó C là hằng số dương.

Vai trò của phân phối χ^2 được thể hiện qua định lý sau:

Định lý. Nếu $X_1, X_2, \ldots, X_n$ là các đại lượng ngẫu nhiên độc lập, có phân phối chuẩn tắc $N(0,1)$ thì

$$X_1^2 + X_2^2 + \cdots + X_n^2$$

có phân phối χ^2 với n bậc tự do.

2. *Phân phối Student*

Đại lượng ngẫu nhiên T được gọi là có phân phối Student với n bậc tự do, nếu hàm mật độ của nó có dạng

$$p(x) = C\big(1 + \frac{x^2}{n}\big)^{-\frac{n+1}{2}},$$

trong đó C là hằng số dương.

Vai trò của phân phối Student được thể hiện qua định lý sau:

Định lý. Nếu $X, X_1, X_2, \ldots, X_n$ là các đại lượng ngẫu nhiên độc lập, có phân phối chuẩn tắc $N(0, 1)$ thì

$$T = \frac{X}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} X_i^2}}$$

có phân phối Student với n bậc tự do.

Cùng với phân phối chuẩn, phân phối χ^2 và phân phối Student được sử dụng nhiều trong thống kê toán học.

2.4 Vectơ ngẫu nhiên

2.5.1 Định nghĩa và ví dụ

Từ trước đến nay, ta chỉ xét các đại lượng ngẫu nhiên riêng lẻ. Trong lý thuyết và ứng dụng, nhiều khi ta cần xét đồng thời một hệ thống có thứ tự gồm nhiều đại lượng ngẫu nhiên. Hệ thống như thế được gọi là vectơ ngẫu nhiên.

Định nghĩa. Giả sử $X_1, X_2, \ldots, X_n$ là các đại lượng ngẫu nhiên. Khi đó, vectơ $\overrightarrow{X} = (X_1, X_2, \ldots, X_n)$ được gọi là *vectơ ngẫu nhiên n chiều.*

Các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$ được gọi là các thành phần của vectơ ngẫu nhiên $\overrightarrow{X}$.

Trên thực tế, vectơ ngẫu nhiên n chiều dùng để nghiên cứu đồng thời n đặc trưng về lượng của các biến cố.

Ví dụ. Khi điều tra chất lượng học tập của một học sinh, nếu ta quan tâm đồng thời đến điểm toán X và điểm văn Y thì ta phải đề cập tới vectơ ngẫu nhiên hai chiều (X, Y). Nếu cần quan tâm đến điểm lý Z nữa thì phải xét vectơ ngẫu nhiên ba chiều (X, Y, Z), ...

Để tránh các ký hiệu cồng kềnh, từ nay về sau, ta chỉ xét vectơ ngẫu nhiên hai chiều.

Phân loại vectơ ngẫu nhiên.

Vectơ ngẫu nhiên (X, Y) được gọi là *vectơ ngẫu nhiên rời rạc* nếu cả hai đại lượng ngẫu nhiên X và Y đều là đại lượng ngẫu nhiên rời rạc.

Vectơ ngẫu nhiên (X, Y) được gọi là *vectơ ngẫu nhiên liên tục* nếu cả hai đại lượng ngẫu nhiên X và Y đều là đại lượng ngẫu nhiên liên tục.

Vectơ ngẫu nhiên (X, Y) được gọi là *vectơ ngẫu nhiên hỗn hợp* nếu một trong hai đại lượng ngẫu nhiên thành phần là đại lượng ngẫu nhiên rời rạc, còn đại lượng ngẫu nhiên kia là đại lượng ngẫu nhiên liên tục.

Trong phạm vi giáo trình này, chúng ta chỉ xét hai loại vectơ ngẫu nhiên: vectơ ngẫu nhiên rời rạc và vectơ ngẫu nhiên liên tục.

2.5. 2 Bảng phân phối của vectơ ngẫu nhiên rời rạc

Định nghĩa. Giả sử X và Y là hai đại lượng ngẫu nhiên rời rạc và

$$X(\Omega) = \{x_1, x_2, \ldots, x_m \ldots\}; \qquad Y(\Omega) = \{y_1, y_2, \ldots, y_n \ldots\}.$$

Ký hiệu $p_{ij} = \mathbb{P}(X = x_i, Y = y_j)$.

Khi đó bảng chữ nhật sau đây được gọi là bảng phân phối của vectơ ngẫu nhiên (X, Y).

X \ Y	y_1	y_2	...	y_j	...
x_1	p_{11}	p_{12}	...	p_{1j}	...
x_2	p_{21}	p_{22}	...	p_{2j}	...
...	...	...	...	...	...
x_i	p_{i1}	p_{i2}	...	p_{ij}	...
...	...	...	...	...	...

Tính chất

1. $\sum_{i,j} p_{ij} = 1$.

Tính chất này hiển nhiên.

2. $\mathbb{P}(X = x_i) = \sum_j p_{ij}; \mathbb{P}(Y = y_j) = \sum_i p_{ij}$.

Thật vậy

$$\mathbb{P}(X = x_i) = \mathbb{P}(X = x_i, Y = y_1) + \mathbb{P}(X = x_i, Y = y_2) +$$

$$\cdots + \mathbb{P}(X = x_i, Y = y_n) + \cdots = \sum_j p_{ij}.$$

Tương tự

$$\mathbb{P}(Y = y_j) = \sum_i p_{ij}.$$

Như vậy, nếu biết bảng phân phối của vectơ (X, Y) thì cũng biết được bảng phân phối của các đại lượng ngẫu nhiên X và Y. Điều ngược lại nói chung không đúng. Tuy nhiên, ta có

3. *Nếu các đại lượng ngẫu nhiên thành phần X và Y độc lập thì khi biết bảng phân phối của chúng, có thể suy ra được bảng phân phối của vectơ ngẫu nhiên (X, Y).*

Thật vậy

$$p_{ij} = \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j).$$

4. $\mathbb{E}(X + Y) = \sum_{i,j}(x_i + y_j)p_{ij}, \quad \mathbb{E}(XY) = \sum_{i,j}(x_i y_j)p_{ij}$.

Tính chất này được suy ra ngay từ công thức tính kỳ vọng của đại lượng ngẫu nhiên rời rạc.

Ví dụ. Gieo ba đồng tiền cân đối A, B và C. Giả sử X, Y là các đại lượng ngẫu nhiên được xác định như sau:

X: Số mặt ngửa trên các đồng tiền A và B.

Y: Số mặt ngửa trên cả ba đồng tiền A, B và C.

Hãy lập bảng phân phối của vectơ ngẫu nhiên (X, Y).

Giải. Rõ ràng

$$X(\Omega) = \{0, 1, 2\}, \qquad Y(\Omega) = \{0, 1, 2, 3\}.$$

Các kết quả của phép thử và giá trị tương ứng của X và Y được cho bởi bảng sau (N: Ngửa, S: Sấp).

A	B	C	X	Y
N	N	N	2	3
N	N	S	2	2
N	S	N	1	2
N	S	S	1	1
S	N	N	1	1
S	N	S	1	2
S	S	N	0	0
S	S	S	0	0

Vậy

$$\mathbb{P}(X=0,Y=0)=\frac{1}{8},\ \ \mathbb{P}(X=0,Y=1)=\frac{1}{8},$$
$$\mathbb{P}(X=0,Y=2)=0,\ \mathbb{P}(X=0,Y=3)=0,$$
$$\mathbb{P}(X=1,Y=0)=0,\ \mathbb{P}(X=1,Y=1)=\frac{2}{8},$$
$$\mathbb{P}(X=1,Y=2)=\frac{1}{8},\ \mathbb{P}(X=1,Y=3)=0,$$
$$\mathbb{P}(X=2,Y=0)=0,\ \mathbb{P}(X=2,Y=1)=0,$$
$$\mathbb{P}(X=2,Y=2)=\frac{1}{8},\ \ \mathbb{P}(X=2,Y=3)=\frac{1}{8}.$$

Do đó, bảng phân phối xác suất của vectơ (X, Y) là

X \ Y	0	1	2	3
0	1/8	1/8	0	0
1	0	2/8	2/8	0
2	0	0	1/8	1/8

2.5.3 Hàm phân phối của vectơ ngẫu nhiên

Định nghĩa. Giả sử (X, Y) là vectơ ngẫu nhiên hai chiều. Khi đó hàm hai biến $F(x, y)$ xác định bởi đẳng thức

$$F(x, y) = \mathbb{P}(X < x, Y < y)$$

được gọi là *hàm phân phối* của (X, Y).

Hàm phân phối của vectơ ngẫu nhiên (X, Y) còn được gọi là *hàm phân phối đồng thời* của các đại lượng ngẫu nhiên X và Y.

Tính chất

1. *$F(x, y)$ là hàm không giảm theo từng đối số, tức là*

$$F(x_1, y) \geqslant F(x_2, y) \text{ nếu } x_1 > x_2;$$

$$F(x, y_1) \geqslant F(x, y_2) \text{ nếu } y_1 > y_2.$$

2. $F(-\infty, -\infty) = \lim_{\substack{x \to -\infty \\ y \to -\infty}} F(x, y) = 0$;
$F(+\infty, +\infty) = \lim_{\substack{x \to +\infty \\ y \to +\infty}} F(x, y) = 1$.
3. $F(x, +\infty) = \lim_{y \to +\infty} F(x, y) = F_X(x)$;
$F(y, +\infty) = \lim_{x \to +\infty} F(x, y) = F_Y(y)$.
Trong đó $F_X(x)$, $F_Y(y)$ là các hàm phân phối của các đại lượng ngẫu nhiên X và Y ($F_X(x)$, $F_Y(y)$ được gọi là hàm phân phối biên duyên của các đại lượng ngẫu nhiên X và Y tương ứng).
4. $\mathbb{P}(a \leqslant X < b, c \leqslant Y < d) = F(b, d) - F(b, c) - F(a, d) + F(a, c)$.
5. Điều kiện cần và đủ để các đại lượng ngẫu nhiên X và Y độc lập là

$$F(x, y) = F_X(x) F_Y(y).$$

Việc chứng minh các tính chất trên khá dễ dàng, xin dành cho bạn đọc.

Ví dụ. Cho vectơ ngẫu nhiên (X, Y) có hàm phân phối

$$F(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y} & \text{nếu } x \geqslant 0, y \geqslant 0, \\ 0 & \text{nếu trái lại.} \end{cases}$$

Tìm hàm phân phối $F_X(x)$ của X.

Giải.

Ta có

$$F_X(x) = \lim_{y\to\infty} F(x,y) = \begin{cases} 1 - e^{-x} & \text{nếu } x \geqslant 0 \\ 0 & \text{nếu } x < 0. \end{cases}$$

Do đó X là đại lượng ngẫu nhiên có phân phối mũ với tham số $\lambda = 1$.

2.5.4 Hàm mật độ của vectơ ngẫu nhiên liên tục

Định nghĩa. Giả sử (X, Y) là vectơ ngẫu nhiên liên tục, có hàm phân phối $F(x, y)$. Khi đó hàm hai biến $p(x, y) \geqslant 0$ thoả mãn đẳng thức

$$F(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} p(x,y)dxdy$$

được gọi là *hàm mật độ xác suất* của vectơ ngẫu nhiên (X, Y).

Hàm mật độ xác suất của vectơ ngẫu nhiên (X, Y) còn được gọi là *hàm mật độ đồng thời* của các đại lượng ngẫu nhiên X và Y.

Tính chất

1. $\int_{-\infty}^{x} \int_{-\infty}^{y} p(x,y)dxdy = 1$.

Thật vậy

$$\int_{-\infty}^{x} \int_{-\infty}^{y} p(x,y)dxdy = \mathbb{P}(X < +\infty, Y < +\infty) = 1.$$

2. $p_X(x) = \int_{-\infty}^{+\infty} p(x,y)dy$; $p_Y(y) = \int_{-\infty}^{+\infty} p(x,y)dx$.

Thật vậy, ta có

$$\int_{-\infty}^{x} (\int_{-\infty}^{+\infty} p(x,y)dy)dx = \mathbb{P}(X < x, Y < +\infty) = \mathbb{P}(X < x).$$

Do đó, hàm số

$$p_X(x) = \int_{-\infty}^{+\infty} p(x,y)dy$$

là hàm mật độ của đại lượng ngẫu nhiên X.

Tương tự, hàm số

$$p_Y(y) = \int_{-\infty}^{+\infty} p(x,y)dx$$

là hàm mật độ của đại lượng ngẫu nhiên Y.

Như vậy, nếu biết hàm mật độ xác suất của (X, Y) thì cũng biết hàm mật độ xác suất của X và Y. Điều ngược lại không đúng, tuy nhiên ta có

3. *Nếu X và Y là các đại lượng ngẫu nhiên độc lập thì*

$$p(x,y) = p_X(x)p_Y(y).$$

Thật vậy, ta có

$$\begin{aligned}\mathbb{P}(X < x, Y < y) &= \mathbb{P}(X < x)\mathbb{P}(Y < y)\\ &= \int_{-\infty}^{x} p_X(x)dx \int_{-\infty}^{y} p_Y(y)dy\\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} p_X(x)p_Y(y)dxdy.\end{aligned}$$

Suy ra

$$p(x,y) = p_X(x)p_Y(y)$$

là hàm mật độ của (X, Y).

Đẳng thức trên cũng là điều kiện đủ để các đại lượng ngẫu nhiên X và Y độc lập (hãy tự chứng minh điều đó).

Như vậy, nếu X và Y là các đại lượng ngẫu nhiên độc lập thì khi biết hàm mật độ xác suất của X và Y, ta cũng sẽ xác định được hàm mật độ xác suất của vectơ (X, Y).

4. $\mathbb{P}(a \leqslant X < b, c \leqslant Y < d) = \mathbb{P}(a < X < b, c < Y < d) = \int_a^b \int_c^d p(x,y)dxdy$.

Tổng quát, nếu $D \subset \mathbb{R}^2$ là một miền đo được bất kỳ thì

$$\mathbb{P}((X,Y) \in D) = \iint\limits_D p(x,y)dxdy.$$

Tính chất này suy trực tiếp từ tính chất tương ứng của hàm phân phối.

5. $\quad \mathbb{E}(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyp(x,y)dxdy$,

$\mathbb{E}(X+Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x+y)p(x,y)dxdy$.

Chúng ta công nhận tính chất này.

Ví dụ. Cho vectơ ngẫu nhiên (X,Y) có hàm mật độ

$$p(x,y) = \begin{cases} C(x^2 + \frac{1}{2}xy) & \text{nếu } 0 < x < 1, 0 < y < 2, \\ 0 & \text{nếu trái lại.} \end{cases}$$

Hãy tìm hằng số C và tính $\mathbb{E}(XY)$.

Giải. Ta có

$$1 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x,y)dxdy = C \int_0^1 \int_0^2 (x^2 + \frac{xy}{2})dxdy = C\frac{7}{6}.$$

Từ đó $C = \dfrac{6}{7}$.

$$\begin{aligned} \mathbb{E}(XY) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyp(x,y)dxdy = C \int_0^1 \int_0^2 xy(x^2 + \frac{xy}{2})dxdy \\ &= C \int_0^1 \int_0^2 (x^3y + \frac{x^2y^2}{2})dxdy = \frac{17}{21}. \end{aligned}$$

2.5.5 Phân phối có điều kiện

Trường hợp rời rạc

Giả sử (X,Y) là vectơ ngẫu nhiên rời rạc

$$X(\Omega) = \{x_1; x_2; \dots; x_m; \dots\}; \qquad Y(\Omega) = \{y_1; y_2; \dots; y_n; \dots\}.$$

Với mỗi x_i, y_j $(i = 1, 2, \dots, m;\ j = 1, 2, \dots, n)$, đặt

$$p(x_i/y_j) = \mathbb{P}[(X = x_i)/(Y = y_j)];\ \ p(y_j/x_i) = \mathbb{P}[(Y = y_j)/(X = x_i)].$$

Khi đó, với mỗi y_j cố định, bảng

X	x_1	x_2	$\cdots$	x_m	$\cdots$
$p(x_i/y_j)$	$p(x_1/y_j)$	$p(x_2/y_j)$	$\cdots$	$p(x_m/y_j)$	$\cdots$

gọi là *bảng phân phối có điều kiện* của đại lượng ngẫu nhiên X với điều kiện $Y = y_j$.

Bảng phân phối có điều kiện của đại lượng ngẫu nhiên X với điều kiện $Y = y_j$ cho biết phân phối xác suất của X trong điều kiện giả thiết rằng biến cố $(Y = y_j)$ đã xảy ra.

Tương tự, ta có thể định nghĩa bảng phân phối có điều kiện của đại lượng ngẫu nhiên Y với điều kiện $X = x_i$.

Nếu biết bảng phân phối của vectơ ngẫu nhiên (X, Y) thì có thể tìm được bảng phân phối có điều kiện của các đại lượng ngẫu nhiên thành phần. Bởi vì

$$p(x_i/y_i) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)};\ \ p(y_i/x_i) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(X = x_i)}.$$

Ví dụ. Vectơ ngẫu nhiên (X, Y) được cho bởi bảng phân phối

Y \ X	x_1	x_2	x_3
y_1	0,10	0,30	0,20
y_2	0,06	0,18	0,16

Tìm bảng phân phối có điều kiện của đại lượng ngẫu nhiên X với điều kiện đại lượng ngẫu nhiên Y nhận giá trị y_1.

Giải. Ta có

$$p(x_1/y_1) = \frac{\mathbb{P}(X = x_1, Y = y_1)}{\mathbb{P}(Y = y_1)} = \frac{0,1}{0,1 + 0,3 + 0,2} = \frac{1}{6}.$$

Tương tự

$$p(x_2/y_1) = \frac{0,3}{0,6} = \frac{1}{2}; \qquad p(x_3/y_1) = \frac{0,2}{0,6} = \frac{1}{3}.$$

Từ đó có phân phối có điều kiện của đại lượng ngẫu nhiên X với điều kiện $Y = y_1$

X	x_1	x_2	x_3
$p(x_i/y_1)$	1/6	1/2	1/3

Trường hợp liên tục

Giả sử (X, Y) là vectơ ngẫu nhiên liên tục có hàm mật độ là $p(x, y)$. Các đại lượng ngẫu nhiên thành phần X và Y tương ứng có hàm mật độ là $p_X(x)$ và $p_Y(y)$. Khi đó các hàm số

$$p(x/y) = \frac{p(x,y)}{p_Y(y)} \text{ và } p(y/x) = \frac{p(x,y)}{p_X(x)}$$

tương ứng được gọi là *hàm mật độ có điều kiện* của đại lượng ngẫu nhiên X với điều kiện đại lượng ngẫu nhiên Y lấy giá trị $Y = y$ và hàm mật độ có điều kiện của đại lượng ngẫu nhiên Y với điều kiện đại lượng ngẫu nhiên X lấy giá trị $X = x$.

Ví dụ. Vectơ ngẫu nhiên (X, Y) có hàm mật độ

$$p(x,y) = \begin{cases} \dfrac{1}{\pi r^2} & \text{nếu } x^2 + y^2 \leqslant r^2, \\ 0 & \text{nếu } x^2 + y^2 > r^2. \end{cases}$$

Dễ dàng tính được

$$p(x/y) = \begin{cases} \dfrac{1}{2\sqrt{r^2 - y^2}} & \text{nếu } |x| \leqslant \sqrt{r^2 - x^2}, \\ 0 & \text{nếu } |x| > \sqrt{r^2 - x^2}. \end{cases}$$

$$p(y/x) = \begin{cases} \dfrac{1}{2\sqrt{r^2 - y^2}} & \text{nếu } |y| \leqslant \sqrt{r^2 - x^2}, \\ 0 & \text{nếu } |y| > \sqrt{r^2 - x^2}. \end{cases}$$

2.5.6 Các số đặc trưng của vectơ ngẫu nhiên

Vectơ kỳ vọng. Giả sử $(X_1, X_2, \ldots, X_n)$ là vectơ ngẫu nhiên n chiều. Khi đó, nếu với mọi i đều tồn tại $\mathbb{E}X_i$ thì vectơ $(\mathbb{E}X_1, \mathbb{E}X_2, \ldots, \mathbb{E}X_n)$ được gọi là *vectơ kỳ vọng* của $(X_1, X_2, \ldots, X_n)$.

Covarian (Moment tương quan). Giả sử $(X_1, X_2, \ldots, X_n)$ là vectơ ngẫu nhiên. Khi đó, với mỗi cặp i, j; số

$$cov(X_i, X_j) = \mathbb{E}[(X_i - \mathbb{E}X_i)(X_j - \mathbb{E}X_j)] = \lambda_{i,j}$$

(nếu tồn tại) được gọi là *Covarian (hay Moment tương quan)* của X_i, X_j. Ma trận

$$\lambda = (\lambda_{i,j}) = \begin{bmatrix} \lambda_{1,1} & \lambda_{1,2} & \ldots & \lambda_{1,n} \\ \lambda_{2,1} & \lambda_{2,2} & \ldots & \lambda_{2,n} \\ \ldots & \ldots & \ldots & \ldots \\ \lambda_{n,1} & \lambda_{n,2} & \ldots & \lambda_{n,n} \end{bmatrix}$$

được gọi là ma trận Covarian của $(X_1, X_2, \ldots, X_n)$.

Dễ thấy, ma trận Covarian là ma trận đối xứng, xác định dương.

Hệ số tương quan. Giả sử (X, Y) là vectơ ngẫu nhiên. Khi đó số

$$\rho(X, Y) = \frac{cov(X, Y)}{\sqrt{DXDY}} = \frac{\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]}{\sqrt{\mathbb{E}(X - \mathbb{E}X)^2 \mathbb{E}(Y - \mathbb{E}Y)^2}}$$

được gọi là *hệ số tương quan* của các đại lượng ngẫu nhiên X và Y.

Nếu $\rho(X, Y) = 0$ thì ta nói X và Y không tương quan.

Hệ số tương quan có những tính chất sau đây:

i) $D(X + Y) = DX + DY$ khi và chỉ khi $\rho(X, Y) = 0$.

ii) $-1 \leqslant \rho(X, Y) \leqslant 1$.

iii) $|\rho(X,Y)| = 1$ khi và chỉ khi X và Y phụ thuộc tuyến tính, tức là tồn tại các số a, b, c sao cho

$$aX + bY + c = 0.$$

Chứng minh.

$i)$ Suy ra từ đẳng thức $D(X+Y) = DX + DY + 2cov(X,Y)$.

$ii)$ Suy từ bất đẳng thức Schwarz $|cov(X,Y)| \leqslant \sqrt{DXDY}$.

$iii)$ Trước hết ta thấy ngay rằng, nếu X, Y phụ thuộc tuyến tính, thì bằng cách tính trực tiếp ta có $|\rho(X,Y)| = 1$. Ngược lại, giả sử $|\rho(X,Y)| = 1$, chẳng hạn $\rho(X,Y) = 1$, ta có

$$D\left(\frac{X - \mathbb{E}X}{\sqrt{DX}} - \frac{Y - \mathbb{E}Y}{\sqrt{DY}}\right) = 2(1 - \rho(X,Y)) = 0$$

Do đó

$$\frac{X - \mathbb{E}X}{\sqrt{DX}} - \frac{Y - \mathbb{E}Y}{\sqrt{DY}} = c \text{ (hằng số)}$$

Từ đó suy ra điều phải chứng minh.

Qua các tính chất trên, ta thấy $\rho(X,Y)$ biểu thị mức độ phụ thuộc tuyến tính giữa X và Y.

2.5 Một số định lý giới hạn

Trong phần này, chúng ta sẽ nghiên cứu một số định lý giới hạn được sử dụng nhiều trong thống kê. Một số kết quả sẽ được công nhận mà không chứng minh chi tiết, độc giả quan tâm có thể tìm đọc các chứng minh này trong các tài liệu tham khảo.

2.5.1 Các dạng hội tụ

Định nghĩa. Ta nói dãy đại lượng ngẫu nhiên $(X_n,\ n \geq 1)$ hội tụ đến đại lượng ngẫu nhiên X (khi $n \to \infty$)

- *Hầu chắc chắn* nếu $\mathbb{P}(\lim\limits_{n\to\infty} |X_n - X| = 0) = 1$.
  Kí hiệu $X_n \xrightarrow{h.c.c} X$.

- *Theo xác suất* nếu với mọi $\varepsilon > 0$ thì

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0.$$

  Kí hiệu $X_n \xrightarrow{P} X$.

- *Theo trung bình cấp p* $(p > 0)$ nếu $\lim\limits_{n\to\infty} \mathbb{E}|X_n - X|^p = 0$.
  Kí hiệu $X_n \xrightarrow{\mathcal{L}_p} X$.

- *Yếu (theo phân phối)* nếu

$$\lim_{n\to\infty} F_n(x) = F(x), \qquad \forall x \in C(F),$$

trong đó $F_n(x)$ và $F(x)$ tương ứng là hàm phân phối của các đại lượng ngẫu nhiên X_n và X; $C(F)$ là tập hợp các điểm mà tại đó $F(x)$ liên tục. Ký hiệu $X_n \xrightarrow{D} X$.

Hội tụ hầu chắc chắn còn được gọi là *hội tụ với xác suất* 1; hội tụ theo trung bình cấp p còn được gọi là *hội tụ trong $\mathcal{L}_p$*.

Ví dụ 1. Cho $a \in \mathbb{R}$ và X_n là đại lượng ngẫu nhiên rời rạc nhận các giá trị 0 và a với các xác suất tương ứng là $1 - \dfrac{1}{n}$ và $\dfrac{1}{n}$. Khi đó $X_n \xrightarrow{P} 0$ và $X_n \xrightarrow{\mathcal{L}_2} 0$ (khi $n \to \infty$).

Thật vậy, ta có với mọi $\varepsilon > 0$

$$\begin{aligned} 0 \leqslant \mathbb{P}\big(|X_n - 0| > \varepsilon\big) &= \mathbb{P}\big(|X_n| > \varepsilon\big) \\ &\leqslant \mathbb{P}\big(|X_n| = |a|\big) \\ &= \mathbb{P}(X_n = a) \\ &= \frac{1}{n} \to 0, \text{ khi } n \to \infty. \end{aligned}$$

Do đó $\lim\limits_{n\to\infty} \mathbb{P}\big(|X_n - 0| > \varepsilon\big) = 0, \forall\varepsilon > 0$, nên $X_n \xrightarrow{P} 0$.
Mặt khác, vì $0 \leqslant \mathbb{E}|X_n - 0|^2 = \mathbb{E}|X_n|^2 = |a|^2 \cdot \dfrac{1}{n} \to 0$ khi $n \to \infty$ nên $X_n \xrightarrow{\mathcal{L}_2} 0$.

Ví dụ 2. Giả sử X là đại lượng ngẫu nhiên rời rạc xác định bởi

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

Dãy đại lượng ngẫu nhiên $(X_n, n \geq 1)$ được xác định như sau

$$X_n = \begin{cases} X & \text{nếu } n \text{ chẵn}, \\ -X & \text{nếu } n \text{ lẻ}. \end{cases}$$

Rõ ràng

$$F_n(x) = F(x) \qquad (\forall x \in \mathbb{R}).$$

Suy ra

$$X_n \xrightarrow{D} X.$$

Tuy nhiên

$$X_n \overset{P}{\not\to} X \qquad (\text{khi } n \to \infty).$$

Thật vậy, với $n = 2m + 1$, ta có

$$\mathbb{P}(|X_{2m+1} - X| > 1) = \mathbb{P}(|2X| > 1) = \mathbb{P}(|X| > \frac{1}{2}) = 1 \not\to 0$$

(khi $n \to \infty$).

Do đó

$$X_{2m+1} \overset{P}{\not\to} 0 \qquad (\text{khi } m \to \infty).$$

Suy ra

$$X_n \overset{P}{\not\to} 0 \qquad (\text{khi } n \to \infty).$$

Tính chất

Định lý 1. $X_n \xrightarrow{h.c.c} X$ *khi và chỉ khi với mọi* $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(\sup_{m\geq n} |X_m - X| > \varepsilon) = 0.$$

Hệ quả 1. *Nếu chuỗi*

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - X| > \varepsilon)$$

hội tụ với mọi $\varepsilon > 0$*, thì*

$$X_n \to X \ h.c.c \ (khi\ n \to \infty).$$

Chứng minh. Điều khẳng định này rút ra từ bất đẳng thức

$$\mathbb{P}(\sup_{m\geqslant n} |X_m - X| > \varepsilon\}) \leqslant \sum_{m=n}^{\infty} \mathbb{P}(|X_m - X| > \varepsilon\}).$$

Trong trường hợp $(X_n,\ n \geq 1)$ là dãy đại lượng ngẫu nhiên độc lập và X là hằng số thì điều ngược lại cũng đúng, cụ thể là ta có

Định lý 2. *Nếu* $(X_n,\ n \geq 1)$ *là dãy đại lượng ngẫu nhiên độc lập và*

$$X_n \to C \ h.c.c \ (khi\ n \to \infty)$$

thì chuỗi

$$\sum_{n=1}^{\infty} \mathbb{P}(|X_n - C| > \varepsilon)$$

hội tụ với mọi $\varepsilon > 0$.

Chứng minh. Đặt $A_n = (|X_n - C| > \varepsilon)$ và chú ý rằng

$$\overline{\lim} A_n \subset (X_n \nrightarrow C).$$

Từ đó, áp dụng Bổ đề Borel- Cantelli được điều phải chứng minh.

Định lý 3. *Nếu $X_n \xrightarrow{h.c.c} X$ hoặc $X_n \xrightarrow{\mathcal{L}_r} X$ thì $X_n \xrightarrow{P} X$.*

Chứng minh. i) Giả sử $X_n \xrightarrow{h.c.c} X$. Khi đó, với mọi $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(\sup_{m\geq n} |X_m - X| > \varepsilon) = 0.$$

Mặt khác

$$0 \leqslant \mathbb{P}(|X_n - X| > \varepsilon) \leqslant \mathbb{P}(\sup_{m\geq n} |X_m - X| > \varepsilon).$$

Nên

$$\lim_{n\to\infty} \mathbb{P}(|X_m - X| > \varepsilon) = 0 \ \ (\forall \varepsilon > 0).$$

Do đó $X_n \xrightarrow{P} X$ khi $n \to \infty$.

ii) Nếu $X_n \xrightarrow{\mathcal{L}_r} X$ thì $X_n \xrightarrow{P} X$. Thật vậy, áp dụng bất đẳng thức Markov cho đại lượng ngẫu nhiên $|X_n - X|$, ta có, với mọi $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) \leqslant \frac{\mathbb{E}|X_n - X|^r}{\varepsilon^r} \to 0.$$

Do đó $X_n \xrightarrow{P} X$ khi $n \to \infty$. Đó là điều cần chứng minh.

Định lý 4. *Nếu $X_n \xrightarrow{P} X$. thì $X_n \xrightarrow{D} X$.*

Chứng minh. Giả sử $x \in C(F)$; $x' < x$. Vậy thì, do

$$X_n \xrightarrow{P} X \qquad \text{(khi } n \to \infty)$$

nên

$$\lim_{n\to\infty} \mathbb{P}(X < x', X_n \geqslant x) \leqslant \lim_{n\to\infty} \mathbb{P}(|X_n - X| > x - x') = 0.$$

Mặt khác

$$F(x') = \mathbb{P}(X < x') \leqslant \mathbb{P}(X_n < x) + \mathbb{P}(X < x', X_n \geqslant x).$$

Suy ra

$$F(x') \leqslant \underline{\lim} F_n(x) \qquad (\forall x' < x).$$

Chứng minh tương tự, ta được

$$F(x'') \geqslant \overline{\lim} F_n(x) \qquad (\forall x'' > x).$$

Khi $x' \uparrow x$; $x'' \downarrow x$ thì $F(x')$ và $F(x'')$ hội tụ về $F(x)$ (vì $x \in C(F)$) nên ta có

$$F(x) = \lim_{n\to\infty} F_n(x) \qquad (\forall x \in C(F)).$$

Qua định lý trên và ví dụ trước đó, ta thấy hội tụ theo phân phối nói chung thực sự yếu hơn hội tụ theo xác suất. Tuy nhiên trong trường hợp đại lượng ngẫu nhiên giới hạn là hằng số thì ta có

Định lý 5. *Nếu* $X_n \xrightarrow{D} X$ *và* $\mathbb{P}(X = C) = 1$ *thì* $X_n \xrightarrow{P} X$..

Chứng minh. Ta có

$$F_X(x) = \begin{cases} 0 & \text{nếu } x \leqslant C, \\ 1 & \text{nếu } x > C. \end{cases}$$

Do đó

$$C(F_X) = R \backslash \{C\}.$$

Giả sử

$$X_n \xrightarrow{D} X \qquad (\text{khi } n \to \infty).$$

Khi đó, với mọi $\varepsilon > 0$

$$\mathbb{P}(|X_n - X| > \varepsilon) = \mathbb{P}(|X_n - C| > \varepsilon) = 1 - \mathbb{P}(|X_n - C| \leqslant \varepsilon)$$

$$= 1 - \mathbb{P}(C - \varepsilon \leqslant X_n \leqslant C + \varepsilon)$$
$$\leqslant 1 - (F_{X_n}(C+\varepsilon) - F_{X_n}(C-\varepsilon)) \to 1 - (F_X(C+\varepsilon) - F_X(C-\varepsilon))$$
$$= 1 - (1 - 0) = 0 \qquad \text{(khi } n \to \infty).$$

Suy ra

$$X_n \xrightarrow{P} X \qquad \text{(khi } n \to \infty).$$

Định lý 6. *Nếu* $X_n \xrightarrow{P} X$ *thì tồn tại dãy con* $(X_{n_k}; k \geq 1) \subset (X_n,\ n \geq 1)$ *sao cho* $X_{n_k} \xrightarrow{h.c.c} X$.

Chúng ta công nhận không chứng minh định lý này.

2.5.2 Luật số lớn

Định nghĩa. Cho dãy $X_1, X_2, \ldots, X_n, \ldots$ các đại lượng ngẫu nhiên bất kỳ có kỳ vọng $\mathbb{E}X_i = a_i$ $(i = 1, 2, \ldots)$. Dãy $(X_n,\ n \geq 1)$ gọi là tuân theo *luật số lớn* nếu

$$\frac{X_1 + \cdots + X_n}{n} - \frac{a_1 + \cdots + a_n}{n} \xrightarrow{\mathrm{P}} 0 \qquad \text{(khi } n \to \infty).$$

Dãy $(X_n,\ n \geq 1)$ gọi là tuân theo *luật số lớn tổng quát* nếu tồn tại dãy số $(b_n),\ 0 < b_n \uparrow \infty$ sao cho

$$\frac{X_1 + \cdots + X_n}{b_n} - \frac{a_1 + \cdots + a_n}{b_n} \xrightarrow{\mathrm{P}} 0 \qquad \text{(khi } n \to \infty).$$

Trước khi thiết lập luật số lớn, ta cần định lý sau đây

Định lý 1. (Bất đẳng thức Chebyshev). Giả sử X là đại lượng ngẫu nhiên bất kỳ. Khi đó nếu tồn tại DX thì với mọi $\varepsilon > 0$, ta có

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) \leqslant \frac{DX}{\varepsilon^2}\,.$$

Chứng minh. áp dụng bất đẳng thức Markov cho đại lượng ngẫu nhiên $Y = |X - \mathbb{E}X|^2 \geqslant 0$ ta được

$$\mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon) = \mathbb{P}(Y \geq \varepsilon^2) \leqslant \frac{\mathbb{E}Y}{\varepsilon^2} = \frac{DX}{\varepsilon^2}.$$

Đó là điều cần chứng minh.

Định lý 2. (Markov) *Nếu* $(X_n,\ n \geq 1)$ *là dãy đại lượng ngẫu nhiên độc lập đôi một và thoả mãn điều kiện*

$$\frac{1}{n^2}\sum_{i=1}^{n} DX_i \to 0 \qquad (khi\ n \to \infty)$$

thì $(X_n,\ n \geq 1)$ *tuân theo luật số lớn.*

Chứng minh. Đặt

$$\overline{X}_n = \frac{\sum X_i}{n}.$$

Ta có

$$\mathbb{E}\overline{X}_n = \frac{a_1 + \cdots + a_n}{n}$$

và

$$D\overline{X}_n = \frac{1}{n^2}D(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} DX_i.$$

Từ đó, áp dụng bất đẳng thức Chebyshev ta có điều phải chứng minh.

Ví dụ. Cho dãy $(X_n,\ n \geq 1)$ các đại lượng ngẫu nhiên độc lập đôi một xác định như sau

$$\mathbb{P}(X_k = \sqrt{k}) = \frac{1}{2\sqrt{k}}, \quad P(X_k = -\sqrt{k}) = \frac{1}{2\sqrt{k}}, \quad P(X_k = 0) = 1 - \frac{1}{\sqrt{k}}\ .$$

Khi đó, dãy $(X_n,\ n \geq 1)$ tuân theo luật số lớn.

Thật vậy, ta có

$$\mathbb{E}X_k = 0, \qquad DX_k = k\frac{1}{2\sqrt{k}} + k\frac{1}{2\sqrt{k}} = \sqrt{k}\ .$$

Do đó

$$\frac{1}{n^2}\sum_{k=1}^{n} DX_k = \frac{1}{n^2}\sum_{k=1}^{n}\sqrt{k} < \frac{1}{n^2}(\sqrt{n} + \cdots + \sqrt{n}) = \frac{1}{\sqrt{n}}\ .$$

Suy ra

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n} DX_k = 0.$$

Từ định lý Markov, ta kết luận dãy $(X_n,\ n \geq 1)$ tuân theo luật số lớn.

Trong trường hợp dãy đại lượng ngẫu nhiên $(X_n,\ n \geq 1)$ độc lập đôi một có cùng kỳ vọng, ta có hệ quả sau

Hệ quả 1. *Giả sử $(X_n,\ n \geq 1)$ là dãy đại lượng ngẫu nhiên độc lập đôi một sao cho $\mathbb{E}X_i = a$ và $DX_i \leqslant C$ với mọi $i = 1, 2, \ldots$ Khi đó $(X_n,\ n \geq 1)$ tuân theo luật số lớn và trung bình cộng $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$ hội tụ theo xác suất tới a.*

Chứng minh. Ta có

$$\mathbb{E}\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} EX_i = a,$$

$$D\overline{X_n} = \frac{1}{n^2}\sum_{i=1}^{n} DX_i \leqslant \frac{nC}{n^2} = \frac{C}{n}.$$

áp dụng định lý Markov được điều phải chứng minh.

Từ hệ quả trên, ta suy ra ngay luật số lớn Chebyshev sau đây

Hệ quả 2. (Luật số lớn Chebyshev) *Giả sử $(X_n,\ n \geq 1)$ là dãy các đại lượng ngẫu nhiên độc lập đôi một, cùng phân phối sao cho $\mathbb{E}X_i = a$ và $DX_i = C < \infty$ với mọi $i = 1, 2, \ldots$ Khi đó $(X_n,\ n \geq 1)$ tuân theo luật số lớn*

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} a \qquad (\textit{khi } n \to \infty)$$

Ví dụ. Gieo một con xúc xắc cân đối. Giả sử X là số nốt xuất hiện ở mặt trên con xúc xắc. Ta có $\mathbb{E}X = 3,5$. Một nhà thống kê đã gieo một con xúc xắc cân đối 1 triệu lần (nhờ sự trợ giúp của máy vi tính) và ghi lại số nốt xuất hiện ở mặt trên con xúc xắc. Số trung bình của 1 triệu lần gieo được tìm thấy là

$$\frac{x_1 + \cdots + x_{10^6}}{10^6} \approx 3,500867 \approx 3,5.$$

Sự kiện này rất phù hợp với luật số lớn.

Hệ quả 3 (Bernoulli) *Tần suất f_n của một biến cố hội tụ theo xác suất* *về xác suất của biến cố đó (khi $n \to \infty$).*

Chứng minh. Xét dãy n phép thử Bernoulli đối với biến cố A và gọi k_n là số lần xảy ra A trong n phép thử đó, $f_n = \dfrac{k_n}{n}$ là tần suất xuất hiện A trong n phép thử đó.

Gọi X_k là đại lượng ngẫu nhiên chỉ số lần xuất hiện A ở phép thử thứ k $(1 \leqslant k \leqslant n)$

$$X_k = \begin{cases} 1 & \text{nếu } A \text{ xảy ra ở phép thử thứ } k, \\ 0 & \text{nếu trái lại.} \end{cases}$$

Khi đó

$$\mathbb{P}(X_k = 1) = p, \qquad \mathbb{P}(X_k = 0) = 1 - p,$$

$$\mathbb{E}X_k = p = \mathbb{P}(A).$$

Vì các phép thử này độc lập nên $X_1, X_2, \ldots, X_n$ độc lập và ta có

$$\frac{X_1 + \cdots + X_n}{n} = \frac{k_n}{n} = f_n$$

Vậy theo luật số lớn Chebyshev, f_n hội tụ theo xác suất tới p.

Ví dụ. 1. Nếu ta gieo một đồng tiền cân đối thì xác suất xuất hiện mặt ngửa là $0,5$. Vào thế kỷ 18, nhà toán học Pháp Buffon gieo một đồng tiền như vậy 4040 lần và ghi được 2048 lần xuất hiện mặt ngửa. Tần suất là $0,507$. Một nhà thống kê người Anh gieo đồng tiền 12.000 lần và thu được 6019 lần xuất hiện mặt ngửa. Tần suất xuất hiện mặt ngửa trong thí nghiệm này là $0,5016$. Trong một thí nghiệm khác, ông gieo đồng tiền 24.000 lần và thu được 12.012 lần xuất hiện mặt ngửa. Tần suất xuất hiện mặt ngửa là $0,5005$.

2. Một cỗ bài 36 con bài gồm 18 con màu đỏ và 18 con màu đen. Chia cỗ bài làm hai phần bằng nhau. Dễ dàng thấy rằng xác suất của

biến cố A "Hai phần đều có số con bài đỏ và đen bằng nhau" là

$$\mathbb{P}(A) = \frac{C_{18}^{9} C_{18}^{9}}{C_{36}^{18}} \approx 0,26.$$

Một nhà thống kê đã thử tiến hành việc chia cỗ bài làm đôi trong 100 lần. Tính đến lần thứ 80 tần suất của biến cố A là $0,25$; tính đến lần thứ 90 và 100 cho tần suất của biến cố A là $0,24$.

2.5.3 Hàm đặc trưng

Trong mục này, chúng ta sẽ nghiên cứu hàm đặc trưng, một công cụ rất có hiệu lực để nghiên cứu nhiều vấn đề của lí thuyết xác suất; đặc biệt là những vấn đề liên quan đến luật phân phối của tổng các đại lượng ngẫu nhiên độc lập và định lý giới hạn trung tâm.

Giả sử X, Y là hai đại lượng ngẫu nhiên. Khi đó X + iY (i là đơn vị ảo) là đại lượng ngẫu nhiên nhận giá trị trên mặt phẳng phức. Ta định nghĩa $\mathbb{E}(X + iY) = \mathbb{E}X + i\mathbb{E}Y$ nếu $\mathbb{E}X, \mathbb{E}Y$ xác định.

Định nghĩa. Giả sử X là đại lượng ngẫu nhiên. Khi đó hàm số $\varphi(t)$ xác định bởi công thức

$$\varphi(t) = \varphi_X(t) = \mathbb{E}(e^{itX}) = \mathbb{E}(\cos tX + i \sin tX) \qquad t \in \mathbb{R}$$

được gọi là *hàm đặc trưng* của X.

Chú ý: 1. Nói chung hàm đặc trưng của một đại lượng ngẫu nhiên nhận các giá trị phức.

2. Nhiều khi để chỉ rõ hàm đặc trưng của đại lượng ngẫu nhiên cụ thể X, người ta viết $\varphi_X(t)$ thay cho $\varphi(t)$.

Như vậy, nếu X là đại lượng ngẫu nhiên rời rạc: $\mathbb{P}(X = x_k) = p_k$; $(k = 1, 2 \ldots, n, \ldots)$ thì

$$\varphi_X(t) = \sum_k e^{itx_k} p_k.$$

Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ là $p(x)$ thì

$$\varphi_X(t) = \int_{-\infty}^{+\infty} e^{itx} p(x) dx.$$

Ví dụ

1. Giả sử X là đại lượng ngẫu nhiên biểu thị số lần xuất hiện biến cố A trong một phép thử (với $\mathbb{P}(A) = p$). Vậy thì

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

Do đó

$$\varphi_X(t) = e^{it0} q + e^{it1} p = pe^{it} + q.$$

2. Giả sử X là đại lượng ngẫu nhiên có phân phối chuẩn $N(0, 1)$. Vậy thì X có hàm mật độ

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}.$$

Do đó

$$\begin{aligned}
\varphi_X(t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx} e^{\frac{-x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{itx - \frac{x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}[(x-it)^2 + t^2]} dx \\
&= e^{-\frac{t^2}{2}} \Big(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du\Big) = e^{-\frac{t^2}{2}}.
\end{aligned}$$

Tính chất

1. $|\varphi(t)| \leqslant 1$, $\varphi(0) = 1$.

Thật vậy

$$|\varphi(t)| = |\mathbb{E} e^{itX}| \leqslant \mathbb{E} |e^{itX}| = \mathbb{E} 1 = 1,$$

$\varphi(0) = \mathbb{E} e^{it0} = \mathbb{E} 1 = 1.$

2. $\varphi(-t) = \overline{\varphi(t)}$.

Thật vậy

$$\begin{aligned}\varphi(-t) &= \mathbb{E}e^{i(-t)X} = \mathbb{E}e^{i(-tX)} = \mathbb{E}(\cos(-tX) + i\sin(-tX))\\ &= \mathbb{E}(\cos(tX) - i\sin(tX)) = \overline{\mathbb{E}(\cos(tX) + i\sin(tX))}\\ &= \overline{\mathbb{E}e^{itX}} = \overline{\varphi(t)}.\end{aligned}$$

3. $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$.

Thật vậy

$$\varphi_{aX+b}(t) = \mathbb{E}e^{i(at)X}e^{itb} = e^{itb}\mathbb{E}e^{itX} = e^{itb}\varphi_X(at).$$

4. Nếu $X_1, X_2, \ldots, X_n$ *là các đại lượng ngẫu nhiên độc lập thì* $\varphi_{\sum_{k=1}^n X_k}(t) = \prod_{k=1}^n \varphi_{X_k}(t)$.

Thật vậy, do $X_1, X_2, \ldots, X_n$ độc lập nên $e^{itX_1}, e^{itX_2}, \ldots, e^{itX_n}$ cũng độc lập. Vì thế

$$\varphi_{\sum_{k=1}^n X_k}(t) = \mathbb{E}e^{it(\sum_{k=1}^n X_k)} = \mathbb{E}\Big(\prod_{k=1}^n e^{itX_k}\Big)$$

$$= \prod_{k=1}^n \mathbb{E}e^{itX_k} = \prod_{k=1}^n \varphi_{X_k}(t).$$

5. (Công thức ngược) Giả sử $\varphi(t)$ *và* $F(x)$ *là hàm đặc trưng và hàm phân phối của đại lượng ngẫu nhiên* X*. Nếu* x*,* y *là các điểm liên tục của* $F(x)$ *thì*

$$F(x) - F(y) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \frac{e^{-ity} - e^{-itx}}{it}\varphi(t)dt.$$

Từ công thức ngược dễ dàng suy ra tính chất sau đây.

6. (Định lý duy nhất). Hàm đặc trưng xác định duy nhất hàm phân phối.

Thật vậy

$$F(x) = \lim_{y\to-\infty}(F(x) - F(y)) = \frac{1}{2\pi}\lim_{y\to-\infty}\int_{-\infty}^{+\infty} \frac{e^{-ity} - e^{-itx}}{it}\varphi(t)dt,$$

trong đó y lấy trên tập các điểm liên tục của $F(x)$.

Ví dụ. Giả sử đại lượng ngẫu nhiên X có hàm đặc trưng $\varphi(t) = \cos(t)$, dùng công thức ngược có thể chứng minh được hàm phân phối của X là

$$F(x) = \begin{cases} 0 & \text{nếu } x \leqslant -1, \\ \dfrac{1}{2} & \text{nếu } -1 < x \leqslant 1, \\ 1 & \text{nếu } x > 1. \end{cases}$$

(Độc giả xem đây như là bài tập).

7. Nếu đại lượng ngẫu nhiên X có moment tuyệt đối cấp n thì hàm đặc trưng có đạo hàm đến cấp n và ta có

$$\varphi^k(0) = i^k \mathbb{E}X^k \quad (\forall k \leqslant n),$$

Thật vậy, ta có

$$\varphi^k(t) = i^k \mathbb{E}X^k e^{itX} \leqslant |i^k \mathbb{E}X^k e^{itX}| \leqslant \mathbb{E}|X^k| < +\infty.$$

Do đó tồn tại $\varphi^k(t)$ và $\varphi^k(0) = i^k EX^k$.

Tính chất trên cho phép ta tính các moment của đại lượng ngẫu nhiên X dựa vào các giá trị của hàm đặc trưng của nó tại $t = 0$.

8. Giả sử dãy đại lượng ngẫu nhiên $(X_n,\ n \geq 1)$ có dãy hàm đặc trưng tương ứng là $(\varphi_n(t))$, còn đại lượng ngẫu nhiên X có hàm đặc trưng là $\varphi(t)$. Vậy thì

$$X_n \xrightarrow{\mathrm{D}} X \Leftrightarrow \varphi_n(t) \to \varphi(t) \qquad \forall t \in \mathbb{R} \text{ (khi } n \to \infty).$$

Tính chất này là một công cụ rất có hiệu quả khi nghiên cứu sự hội tụ theo phân phối của dãy các đại lượng ngẫu nhiên.

Hàm đặc trưng của một số phân phối xác suất thường gặp

1. Phân phối nhị thức $B(n; p)$

Giả sử A là biến cố của phép thử G với $\mathbb{P}(A) = p$. Thực hiện n phép thử Bernoulli từ G. Gọi X_i là số lần xuất hiện A trong phép thử thứ k $(k = 1, 2, \ldots, n)$, còn X là số lần xuất hiện A trong n phép thử Bernoulli đã cho. Vậy thì $X_1, X_2, \ldots, X_n$ độc lập, cùng phân phối

$$\mathbb{P}(X_k = 0) = 1 - p = q, \quad \mathbb{P}(X_k = 1) = p,$$

$$X = \sum_{k=1}^{n} X_k, \quad X \sim B(n; p).$$

Do đó

$$\varphi_k(t) = pe^{it} + q \qquad (k = 1, 2, \ldots, n)$$

và

$$\varphi_X(t) = \varphi_{\sum_{k=1}^n X_k}(t) = \Pi_{k=1}^n \varphi_{X_{k\varphi}}(t) = (\varphi_{X_1}(t))^n = (pe^{it} + q)^n.$$

2. Phân phối đều

Giả sử đại lượng ngẫu nhiên X có phân phối đều trên đoạn $[a, b]$. Khi đó X có hàm mật độ

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

Do đó

$$\varphi_X(t) = \int_a^b \frac{e^{itx}}{b-a} dx = \frac{1}{it(b-a)}(e^{itb} - e^{ita}).$$

Đặc biệt, nếu X có phân phối đều trên đoạn $[-a, a]$ thì

$$\varphi_X(t) = \frac{1}{2ita}(e^{ita} - e^{-ita})$$

$$= \frac{1}{2ita}(\cos ta - i \sin ta - \cos(-ta) - i \sin(-ta))$$

$$= \frac{1}{2ita} 2i \sin ta = \frac{\sin at}{at}.$$

3. Phân phối chuẩn $N(\mu, \sigma^2)$

Giả sử $X \sim N(\mu, \sigma^2)$. Vậy thì

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

Do đó

$$\varphi_Y(t) = e^{-\frac{t^2}{2}}.$$

Mặt khác

$$X = \sigma Y + \mu.$$

Suy ra

$$\varphi_X(t) = e^{i\mu t}\varphi_Y(\sigma t) = e^{i\mu t - \frac{(\sigma t)^2}{2}}.$$

Từ đó, áp dụng tính chất về hàm đặc trưng của tổng các đại lượng ngẫu nhiên độc lập có thể chứng minh được rằng: Nếu các đại lượng ngẫu nhiên độc lập $X_1, X_2, \ldots, X_n$ có phân phối chuẩn với kì vọng và phương sai tương ứng là μ_k và σ_k^2 thì tổng của chúng là $\sum_{k=1}^n X_k$ cũng có phân phối chuẩn với kỳ vọng là $\sum_{k=1}^n \mu_k$ và phương sai là $\sum_{k=1}^n \sigma_k^2$.

Thật vậy, với $n = 2$, giả sử $X_1 \sim N(\mu_1, \sigma_1)$, $X_2 \sim N(\mu_2, \sigma_2)$, X_1, X_2 độc lập. Vậy thì

$$(X_1 + X_2) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Thật vậy, ta có

$$\varphi_{X_1}(t) = e^{i\mu_1 t - \frac{1}{2}\sigma_1^2 t^2}, \qquad \varphi_{X_2}(t) = e^{i\mu_2 t - \frac{1}{2}\sigma_2^2 t^2}.$$

Suy ra

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t) = e^{i(\mu_1+\mu_2)t - \frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}.$$

Nghĩa là $\varphi_{X_1+X_2}(t)$ là hàm đặc trưng của đại lượng ngẫu nhiên có phân phối chuẩn $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. Từ định lý duy nhất suy ra rằng $(X_1 + X_2) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Trường hợp $n > 2$, phép chứng minh hoàn toàn tương tự.

Tiếp theo, bằng cách sử dụng công cụ hàm đặc trưng, chúng ta sẽ nghiên cứu một số dạng đặc biệt của định lý giới hạn trung tâm và định lý Poisson; từ đó rút ra các công thức xấp xỉ phân phối nhị thức $B(n;p)$ bởi phân phối chuẩn và phân phối Poisson khi n khá lớn.

2.5.4 Định lý giới hạn trung tâm

Định lý. *Giả sử* $(X_n,\ n \geq 1)$ *là dãy đại lượng ngẫu nhiên độc lập, có cùng phân phối,* $\mathbb{E}X_1 = 0$, $DX_1 = 1$. *Đặt*

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}.$$

Khi đó

$$Y_n \xrightarrow{D} N(0, 1).$$

Chứng minh. Giả sử $\varphi_n(t)$ là hàm đặc trưng của đại lượng ngẫu nhiên Y_n. Ta cần chứng minh $\lim_{n\to\infty} \varphi_n(t) = e^{-\frac{t^2}{2}}$.

Đặt

$$Z_n = X_1 + X_2 + \cdots + X_n, \quad \varphi_{X_i}(t) = \varphi(t).$$

Vậy thì

$$\varphi_{Z_n}(t) = [\varphi(t)]^n.$$

Mặt khác

$$\varphi'(0) = 0, \quad \varphi''(0) = -1.$$

Do đó $\varphi(t)$ có khai triển Maclaurin:

$$\varphi(t) = \varphi(0) + \frac{\varphi'(0)}{1!}t + \frac{\varphi''(0)}{2!}t^2 + R(t) = 1 - \frac{t^2}{2} + R(t),$$

trong đó

$$R(t) = o(t^2) \text{ khi } t \to 0.$$

Suy ra

$$\varphi_n(t) = \varphi_{Z_n}(\frac{t}{\sqrt{n}}) = [\varphi(\frac{t}{\sqrt{n}})]^n = [1 - \frac{t^2}{2n} + R(\frac{t}{\sqrt{n}})]^n.$$

Nên

$$\lim_{n\to\infty} \varphi_n(t) = \lim_{n\to\infty} [1 - \frac{t^2}{2n} + R(\frac{t}{\sqrt{n}})]^n$$

$$= \lim_{n\to\infty}(1 - \frac{t^2}{2n})^n = e^{-\frac{t^2}{2}} = \varphi_X(t)$$

với $X \sim N(0,1)$.

Vậy

$$Y_n \xrightarrow{\mathrm{D}} N(0,1).$$

Nhận xét. Từ định lý trên, suy ra rằng nếu $X_1, X_2, \ldots$ là các đại lượng ngẫu nhiên độc lập, có cùng phân phối, $\mathbb{E}X_1 = 0$, $DX_1 = 1$, thì

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{\sqrt{n}}$$

có phân phối xấp xỉ chuẩn $N(0,1)$. Do đó nếu $X_1, X_2, \ldots$ là các đại lượng ngẫu nhiên độc lập, có cùng phân phối, $\mathbb{E}X_1 = \mu$, $DX_1 = \sigma^2$, thì trung bình cộng

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

có phân phối xấp xỉ chuẩn $N(\mu, \sigma^2)$. Kết quả này sẽ được dùng nhiều trong thống kê. **Hệ quả** (Định lý Moivre - Laplace). *Giả sử đại lượng ngẫu nhiên S_n có phân phối nhị thức $B(n;p)$.*

Đặt $Y_n = \dfrac{S_n - np}{\sqrt{npq}}$. Khi đó

$$Y_n \xrightarrow{D} N(0,1).$$

Chứng minh. Giả sử A là biến cố liên quan đến phép thử nào đó với $\mathbb{P}(A) = p$.

Thực hiện n phép thử Bernoulli từ phép thử đã cho. Gọi X_k là số lần xuất hiện A trong lần thử thứ k. Vậy thì

$$\mathbb{P}(X_k = 0) = 1 - p \qquad \mathbb{P}(X_k = 1) = p.$$

Đồng thời các đại lượng ngẫu nhiên $X_1, X_2, \ldots, X_n$, độc lập cùng phân phối

$$\mathbb{E}X_k = p, \qquad DX_k = pq.$$

Hơn nữa

$$S_n = X_1 + X_2 + \cdots + X_n.$$

Có phân phối nhị thức $B(n;p)$.

Đặt

$$Z_k = \frac{X_k - p}{\sqrt{pq}}.$$

Lúc đó dãy $(Z_n; n \geq 1)$ thoả mãn định lý giới hạn trung tâm. Do đó:

$$Y_n = \frac{X - np}{\sqrt{npq}} = \frac{Z_1 + Z_2 + \cdots + Z_n}{\sqrt{n}} \xrightarrow{\text{D}} N(0,1).$$

ứng dụng. Từ hệ quả trên suy ra rằng nếu $X \sim B(n;p)$ thì $\dfrac{X-np}{\sqrt{npq}}$ sẽ có phân phối xấp xỉ phân phối chuẩn khi n khá lớn. Do đó

$$\mathbb{P}(k_1 < X < k_2) = \mathbb{P}(\frac{k_1 - np}{\sqrt{npq}} < \frac{X - np}{\sqrt{npq}} < \frac{k_2 - np}{\sqrt{npq}})$$

$$\approx \Phi(\frac{k_2 - np}{\sqrt{npq}}) - \Phi(\frac{k_1 - np}{\sqrt{npq}}),$$

trong đó $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$ là hàm phân phối của phân phối chuẩn $N(0,1)$.

Ví dụ. Tung đồng tiền 10000 lần. Tìm xác suất để có ít hơn 5000 lần xuất hiện mặt sấp.

Giải. Gọi X là số lần xuất hiện mặt sấp khi tung đồng tiền 10000 lần. Vậy thì

$$X \sim B(10000; \frac{1}{2}).$$

Do đó

$$\mathbb{P}(0 \leqslant X < 5000) \approx \mathbb{P}(0 < X < 5000) \approx \Phi(0) - \Phi(\frac{0 - 5000}{\sqrt{1000.\frac{1}{2}\frac{1}{2}}}) \approx 0,5.$$

2.5.5 Định lý Poisson

Như vậy, khi n khá lớn có thể xấp xỉ phân phối nhị thức bởi phân phối chuẩn. Tuy nhiên, xấp xỉ đó chỉ tốt khi np và nq lớn hơn 5 hoặc khi npq lớn hơn 25. Trong trường hợp npq khá bé ta có thể xấp xỉ phân phối nhị thức bởi phân phối Poisson.

Định lý. (Poisson) *Giả sử với mỗi* $n = 1, 2, \ldots$, (X_{nk}), $(k = 1, 2, \ldots, n)$ *là dãy đại lượng ngẫu nhiên độc lập cùng phân phối, lấy hai giá trị* 0 *và* 1 *với*

$$\mathbb{P}(X_{nk} = 1\} = \frac{\lambda}{n} + o(\frac{1}{n}) \qquad k = 1, 2, \ldots, n.$$

Ký hiệu $Y_n = \sum_{k=1}^{n} X_{nk}$. *Khi đó, với mọi* $m = 0, 1, \ldots$ *ta có công thức tiệm cận*

$$\lim_{n\to\infty} \mathbb{P}(Y_n = m\} = e^{-\lambda}\frac{\lambda^m}{m!}.$$

Chứng minh. Ta có hàm đặc trưng của Y_m là

$$\mathbb{E}e^{itY_n} = \Pi_{k=1}^{n}\mathbb{E}e^{itX_{nk}} = [1 + \frac{\lambda}{n}(e^{it} - 1) + o(\frac{1}{n})]^n \to \exp\{\lambda(e^{it} - 1)\}$$

(*khi* $n \to \infty$). Vì $\exp\{\lambda(e^{it} - 1)\}$ là hàm đặc trưng của phân phối Poisson với tham số λ, nên ta suy ra điều phải chứng minh.

Định lý Poisson có thể minh hoạ như sau:

Xét dãy phép thử độc lập. Trong phép thử thứ n xác suất xuất hiện biến cố A là p_n và $np_n \to \lambda$, $n \to \infty$. Khi đó phân phối giới hạn của số lần xuất hiện biến cố A trong dãy các phép thử tuân theo luật Poisson với tham số λ.

Ta xét ví dụ.

b) Ví dụ. Giả sử trong 2000 chiếc bánh nướng có 16000 hạt lạc. Tìm phân phối xác suất số hạt lạc có trong một chiếc bánh được chọn ngẫu nhiên.

Giải. Giả sử các hạt lạc được đánh số từ 1 đến 16000. Ta xem việc chia một hạt lạc vào bánh là một phép thử ngẫu nhiên. Ta có dãy gồm $n = 16000$ phép thử. Nói chung các phép thử này không độc lập, nhưng nếu các hạt lạc xáo trộn tốt thì xác suất để hạt thứ k rơi vào chiếc bánh đã chọn bằng $\frac{1}{2000}$ và thực tế có thể xem các phép thử là độc lập. Khi đó số những hạt lạc có trong chiếc bánh đã chọn là đại lượng ngẫu nhiên tuân theo luật Poisson với tham số $\lambda = np = 8$, tức là

$$\mathbb{P}(Y = k) = e^{-8}\frac{8^k}{k!}.$$

BÀI TẬP

1. Một nhóm có 10 người gồm 6 nam và 4 nữ. Chọn ngẫu nhiên ra 3 người. Gọi X là số nữ trong nhóm. Hãy lập bảng phân phối xác suất của X và tính $\mathbb{E}X$, DX, $\mod X$.

2. Một túi chứa 10 thẻ đỏ và 6 thẻ xanh. Chọn ngẫu nhiên ra 3 tấm thẻ (không hoàn lại). Gọi X là số thẻ đỏ được chọn. Lập bảng phân phối xác suất của X. Tính $\mathbb{E}X$ và DX.

Giả sử rút mỗi thẻ đỏ được 5 điểm và rút mỗi thẻ xanh được 8 điểm. Gọi Y là số điểm tổng cộng trên 3 thẻ rút ra. Lập bảng phân phối xác suất của Y và tính $\mathbb{E}Y$, DY.

3. Hai xạ thủ A và B tập bắn, mỗi người bắn 2 phát. Xác suất bắn trúng đích của A trong mỗi phát là $0,4$, còn của B là $0,5$. Gọi X là số phát trúng của A trừ đi số phát trúng của B. Tìm bảng phân phối xác suất của X và bảng phân phối xác suất của $Y = |X|$.

4. Giả sử $X \sim B(2; 0,4)$; $Y \sim B(2; 0,7)$; X và Y độc lập. Tìm bảng phân phối xác suất của $X + Y$. Chứng minh rằng $X + Y$ không có phân phối nhị thức.

5. Cho X và Y là hai đại lượng ngẫu nhiên độc lập.

a) Giả sử $X \sim B(1; \frac{1}{5})$, $Y \sim B(2; \frac{1}{5})$. Viết bảng phân phối xác suất của X, Y. Từ đó tìm bảng phân phối xác suất của $X + Y$. Kiểm tra rằng $X + Y \sim B(3; \frac{1}{5})$.

b) Giả sử $X \sim B(1; \frac{1}{2})$; $Y \sim B(2; \frac{1}{5})$. Tìm bảng phân phối xác suất của $X + Y$. Chứng minh rằng $X + Y$ không có phân phối nhị thức.

6. Trong một thành phố nhỏ, trung bình một tuần có 2 người chết. Tính

xác suất để:

a) Không có người nào chết trong vòng 1 ngày.

b) Có ít nhất 3 người chết trong vòng hai ngày.

7. Một trạm cho thuê xe tắcxi có 3 chiếc xe. Hàng ngày trạm phí nộp thuế là 8 USD cho một chiếc xe (dù xe đó có được thuê hay không). Mỗi chiếc xe cho thuê với giá 20 USD. Giả sử số yêu cầu thuê xe của trạm trong một ngày là đại lượng ngẫu nhiên X có phân phối Poát Xông với tham số $\lambda = 2,8$. Gọi Y là số tiền thu được trong một ngày của trạm (nếu không có ai thuê thì số tiền thu được là -24 USD). Tìm bảng phân phối xác suất của Y. Từ đó tính số tiền trung bình thu được của trạm trong một ngày. Giải bài toán trên trong trường hợp trạm có 4 chiếc xe. Trạm nên có 3 hay 4 chiếc xe.

8. Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ

$$p(x) = \begin{cases} cx^2(1-x) & \text{nếu } x \in [0,1], \\ 0 & \text{nếu } x \notin [0,1]. \end{cases}$$

Tìm hằng số c; $modX$ và $\mathbb{P}(0,4 < X < 0,6)$.

9. a) Cho đại lượng ngẫu nhiên X có phân phối đều trên $[1,2]$. Tính $\mathbb{P}(2 < X < 5)$.

b) Cho đại lượng ngẫu nhiên X có phân phối đều trên $[-1,3]$. Tính $\mathbb{P}(X^2 > 2)$.

10. Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ

$$p(x) = \begin{cases} kx^2 & \text{nếu } 0 \leqslant x \leqslant 3, \\ 0 & \text{nếu trái lại.} \end{cases}$$

Tìm hằng số k; tính $P\{X > 2\}$, tìm median của X và xác định a để $P\{X < a\} = \dfrac{3}{4}$.

11. Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ

$$p(x) = \begin{cases} \dfrac{x}{4} + \dfrac{1}{2} & \text{nếu } -2 \leqslant x \leqslant 0, \\ -\dfrac{x}{4} + \dfrac{1}{2} & \text{nếu } 0 \leqslant x \leqslant 2, \\ 0 & \text{nếu trái lại.} \end{cases}$$

Tìm kỳ vọng và phương sai của X.

12. Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ

$$p(x) = \begin{cases} kx & \text{nếu } 0 \leqslant x \leqslant 1, \\ k & \text{nếu } 1 \leqslant x \leqslant 4, \\ 0 & \text{nếu trái lại.} \end{cases}$$

Tìm hằng số k, kỳ vọng, phương sai và median của X.

13. Đại lượng ngẫu nhiên liên tục X có hàm mật độ

$$p(x) = \begin{cases} \dfrac{3}{4}(1 - x^2) & \text{nếu } |x| \leqslant 1, \\ 0 & \text{nếu trái lại.} \end{cases}$$

Tìm kỳ vọng và phương sai của $Y = 2X^2$.

14. Cho đại lượng ngẫu nhiên liên tục X có hàm mật độ

$$p(x) = \begin{cases} 3x^2 & \text{nếu } 0 \leqslant x \leqslant 1, \\ 0 & \text{nếu trái lại.} \end{cases}$$

Xét $Y = 2\sqrt{X}$. Tìm $\mathbb{P}(0,2 < Y < 1,5)$ và $\mathbb{P}(Y > 1)$.

15. Cho hàm mật độ của một đại lượng ngẫu nhiên X có dạng

$$p(x) = \begin{cases} \dfrac{a}{x^4} & \text{nếu } x \geqslant 1, \\ 0 & \text{nếu } x < 1. \end{cases}$$

$a)$ Xác định hằng số a.

$b)$ Tính $\mathbb{P}(0 < X < 5)$.

$c)$ Tính phương sai của X.

16. Cho đại lượng ngẫu nhiên X thoả mãn:

$$\mathbb{P}(X \geqslant x) = \begin{cases} e^{a(x-3)} & \text{nếu } x \geqslant 3, \\ 1 & \text{nếu } x < 3. \end{cases}$$

$a)$ Xác định hằng số a.

$b)$ Tính kỳ vọng của X.

17. Cho hàm mật độ của một đại lượng ngẫu nhiên X có dạng:

$$p(x) = \begin{cases} ae^{-2x^2} & \text{nếu } x \geqslant 0, \\ 0 & \text{nếu } x < 0. \end{cases}$$

a) Xác định hằng số a.

b) Tính $\mathbb{P}(-3 < X < 2)$.

c) Tính kỳ vọng của X.

18. Một người hàng ngày đi bộ từ nhà đến nơi làm việc với quãng đường $600m$ với vận tốc đều v(m/giây). Biết rằng thời gian đi bộ của người đó là một đại lượng ngẫu nhiên phân phối đều trong khoảng từ 6 phút đến 10 phút. Tìm kỳ vọng, độ lệch chuẩn và median của v.

19. Trọng lượng của một con bò là một đại lượng ngẫu nhiên phân phối chuẩn với kỳ vọng là 250 kg và độ lệch tiêu chuẩn 40 kg. Tìm xác suất để một con bò có trọng lượng:

a) Nặng hơn 300 kg.

b) Nhẹ hơn 175 kg.

c) Trong khoảng 260 kg đến 270 kg.

20. Thời gian từ nhà đi đến trường của sinh viên Bình là một đại lượng

ngẫu nhiên X có phân phối chuẩn. Biết rằng 65% số ngày Bình đến trường mất hơn 20 phút còn 8% số ngày mất hơn 30 phút.
Tìm thời gian trung bình và độ lệch tiêu chuẩn của thời gian đến trường.
Bình xuất phát từ nhà trước giờ vào học bao nhiêu phút để khả năng bị muộn học là bé hơn $0,02$.
21. Chiều cao của một loại cây là một đại lượng ngẫu nhiên có phân phối chuẩn. Trong một mẫu gồm 640 cây có 25 cây thấp hơn 18m và 110 cây cao hơn 24m.
Tìm kỳ vọng và độ lệch tiêu chuẩn của chiều cao của cây.
Ước lượng số cây có độ cao trong khoảng từ 16m đến 20m trong mẫu nói trên.
22. Cho X là đại lượng ngẫu nhiên có phân phối mũ với tham số $\lambda = 2$. Tìm kỳ vọng và độ lệch tiêu chuẩn của e^{-X}.
23. Cho X là đại lượng ngẫu nhiên có phân phối mũ với tham số $\lambda = 1$ và $Y = 2X^2$. Tính $\mathbb{P}(2 < Y < 18)$ và $\mathbb{P}(Y < 4)$.
24. Cho vectơ ngẫu nhiên liên tục (X, Y) có hàm mật độ là

$$p(x,y) = \begin{cases} Cx & \text{nếu } (x,y) \text{ thoả mãn } 0 < y < x < 1, \\ 0 & \text{nếu trái lại.} \end{cases}$$

a) Tìm C.

b) Tìm các hàm mật độ của X và của Y.

c) X và Y có độc lập hay không.

25. Vectơ ngẫu nhiên liên tục (X, Y) có hàm mật độ là

$$p(x,y) = \begin{cases} C & \text{nếu } (x,y) \text{ thoả mãn } \dfrac{x^2}{9} + \dfrac{y^2}{4} < 1, \\ 0 & \text{nếu trái lại.} \end{cases}$$

a) Tìm C.

b) Tìm hàm mật độ của X và Y.

26. Cho vectơ ngẫu nhiên liên tục (X, Y) có hàm mật độ

$$p(x, y) = \frac{C}{(1+x^2)(1+y^2)}.$$

a) Tìm C.

b) Tìm hàm phân phối của (X, Y).

c) X và Y có độc lập không?

d) Tìm xác suất để điểm (X, Y) rơi vào hình chữ nhật với các đỉnh là $A(1, 1)$, $B(\sqrt{3}, 1)$, $C(1, 0)$ và $D(\sqrt{3}, 0)$.

27. Một điểm A rơi ngẫu nhiên vào một hình vuông D có cạnh bằng 1. Giả sử (X, Y) là toạ độ của A. Biết rằng hàm mật độ $p(x, y)$ của (X, Y) là

$$p(x, y) = \begin{cases} 1 & \text{nếu } (x, y) \in D \ , \\ 0 & \text{nếu } (x, y) \notin D. \end{cases}$$

Tính xác suất để khoảng cách từ A đến cạnh gần nhất nó bé hơn hay bằng $0,3$.

28. Giả sử X và Y là hai đại lượng ngẫu nhiên độc lập có các hàm phân phối tương ứng là $F(x)$ và $G(x)$. Tìm hàm phân phối của các đại lượng ngẫu nhiên sau đây

$$a) \max\{X, Y\}, \quad b) \min\{X, Y\}, \quad c) \max\{2X, Y\}, \quad d) \min\{X, Y^3\}.$$

29. Giả sử X và Y là hai đại lượng ngẫu nhiên độc lập có cùng phân phối mũ với tham số λ. Tìm hàm phân phối và hàm mật độ của các đại lượng ngẫu nhiên sau đây:

$$a) X^3; \quad b) \max\{X, Y^3\}; \quad c) \min\{X, Y^3\}.$$

30. Giải bài toán trên với giả thiết là X và Y có phân phối đều trên đoạn [0, 1].

31. Giải sử X là đại lượng ngẫu nhiên có phân phối mũ với tham số λ. $f(x)$ là hàm nhận giá trị dương, khả vi và tăng nghiêm ngặt. Tìm hàm mật độ của đại lượng ngẫu nhiên $f(X)$.

32. Chứng minh rằng hàm phân phối của một đại lượng ngẫu nhiên có không quá đếm được điểm gián đoạn.

33. Chứng minh rằng nếu hàm phân phối của một đại lượng ngẫu nhiên liên tục trên toàn trục số, thì nó liên tục đều trên đó.

34. Giả sử $X_1, \ldots, X_n$ là các đại lượng ngẫu nhiên có kỳ vọng hữu hạn. Chứng minh rằng

$$\mathbb{E}\max\{X_1, \ldots, X_n\} \geq \max\{\mathbb{E}X_1, \ldots, \mathbb{E}X_n\}$$

và

$$\mathbb{E}\min\{X_1, \ldots, X_n\} \leqslant \min\{\mathbb{E}X_1, \ldots, \mathbb{E}X_n\}.$$

35. Giả sử X là đại lượng ngẫu nhiên có phương sai hữu hạn. Chứng minh rằng

$$DX = \min_{a\in R}\mathbb{E}(X-a)^2.$$

36. Giả sử $X_n \xrightarrow{\mathrm{P}} X$, $X_n \xrightarrow{\mathrm{P}} Y$. Chứng minh rằng

$$\mathbb{P}(X = Y) = 1.$$

37. Giả sử $X_n \xrightarrow{\mathrm{P}} X$, $Y_n \xrightarrow{\mathrm{P}} Y$. Chứng minh rằng

$$aX_n + bY_n \xrightarrow{\mathrm{P}} aX + bY, \qquad a, b \in \mathbb{R}.$$

38. Chứng minh rằng nếu $\lim_{n\to\infty} \mathbb{P}(X_n \neq Y_n) = 0$ thì

$$X_n \xrightarrow{\mathrm{P}} X \Leftrightarrow Y_n \xrightarrow{\mathrm{P}} X \qquad (\text{khi } n \to \infty).$$

39. Giả sử $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$ và $\mathbb{P}(X = Y) = 1$. Chứng minh rằng với mọi $\varepsilon > 0$ thì

$$\lim_{n\to\infty} \mathbb{P}(|X_n - Y_n| > \varepsilon) \to 0.$$

40. Chứng minh rằng $X_n \xrightarrow{P} X$ khi và chỉ khi mọi dãy con của nó đều chứa một dãy con hội tụ h.c.c đến X.

41. Giả sử $(X_n - X)^2 \xrightarrow{P} 0$. Chứng minh rằng $X_n^2 \xrightarrow{P} X^2$.

42. Giả sử $(X_n, n \geq 1)$ và $(Y_n, n \geq 1)$ là các dãy đại lượng ngẫu nhiên thỏa mãn

$$\sum_{n=1}^{\infty} \mathbb{P}(X_n \neq Y_n) < \infty.$$

Chứng minh rằng

a) Nếu $X_n \xrightarrow{\text{h.c.c}} X$ thì $Y_n \xrightarrow{\text{h.c.c}} X$ (khi $n \to \infty$).

b) Nếu chuỗi $\sum_{n=1}^{\infty} X_n$ hội tụ hầu chắc chắn, thì chuỗi $\sum_{n=1}^{\infty} Y_n$ hội tụ hầu chắc chắn.

43. Giả sử $X_n \xrightarrow{D} X$, $Y_n \xrightarrow{P} 0$. Chứng minh rằng $X_n + Y_n \xrightarrow{D} X$, $X_n Y_n \xrightarrow{P} 0$.

44. Giả sử $\mathbb{P}(X_{n+1} \leqslant X_n) = 1$, với mọi $n = 1, 2, \ldots$ Chứng minh rằng nếu $X_n \xrightarrow{P} X$ thì $X_n \xrightarrow{\text{h.c.c}} X$ (khi $n \to \infty$).

45. Giả sử $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} X$ và $\mathbb{P}(X_n \leqslant Z_n \leqslant Y_n) = 1$, với mọi $n = 1, 2, \ldots$ Chứng minh rằng $Z_n \xrightarrow{P} X$.

46. Giả sử $X_n \xrightarrow{P} X$, $Y_n \xrightarrow{P} Y$ và $\mathbb{P}(X_n \leqslant Y_n) = 1$, với mọi $n = 1, 2, \ldots$ Chứng minh rằng $\mathbb{P}(X \leqslant Y) = 1$.

47. Gieo một con xúc xắc cân đối n lần. Gọi X là số lần xuất hiện mặt 6 chấm. Chứng minh rằng:

$$\mathbb{P}(\frac{n}{6} - \sqrt{n} < X < \frac{n}{6} + \sqrt{n}) \geqslant \frac{31}{36}.$$

48. Giả sử X là đại lượng ngẫu nhiên có $\mathbb{E}X = 5$, $DX = 0,16$. Chứng minh rằng:

$a)\mathbb{P}(3 < X < 7) \geqslant 0,96.$

$b)\mathbb{P}(2 < X < 8) \geqslant 0,98.$

$c)\mathbb{P}(3 < \dfrac{X_1 + X_2 + \cdots + X_9}{9} < 7) \geqslant 0,99,$

trong đó $X_1; \ldots; X_9$ là các đại lượng ngẫu nhiên độc lập, có cùng phân phối với X.

49. Giả sử X nhận giá trị -1 và 1 với $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \dfrac{1}{2}$. Tìm hàm đặc trưng của X.

50. Giả sử $X_1, X_2, \ldots, X_n$ là các đại lượng ngẫu nhiên độc lập cùng phân phối và

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

Xác định hàm đặc trưng của $S_n = X_1 + \cdots + X_n$.

51. a) Đại lượng ngẫu nhiên X có mật độ

$$p(x) = \begin{cases} \dfrac{1}{a}(1 - \dfrac{|x|}{a}) & \text{với } |x| \leqslant a, \\ 0 & \text{với } |x| > a. \end{cases}$$

Chứng minh rằng hàm đặc trưng của X là

$$\varphi(t) = 2\frac{1 - \cos at}{a^2 t^2}.$$

b) Đại lượng ngẫu nhiên Y có mật độ

$$g(y) = \frac{1}{\pi}.\frac{1 - \cos ay}{a^2 y^2}.$$

Chứng minh rằng hàm đặc trưng của Y là

$$\varphi(t) = \begin{cases} 1 - \dfrac{|t|}{a} & \text{với } |t| \leqslant a \\ 0 & \text{với } |t| > a \end{cases}$$

52. Giả sử X_1, X_2 là các đại lượng ngẫu nhiên độc lập. Bằng phương pháp hàm đặc trưng chứng minh rằng:

a) $X_i \sim B(n_i;p), i = 1,2 \Rightarrow X_1 + X_2 \sim B(n_1 + n_2;p)$.

b) $Xi \sim P(\lambda_i), i = 1,2 \Rightarrow X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$.

c) $X_i \sim N(a_i, \sigma_i^2), i = 1,2 \Rightarrow X_1 + X_2 \sim N(a_1 + a_2, \sigma_i^2 + \sigma_2^2)$.

53. Đặt $S_n = X_1 + \cdots + X_n$. Chứng minh rằng với mọi dãy đại lượng ngẫu nhiên $(X_n,\ n \geq 1)$ thì

a) $X_n \to 0 \quad$ h.c.c $\Rightarrow \dfrac{S_n}{n} \to 0 \quad$ h.c.c.

b) $X_n \xrightarrow{L^2} 0 \Rightarrow \dfrac{S_n}{n} \xrightarrow{L^2} 0$.

c) $\max_{k \leqslant n} |X_k| \xrightarrow{\text{P}} 0 \Rightarrow \dfrac{S_n}{n} \xrightarrow{\text{P}} 0$.

d) $\dfrac{S_n}{n} \xrightarrow{\text{P}} 0 \Rightarrow X_n \xrightarrow{\text{p}} 0$.

54. Giả sử $(X_n, n \geqslant 1)$ là dãy đại lượng ngẫu nhiên độc lập và $\mathbb{P}[X_n = 2^n] = \mathbb{P}[X_n = -2^n] = \dfrac{1}{2}$. Chứng minh rằng $(X_n,\ n \geq 1)$ không tuân theo luật số lớn.

55. Giả sử $(X_n,\ n \geq 1)$ là dãy đại lượng ngẫu nhiên độc lập. Trong các trường hợp sau, xét xem dãy $(X_n,\ n \geq 1)$ có tuân theo luật số lớn không?

a) $\mathbb{P}[X_n = \sqrt{n}] = \mathbb{P}[X_n = -\sqrt{n}] = \dfrac{1}{2n}, \quad \mathbb{P}[X_n = 0] = 1 - \dfrac{1}{n}$.

b) $\mathbb{P}[X_n = n] = \mathbb{P}[X_n = -n] = \dfrac{1}{2n^2} \quad \mathbb{P}[X_n = 0] = 1 - \dfrac{1}{n^2}$.

c) $\mathbb{P}[X_n = 2^n] = \mathbb{P}[X_n = -2^n] = \dfrac{1}{2^{2n+1}} \quad \mathbb{P}[X_n = 0] = 1 - \dfrac{1}{2^{2n}}$.

d) $\mathbb{P}[X_n = n^\alpha] = \mathbb{P}[X_n = -n^\alpha] = \dfrac{1}{2}\ (0 < \alpha < 1/2)$.

56. Giả sử $(X_n,\ n \geq 1)$ là dãy đại lượng ngẫu nhiên độc lập, cùng phân phối, có các phương sai dương, hữu hạn. Chứng minh rằng

a) $\lim_{n\to\infty} \mathbb{P}(a \leqslant S_n \leqslant b) = 0 \ \ (\forall a, b \in R)$.

b) $\lim_{n\to\infty} \mathbb{P}(S_n \leqslant a) = 0 \ \ (a \in R)$ hoặc bằng 0; hoặc bằng 1; hoặc bằng 1/2.

Trong đó $S_n = \sum_{k=1}^{n} X_k$.

57. Giả sử $(X_n,\ n \geq 1)$ là dãy đại lượng ngẫu nhiên có các phương sai

hữu hạn và có một dãy số $\alpha(0), \alpha(1), \ldots$ không âm sao cho

$$\rho(X_i, X_j) \leqslant \alpha(|i-j|) \quad \forall (i,j),$$

$$n^{-2}[\alpha(0) + \alpha(1) + \cdots + \alpha(n-1)][\sigma_1^2 + \cdots + \sigma_n^2] \to 0, \quad n \to \infty.$$

Chứng minh rằng dãy $(X_n,\ n \geq 1)$ tuân theo luật số lớn.

PHẦN II. THỐNG KÊ

Thống kê là khoa học về thu thập và xử lý số liệu để từ đó đưa ra các kết luận khoa học và thực tiễn. Sơ đồ tiến hành như sau:

Thu thập số liệu $\Rightarrow$ Tổng hợp số liệu $\Rightarrow$ Chuyển hóa về mô hình toán $\Rightarrow$ Xử lý $\Rightarrow$ Đưa ra kết luận.

CHƯƠNG 3

LÝ THUYẾT MẪU

3.1 Khái niệm mẫu và phương pháp lấy mẫu

Trong thực tế, nhiều khi ta cần quan tâm đến một số đặc điểm (định tính hoặc định lượng) của các phần tử thuộc về một tập hợp nào đó, chẳng hạn tuổi thọ của một loại đĩa cứng, giá thành bán lẻ của một loại mặt hàng nào đó, tỉ lệ nẩy mầm của một giống lúa, ... Tập hợp các phần tử cần nghiên cứu này được gọi là *đám đông*, ký hiệu là $\mathcal{C}$.

Việc tiến hành thu thập thông tin trên các phần tử của đám đông được gọi là *quan sát*. Đặc điểm cần quan tâm đó thay đổi từ phần tử này sang phần tử khác khi ta thực hiện các quan sát ngẫu nhiên trên một số phần tử của đám đông. Đặc điểm thay đổi đó của đám đông được coi như một đại lượng ngẫu nhiên, ký hiệu là X và được gọi là *đại lượng ngẫu nhiên gốc đám đông* $\mathcal{C}$. Quá trình đi nghiên cứu đám đông của $\mathcal{C}$ thực chất là quá trình đi tìm quy luật phân phối của đại lượng ngẫu nhiên X, nhiều khi đó là quá trình đi tìm các số đặc trưng của X. Nếu không gây nhầm lẫn ta có thể gọi ngắn gọn là đám đông X.

Đặc điểm của đám đông thường được nghiên cứu dưới hai phương diện:

$\diamond$ Phương diện định lượng: Khi ta cần quan tâm đến các giá trị về lượng của đại lượng ngẫu nhiên X như: trọng lượng, năng suất, tuổi thọ, ... và ta thường quan tâm đến hai đặc trưng:

- Kỳ vọng $\mathbb{E}X = \mu$: đặc trưng giá trị trung bình của đặc điểm định lượng cần quan tâm trên đám đông $\mathcal{C}$.

- Phương sai $\mathbb{D}X = \sigma^2$: đặc trưng cho mức độ biến động giá trị của đặc điểm định lượng cần quan tâm trên đám đông $\mathcal{C}$.

$\diamond$ Phương diện định tính: Khi ta cần quan tâm đến một tính chất A nào đó trên đám đông, các phần tử của đám đông hoặc có tính chất A hoặc không có tính chất A như: chất lượng sản phẩm, sự nẩy mầm của một giống lúa, chất độc hại trong nguồn nước, ... Giá trị mà đại lượng ngẫu nhiên X có thể nhận được

$$X = \begin{cases} 1 & \text{khi phần tử đó có tính chất } A\,; \\ 0 & \text{khi phần tử đó không có tính chất } A\,, \end{cases}$$

và ta thường quan tâm đến xác suất $\mathbb{E}X = p$.

3.1.1 Khái niệm mẫu

Chúng ta khó có thể quan sát hết tất cả các phần tử của đám đông vì những lý do như thời gian, chi phí tốn kém, ... Chính vì vậy, người ta chỉ lấy ra một số phần tử đại diện cho đám đông và nghiên cứu trên tập phần tử này, tập hợp các phần tử đại diện cho đám đông đó được gọi là *mẫu*. Phương pháp nghiên cứu trên mẫu đại diện cho đám đông được gọi là *phương pháp mẫu* và cách thức thực hiện quá trình lấy mẫu được gọi là *phương pháp lấy mẫu*.

Khi cần quan tâm đến đặc điểm là đại lượng ngẫu nhiên X của đám đông $\mathcal{C}$, ta chọn ra mẫu có n phần tử, trong đó việc chọn phần tử thứ i là quá trình thực hiện một phép thử rút ngẫu nhiên một phần tử của đám đông $\mathcal{C}$, giá trị ngẫu nhiên này được gán cho đại lượng ngẫu nhiên X_i. Với cách chọn này, các đại lượng ngẫu nhiên X_i độc lập với nhau và có cùng luật phân phối với đại lượng ngẫu nhiên X. Mẫu này được gọi là *mẫu ngẫu nhiên* có kích thước n của đám đông $\mathcal{C}$, ký hiệu $(X_1,\ X_2,\ ...,\ X_n)$. Tại lần lấy mẫu thứ i, giá trị mà X_i nhận được là x_i, bộ số $(x_1, x_2, ..., x_n)$ được gọi là *một mẫu cụ thể*.

Ví dụ 1. Thống kê về số chấm của một con xúc xắc khi gieo 5 lần.

Mẫu ngẫu nhiên: $(X_1,\ X_2, ...,\ X_5)$; mẫu cụ thể: $(2,\ 3,\ 1,\ 6,\ 2)$.

3.1.2 Các phương pháp lấy mẫu

Việc lấy mẫu được coi là tốt nếu như thông tin thu được từ mẫu phản ánh càng gần với đặc điểm của đám đông (tính chất đại diện cao). Chính vì vậy, trong thống kê việc lấy mẫu là một công việc hết sức quan trọng. Người ta thường sử dụng một số phương pháp lấy mẫu như sau:

Lấy mẫu ngẫu nhiên đơn giản

Là phương pháp lấy mẫu thỏa mãn các điều kiện: mỗi lần chỉ được chọn một phần tử từ đám đông, khả năng được chọn của tất cả các phần tử trong đám đông đều như nhau. Có hai cách thức tiến hành chọn, đó là chọn hoàn lại và chọn không hoàn lại, tuy nhiên khi kích thước của đám đông lớn hơn nhiều so với kích thước mẫu thì có thể coi hai phương pháp chọn này là giống nhau.

Phương pháp lấy mẫu ngẫu nhiên đơn giản ở trên có tính chất đại diện cho đám đông cao, tuy nhiên nó khó thực hiện và cần nhiều thời gian cũng như kinh phí. Ta có thể xem phương pháp lấy mẫu này là hoàn toàn ngẫu nhiên hay ngẫu nhiên không có định hướng.

Lấy mẫu ngẫu nhiên có định hướng

⋄ Lấy mẫu theo nhóm: là phương pháp chia đám đông thành các nhóm thuần nhất, từ mỗi nhóm này ta lấy ra một mẫu ngẫu nhiên đơn giản với một kích thước tương ứng. Tập hợp tất cả các phần tử thu được từ các mẫu ngẫu nhiên đơn giản đó lập nên mẫu ngẫu nhiên theo nhóm.

⋄ Lấy mẫu theo chùm: là phương pháp chia đám đông thành nhiều chùm (đám đông con) sao cho giữa các chùm có sự đồng đều về quy mô, từ các chùm đó ta lấy một mẫu ngẫu nhiên đơn giản. Tập hợp tất cả phần tử thu được từ các mẫu ngẫu nhiên đơn giản của các chùm lập nên mẫu ngẫu nhiên theo chùm.

Phương pháp này dễ quy hoạch, có thể tiết kiệm được thời gian và kinh phí nhưng sai số chọn mẫu cao hơn các phương pháp nói trên.

Ví dụ 2. Chúng ta muốn đi tìm hiểu về tổng thu nhập trong một năm của toàn bộ cán bộ công chức của một tỉnh.

- Chia đám đông này thành các nhóm theo từng cơ cấu ngành nghề: quốc phòng, an ninh, giáo dục, y tế, kinh doanh, ... Trong mỗi cơ cấu ngành nghề có sự thuần nhất về mức lương (nếu có sự sai khác về thu nhận chủ yếu là do thâm niên và chức vụ công tác). Như vậy, phương pháp lấy mẫu bằng việc gom lại các mẫu ngẫu nhiên đơn giản của từng nhóm ngành nghề chính là phương pháp lấy mẫu theo nhóm.

- Chia đám đông này theo các huyện trong tỉnh A. Giữa các huyện, có sự đồng đều về quy mô (đầy đủ các thành phần) và phương pháp lấy mẫu bằng việc gom lại các mẫu ngẫu nhiên đơn giản của từng huyện chính là phương pháp lấy mẫu theo chùm.

3.2 Cách biểu diễn mẫu

3.2.1 Bảng tần số và bảng tần suất

Ta thực hiện n lần quan sát trên đám đông $\mathcal{C}$, khi đó ta sẽ thu được mẫu cụ thể gồm k giá trị khác nhau $(x_1, x_2, ..., x_k)$, $k \leqslant n$. Giá trị x_i có n_i lần xuất hiện, n_i được gọi là *tần số xuất hiện* của x_i và tỉ số $\dfrac{n_i}{n}$ được gọi là *tần suất xuất hiện* của x_i, ký hiệu là f_i. Ta có biểu diễn kết quả của mẫu bằng bảng tần số và tần suất như sau

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

x_i	x_1	x_2	...	x_k
f_i	f_1	f_2	...	f_k

trong đó

$$n = \sum_{i=1}^{k} n_i; \quad \sum_{i=1}^{k} f_i = 1.$$

Ví dụ 1. Thống kê điểm số kết thúc học phần của một lớp gồm 40 sinh viên, ta có

x_i	4	5	6	7	8
n_i	5	10	12	8	5

x_i	4	5	6	7	8
f_i	5/40	10/40	12/40	8/40	5/40

Trong trường hợp mẫu cụ thể $(x_1, x_2, ..., x_n)$ có nhiều giá trị khác nhau, khi đó ta thực hiện việc ghép lớp. Nguyên tắc ghép lớp được tiến hành như sau

- Số lớp chia k được xác định trên cơ sở $k = \min\{l : 2^l > n\}$.
- Độ dài mỗi lớp: $l = \dfrac{\text{giá trị lớn nhất - giá trị nhỏ nhất}}{k}$.
- Trong 2 lớp liền nhau $x_{i-1} \to x_i$, $x_i \to x_{i+1}$ thì x_i thuộc lớp $x_{i-1} \to x_i$.

Ngoài phương pháp ghép lớp đã trình bày ở trên, còn có một số phương pháp ghép lớp khác, với những mẫu cụ thể rời rạc người ta có thể chia thành các lớp có độ dài khác nhau, các lớp được chia rời nhau. Trong phạm vi giáo trình này, chúng ta không đề cập cụ thể các kiểu ghép lớp này.

Ví dụ 2. Thống kê về chiều cao của 30 sinh viên với chiều cao nằm trong khoảng từ 1m50 đến 1m75.

Nhận thấy $2^5 > 30$ và $2^4 < 30$ nên ta chọn $k = 5$. Bảng tần số, tần suất như sau:

Lớp	Giá trị	Tần số	Tần suất
150-155	152,5	4	4/30
155-160	157,5	7	7/30
160-165	162,5	6	6/30
165-170	167,5	10	10/30
170-175	172,5	3	3/30

3.2.2 Đa giác tần số và tổ chức đồ

Đối với số liệu chưa ghép lớp

- Chấm trên mặt phẳng các điểm $(x_i, n_i), i = 1, 2, \ldots, n$.

- Nối các điểm $(x_i, 0)$ với các điểm (x_i, n_i), ta được *biểu đồ tần số hình gậy*.

- Nối liên tiếp điểm (x_i, n_i) với các điểm (x_{i+1}, n_{i+1}) ta được *biểu đồ đa giác tần số*.

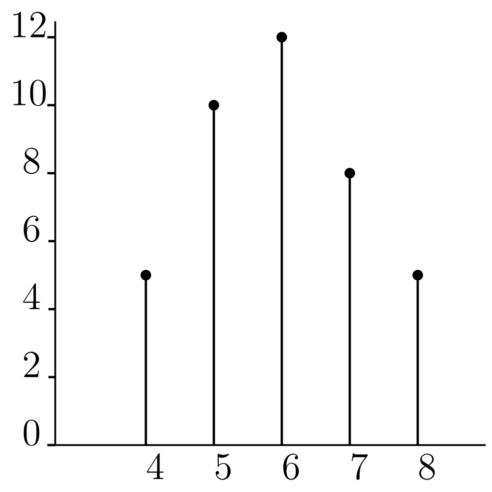
Hoàn toàn tương tự đối với tần suất:

- Chấm trên mặt phẳng các điểm $(x_i, f_i), i = 1, 2, \ldots, n$.

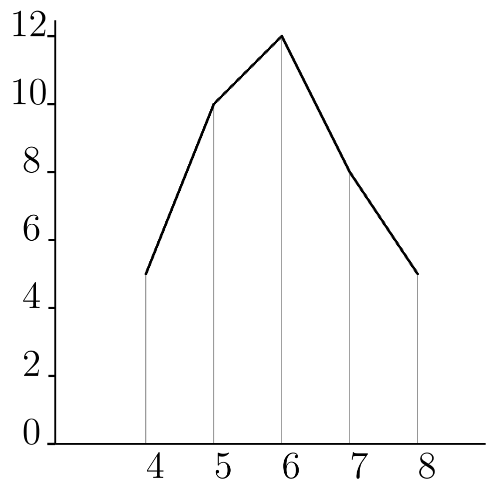
- Nối các điểm $(x_i, 0)$ với các điểm (x_i, f_i), ta được *biểu đồ tần suất hình gậy*.

- Nối liên tiếp điểm (x_i, f_i) với các điểm (x_{i+1}, f_{i+1}) ta được *biểu đồ đa giác tần suất*.

Ví dụ 3. Minh họa số liệu của ví dụ thống kê điểm



Biểu đồ tần số hình gậy



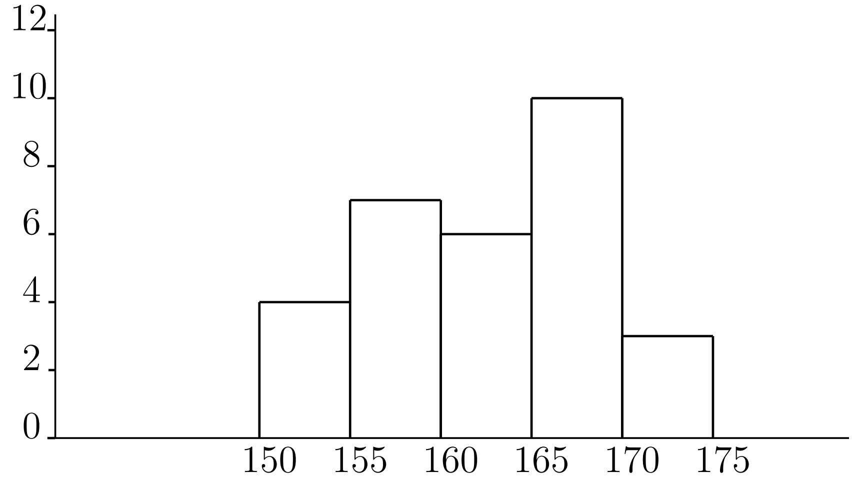
Biểu đồ đa giác tần số

Đối với số liệu đã ghép lớp.

- Trên mỗi lớp ta dựng hình chữ nhật có chiều cao bằng tần số (hay tần suất) tương ứng với lớp đó.

- Tô đậm hoặc kẻ chéo bằng các đường song song các hình chữ nhật này ta thu được *tổ chức đồ tần số* (hay *tổ chức đồ tần suất*).

Ví dụ 4. Minh họa số liệu của Ví dụ 2.



Biểu đồ đa giác tần số

3.3 Các đặc trưng của mẫu

Trong nội dung Chương 2 chúng ta đã được làm quen với việc tính các đặc trưng của đại lượng ngẫu nhiên thông qua phân phối xác suất đã biết trước.

Tuy nhiên, trong thực tế thật khó khăn để xác định được tường minh phân phối xác suất của một đại lượng ngẫu nhiên gốc đám đông. Chính vì vậy, trên cơ sở của các thông tin thu thập được từ các mẫu, người ta đem ra một số công thức giúp chúng ta tính được các đặc trưng của mẫu.

Các giá trị này rất quan trọng và có sự tương ứng với những số đặc trưng của đại lượng ngẫu nhiên đã trình bày ở phần trước.

3.3.1 Hàm phân phối mẫu

X là đại lượng ngẫu nhiên gốc đám đông có hàm phân phối xác suất $F(x)$ chưa biết. Khi ta thực hiện n quan sát, gọi hàm $F_n(x) = \dfrac{m_x}{n}$ với m_x: là số quan sát có giá trị x_i bé hơn x ($i = \overline{1,n}$) là *hàm phân phối mẫu*.

Tính chất của hàm phân phối mẫu $F_n(x)$:

+ $0 \leqslant F_n(x) \leqslant 1$,

+ $F_n(x)$ là hàm đơn điệu tăng,

+ $F_n(x)$ là hàm liên tục bên trái.

Khi kích thước mẫu lớn thì phân phối mẫu $F_n(x)$ càng gần với phân phối xác suất của đại lượng ngẫu nhiên X. Khi n đủ lớn, ta có thể dùng $F_n(x)$ thay thế cho $F(x)$ chưa biết hoặc dựa vào $F_n(x)$ ta có thể sơ lược về dáng điệu của $F(x)$ và đưa ra những dự đoán về dạng của $F(x)$ cũng như tính toán các số đặc trưng có liên quan.

Ví dụ 1. Bảng tần số từ ví dụ thống kê điểm

x_i	4	5	6	7	8
n_i	5	10	12	8	5

Hàm phân phối mẫu

$$F_n(x) = \begin{cases} 0 & \text{với } x \leqslant 4, \\ \dfrac{5}{40} & \text{với } 4 < x \leqslant 5, \\ \dfrac{15}{40} & \text{với } 5 < x \leqslant 6, \\ \dfrac{27}{40} & \text{với } 6 < x \leqslant 7, \\ \dfrac{35}{40} & \text{với } 7 < x \leqslant 8, \\ 1 & \text{với } x > 8. \end{cases}$$

3.3.2 Trung bình mẫu

Định nghĩa. Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên có kích thước n của đám đông X, khi đó $\frac{1}{n}\sum_{i=1}^{n} X_i$ được gọi là *trung bình mẫu* và ký hiệu là $\overline{X}$.

Trong thực hành tính toán

Đối với một mẫu cụ thể $(x_1, x_2, ..., x_n)$ thì trung bình mẫu thực nghiệm được xác định như sau $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Trường hợp mẫu cụ thể đã được ghép bộ có bảng tần số

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

thì trung bình mẫu thực nghiệm là $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i$.

Ví dụ 2. Bảng tần số từ ví dụ thống kê điểm

x_i	4	5	6	7	8
n_i	5	10	12	8	5

Khi đó $\overline{x} = \dfrac{1}{40}\sum_{i=1}^{5} n_i\, x_i = \dfrac{238}{40} = 5{,}95.$

Nhận xét. Công thức tính trung bình mẫu ở trên là dạng tổng quát, tuy nhiên do đặc trưng số nên ta thường dùng khi nghiên cứu về một đặc điểm định lượng nào đó của đám đông. Đối với đặc điểm định tính A ta có khái niệm tỉ lệ mẫu

$$F = \frac{1}{n}\sum_{i=1}^{n} X_i$$

trong đó X_i chỉ nhận 2 giá trị là 0 và 1 (bằng 1 nếu quan sát đó có tính chất A, bằng 0 nếu quan sát đó không có tính chất A). Với $m = \sum_{i=1}^{n} X_i$ chính là số quan sát có tính chất A, công thức tính tỉ lệ mẫu là $F = \dfrac{m}{n}$.

3.3.3 Phương sai mẫu và phương sai hiệu chỉnh mẫu

Định nghĩa. Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên có kích thước n của đám đông X, khi đó $\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ được gọi là *phương sai mẫu* và ký hiệu là $\hat{S}^2$.

Ngoài ra, chúng ta thường dùng một đặc trưng mẫu khá quan trọng là *phương sai hiệu chỉnh mẫu*, ký hiệu là S^2, được xác định $S^2 = \dfrac{n}{n-1}\hat{S}^2$.

Mệnh đề. *Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên có kích thước n của đám đông X. Ta có*

$$\hat{S}^2 = \overline{X^2} - (\overline{X})^2 \quad \textit{trong đó} \quad \overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Chứng minh.

$$\begin{aligned}\hat{S}^2 =& \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i^2 - 2X_i\overline{X} + (\overline{X})^2 \\ =& \overline{X^2} - \frac{2}{n}\overline{X}\sum_{i=1}^{n}X_i + (\overline{X})^2 = \overline{X^2} - (\overline{X})^2.\end{aligned}$$

□

Trong thực hành tính toán

Đối với một mẫu cụ thể đã được ghép bộ có bảng tần số

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

thì *phương sai mẫu thực nghiệm* và *phương sai hiệu chỉnh mẫu thực nghiệm* được xác định như sau

$$\hat{s}^2 = \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i - \overline{x}\right)^2 = \overline{x^2} - \left(\overline{x}\right)^2;$$

$$s^2 = \frac{n}{n-1}\,\hat{s}^2 = \frac{n}{n-1}\left(\overline{x^2} - \left(\overline{x}\right)^2\right).$$

s được gọi là *độ lệch chuẩn mẫu.*

Việc đưa ra các khái niệm trung bình mẫu thực nghiệm (phương sai mẫu thực nghiệm, phương sai hiệu chỉnh mẫu thực nghiệm) chỉ nhằm nhấn mạnh đó là giá trị bằng số cụ thể, được xác định từ thực nghiệm.

Ví dụ 3. Bảng tần số từ ví dụ thống kê điểm

x_i	4	5	6	7	8
n_i	5	10	12	8	5

x_i	n_i	$n_i\,x_i$	$n_i\,x_i^2$
4	5	20	80
5	10	50	250
6	12	72	432
7	8	56	392
8	5	40	320
Tổng	**40**	**238**	**1474**

Ta có $\overline{x} = \dfrac{238}{40} = 5{,}95;\ \overline{x}^2 = \dfrac{1474}{40} = 36{,}85.$

$\hat{s}^2 = 36{,}85 - 5{,}95^2 = 1{,}4475;\ s^2 \approx 1{,}485.$

Chú ý. Đối với mẫu được ghép lớp, việc tính các số đặc trưng của mẫu cũng theo trình tự tiến hành như trên, trong mỗi lớp ta sử dụng giá trị trung điểm $x_i' = \dfrac{x_i + x_{i+1}}{2}$ của lớp.

Các phân phối xác suất của các đặc trưng mẫu

◇ Trường hợp đám đông X có phân phối chuẩn $\mathcal{N}(\mu,\, \sigma^2)$ và σ đã biết

$$\overline{X} \sim \mathcal{N}(\mu,\, \frac{\sigma^2}{n})\ ;\quad \frac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,\, 1).$$

◇ Trường hợp đám đông X có phân phối chuẩn $\mathcal{N}(\mu,\, \sigma^2)$, σ chưa biết và $n < 30$

$$\frac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1).$$

◇ Trường hợp đám đông X không có phân phối chuẩn và $n \geqslant 30$

- Khi σ^2 đã biết: $\dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} \simeq \mathcal{N}(0,\, 1).$

- Khi σ^2 chưa biết: $\dfrac{\overline{X} - \mu}{S}\sqrt{n} \simeq \mathcal{N}(0,\, 1).$

- Khi p đã biết và $np \geqslant 5$; $n(1-p) \geqslant 5$ đủ lớn: $\dfrac{F-p}{\sqrt{p\,(1-p)}}\sqrt{n} \simeq \mathcal{N}(0,\, 1)$.

- Khi p chưa biết và n đủ lớn: $\dfrac{F-p}{\sqrt{F(1-F)}}\sqrt{n} \simeq \mathcal{N}(0,\, 1)$.

BÀI TẬP CHƯƠNG 3

1. Cho ví dụ về đám đông, một số đặc điểm có thể nghiên cứu và các phương pháp thực hiện việc lấy mẫu trên đám đông đó.

2. Phân biệt sự khác nhau giữa mẫu ngẫu nhiên và mẫu cụ thể, cho ví dụ minh họa.

3. Phân biệt sự khác nhau giữa đặc điểm định lượng và đặc điểm định tính. Cho ví dụ về hai đặc điểm cùng nghiên cứu trên một đám đông.

4. Khi đo độ dài của 36 chi tiết được lấy ngẫu nhiên từ một loại sản phẩm, người ta thu được bảng số liệu sau đây:

   15 14 16 14 15 12 13 16 13 12 15 13 16 13 15

   13 16 13 16 13 15 12 15 15 14 14 15 15 16 15

   a. Lập bảng tần số và bảng tần suất.

   b. Vẽ biểu đồ đa giác tần số và tần suất.

   c. Tìm hàm phân phối mẫu.

5. Dưới đây là số liệu được lấy ngẫu nhiên về thời gian đợi của các khách hành (tính bằng giây) tại quầy thanh toán tiền ở một siêu thị đối với 48 khách hàng:

3	24	34	5	14	22	3	19	13	32	19	4	24	30	48	24
14	16	3	4	5	14	19	41	43	16	48	4	58	13	10	60
12	14	14	22	3	16	14	4	34	32	4	19	12	24	13	26

a. Lập bảng tần số ghép lớp và bảng tần suất ghép lớp.

b. Vẽ bảng tổ chức đồ tần số và tần suất.

c. Tính trung bình mẫu, phương sai mẫu và phương sai hiệu chỉnh mẫu.

6. Mẫu điều tra có kích thước 35 đối với hai đặc điểm X và Y của một loại sản phẩm được kết quả cho bởi bảng số liệu dưới đây:

$X \backslash Y$	64	65	66
6-10	3	8	3
10-14	0	5	2
14-16	6	1	0
16-20	0	3	4

a. Lập bảng tần số, tần suất của Y.

b. Những sản phẩm được gọi là đạt chất lượng nếu $X \leqslant 16$ và $Y > 64$. Tính tỉ lệ sản phẩm đạt chất lượng.

c. Lập bảng tần số và tính trung bình mẫu của chỉ tiêu Y đối với các sản phẩm có $X > 10$.

7. Cơ quan quản lý thị trường lấy số liệu về giá thành bán lẻ của một loại sản phẩm tại 40 đại lý (đơn vị: ngàn), người ta thu được bảng tần số như sau:

x_i	19	20	21	22
n_i	8	16	6	10

a. Tìm hàm phân phối mẫu.

b. Tính trung bình mẫu và độ lệch chuẩn mẫu.

8. Tìm hàm phân phối mẫu, trung bình mẫu, phương sai hiệu chỉnh mẫu đối với hai mẫu cụ thể sau:

a.

x_i	19,2	19,8	20,1	20,3	20,7
n_i	6	2	4	2	6

b.

x_i	460	480	490	505
n_i	5	6	10	4

9. Điều tra ngẫu nhiên ý kiến của 2500 số khách hàng thường xuyên đi xe taxi về chất lượng phục vụ của 3 hãng taxi thu được kết quả sau đây:

Chất lượng phục vụ	Hãng taxi		
	A	B	C
Rất tốt	140	110	205
Khá	230	150	350
Bình thường	350	225	520
Kém	80	15	125

Hãy tính đặc trưng mẫu cho từng hãng taxi và nêu đánh giá sơ bộ từ số liệu điều tra trên.

CHƯƠNG 4

ƯỚC LƯỢNG THAM SỐ

Giả sử đại lượng ngẫu nhiên X có luật phân phối phụ thuộc vào một tham số hoặc một véctơ tham số θ chưa biết. Khi đó để xác định hoàn toàn phân phối xác suất của X ta phải xác định được giá trị tham số θ. Đây chính là *bài toán ước lượng tham số*. Chẳng hạn biết X là đại lượng ngẫu nhiên có phân phối Poisson nhưng chưa biết tham số λ là bao nhiêu hoặc Y là đại lượng ngẫu nhiên có phân phối chuẩn nhưng chưa xác định được (μ, σ), ...

Chính vì vậy bài toán ước lượng tham số của đại lượng ngẫu nhiên là rất cần thiết.

4.1 Ước lượng điểm

4.1.1 Định nghĩa

Giả sử X là đại lượng ngẫu nhiên gốc đám đông $\mathcal{C}$, có tham số θ cần ước lượng. Thực hiện n lần quan sát độc lập ta thu được mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$, để ước lượng tham số θ ta phải tìm ra một hàm mẫu thống kê $\hat{\theta}(X_1, X_2, ..., X_n)$ "đủ tốt", chỉ phụ thuộc vào các quan sát mà không phụ thuộc vào θ được gọi là *bài toán ước lượng điểm của θ* và $\hat{\theta}$ được gọi là *ước lượng điểm của θ*.

Do giá trị đúng của θ là chưa biết, nên ta không thể so sánh trực tiếp giá trị của $\hat{\theta}$ và θ mà chỉ đưa ra một số tiêu chuẩn để đánh giá ước lượng. Trong các loại ước lượng điểm, ta thường quan tâm đến bốn loại ước lượng sau đây:

◇ Ước lượng $\hat{\theta}(X_1, X_2, ..., X_n)$ được gọi là *ước lượng không chệch* của θ, nếu thỏa mãn $\mathbb{E}\hat{\theta} = \theta$.

◇ Ước lượng $\hat{\theta}(X_1, X_2, ..., X_n)$ được gọi là *ước lượng vững* của θ, nếu với n lớn vô hạn thì $\hat{\theta}$ hội tụ theo xác suất về θ, nghĩa là với mọi $\varepsilon > 0$ tùy ý thì

$$\lim_{n\to\infty} \mathbb{P}[|\hat{\theta} - \theta| < \varepsilon] = 1.$$

◇ Ước lượng $\hat{\theta}(X_1, X_2, ..., X_n)$ được gọi là *ước lượng hợp lý tối đa* của θ, nếu

$$L(x, \theta) = \prod_{i=1}^{n} p(X_i, \theta)$$

đạt cực đại tại $\hat{\theta}$. $L(x, \theta)$ được gọi là *hàm hợp lý* của X, trong đó $p(x, \theta)$ là hàm mật độ xác suất hoặc là hàm tính xác suất của đại lượng ngẫu nhiên X.

◇ Ước lượng $\hat{\theta}(X_1, X_2, ..., X_n)$ được gọi là *ước lượng hiệu quả* của θ, nếu như nó là ước lượng không chệch và có phương sai bé nhất trong tất cả các ước lượng không chệch của θ.

Nếu hàm mật độ xác suất của đại lượng ngẫu nhiên X thỏa mãn thêm một số điều kiện nhất định thì ta có bất đẳng thức Cramer-Rao

$$D(\theta^*) \geqslant \frac{1}{n\mathbb{E}\Big(\dfrac{\partial \ln p(X,\, \theta)}{\partial \theta}\Big)^2} \; ; \;\; \forall \theta^* \; : \;\; \mathbb{E}(\theta^*) = \theta.$$

do đó, ước lượng không chệch $\hat{\theta}$ là ước lượng hiệu quả của θ khi

$$V(\hat{\theta}) = \frac{1}{n\mathbb{E}\Big(\dfrac{\partial \ln p(X,\, \theta)}{\partial \theta}\Big)^2}.$$

Từ bất đẳng thức Cramer-Rao, ta thấy một điều lý thú đó là: đã là ước lượng thì phải chấp nhận sai số, bất đẳng thức cho ta cận dưới của sai số.

4.1.2 Ước lượng điểm cho kỳ vọng, xác suất và phương sai

Ước lượng điểm cho kỳ vọng

Mệnh đề. *Giả sử X là đại lượng ngẫu nhiên gốc đám đông $\mathcal{C}$, có kỳ vọng μ cần ước lượng, khi đó trung bình mẫu $\overline{X}$ chính là ước lượng không chệch của μ.*

Chứng minh. Thật vậy, vì $X_i, i = \overline{1, n}$ có cùng phân phối với đại lượng ngẫu nhiên X nên

$$\mathbb{E}\overline{X} = \frac{\mathbb{E}X_1 + \mathbb{E}X_2 + ... + \mathbb{E}X_n}{n} = \mathbb{E}X = \mu.$$

□

Ngoài ra, người ta còn chứng minh được trung bình mẫu $\overline{X}$ đồng thời còn là ước lượng vững và ước lượng hiệu quả của μ.

Ví dụ 1. Nếu X là đại lượng ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu,\, \sigma^2)$ thì $\overline{X}$ là ước lượng hiệu quả của μ.

Giải. Vì $\overline{X}$ là ước lượng không chệch của μ nên ta chỉ cần chứng minh nó có phương sai bé nhất trong các ước lượng không chệch khác của μ.

Ở đây $\theta = \mu$ cần ước lượng, hàm mật độ của phân phối chuẩn tắc có dạng

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

Khi đó

$$\frac{\partial \ln p(x, \mu)}{\partial \mu} = \frac{\partial}{\partial \mu}\Big(- \ln(\sigma\sqrt{2\pi}) - \frac{(x-\mu)^2}{2\sigma^2}\Big) = \frac{x-\mu}{\sigma^2}.$$

Vì vậy

$$\frac{1}{n\mathbb{E}\Big(\dfrac{\partial \ln p(X,\, \mu)}{\partial \mu}\Big)^2} = \frac{1}{n\mathbb{E}\big(\frac{X-\mu}{\sigma^2}\big)^2} = \frac{\sigma^4}{n\mathbb{D}X} = \frac{\sigma^2}{n} = \mathbb{D}\overline{X}.$$

Sử dụng bất đẳng thức Cramer-Rao, ta suy ra $\overline{X}$ là ước lượng hiệu quả của μ.

Ước lượng điểm cho phương sai

Mệnh đề. *Giả sử X là đại lượng ngẫu nhiên gốc đám đông $\mathcal{C}$, có phương sai $\mathbb{D}X = \sigma^2$ cần ước lượng, khi đó phương sai hiệu chỉnh mẫu S^2 chính là ước lượng không chệch của σ^2.*

Chứng minh. Thật vậy, vì $X_i, i = \overline{1, n}$ có cùng phân phối với đại lượng ngẫu nhiên X nên

$$\mathbb{E}S^2 = \mathbb{E}\Big(\frac{n}{n-1}\big(\overline{X^2} - (\overline{X})^2\big)\Big) = \frac{1}{n-1}\big(\sum_{i=1}^{n} \mathbb{E}X_i^2 - n\mathbb{E}(\overline{X})^2\big).$$

Mặt khác

$$\mathbb{E}X_i^2 = \mathbb{D}X_i + (\mathbb{E}X_i)^2 = \sigma^2 + \mu^2;$$
$$\mathbb{E}(\overline{X})^2 = \frac{1}{n^2}\big(\sum_{i=1}^{n} \mathbb{E}X_i^2 + \sum_{i \neq j} \mathbb{E}X_i\mathbb{E}X_j\big) = \frac{\sigma^2+\mu^2}{n} + \frac{n-1}{n}\mu^2,$$

suy ra $\mathbb{E}S^2 = \dfrac{1}{n-1}\big[n(\sigma^2+\mu^2) - (\sigma^2+\mu^2+(n-1)\mu^2)\big] = \sigma^2.$ □

Như vậy S^2 là ước lượng không chệch của σ^2. Mặt khác $\hat{S}^2 = \dfrac{n-1}{n} S^2$ nên $\hat{S}^2$ không phải là ước lượng không chệch của σ^2. Tuy nhiên người ta chứng minh được rằng cả S^2 và $\hat{S}^2$ đều là ước lượng vững của σ^2.

Ước lượng hợp lý tối đa được xác định cho từng trường hợp cụ thể. Ví dụ sau là dạng ước lượng kỳ vọng và phương sai cho đại lượng ngẫu nhiên có phân phối chuẩn.

Ví dụ 2. Nếu X là đại lượng ngẫu nhiên có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ thì $\overline{X}$ và $\hat{S}^2$ lần lượt là ước lượng hợp lý tối đa của μ và σ^2.

Giải. Hàm hợp lý

$$L(x, \theta) = \frac{1}{(\sigma\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i-\mu)^2},$$

suy ra

$$\ln L(x, \theta) = -\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu)^2 - \frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2.$$

Việc tìm cực đại hàm $\ln L(x, \theta)$ dẫn đến hệ phương trình

$$\begin{cases} \dfrac{\partial \ln L(x,\theta)}{\partial \mu} = \displaystyle\sum_{i=1}^n \frac{X_i - \mu}{\sigma^2} = 0; \\ \dfrac{\partial \ln L(x,\theta)}{\partial \sigma^2} = \displaystyle\sum_{i=1}^n \frac{(X_i - \mu)^2}{2\sigma^4} - \frac{n}{2\sigma^2} = 0. \end{cases}$$

Do đó $\overline{X}$ và $\hat{S}^2$ lần lượt là ước lượng hợp lý tối đa của μ và σ^2.

Ước lượng điểm cho xác suất

Mệnh đề. *Giả sử X là đại lượng ngẫu nhiên gốc đám đông $\mathcal{C}$, ta cần quan tâm đến một tính chất A có xác suất $p = \mathbb{P}(A) = \mathbb{E}X$ cần ước lượng, khi đó tỉ lệ mẫu F chính là ước lượng không chệch của xác suất p.*

Khẳng định trên là hiển nhiên vì thực chất tỉ lệ mẫu cũng là trung bình mẫu khi đặc điểm định tính được số hóa dưới dạng

$$X_i = \begin{cases} 1 & \text{khi phần tử đó có tính chất } A\,; \\ 0 & \text{khi phần tử đó không có tính chất } A\,, \end{cases}$$

và $\mathbb{E}F = \mathbb{E}\overline{X} = \mathbb{E}X = p$.

Ngoài ra người ta còn chứng minh được F cũng chính là ước lượng vững của xác suất p.

4.2 Ước lượng khoảng

Trong nội dung của phần trước, chúng ta đã đề cập đến ước lượng điểm của tham số. Do θ là tham số chưa biết nên ước lượng điểm chỉ cho ta một cách nhìn hết sức tương đối và có phần chưa thỏa đáng. Sau đây chúng ta sẽ suy nghĩ đến một cách tiếp cận khác để tìm ra miền giá trị của θ.

4.2.1 Khái niệm về khoảng tin cậy

Cho X là đại lượng ngẫu nhiên gốc đám đông $\mathcal{C}$, có tham số θ cần ước lượng. Căn cứ vào mẫu ngẫu nhiên từ n quan sát độc lập $(X_1,\, X_2,\, ...,\, X_n)$, ta cần đưa ra khoảng $(\theta_1,\, \theta_2)$ chứa được hầu hết các giá trị θ với xác suất lớn, nghĩa là

$$\mathbb{P}(\theta_1 < \theta < \theta_2) = 1 - \alpha\,.$$

Một số khái niệm

◇ $(\theta_1,\, \theta_2)$: được gọi là *khoảng tin cậy* của ước lượng.

◇ $\theta_1 - \theta_2 = 2\varepsilon$: được gọi là *độ dài khoảng tin cậy* của ước lượng.

◇ ε: được gọi là *độ chính xác* của ước lượng.

◇ $1-\alpha$: được gọi là *độ tin cậy* của của ước lượng.

◇ Bài toán đi tìm khoảng tin cậy cho tham số θ với độ tin cậy $1-\alpha$ được gọi là *bài toán ước lượng khoảng tin cậy.*

4.2.2 Khoảng tin cậy cho giá trị trung bình

Cho X là đại lượng ngẫu nhiên gốc đám đông $\mathcal{C}$, có trung bình $\mathbb{E}X=\mu$ cần ước lượng và phương sai $\mathbb{D}X=\sigma^2$ (đã biết trước hoặc chưa biết), từ mẫu ngẫu nhiên $(X_1, X_2, ..., X_n)$ ta xác định được $\overline{X}$.

a. Ước lượng hai phía

Vấn đề đặt ra ở đây là với độ tin cậy $1-\alpha$ cho trước, tìm khoảng ước lượng $(\overline{X}-\varepsilon, \overline{X}+\varepsilon)$ của μ để

$$\mathbb{P}[\overline{X}-\varepsilon<\mu<\overline{X}+\varepsilon]=1-\alpha.$$

Ta chia bài toán thành 3 trường hợp để giải quyết:

Trường hợp 1. Phương sai σ^2 đã biết.

Khi đó $\dfrac{\overline{X}-\mu}{\sigma}\sqrt{n}\simeq\mathcal{N}(0,1)$, đặt $t_{\alpha/2}=\varphi^{-1}(1-\frac{\alpha}{2})$, trong đó φ là hàm phân phối chuẩn $\mathcal{N}(0,1)$ và $t_{\alpha/2}$ là mức phân vị $\alpha/2$ cho phân phối chuẩn. Ta có

$$\begin{aligned}\mathbb{P}\big[-t_{\alpha/2}<\frac{\overline{X}-\mu}{\sigma}\sqrt{n}<t_{\alpha/2}\big]&=\varphi(t_{\alpha/2})-\varphi(-t_{\alpha/2})\\&=\varphi(t_{\alpha/2})-(1-\varphi(t_{\alpha/2}))=1-\alpha,\end{aligned}$$

hay $\mathbb{P}\Big[\overline{X}-t_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}<\mu<\overline{X}+t_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\Big]=1-\alpha.$

Quy tắc thực hành

◇ Xác định mức phân vị $t_{\alpha/2}$

Tính giá trị $1 - \frac{\alpha}{2}$, tra bảng hàm phân phối $\mathcal{N}(0,1)$ (xem bảng 4 phần phụ lục), tra từ giữa ra hai biên.

◇ Xác định khoảng ước lượng $(\overline{x} - \varepsilon, \overline{x} + \varepsilon)$ với độ chính xác của ước lượng

$$\varepsilon = t_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Chú ý. Nếu như kích thước mẫu $n < 30$ cần bổ sung thêm điều kiện X tuân theo luật phân phối chuẩn, khi đó $\dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)$.

Ví dụ 1. Tìm khoảng ước lượng cho giá trị trung bình với độ tin cậy 95% từ mẫu của một đám đông tuân theo luật phân phối chuẩn, $\sigma^2 = 16$. Biết mẫu đó có kích thước 16 và trung bình mẫu là 15.

Giải. $\sigma^2 = 16$, $n = 15$; $\overline{x} = 15$; $\alpha =$0,05 tra bảng hàm phân phối chuẩn ứng với $1 - \alpha/2 =$0,975 được $t_{\alpha/2} =$1,96. Độ chính xác của ước lượng

$$\varepsilon = t_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 1{,}96\frac{4}{\sqrt{16}} = 1{,}96.$$

Khoảng ước lượng cho giá trị trung bình:

$$(15 - 1{,}96 < \mu < 15 + 1{,}96) \text{ hay } (13{,}04 < \mu < 16{,}96).$$

Trường hợp 2. Phương sai σ^2 chưa biết và $n \geqslant 30$.

Khi đó $\dfrac{\overline{X} - \mu}{S}\sqrt{n} \simeq \mathcal{N}(0,1)$, việc thiết lập tương tự như ở trường hợp 1, ta được

$$\mathbb{P}\Big[\overline{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}\Big] = 1 - \alpha\,.$$

Như vậy, với một mẫu cụ thể, ta sẽ xác định được độ chính xác của ước lượng $\varepsilon = t_{\alpha/2}\dfrac{s}{\sqrt{n}}$ và khoảng ước lượng

$$(\overline{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}).$$

Ví dụ 2. Để ước lượng khối lượng trung bình mỗi bao xi măng của nhà máy, người ta kiểm tra ngẫu nhiên 49 bao thu được khối lượng trung bình là 49,7kg và độ lệch chuẩn mẫu 0,5kg. Với độ tin cậy là 94%, hãy ước lượng khoảng khối lượng trung bình của một bao xi măng.

Giải. α =0,06, $t_{\alpha/2}$ =1,88. Độ chính xác của ước lượng

$$\varepsilon = t_{\alpha/2}\frac{s}{\sqrt{n}} = 1{,}88\frac{0{,}5}{\sqrt{49}} = 0{,}13.$$

Khoảng ước lượng cho giá trị trung bình: (49,57 $< \mu <$ 49,83).

Trường hợp 3. Phương sai σ^2 chưa biết và $n < 30$.

Nếu $X \sim \mathcal{N}(\mu, \sigma^2)$ thì $\dfrac{\overline{X} - \mu}{S}\sqrt{n} \sim t(n-1)$. Mức phân vị $\alpha/2$ cho phân phối Student với $n-1$ bậc tự do ký hiệu là $t_{(n-1,\alpha/2)}$ là giá trị thỏa mãn $\mathbb{P}(\dfrac{\overline{X} - \mu}{S}\sqrt{n} > t_{(n-1,\alpha/2)}) = \alpha/2$. Khi đó

$$\begin{aligned}
&\mathbb{P}\big[-t_{(n-1,\alpha/2)} < \frac{\overline{X} - \mu}{S}\sqrt{n} < t_{(n-1,\alpha/2)}\big]\\
&= \mathbb{P}\big[t_{(n-1,1-\alpha/2)} < \frac{\overline{X} - \mu}{S}\sqrt{n} < t_{(n-1,\alpha/2)}\big]\\
&= 1 - \alpha/2 - \alpha/2 = 1 - \alpha.
\end{aligned}$$

Quy tắc thực hành

$\diamond$ Xác định mức phân vị $t_{(n-1,\alpha/2)}$.

Tra bảng phân phối Student (xem bảng 5 phần phụ lục), $t_{(n-1,\alpha/2)}$ là giá trị trong bảng ứng với giá trị hàng là $n-1$ và cột là $\alpha/2$.

$\diamond$ Xác định khoảng ước lượng $(\overline{x} - \varepsilon, \overline{x} + \varepsilon)$ với độ chính xác của ước lượng

$$\varepsilon = t_{(n-1,\alpha/2)}\frac{s}{\sqrt{n}}.$$

Ví dụ 3. Độ chịu lực của mỗi tấm bê tông tuân theo luật phân phối chuẩn. Đo độ chịu lực của 20 tấm bê tông cùng loại người ta thu được trung bình mẫu độ chịu lực 220kg/cm^2 và độ lệch chuẩn mẫu 32,4kg/cm^2. Với độ tin cậy 90%, tìm khoảng ước lượng trung bình độ chịu lực của mỗi tấm bê tông.

Giải. Tra bảng hàm phân phối Student ta được $t_{(19;0,05)}$ =1,729. Độ chính xác của ước lượng

$$\varepsilon = t_{(n-1,\alpha/2)}\frac{s}{\sqrt{n}} \approx 12{,}5.$$

Khoảng ước lượng cho giá trị trung bình: $(187{,}5 < \mu < 212{,}5)$.

Các dạng toán phát sinh

Xuất phát từ các công thức tương ứng với từng trường hợp

$$\varepsilon = t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ; \quad \varepsilon = t_{\alpha/2}\frac{s}{\sqrt{n}}\ ; \quad \varepsilon = t_{(n-1,\,\alpha/2)}\frac{s}{\sqrt{n}}\,.$$

◇ Cho $1-\alpha$ và n. Tìm độ chính xác của ước lượng ε.

◇ Cho $1-\alpha$ và ε. Tìm kích thước mẫu n.

◇ Cho ε và n. Tìm độ tin cậy của ước lượng $1-\alpha$.

Một số trong số các vấn đề này sẽ được đề cập ở phần sau.

b. Ước lượng một phía

Vấn đề đặt ra ở đây là với độ tin cậy $1-\alpha$ cho trước, tìm khoảng ước lượng một phía:

◇ Khoảng ước lượng bên trái $(-\infty, \overline{X}+\varepsilon)$: $\mathbb{P}[-\infty < \mu < \overline{X}+\varepsilon] = 1-\alpha$.

◇ Khoảng ước lượng bên phải $(\overline{X}-\varepsilon, +\infty)$: $\mathbb{P}[\overline{X}-\varepsilon < \mu < +\infty] = 1-\alpha$.

Nhận xét. Khoảng tin cậy bên trái cho ta biết giá trị tối đa, khoảng tin cậy bên phải cho ta biết giá trị tối thiểu của μ với độ tin cậy $1-\alpha$.

Ta cũng chia thành 3 trường hợp, điểm khác biệt là thay thế $\alpha/2$ bởi α.

Trường hợp 1. Phương sai σ^2 đã biết.

Đặt $t_\alpha = \varphi^{-1}(1-\alpha)$, ta có

$$\mathbb{P}\Big[-t_\alpha < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < +\infty\Big] = 1 - \varphi(-t_\alpha) = 1-\alpha,$$

$$\mathbb{P}\Big[-\infty < \frac{\overline{X}-\mu}{\sigma}\sqrt{n} < t_\alpha\Big] = \varphi(t_\alpha) = 1-\alpha,$$

hay $\mathbb{P}\Big[-\infty < \mu < \overline{X} + t_\alpha\dfrac{\sigma}{\sqrt{n}}\Big] = \mathbb{P}\Big[\overline{X} - t_\alpha\dfrac{\sigma}{\sqrt{n}} < \mu < +\infty\Big] = 1-\alpha.$

Như vậy, với một mẫu cụ thể, khoảng ước lượng bên trái và bên phải lần lượt là $(-\infty, \overline{x}+\varepsilon)$, $(\overline{x}-\varepsilon, +\infty)$ trong đó $\varepsilon = t_\alpha\dfrac{\sigma}{\sqrt{n}}$.

Trường hợp 2. Phương sai σ^2 chưa biết và $n \geqslant 30$.

Lý luận hoàn toàn tương tự, khoảng ước lượng bên trái và bên phải lần lượt là $(-\infty, \overline{x}+\varepsilon)$, $(\overline{x}-\varepsilon, +\infty)$ trong đó $\varepsilon = t_\alpha\dfrac{s}{\sqrt{n}}$.

Trường hợp 3. Phương sai σ^2 chưa biết và $n < 30$.

Khoảng ước lượng bên trái và bên phải lần lượt là $(-\infty, \overline{x}+\varepsilon)$, $(\overline{x}-\varepsilon, +\infty)$ trong đó $\varepsilon = t_{(n-1,\alpha)}\dfrac{s}{\sqrt{n}}$.

Ước lượng khoảng cho giá trị trung bình ứng với 3 trường hợp trên được mô tả qua bảng tổng hợp sau

<table>
<tr><td rowspan="2">Loại ước lượng</td><td rowspan="2">ε</td><td colspan="3">Độ chính xác của ước lượng: ε</td></tr>
<tr><td>TH1</td><td>TH2</td><td>TH3</td></tr>
<tr><td>Hai phía</td><td>$(\overline{x}-\varepsilon, \overline{x}+\varepsilon)$</td><td>$t_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$</td><td>$t_{\alpha/2}\dfrac{s}{\sqrt{n}}$</td><td>$t_{(n-1,\alpha/2)}\dfrac{s}{\sqrt{n}}$</td></tr>
<tr><td>Bên trái</td><td>$(-\infty, \overline{x}+\varepsilon)$</td><td rowspan="2">$t_\alpha\dfrac{\sigma}{\sqrt{n}}$</td><td rowspan="2">$t_\alpha\dfrac{s}{\sqrt{n}}$</td><td rowspan="2">$t_{(n-1,\alpha)}\dfrac{s}{\sqrt{n}}$</td></tr>
<tr><td>Bên phải</td><td>$(\overline{x}-\varepsilon, +\infty)$</td></tr>
</table>

Ví dụ 4. Để đánh giá về mức doanh thu hàng tháng tại các đại lý nhỏ trên một địa bàn, người ta lấy mẫu gồm 36 đại lý. Kết quả thu được như sau: doanh thu trung bình là 155,3 triệu đồng và độ lệch chuẩn mẫu là 16 triệu đồng. Với độ tin cậy 99%, ước lượng doanh thu trung bình tối đa và tối thiểu của mỗi đại lý.

Giải. $1-\alpha = 0{,}99$; t_α =2,33. Độ chính xác của ước lượng

$$\varepsilon = t_\alpha \frac{s}{\sqrt{n}} = 2{,}33\frac{16}{\sqrt{36}} \approx 6{,}21.$$

Doanh thu tối thiểu: $\overline{x} - \varepsilon$ =149,09;

Doanh thu tối đa: $\overline{x} + \varepsilon$ =161,51.

4.2.3 Khoảng tin cậy cho tỉ lệ

a. Ước lượng hai phía

Đám đông X có tỉ lệ p cần ước lượng, từ mẫu ngẫu nhiên chúng ta xác định được tỉ lệ F, vấn đề đặt ra ở đây là với độ tin cậy $1-\alpha$ cho trước, tìm khoảng ước lượng $(F-\varepsilon,\ F+\varepsilon)$ của p để

$$\mathbb{P}[F-\varepsilon < p < F+\varepsilon] = 1-\alpha.$$

Khi n đủ lớn $\dfrac{F-p}{\sqrt{F(1-F)}}\sqrt{n} \simeq \mathcal{N}(0,\,1)$, đặt $t_{\alpha/2} = \varphi^{-1}(1-\frac{\alpha}{2})$, ta có

$$\begin{aligned}\mathbb{P}\Big[-t_{\alpha/2} < \frac{F-p}{\sqrt{F(1-F)}}\sqrt{n} < t_{\alpha/2}\Big] &= \varphi(t_{\alpha/2}) - \varphi(-t_{\alpha/2})\\ &= \varphi(t_{\alpha/2}) - (1-\varphi(t_{\alpha/2})) = 1-\alpha,\end{aligned}$$

hay $\mathbb{P}\Big[F - t_{\alpha/2}\sqrt{\frac{F(1-F)}{n}} < p < F + t_{\alpha/2}\sqrt{\frac{F(1-F)}{n}}\ \Big] = 1-\alpha.$

Quy tắc thực hành: khi $nf \geqslant 10$ và $n(1-f) \geqslant 10$.

◇ Xác định mức phân vị $t_{\alpha/2}$.

◇ Xác định khoảng ước lượng $(f-\varepsilon, f+\varepsilon)$ với độ chính xác của ước lượng

$$\varepsilon = t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}.$$

Ví dụ 5. Để ước lượng tỉ lệ phế phẩm của một kho hàng. Người ta kiểm tra 100 sản phẩm, phát hiện có 20 sản phẩm là phế phẩm. Với độ tin cậy 95%, hãy ước lượng khoảng tỉ lệ phế phẩm của kho hàng.

Giải. $t_{\alpha/2}$ =1,96; f =0,2; n =100. Độ chính xác của ước lượng

$$\varepsilon = t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}} = 0{,}0784.$$

Khoảng ước lượng cho tỉ lệ phế phẩm: $(0{,}1216 < p < 0{,}2784)$.

b. Ước lượng một phía

Với các bước thiết lập tương tự ta thu được khoảng ước lượng của p bên trái là $p < f+\varepsilon$ và bên phải là $p > f-\varepsilon$, trong đó $\varepsilon = t_\alpha\sqrt{\frac{f(1-f)}{n}}$.

Ví dụ 6. Cho giả thiết như Ví dụ 5. Hãy ước lượng tỉ lệ phế phẩm tối đa và tối thiểu.

Giải. t_α =1,64; f =0,2; n =100. Độ chính xác của ước lượng

$$\varepsilon = t_\alpha\sqrt{\frac{f(1-f)}{n}} = 0{,}0656.$$

Tỉ lệ sản phẩm tối thiểu: $f-\varepsilon$ =0,1344;

Tỉ lệ sản phẩm tối đa: $f+\varepsilon$ =0,2656.

Ví dụ 7. Một lô hàng nhập cảng gồm 5.000 thiết bị điện tử đã qua sử dụng. Cơ quan quản lý kiểm tra ngẫu nhiên 100 thiết bị từ lô hàng thì có 82 thiết bị có thể tiếp tục sử dụng được. Với độ tin cậy 90%, lô hàng có tối thiểu bao nhiêu thiết bị có thể tiếp tục sử dụng được?

Giải. $t_\alpha = 1,28$; $f = 0,82$; $n = 100$; $N = 5.000$. Độ chính xác xủa ước lượng

$$\varepsilon = t_\alpha\sqrt{\frac{f(1-f)}{n}} = 0{,}0492.$$

Tỉ lệ sản phẩm tối thiểu: $f - \varepsilon =$0,7708.

Vậy, số thiết bị tối thiểu có thể tiếp tục sử dụng được: $N(f-\varepsilon) =$ 4864.

Các dạng toán phát sinh

Xuất phát từ công thức

$$\varepsilon = t_{\alpha/2}\sqrt{\frac{f(1-f)}{n}}\,.$$

◇ Cho $1-\alpha$ và n. Tìm độ chính xác của ước lượng ε.

◇ Cho $1-\alpha$ và ε. Tìm kích thước mẫu n.

◇ Cho ε và n. Tìm độ tin cậy của ước lượng $1-\alpha$.

4.2.4 Độ chính xác của ước lượng

Trong các nội dung trước chúng ta đã giải quyết bài toán xây dựng ước lượng khoảng cho trung bình và ước lượng khoảng cho tỉ lệ, nghĩa là từ mẫu cụ thể, độ tin cậy $1-\alpha$ ta sẽ xác định được khoảng ước lượng cho tham số θ là (θ_1, θ_2) trong đó độ chính xác của ước lượng $\varepsilon = \frac{\theta_2-\theta_1}{2}$.

Trong các trường hợp đã trình bày thì ε phụ thuộc vào kích thước mẫu n. Bây giờ ta đặt ra bài toán ngược: với độ tin cậy $1-\alpha$ đã biết, cho độ chính xác của ước lượng ε, tìm kích thước mẫu n cần thiết để nhận được ước lượng với độ chính xác đã cho. Chúng ta sẽ giải quyết bài toán này đối với trường hợp 1 của bài toán ước lượng khoảng trung bình. Các trường hợp còn lại là hoàn toàn tương tự (giành cho bạn đọc).

Trong trường hợp này, khoảng ước lượng là $(\overline{x}-\varepsilon, \overline{x}+\varepsilon)$ và công thức xác định độ chính xác của ước lượng $\varepsilon = t_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$. Kích thước mẫu điều tra cần thiết nếu độ chính xác của ước lượng ε_0 là

$$n = \left[\frac{t^2_{\alpha/2}\,\sigma^2}{\varepsilon_0^2}\right] + 1,$$

trong đó ký hiệu $[x]$ là phần nguyên của x, chẳng hạn $[20,36] = 20$.

Ví dụ 8. Với giả thiết như ở Ví dụ 1: $\sigma^2 = 16;\;\; 1-\alpha =$0,95. Muốn có ước lượng có độ chính xác là 1 thì phải điều tra mẫu có kích thước bao nhiêu?

Giải. Như vậy $\varepsilon_0 = 1$, khi đó

$$n = \left[\frac{t^2_{\alpha/2}\,\sigma^2}{\varepsilon_0^2}\right] + 1 = 62.$$

Ngoài ra, chúng ta còn giải quyết được bài toán ngược dạng tìm độ tin cậy của ước lượng khi biết độ chính xác của ước lượng và kích thước mẫu. Vấn đề này được đề cập trong ví dụ sau đây:

Ví dụ 9. Một mẫu thống kê có kích thước $n = 36$, có trung bình mẫu là 100 và độ lệch chuẩn mẫu là 5. Tìm độ tin cậy của ước lượng nếu khoảng ước lượng là (99; 101).

Giải. Tính mức phân vị: $t_{\frac{\alpha}{2}} = \dfrac{\varepsilon\sqrt{n}}{s} = 2$. Độ tin cậy của ước lượng

$$1-\alpha = 2\varphi(t_{\alpha/2}) = 0{,}955.$$

BÀI TẬP CHƯƠNG 4

1. Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên kích thước n của đám

đông X có $\mathbb{E}X = \mu$. Chứng minh rằng

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \quad \text{và} \quad \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

đều là các ước lượng không chệch của phương sai $\mathbb{D}X$.

2. Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên kích thước n từ phân phối với hàm mật độ là:

$$p(x,\theta) = \begin{cases} \dfrac{1}{\theta}\, e^{-\frac{x}{\theta}} & \text{với } x > 0,\ \theta > 0, \\ 0 & \text{với } x \leqslant 0. \end{cases}$$

Tìm ước lượng hiệu quả của θ.

3. Giả sử $(X_1, X_2, ..., X_n)$ là mẫu ngẫu nhiên kích thước n từ phân phối Poisson với tham số $\mathbb{E}X = \mathbb{D}X = \lambda > 0$. Tìm ước lượng hợp lý tối đa của λ.

4. Để xác định độ chính xác của một chiếc cân, người ta tiến hành cân một quả tạ. Kết quả thu được sau 7 lần cân như sau:

159,8   159,7   160,2   159,6   160,4   159,5   160,6   (kg)

   (a) Tìm ước lượng không chệch của khối lượng quả cân.

   (b) Tìm ước lượng không chệch của phương sai số đo trong hai trường hợp:

      - Biết khối lượng quả cân là 160 kg.

      - Chưa biết khối lượng của quả cân.

5. Cơ quan quản lý thị trường lấy số liệu về giá thành bán lẻ của một loại sản phẩm tại 40 đại lý, người ta thu được bảng tần số như sau: (đơn vị: ngàn đồng)

x_i	39	40	41	42
n_i	8	16	4	12

(a) Tính trung bình mẫu $\overline{x}$ và phương sai mẫu hiệu chỉnh s^2.

(b) Với độ tin cậy 95%, ước lượng khoảng giá thành bán lẻ trung bình mỗi sản phẩm.

6. Một dây chuyền sản xuất những thanh kim loại có chiều dài tuân theo luật phân phối chuẩn. Người ta chọn ngẫu nhiên ra một số thanh và đo chiều dài (đơn vị: cm) của chúng, thu được dãy số liệu sau:

149; 151; 148; 152; 151; 152; 149; 148; 149; 151; 152; 149; 151; 149; 152.

(a) Tính trung bình mẫu $\overline{x}$ và phương sai mẫu hiệu chỉnh s^2.

(b) Với độ tin cậy 90%, ước lượng khoảng độ dài trung bình của mỗi thanh kim loại.

7. Một dây chuyền tự động đóng gói một loại bao gạo có khối lượng tuân theo luật phân phối chuẩn với độ lệch chuẩn là 0,5. Người ta cân kiểm tra 20 bao gạo, thu được bảng tần số như sau: (đơn vị: kg)

x_i	49,3	49,5	49,9	50,2
n_i	6	2	4	8

(a) Tính trung bình mẫu $\overline{x}$ và phương sai mẫu hiệu chỉnh s^2.

(b) Với độ tin cậy 98%, ước lượng khoảng khối lượng trung bình của mỗi bao gạo.

8. Nhà sản xuất muốn ước lượng khối lượng sắt trong mỗi cuộn được sản xuất từ một dây chuyền công nghệ quốc gia. Theo tiêu chuẩn của công nghệ, độ lệch chuẩn là 8 kg. Điều tra một mẫu 50 cuộn được khối lượng sắt trung bình là 97kg.

   (a) Với độ tin cậy là 99%, ước lượng khối lượng sắt trung bình của một cuộn.

   (b) Với độ tin cậy là 99%, ước lượng khối lượng sắt trung bình tối thiểu của một cuộn.

   (c) Nếu nhà sản xuất muốn ước lượng khối lượng sắt trung bình của mỗi cuộn đảm bảo độ chính xác là 2 kg thì cần điều tra thêm bao nhiêu cuộn nữa.

9. Một công ty có 500 đại lý, để đánh giá về mức doanh thu, người ta lấy mẫu gồm 36 đại lý. Kết quả thu được như sau: doanh thu trung bình là 84,5 triệu đồng và độ lệch chuẩn mẫu là 3 triệu đồng. Với độ tin cậy 99%, hãy ước lượng doanh thu tối thiểu và tối đa của công ty.

10. Người ta đo chiều sâu của biển bằng một loại thiết bị điện tử, kết quả đo tuân theo luật phân phối chuẩn có phương sai $400m^2$. Với độ tin cậy là 95%, cần phải đo ít nhất bao nhiêu lần để kết quả có sai số không vượt quá $15m$.

11. Một mẫu thống kê có kích thước $n = 64$, tuân theo luật phân phối chuẩn với trung bình mẫu là 200, độ lệch chuẩn mẫu là 3. Tìm độ tin cậy của ước lượng nếu khoảng ước lượng là $(199, 201)$.

12. Để đánh giá hiệu quả của một loại thuốc, người ta đem sử dụng cho 1000 bệnh nhân thì có 820 người khỏi bệnh. Với độ tin cậy 96 %,

    (a) Hãy ước lượng khoảng cho tỉ lệ chữa khỏi bệnh của loại thuốc trên.

    (b) Hãy ước lượng tỉ lệ chữa bệnh tối đa và tối thiểu của loại thuốc trên.

13. Tỉ lệ chính phẩm của một nhà máy là 90%. Với độ tin cậy 95%, muốn ước lượng tỉ lệ chính phẩm của nhà máy với độ dài khoảng tin cậy không quá 0,02 thì phải kiểm tra ít nhất bao nhiêu sản phẩm?

14. Một kho hàng tồn gồm 10.000 chiếc bút bi. Lấy mẫu gồm 100 chiếc bút từ kho hàng ra kiểm tra thì có 75 chiếc đạt chất lượng. Với độ tin cậy 95%, hãy ước lượng khoảng tỉ lệ số bút không đạt chất lượng và suy ra khoảng tin cậy số bút không đạt chất lượng của kho.

15. Tại một bang có 4 triệu người tham gia bầu cử, người ta phỏng vấn ngẫu nhiên 1000 cử tri thì có 720 cử tri ủng hộ một ứng cử viên A. Với độ tin cậy là 95%, có ít nhất bao nhiêu cử tri của bang đó đã ủng hộ ứng cử viên A?

16. Để đánh giá trữ lượng cá trong một hồ nuôi, người ta bắt 1000 con cá và đánh dấu chúng, sau đó thả lại hồ. Lần thứ hai người ta bắt 200 con thì thấy có 30 con được đánh dấu. Với độ tin cậy là 95%,

    (a) Hãy ước lượng trữ lượng cá trong hồ.

    (b) Nếu muốn sai số của ước lượng giảm đi một nửa thì cần phải bắt bao nhiêu con cá.

CHƯƠNG 5

KIỂM ĐỊNH GIẢ THIẾT

Tuần 14, 15

5.1 Đặt vấn đề

Trong thực tế cuộc sống chúng ta thường gặp 2 quan điểm trái ngược nhau về một vấn đề nào đó. Chẳng hạn, các nhà sản xuất cho rằng có 95% sản phẩm của công ty đảm bảo tiêu chuẩn, trong khi đó các nhà quản lý thị trường lại cho rằng không phải như vậy thực tế thấp hơn nhiều; trước cuộc bầu cử tổng thống đảng phái A cho rằng có 65% cử tri ủng hộ UWCV của đảng phái họ, trong khi đó đảng đối lập lại cho rằng thực tế thấp hơn nhiều.

Vấn đề đặt ra là, thông qua số liệu thống kê hãy chỉ ra chấp nhận ý kiến nào trong 2 ý kiến đó với một mức ý nghĩa α cho trước.

Tổng quát: Chúng ta thường có bài toán

$$\begin{cases} H: & \text{(Giả thiết)} \quad \text{có tính chất A} \\ K: & \text{(Giả thiết)} \quad \text{không có tính chất A} \end{cases}$$

Từ số liệu thống kê hãy đưa ra kết luận cho bài toán trên.

Trong kiểm định giả thiết thường gặp 2 loại sai lầm:

- Sai lầm loại 1: Bác bỏ H trong khi H đúng

- Sai lầm loại 2: Chấp nhận H trong khi H sai.

Mục đích của các nhà thống kê là làm giảm cả 2 loại sai lầm này. Tuy vậy điều này không thể vì giảm sai lầm này thì khả năng mắc sai lầm loại kia tăng lên.

Trong thực tế thống kê người ta thấy mỗi loại sai lầm sẽ gây ra một tác hại khác nhau. Tuy vậy người ta thấy cần phải giảm sai lầm loại 1 với một xác suất xãy ra bé. Chẳng hạn như trong xã hội hiện đại người ta cho rằng "*Kết án người vô tội nguy hiểm hơn rất nhiều so với việc tha bổng một người có tội*". .. Do đó, Neyman- Pearson đã cho rằng chúng ta chỉ xét những bài toán thống kê với

$\mathbb{P}$(Sai lầm loại 1) = $\mathbb{P}$(Bác bỏ H| H đúng) $\leq \alpha$ (αlà một số bé và gọi là mức ý n

Thông thường $\alpha \leq 10\%$.

5.2 Kiểm định giả thiết về giá trị trung bình và về tỉ lệ

5.2.1 Kiểm định giả thiết về giá trị trung bình

Đây là một dạng bài toán kiểm định số đặc trưng $\mathbb{E}X = \mu$ của biến ngẫu nhiên gốc đám đông X (so sánh giá trị kỳ vọng của đại lượng ngẫu nhiên X với giá trị μ_0 cho trước). Có 2 dạng bài toán kiểm định giả thiết về giá trị trung bình. **a. Kiểm định hai phía**

Vấn đề đặt ra ở đây là với mức ý nghĩa α và một giá trị μ_0 cho trước, đánh giá về cặp giả thiết thống kê

$$H: \ \mu = \mu_0 \ ; \quad K: \ \mu \neq \mu_0.$$

Trường hợp 1. Phương sai σ^2 đã biết.

Khoảng ước lượng của μ với độ tin cậy $1-\alpha$

$$(\overline{x} - t_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}).$$

Chấp nhận giả thiết H khi $\mu_0 \in [\overline{x} - t_{\alpha/2}\frac{\sigma}{\sqrt{n}}\ ;\ \overline{x} + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$ hay

$$\overline{x} - t_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqslant \mu_0 \leqslant \overline{x} + t_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

tương đương với $\left|\dfrac{\overline{x}-\mu_0}{\sigma}\sqrt{n}\right| \leqslant t_{\alpha/2}$.

Quy tắc thực hành

◇ Từ mẫu cụ thể xác định giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{\overline{x}-\mu_0}{\sigma}\sqrt{n}.$$

◇ So sánh giá trị của t_{tn} với mức phân vị $t_{\alpha/2}$, nếu

$|t_{tn}| \leqslant t_{\alpha/2}$: chấp nhận H.

$|t_{tn}| > t_{\alpha/2}$: bác bỏ H, chấp nhận K.

Chú ý

- Nếu như kích thước mẫu $n < 30$ thì ta cần bổ sung thêm điều kiện X tuân theo luật phân phối chuẩn.

- Như vậy miền bác bỏ $W_\alpha = (-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, +\infty)$, điều này là hợp lý. Giả sử $H:\ \mu = \mu_0$ đúng thì $T = \dfrac{\overline{X}-\mu_0}{\sigma}\sqrt{n} \simeq \mathcal{N}(0,1)$, khi đó

$$\begin{aligned}\mathbb{P}[T \in W_\alpha | H \text{ đúng}] &= \mathbb{P}\left[\left|\frac{\overline{X}-\mu_0}{\sigma}\sqrt{n}\right| > t_{\alpha/2}\right]\\ &= 1 - \big(\varphi(t_{\alpha/2}) - \varphi(-t_{\alpha/2})\big) = \alpha.\end{aligned}$$

Nghĩa là xác suất phạm sai lầm loại 1 được ấn định bởi một giá trị tương đối nhỏ α nào đó, việc chứng minh xác suất phạm sai lầm loại 2 cực tiểu bạn đọc tham khảo tài liệu [5].

Ví dụ 1. Một máy tiện tự động cho ra những trục máy có đường kính là 120mm và độ lệch chuẩn cho phép là 3mm. Kiểm tra ngẫu nhiên 50 trục máy, kết quả thu được đường kính trung bình là 119,2mm. Với mức ý nghĩa là 10%, máy tiện trên có hoạt động bình thường không?

Giải. Máy tiện được gọi là hoạt động bình thường khi nó sản xuất những trục máy với sai số không vượt quá mức cho phép. Cặp giả thiết thống kê

$$H: \ \mu = \mu_0 = 120 \ ; \quad K: \ \mu \neq \mu_0.$$

$\mu_0 = 120$; $\sigma = 3$; $\alpha = 0{,}1$; $t_{\alpha/2} = 1{,}64$; $n = 50$; $\overline{x} = 119{,}2$. Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{119{,}2 - 120}{3}\sqrt{50} \approx -1{,}89,$$

Vì $|t_{tn}| > t_{\alpha/2}$ nên ta bác bỏ H, chấp nhận K. Nghĩa là chấp nhận khẳng định cho rằng máy tiện trên hoạt động không bình thường.

Trường hợp 2. Phương sai σ^2 chưa biết và $n \geqslant 30$.

Tương tự như ở trường hợp 1, đặt $t_{tn} = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$, khi đó

$|t_{tn}| \leqslant t_{\alpha/2}$: chấp nhận H.

$|t_{tn}| > t_{\alpha/2}$: bác bỏ H, chấp nhận K.

Trường hợp 3. Phương sai σ^2 chưa biết và $n < 30$.

Giả sử X tuân theo luật phân phối chuẩn $\mathcal{N}(0, 1)$. Đặt $t_{tn} = \dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$, khi đó

$|t_{tn}| \leqslant t_{(n-1,\alpha/2)}$: chấp nhận H.

$|t_{tn}| > t_{(n-1,\alpha/2)}$: bác bỏ H, chấp nhận K.

Ví dụ 2. Thể tích sơn chứa trong mỗi thùng sơn nước nhãn hiệu A là đại lượng ngẫu nhiên tuân theo luật phân phối chuẩn với trung bình 18 lít. Kiểm tra ngẫu nhiên 25 thùng thu được kết quả: thể tích trung bình là 17,92 lít và độ lệch chuẩn mẫu là 0,24 lít. Với mức ý nghĩa 5%, thể tích sơn trong các thùng sơn có đúng tiêu chuẩn không?

Giải. Cặp giả thiết thống kê

$$H:\ \mu = \mu_0 = 18\ ;\quad K:\ \mu \neq \mu_0.$$

$\mu_0 = 18;\ s = 0{,}24;\ \alpha = 0,05;\ t_{(n-1,\alpha/2)} = 2{,}11;\ n = 25;\ \overline{x} = 17{,}92.$

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{17{,}92 - 18}{0{,}24}\sqrt{25} \approx -1{,}67,$$

Vì $|t_{tn}| \leqslant t_{\alpha/2}$ nên ta chấp nhận H. Nghĩa là chấp nhận khẳng định cho rằng thể tích sơn trong các thùng sơn đúng tiêu chuẩn.

b. Kiểm định một phía

Trong thực tế xuất hiện một số dạng toán về kiểm định như:

- Sau chiến dịch quảng cáo, doanh số bán ra một loại hàng có tăng lên hay không? (kiểm định lớn hơn)

- Kiểm tra xem khối lượng đóng gói các bao gạo của một kho có nhỏ hơn giá trị in trên bao bì hay không? (kiểm định nhỏ hơn)

Các dạng bài toán này được gọi là *bài toán kiểm định một phía.*

⋄ Kiểm định lớn hơn: $H:\ \mu = \mu_0\ ;\quad K:\ \mu > \mu_0.$

⋄ Kiểm định nhỏ hơn: $H:\ \mu = \mu_0\ ;\quad K:\ \mu < \mu_0.$

Giải quyết bài toán kiểm định một phía được phân chia các trường hợp giống như trong bài toán kiểm định hai phía. Tiêu chuẩn kiểm định ứng với 3 trường hợp của bài toán kiểm định giá trị trung bình được mô tả qua bảng tổng hợp sau đây

Trường hợp	t_{tn}	Điều kiện chấp nhận $H: \ \mu = \mu_0$		
		$K: \ \mu = \mu_0$	$K: \ \mu > \mu_0$	$K: \ \mu < \mu_0$
TH1	$\dfrac{\overline{x} - \mu_0}{\sigma}\sqrt{n}$	$\lvert t_{tn}\rvert \leqslant t_{\alpha/2}$	$t_{tn} \leqslant t_\alpha$	$t_{tn} \geqslant -t_\alpha$
TH2	$\dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$	$\lvert t_{tn}\rvert \leqslant t_{\alpha/2}$	$t_{tn} \leqslant t_\alpha$	$t_{tn} \geqslant -t_\alpha$
TH3	$\dfrac{\overline{x} - \mu_0}{s}\sqrt{n}$	$\lvert t_{tn}\rvert \leqslant t_{(n-1,\alpha/2)}$	$t_{tn} \leqslant t_{(n-1,\alpha)}$	$t_{tn} \geqslant -t_{(n-1,\alpha)}$

Ví dụ 3. Một nhà máy cung cấp nước sạch cho rằng khối lượng trung bình của một loại chất độc hại trong một lít nước của nhà máy là 14mg. Người ta nghi ngờ số liệu này thấp hơn thực tế. Kiểm tra ngẫu nhiên với 64 mẫu nước thu được kết quả: $\bar{x} = 14{,}2$ và $s = 0{,}24$. Hãy cho kết luận về nghi ngờ nói trên với mức ý nghĩa 8%.

Giải. Cặp giả thiết thống kê: $H: \ \mu = \mu_0 = 120$; $\quad K: \ \mu > \mu_0$.

$\mu_0 = 14$; $s = 0{,}24$; $\alpha = 0{,}08$; $t_\alpha = 1{,}4$; $n = 64$; $\overline{x} = 14{,}2$.

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{14{,}2 - 14}{0{,}24}\sqrt{64} \approx 6{,}67,$$

Vì $t_{tn} > t_\alpha$ nên ta bác bỏ H, chấp nhận K. Nghĩa là chấp nhận nghi ngờ trên.

5.2.2 Kiểm định giả thiết về tỉ lệ

Đây là dạng bài so sánh giá trị tỉ lệ p của đám đông X với giá trị p_0 cho trước. Có hai dạng bài toán kiểm định giả thiết về tỉ lệ.

a. Kiểm định hai phía

Vấn đề đặt ra ở đây là với mức ý nghĩa α và một giá trị p_0 cho trước, đánh giá về cặp giả thiết thống kê

$$H:\ p = p_0\ ;\quad K:\ p \neq p_0.$$

Với n đủ lớn và $H:\ p = p_0$ đúng thì $T = \dfrac{F - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \simeq \mathcal{N}(0,1)$, khi đó

$$\mathbb{P}[T \in W_\alpha | H \text{ đúng}] = \mathbb{P}\Big[\Big|\frac{F - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}\Big| > t_{\alpha/2}\Big] = \alpha,$$

trong đó miền bác bỏ $W_\alpha = (-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, +\infty)$.

Quy tắc thực hành: Khi $np_0 \geqslant 5$; $n(1-p_0) \geqslant 5$.

◇ Xác định giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}.$$

◇ So sánh giá trị của t_{tn} với mức phân vị $t_{\alpha/2}$, nếu

$|t_{tn}| \leqslant t_{\alpha/2}$: chấp nhận H.

$|t_{tn}| > t_{\alpha/2}$: bác bỏ H, chấp nhận K.

Ví dụ 4. Một hãng sản xuất đĩa cứng công bố rằng: có 10% đĩa cứng của hãng phải bảo hành trong thời gian 2 năm đầu sử dụng. Người ta điều tra ngẫu nhiên 200 khách hàng đã sử dụng đĩa cứng của hãng thì có 29 đĩa cứng phải bảo hành trong thời gian 2 năm đầu sử dụng. Với mức ý nghĩa 5%, tỉ lệ trong công bố trên có đúng với thực tế không?

Giải. Cặp giả thiết thống kê: $H:\ p = p_0 = 0{,}1\ ;\quad K:\ p \neq p_0$.

$n = 200$; $f = 0{,}145$; $p_0 = 0{,}1$; $\alpha = 0,05$; $t_{\alpha/2} = 1{,}96$.

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{0,145 - 0,1}{\sqrt{0,1 \times 0,9}}\sqrt{200} \approx 2{,}12.$$

Vì $|t_{tn}| > t_{\alpha/2}$ nên ta bác bỏ H, chấp nhận K. Nghĩa là chấp nhận khẳng định cho rằng tỉ lệ trong công bố trên không đúng với thực tế.

b. Kiểm định một phía

Tương tự như bài toán kiểm định về giá trị trung bình, bài toán kiểm định tỉ lệ cũng có hai dạng kiểm định một phía như sau:

◇ Kiểm định lớn hơn: $H: \ p = p_0 \ ; \quad K: \ p > p_0.$

◇ Kiểm định nhỏ hơn: $H: \ p = p_0 \ ; \quad K: \ p < p_0.$

Bảng dưới đây sẽ trình bày các tiêu chuẩn kiểm định của bài toán kiểm định tỉ lệ

t_{tn}	Điều kiện chấp nhận $H: \ p = p_0$		
	$K: \ p = p_0$	$K: \ p > p_0$	$K: \ p < p_0$
$\dfrac{f - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}$	$\|t_{tn}\| \leqslant t_{\alpha/2}$	$t_{tn} \leqslant t_\alpha$	$t_{tn} \geqslant -t_\alpha$

Ví dụ 5. Một trung tâm đào tạo nghề báo cáo rằng tỷ lệ người học tại trung tâm kiếm được việc làm ngay sau khi tốt nghiệp là 70%. Một mẫu ngẫu nhiên gồm 200 người đã tốt nghiệp ở trung tâm cho thấy có 130 người kiếm được việc làm ngay sau khi tốt nghiệp. Với mức ý nghĩa 5%, kiểm định xem phải chăng tỉ lệ trong báo cáo của trung tâm là cao hơn thực tế.

Giải. Cặp giả thiết thống kê: $H: \ p = p_0 = 0{,}7 \ ; \quad K: \ p < p_0.$

$n = 200; \ f = 0{,}65; \ p_0 = 0{,}7; \ \alpha = 0,05; \ t_\alpha = 1{,}64.$

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{0,65 - 0,7}{\sqrt{0,7 \times 0,3}}\sqrt{200} \ \approx -1{,}54.$$

Vì $t_{tn} > -t_\alpha$ nên ta chấp nhận H. Nghĩa là chấp nhận ý kiến cho rằng tỉ lệ trong báo cáo của trung tâm là đúng thực tế.

5.3 So sánh các giá trị trung bình và các giá trị tỉ lệ

Giả sử chúng ta có hai đám đông $\mathcal{C}_1$ và $\mathcal{C}_2$ có chung một đặc điểm cần nghiên cứu nào đó; hai đại lượng ngẫu nhiên gốc đám đông tương ứng lần lượt là X_1 và X_2. Trong mục này chúng ta đề cập đến dạng bài toán so sánh hai giá trị đặc trưng của hai đại lượng ngẫu nhiên này.

5.3.1 So sánh hai giá trị trung bình

Hai đám đông $\mathcal{C}_1$ và $\mathcal{C}_2$ có hai giá trị trung bình là $\mathbb{E}X_1 = \mu_1$ và $\mathbb{E}X_2 = \mu_2$ cần so sánh. Vấn đề đặt ra ở đây là với mức ý nghĩa α cho trước, đánh giá về cặp giả thiết thống kê

$$H: \ \mu_1 = \mu_2 \ ; \quad K: \ \mu_1 \neq \mu_2.$$

Giả sử $\mathbb{D}X_1 = \sigma_1^2$, $\mathbb{D}X_2 = \sigma_2^2$. Từ hai mẫu cụ thể $(x_1, x_2, ..., x_{n_1})$ của đám đông $\mathcal{C}_1$ và $(y_1, y_2, ..., y_{n_2})$ của đám đông $\mathcal{C}_2$ chúng ta xác định được trung bình mẫu và phương sai hiệu chỉnh mẫu lần lượt là $\overline{x}_1, \overline{x}_2, s_1^2, s_2^2$.

Quy tắc thực hành

Trường hợp 1. σ_1^2, σ_2^2 đã biết.

◇ Xác định giá trị kiểm định từ thực nghiệm

$$t_{tn} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}.$$

◇ So sánh giá trị của t_{tn} với mức phân vị $t_{\alpha/2}$, nếu

$|t_{tn}| \leqslant t_{\alpha/2}$: chấp nhận H.

$|t_{tn}| > t_{\alpha/2}$: bác bỏ H, chấp nhận K.

Chú ý. Nếu như kích thước mẫu $n_1 < 30$ hoặc $n_2 < 30$ thì ta cần bổ sung thêm điều kiện X_1, X_2 tuân theo luật phân phối chuẩn.

Ví dụ 1. Người ta muốn so sánh tuổi thọ của hai loại thiết bị điện tử (trong điều kiện hoạt động liên tục) được sản xuất bởi hai công nghệ khác nhau. Biết rằng độ lệch chuẩn tuổi thọ của thiết bị được sản xuất từ công nghệ thứ nhất và công nghệ thứ hai tương ứng là 120 giờ và 125 giờ. Thử nghiệm 50 thiết bị cho mỗi công nghệ trên thu được tuổi thọ trung bình của chúng tương ứng là 264 giờ và 245 giờ. Với mức ý nghĩa 5%, tuổi thọ của hai loại thiết bị điện tử được sản xuất từ hai công nghệ trên có khác nhau không?

Giải. Cặp giả thiết thống kê: $H: \mu_1 = \mu_2$; $K: \mu_1 \neq \mu_2$.

$\sigma_1 = 120$; $\sigma_2 = 125$; $n_1 = n_2 = 50$; $\overline{x}_1 = 264$; $\overline{x}_2 = 245$;

$\alpha = 0{,}05$; $t_{\alpha/2} = 1{,}96$.

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{264 - 245}{\sqrt{\dfrac{120^2}{50} + \dfrac{125^2}{50}}} \approx 0{,}78.$$

Vì $|t_{tn}| \leqslant t_{\alpha/2}$ nên ta chấp nhận H. Nghĩa là chấp nhận khẳng định rằng tuổi thọ của hai loại thiết bị điện tử được sản xuất từ hai công nghệ trên là giống nhau.

Trường hợp 2. σ_1^2, σ_2^2 chưa biết và $n_1 \geqslant 30, n_2 \geqslant 30$.

◇ Xác định giá trị kiểm định từ thực nghiệm

$$t_{tn} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}.$$

◇ So sánh giá trị của t_{tn} với mức phân vị $t_{\alpha/2}$, nếu

$|t_{tn}| \leqslant t_{\alpha/2}$: chấp nhận H.

$|t_{tn}| > t_{\alpha/2}$: bác bỏ H, chấp nhận K.

Trường hợp 3. X_1, X_2 có phân phối chuẩn, $\sigma_1^2 = \sigma_2^2$ chưa biết và $n_1 < 30, n_2 < 30$.

◇ Xác định giá trị kiểm định từ thực nghiệm

$$t_{tn} = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{s^2\Big(\dfrac{1}{n_1} + \dfrac{1}{n_2}\Big)}}, \quad \text{trong đó} \quad s^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

◇ So sánh giá trị của t_{tn} với mức phân vị $t_{(n_1+n_2-2,\alpha/2)}$, nếu

$|t_{tn}| \leqslant t_{(n_1+n_2-2,\alpha/2)}$: chấp nhận H.

$|t_{tn}| > t_{(n_1+n_2-2,\alpha/2)}$: bác bỏ H, chấp nhận K.

Ví dụ 2. Hai máy tự động dùng cắt những thanh kim loại với cùng một yêu cầu. Từ máy thứ nhất lấy ra 12 sản phẩm thu được chiều dài trung bình là 55cm và độ lệch chuẩn mẫu là 0,3cm, từ máy thứ 2 lấy ra 18 sản phẩm có các kết quả tương ứng là : 55,2cm và 0,2cm. Với mức ý nghĩa là 0,1, đánh giá về nhận định: hai máy đó sản xuất ra các thiết bị cùng kích cỡ. Giả sử rằng kích cỡ sản phẩm từ 2 máy có phân phối chuẩn và có cùng phương sai.

Giải. Cặp giả thiết thống kê: $H: \ \mu_1 = \mu_2$; $K: \ \mu_1 \neq \mu_2$.

$s_1 =$ 0,3cm; $s_2 =$ 0,2; $n_1 = 12$; $n_2 = 18$; $\overline{x_1} =$ 55cm; $\overline{x_2} =$ 55,2;

$\alpha = 0{,}1;\ t_{(28;0{,}05)} = 1,701.$

Giá trị kiểm định thực nghiệm

$$s^2 = \frac{11 \times 0,3^2 + 17 \times 0,2^2}{28} \approx 0,06;$$
$$t_{tn} = \frac{55 - 55,2}{\sqrt{0,06\left(\frac{1}{12} + \frac{1}{18}\right)}} \approx -2,2.$$

Vì $|t_{tn}| > t_{(28;0{,}05)}$ nên ta bác bỏ H, chấp nhận K. Nghĩa là chấp nhận nhận định cho rằng hai máy đó sản xuất ra các thiết bị không cùng kích cỡ.

Đối với bài toán so sánh 2 giá trị trung bình, có hai dạng bài toán kiểm định một phía như sau:

⋄ Kiểm định lớn hơn: $H:\ \mu_1 = \mu_2\ ;\quad K:\ \mu_1 > \mu_2.$

⋄ Kiểm định nhỏ hơn: $H:\ \mu_1 = \mu_2\ ;\quad K:\ \mu_1 < \mu_2.$

Giải quyết bài toán kiểm định một phía được phân chia các trường hợp giống như trong bài toán kiểm định hai phía. Tiêu chuẩn kiểm định ứng với 3 trường hợp được mô tả qua bảng tổng hợp sau:

Trường hợp	t_{tn}	Điều kiện chấp nhận $H:\ \mu_1 = \mu_2$	
		$K:\ \mu_1 > \mu_2$	$K:\ \mu_1 < \mu_2$
TH1	$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	$t_{tn} \leqslant t_\alpha$	$t_{tn} \geqslant -t_\alpha$
TH2	$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$	$t_{tn} \leqslant t_\alpha$	$t_{tn} \geqslant -t_\alpha$
TH3	$\dfrac{\overline{x}_1 - \overline{x}_2}{\sqrt{s^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	$t_{tn} \leqslant t_{(n_1+n_2-2,\alpha)}$	$t_{tn} \geqslant -t_{(n_1+n_2-2,\alpha)}$

Ví dụ 3. Với giả thiết như ở Ví dụ 2: $s_1 = 0{,}3$cm; $s_2 = 0{,}2$; $n_1 = 12$; $n_2 = 18$; $\overline{x_1} = 55$cm; $\overline{x_2} = 55{,}2$. Đánh giá nhận định: máy thứ hai sản xuất ra thiết bị có kích cỡ lớn hơn máy thứ nhất.

Giải. Cặp giả thiết thống kê: $H:\ \mu_1 = \mu_2$; $K:\ \mu_1 < \mu_2$.

$\alpha = 0{,}1$; $t_{(28;0{,}1)}$ =1,313. Giá trị kiểm định thực nghiệm

$$s^2 \approx 0,06;\ t_{tn} \approx -2,2.$$

Vì $t_{tn} < -t_{(28;0{,}1)}$ nên ta bác bỏ H, chấp nhận K. Nghĩa là chấp nhận nhận định cho rằng máy thứ hai sản xuất ra thiết bị có kích cỡ lớn hơn máy thứ nhất.

5.3.2 So sánh hai tỉ lệ

Hai đám đông $\mathcal{C}_1$ và $\mathcal{C}_2$ có hai tỉ lệ p_1 và p_2 cần so sánh. Vấn đề đặt ra ở đây là với mức ý nghĩa α cho trước, đánh giá về cặp giả thiết thống kê

$$H:\ p_1 = p_2\ ;\quad K:\ p_1 \neq p_2.$$

Từ mẫu cụ thể kích thước n_1 của đám đông $\mathcal{C}_1$ ta xác định được k_1 phần tử có đặc điểm cần nghiên cứu, do đó tỉ lệ mẫu là $f_1 = k_1/n_1$; tương tự đối với mẫu kích thước n_2 của đám đông $\mathcal{C}_2$ ta xác định được k_2 và $f_2 = k_2/n_2$.

Quy tắc thực hành: Khi n_1, n_2 đủ lớn.

◇ Xác định giá trị kiểm định từ thực nghiệm

$$t_{tn} = \frac{f_1 - f_2}{\sqrt{f(1-f)\Big(\dfrac{1}{n_1} + \dfrac{1}{n_2}\Big)}}, \quad \text{trong đó} \quad f = \frac{k_1 + k_2}{n_1 + n_2}.$$

◇ So sánh giá trị của t_{tn} với mức phân vị $t_{\alpha/2}$, nếu

$|t_{tn}| \leqslant t_{\alpha/2}$: chấp nhận H.

$|t_{tn}| > t_{\alpha/2}$: bác bỏ H, chấp nhận K.

Chú ý. Khi kích thước mẫu điều tra càng lớn thì kết quả kiểm định càng chính xác, ở mức độ tương đối khái niệm n_1, n_2 đủ lớn ở đây được hiểu là thỏa mãn hai điều kiện: $(n_1 + n_2)f \geqslant 10$, $(n_1 + n_2)(1 - f) \geqslant 10$.

Ví dụ 4. Người ta kiểm tra ngẫu nhiên 400 sản phẩm từ dây chuyền thứ nhất thì có 24 phế phẩm, kiểm tra 800 sản phẩm từ dây chuyền thứ hai thấy có 42 phế phẩm. Với mức ý nghĩa $\alpha = 0{,}05$, tỉ lệ phế phẩm của 2 dây chuyền trên có như nhau hay không?

Giải. Cặp giả thiết thống kê: $H:\ p_1 = p_2$; $K:\ p_1 \neq p_2$.

$t_{\alpha/2} = 1,96$; $f_1 = 0{,}06$; $f_2 = 0{,}0525$; $f = 0{,}055$.

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{0,06 - 0,0525}{\sqrt{0,055 \times 0,945\big(1/400 + 1/800\big)}} \approx 0,537.$$

Vì $|t_{tn}| < t_{\alpha/2}$ nên ta chấp nhận H. Nghĩa là chấp nhận khẳng định cho rằng tỉ lệ phế phẩm của 2 dây chuyền trên là như nhau.

Tương tự như bài toán kiểm định về giá trị trung bình, bài toán kiểm định tỉ lệ cũng có hai dạng kiểm định một phía như sau:

◇ Kiểm định lớn hơn: $H:\ p_1 = p_2\ ;\quad K:\ p_1 > p_2$.

◇ Kiểm định nhỏ hơn: $H:\ p_1 = p_2\ ;\quad K:\ p_1 < p_2$.

Bảng dưới đây sẽ trình bày các tiêu chuẩn kiểm định của bài toán kiểm định tỉ lệ:

t_{tn}	Điều kiện chấp nhận $H:\ p_1 = p_2$		
	$K:\ p_1 \neq p_2$	$K:\ p_1 > p_2$	$K:\ p_1 < p_2$
$\dfrac{f_1 - f_2}{\sqrt{f(1-f)\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$	$\lvert t_{tn}\rvert \leqslant t_{\alpha/2}$	$t_{tn} \leqslant t_\alpha$	$t_{tn} \geqslant -t_\alpha$

Ví dụ 5. Dùng thuốc A cho 200 bệnh nhân thì có 160 người khỏi bệnh. Dùng thuốc B cho 300 bệnh nhân thì có 210 người khỏi bệnh. Với mức ý nghĩa $\alpha = 0{,}04$, hiệu quả của thuốc A có cao hơn thuốc B hay không?

Giải. Cặp giả thiết thống kê: $H:\ p_1 = p_2;\ \ K:\ p_1 > p_2$.

$t_\alpha = 1{,}75;\ f_1 = 0{,}8;\ f_2 = 0{,}7;\ f = 0{,}74.$

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{(0,8 - 0,7)}{\sqrt{0,74 \times 0,26\big(1/200 + 1/300\big)}} \approx 2,497.$$

Vì $t_{tn} > t_\alpha$ nên ta bác bỏ H, chấp nhận K. Nghĩa là chấp nhận khẳng định cho rằng hiệu quả của thuốc A cao hơn thuốc B.

5.4 Kiểm tra tính độc lập

Giả sử trên cùng một đám đông $\mathcal{C}$ có hai đặc điểm cần nghiên cứu nào đó; hai đại lượng ngẫu nhiên gốc đám đông tương ứng lần lượt là X và Y, trong mục này chúng ta sẽ đề cập đến dạng bài toán kiểm định tính độc lập của hai đặc điểm này. Như vậy, vấn đề đặt ra ở đây là với mức ý nghĩa α cho trước, đánh giá về cặp giả thiết thống kê

$$H : X \text{ và } Y \text{ độc lập}; \quad K : X \text{ và } Y \text{ phụ thuộc}.$$

Thực hiện n lần quan sát trên đám đông $\mathcal{C}$, đối với mỗi phần tử của mẫu ta quan sát cả hai đặc điểm cần kiểm định. Giả sử ở đặc điểm thứ nhất có k giá trị khác nhau, ở đặc điểm thứ hai có l giá trị khác nhau và n_{ij} là số quan sát có cùng đặc điểm (x_i, y_j). Các số liệu này được sắp xếp thành bảng hai chiều, bảng này được gọi là *bảng tương liên*

X\Y	y_1	...	y_j	...	y_l	$\sum$
x_1	n_{11}	...	n_{1j}	...	n_{1l}	m_1
...	...	...	...	...	...	...
x_i	n_{i1}	...	n_{ij}	...	n_{il}	m_i
...	...	...	...	...	...	...
x_k	n_{k1}	...	n_{kj}	...	n_{kl}	m_k
$\sum$	n_1	...	n_j	...	n_l	n

$$m_i = \sum_{j=1}^{l} n_{ij};\ n_j = \sum_{i=1}^{k} n_{ij};\ n = \sum_{i=1}^{k} m_i = \sum_{j=1}^{l} n_j.$$

Với n đủ lớn và H đúng thì

$$T = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{\left(n_{ij} - \dfrac{m_i\, n_j}{n}\right)^2}{\dfrac{m_i\, n_j}{n}} \simeq \chi^2(k-1)(l-1),$$

trong đó $\chi^2(k-1)(l-1)$ là phân phối "khi bình phương" với $(k-1)(l-1)$ bậc tự do. Suy ra

$$\mathbb{P}[T \in W_\alpha | H \text{ đúng}] = \mathbb{P}\Big[T > \chi^2\big((k-1)(l-1), \alpha\big)\Big] = \alpha,$$

trong đó miền bác bỏ $W_\alpha = \big(\chi^2((k-1)(l-1), \alpha), +\infty\big)$.

Quy tắc thực hành

◇ Xác định các giá trị tần số lý thuyết: $t_{ij} = \dfrac{m_i\, n_j}{n}$.

◇ Xác định giá trị kiểm định từ thực nghiệm

$$t_{tn} = \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{(n_{ij} - t_{ij})^2}{t_{ij}} = n \sum_{i=1}^{k}\sum_{j=1}^{l} \frac{(n_{ij} - t_{ij})^2}{m_i\, n_j}.$$

◇ So sánh giá trị của t_{tn} và $\chi^2\big((k-1)(l-1), \alpha\big)$, nếu

$t_{tn} \leqslant \chi^2\big((k-1)(l-1), \alpha\big)$: chấp nhận H.

$t_{tn} > \chi^2\big((k-1)(l-1), \alpha\big)$: bác bỏ H, chấp nhận K.

Tra bảng phân phối χ^2 (xem bảng 5 phần phụ lục): $\chi^2(n, \alpha)$ là giá trị trong bảng ứng với giá trị hàng là n và cột là α.

Ví dụ 1. Theo dõi huyết áp của người không hút thuốc lá, hút ít và nghiện nặng thu được bảng sau

Huyết áp\ Hút thuốc	không hút thuốc	hút ít	nghiện nặng
Bình thường	46	26	20
Huyết áp cao	23	35	30

Với mức ý nghĩa $\alpha = 0{,}05$, việc hút thuốc lá có ảnh hưởng đến huyết áp không?

Giải. Cặp giả thiết thống kê

H : hút thuốc không ảnh hưởng đến huyết áp; K : hút thuốc ảnh hưởng đến huyết áp.

$n = 180; n_1 = 69; n_2 = 61; n_3 = 50; m_1 = 92; m_2 = 88.$

$$t_{tn} = 180\Big(\frac{(46 - \frac{69\cdot 92}{180})^2}{69\cdot 92} + \frac{(26 - \frac{61\cdot 92}{180})^2}{61\cdot 92} + \frac{(20 - \frac{50\cdot 92}{180})^2}{50\cdot 92} + \frac{(23 - \frac{69\cdot 88}{180})^2}{69\cdot 88} + \frac{(35 - \frac{61\cdot 88}{180})^2}{61\cdot 88} + \frac{(30 - \frac{50\cdot 88}{180})^2}{50\cdot 88}\Big) = 10{,}91.$$

Vì $t_{tn} > \chi^2(2; 0{,}05) = 5{,}991$ nên ta bác bỏ H, chấp nhận K. Nghĩa là chấp nhận khẳng định cho rằng việc hút thuốc có ảnh hưởng đến huyết áp.

Chú ý. Đối với mẫu có kích thước lớn hơn, chúng ta nên lập thêm bảng tần số lý thuyết làm bảng trung gian, điều này khá hợp lý khi sử dụng Excel để tính toán.

Ví dụ 2. Người ta lấy số liệu một số hộ gia đình vay tiền làm kinh tế vườn, bảng sau đây thu thập thông tin theo hai phương diện: lứa tuổi chủ hộ và hiệu quả sử dụng tiền vay

Hiệu quả / Lứa tuổi	Thành công	Bình thường	Thất bại
Từ 20 đến 25	100	64	36
Từ 26 đến 35	156	88	56
Từ 36 đến 45	110	56	34
Trên 45	54	32	14

Với mức ý nghĩa $\alpha = 0{,}01$, hiệu quả việc sử dụng tiền vay có phụ thuộc vào lứa tuổi của chủ hộ không?

Giải. Cặp giả thiết thống kê

H : hiệu quả sử dụng tiền vay không phụ thuộc vào lứa tuổi;

K : hiệu quả sử dụng tiền vay phụ thuộc vào lứa tuổi.

Bảng tần số lý thuyết: $t_{ij} = \dfrac{m_i n_j}{n}$.

Hiệu quả Lứa tuổi	Thành công	Bình thường	Thất bại	m_i
Từ 20 đến 25	105	60	35	200
Từ 26 đến 35	157.5	90	52.5	300
Từ 36 đến 45	105	60	35	200
Trên 45	52.5	30	17.5	100
n_j	420	240	140	800

Xác định giá trị kiểm định từ thực nghiệm

$$t_{tn} = \sum_{i=1}^{4}\sum_{j=1}^{3}\frac{(n_{ij} - t_{ij})^2}{t_{ij}} = 2{,}235.$$

Vì $t_{tn} < \chi^2(6; 0{,}01) = 16{,}812$ nên ta chấp nhận H. Nghĩa là chấp nhận khẳng định cho rằng hiệu quả việc sử dụng tiền vay không phụ thuộc vào lứa tuổi của chủ hộ.

BÀI TẬP CHƯƠNG 5

1. Quy định của một thiết bị phải có chiều dài là 300cm và độ lệch chuẩn là 3cm. Từ một lô hàng người ta lấy ra 40 chiếc, kết quả thu được độ dài trung bình là 301,2cm. Với mức ý nghĩa 5%, lô hàng trên có đạt tiêu chuẩn hay không?

2. Trong điều kiện chăn nuôi bình thường, lượng sữa thu được trung bình hàng ngày của một loại giống bò sữa là 19,4 (đơn vị: kg/ngày). Lấy mẫu 49 con bò sữa ở một trang trại thu được lượng sữa trung bình của một con trong một ngày là 18,9 và độ lệch chuẩn mẫu là 3,24. Với mức ý nghĩa $\alpha = 0{,}08$, lượng sữa thu được hàng ngày từ bò sữa của trang trại có đúng chuẩn không?

3. Khối lượng chuẩn của một bao gạo được đóng gói bằng dây chuyền tự động là đại lượng ngẫu nhiên có phân phối chuẩn với khối lượng mỗi bao là 50 kg. Sau một thời gian hoạt động người ta nghi ngờ khối lượng đó có xu hướng giảm sút. Cân 28 bao gạo thu được khối lượng trung bình mỗi bao là 49,8 kg và độ lệch chuẩn mẫu là 0,6 kg. Với mức ý nghĩa 1%, hãy kết luận về nghi ngờ nói trên.

4. Thời gian trước đây, số tiền gửi tiết kiệm trung bình của mỗi khách hàng vào ngân hàng A là 1000 USD. Sau đợt tăng lãi suất tiết kiệm, kiểm tra ngẫu nhiên 36 khách hàng thu được kết quả: số tiền gửi trung bình là 1060 USD và độ lệch chuẩn mẫu là 100 USD. Với mức ý nghĩa 4%, việc tăng lãi suất có làm tăng lượng tiền gửi tiết kiệm của mỗi khách hàng không?

5. Một kênh truyền thông tuyên bố rằng 30% khán giả truyền hình yêu thích các chương trình phát sóng của họ. Thăm dò ý kiến ngẫu nhiên qua mạng đối với 800 người xem truyền hình thì có 192 người yêu thích các chương trình của kênh truyền thông đó. Với mức ý nghĩa 0,08, tỉ lệ trong tuyên bố trên có đúng với thực tế không?

6. Tỉ lệ phế phẩm của một nhà máy trước đây là 10%. Sau khi cải tiến kỹ thuật, kiểm tra 400 sản phẩm thì thấy có 38 phế phẩm. Với mức ý nghĩa là 1%, kiểm tra xem việc cải tiến kỹ thuật có mang lại hiệu quả không?

7. Tỉ lệ người chữa khỏi một loại bệnh bằng loại thuốc cũ là 80%. Người ta thay thế bằng loại thuốc mới để chữa bệnh cho 1000 người thì có 820 người khỏi bệnh. Với mức ý nghĩa 1%, có thể kết luận thuốc mới tốt hơn thuốc cũ không?

8. Hai giống vịt được nuôi sau 4 tháng. Lấy mẫu $n_1 = 50$ ở giống vịt thứ nhất, được $\overline{x}_1 = 1.9kg$ và $s_1^2 = 1$. Lấy mẫu $n_2 = 80$ ở giống vịt thứ hai, được $\overline{x}_2 = 2kg$ và $s_2^2 = 0.8$. Với mức ý nghĩa $\alpha = 10\%$, hai giống vịt này có trọng lượng trung bình bằng nhau không?

9. Chọn ngẫu nhiên 20 đại lý có áp dụng khuyến mãi thu được số lượng bán trung bình mỗi ngày là 140 sản phẩm và độ lệch chuẩn mẫu là 12; còn tại 20 đại lý không có khuyến mãi được 2 số liệu tương ứng là 120 và 10. Giả sử lượng hàng bán được có phân phối chuẩn, có cùng phương sai. Với mức ý nghĩa 5%, hình thức khuyến mãi có làm tăng số lượng hàng bán không?

10. Một công ty bán hàng muốn kiểm tra hiệu quả từ việc thay đổi kiểu đóng gói. Chọn 2 mẫu: mẫu 1 là 35 đại lý bán hàng theo loại gói cũ và mẫu 2 là 35 đại lý bán hàng theo loại gói mới để thống kê về số gói hàng bán ra sau một tháng, thu được 2 giá trị đặc trưng cho 2 mẫu tương ứng như sau: loại gói cũ: $\overline{x_1} = 560$ gói, với $s_1 = 20$; loại gói mới: $\overline{x_2} = 580$ gói, với $s_2 = 30$. Với mức ý nghĩa 1%, hãy đánh giá việc thay đổi kiểu đóng gói có đem lại hiệu quả hay không?

11. Để so sánh tỉ lệ nảy mầm của hai giống cây trong điều kiện độ ẩm thấp. Người ta đem gieo 200 hạt giống loại I thì có 150 hạt nảy mầm, gieo 300 hạt giống loại II thấy có 210 hạt nảy mầm. Với mức ý nghĩa $\alpha = 0{,}05$, tỉ lệ nảy mầm trong điều kiện độ ẩm thấp của 2 giống cây trên có như nhau không ?

12. Lấy số liệu thực tế từ các hộ gia đình vay vốn của ngân hàng nông nghiệp đối với 2 huyện. Huyện A: có 2000 hộ vay thì có 1692 hộ sử dụng tiền vay có hiệu quả; huyện B: có 1000 hộ vay thì có 810 hộ

sử dụng tiền vay có hiệu quả. Với mức ý nghĩa 5%, tỉ lệ hộ sử dụng tiền vay có hiệu quả của huyện A có cao hơn ở huyện B không?

13. Để đánh giá về chất lượng sản phẩm của nhà máy do 2 dây chuyền sản xuất. Người ta kiểm tra ngẫu nhiên 200 sản phẩm từ dây chuyền thứ nhất thì có 15 phế phẩm, kiểm tra 300 sản phẩm từ dây chuyền thứ hai thấy có 21 phế phẩm. Từ số liệu thu được có thể đánh giá sơ bộ dây chuyền nào làm việc tốt hơn. Với mức ý nghĩa $\alpha = 0{,}08$, kiểm định đánh giá sơ bộ đó.

14. Có hai phương pháp gieo một loại hạt giống: theo phương pháp A, gieo 125 hạt thấy có 90 hạt nảy mầm; theo phương pháp B, gieo 100 hạt thấy có 85 hạt nảy mầm. Từ số liệu thu được có thể đánh giá sơ bộ phương pháp gieo nào tốt hơn. Với mức ý nghĩa $\alpha = 0{,}05$, kiểm định đánh giá sơ bộ đó.

15. Tại một nhà máy làm việc theo chế độ 3 ca: buổi sáng, buổi chiều và buổi tối, chọn ngẫu nhiên một số sản phẩm để kiểm tra chất lượng, thu được bảng số liệu sau

Chất lượng	Ca		
	Sáng	Chiều	Tối
Chính phẩm	84	64	70
Phế phẩm	2	8	2

Với mức ý nghĩa $\alpha = 0{,}05$, có thể kết luận chất lượng sản phẩm phụ thuộc vào ca làm việc không?

16. Tại một nhà máy có 4 phân xưởng: I, II, III, IV; cùng sản xuất ra một loại sản phẩm với 3 tiêu chí đánh giá chất lượng: Loại A (tốt),

loại B (đạt), loại C (chưa đạt). Kiểm tra 1000 sản phẩm khi nhập tổng kho, thu được bảng số liệu sau

Chất lượng Xưởng	Loại A	Loại B	Loại C
I	105	90	25
II	135	102	13
III	124	100	6
IV	146	138	16

Với mức ý nghĩa $\alpha = 0{,}01$, có thể kết luận chất lượng sản phẩm phụ thuộc vào phân xưởng sản xuất không?

CHƯƠNG 6

HỒI QUY VÀ TƯƠNG QUAN

6.1 Hệ số tương quan mẫu

6.1.1 Mở đầu

Trên cùng một đám đông $\mathcal{C}$ có hai đặc điểm định lượng cần nghiên cứu, hai đại lượng ngẫu nhiên gốc đám đông tương ứng lần lượt là X và Y. Bài toán đặt ra ở đây là tìm hiểu mức độ phụ thuộc giữa hai đại lượng ngẫu nhiên và tìm biểu thức biểu diễn sự liên hệ giữa chúng.

Đây là một vấn đề hoàn toàn thực tế, sự phụ thuộc của hai đại lượng ngẫu nhiên X và Y có thể phân thành ba loại:

⋄ Sự phụ thuộc hàm số: tồn tại hàm φ để $Y = \varphi(X)$.

⋄ Sự phụ thuộc thống kê: khi X thay đổi thì phân phối xác suất của Y cũng thay đổi.

⋄ Sự phụ thuộc tương quan: X thay đổi thì kỳ vọng có điều kiện $\mathbb{E}(Y|X)$ cũng thay đổi, nghĩa là $\mathbb{E}(Y|X) = \varphi(X) \neq$ hằng số.

Nếu $\varphi(X) = AX + B$ thì ta nói X và Y có *tương quan tuyến tính*, trong trường hợp ngược lại thì ta nói X và Y có *tương quan phi tuyến.*

Phụ thuộc tương quan là trường hợp riêng của phụ thuộc thống kê, nghĩa là nếu phụ thuộc tương quan thì có sự phụ thuộc về phân phối xác suất. Khi phân tích độ phụ thuộc tương quan giữa hai đại lượng ngẫu nhiên X và Y thì ta không cần xét đến trường hợp nó độc lập với nhau.

6.1.2 Hệ số tương quan mẫu

Chúng ta đã được làm quen với khái niệm hệ số tương quan giữa hai đại lượng ngẫu nhiên X và Y

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\mathbb{D}X\,\mathbb{D}Y}} = \frac{\mathbb{E}(XY) - \mathbb{E}X\mathbb{E}Y}{\sqrt{\mathbb{D}X\,\mathbb{D}Y}}.$$

Đó là số đo mức độ phụ thuộc tuyến tính giữa hai đại lượng ngẫu nhiên X và Y, nhưng nếu chưa biết được phân phối xác suất thì hệ số tương quan lý thuyết $\rho(X, Y)$ chưa xác định được. Do đó ta tìm cách ước lượng $\rho(X, Y)$ bởi một giá trị thu được từ mẫu quan sát, giá trị đó được gọi là *hệ số tương quan mẫu.*

Giả sử ta có n cặp quan sát $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ của (X, Y), *hệ số tương quan mẫu* được tính theo công thức

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}}.$$

Do vậy

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2}\ \sqrt{n\sum_{i=1}^{n} y_i^2 - \Big(\sum_{i=1}^{n} y_i\Big)^2}} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\hat{s}_X\,\hat{s}_Y},$$

trong đó $\overline{xy} = \dfrac{1}{n}\sum_{i=1}^{n} x_i y_i.$

Tương tự như hệ số tương quan, hệ số tương quan mẫu cũng có tính chất $|r| \leqslant 1$. Biểu diễn các cặp (x_i, y_i) của mẫu lên một mặt phẳng tọa độ tạo thành đám mây điểm. Hình ảnh của đám mây điểm thể hiện mối quan hệ giữa X và Y. Nếu đám mây điểm có xu hướng tập trung quanh một đường thẳng nào đó (có hệ số góc khác 0) thì $|r|$ càng gần 1 và ta có thể kết luận X, Y có quan hệ gần với quan hệ tuyến tính (tương quan tuyến tính), còn nếu nó phân tán thành hình tròn hay hình vuông thì $|r|$ gần bằng 0.

Ví dụ 1. Bảng số liệu sau đây là kết quả thu thập từ một công ty về doanh thu (X) và số tiền dành cho quảng cáo (Y) của một số tháng như sau:

X (tỉ đồng)	5	7	8	11	9
Y (triệu đồng)	45	60	75	90	80

Hãy xác định hệ số tương quan mẫu.

Giải. Bảng tính

x_i	y_i	x_iy_i	x_i^2	y_i^2
5	45	225	25	2025
7	60	420	49	3600
8	75	600	64	5625
11	90	990	121	8100
9	80	720	81	6400
40	**350**	**2955**	**340**	**25750**

Hệ số tương quan mẫu

$$r = \frac{5 \cdot 2955 - 40 \cdot 350}{\sqrt{5 \cdot 340 - 40^2}\ \sqrt{5 \cdot 25750 - 350^2}} \approx 0{,}98.$$

Chú ý. Trường hợp số liệu thu thập có kích thước lớn, dạng bảng có tần số chúng ta cũng lập bảng tính trung gian như trên sau đó sử dụng công thức: $r = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\hat{s}_X\,\hat{s}_Y}$ (xem bài tập 4 và 5).

6.2 Phương trình hồi quy bình phương trung bình tuyến tính thực nghiệm

6.2.1 Phương trình hồi quy

Mệnh đề. *Trong tất cả các hàm $h(X)$ dùng để ước lượng Y thì $\varphi(X) = \mathbb{E}(Y|X)$ là hàm có sai số bình phương trung bình nhỏ nhất. Nghĩa là*

$$\mathbb{E}\big(Y - \mathbb{E}(Y|X)\big)^2 \leqslant \mathbb{E}\big(Y - h(X)\big)^2.$$

Chứng minh.

$$\begin{aligned}\mathbb{E}\big(Y - h(X)\big) &= \mathbb{E}\big(Y - \mathbb{E}(Y|X) + \mathbb{E}(Y|X) - h(X)\big)^2\\ &= \mathbb{E}\big(Y - \mathbb{E}(Y|X))^2 + \mathbb{E}\big(\mathbb{E}(Y|X) - h(X)\big)^2 +\\ &\quad 2\mathbb{E}\Big[\big(Y - \mathbb{E}(Y|X)\big)\big(\mathbb{E}(Y|X) - h(X)\big)\Big].\end{aligned}$$

Với mọi hàm $k(X)$ ta luôn có

$$\begin{aligned}\mathbb{E}\big(k(X)\,E(Y|X)\big) &= \int\Big[k(x)\int y\,p(y|x)\,dy\Big]p_X(x)\,dx\\ &= \int\int k(x)\,y\,p(y|x)p_X(x)\,dx\,dy\\ &= \int\int k(x)\,y\,p(x,\,y)\,dx\,dy = \mathbb{E}\big(k(X)\,Y\big).\end{aligned}$$

Đặt $k(X) = \mathbb{E}(Y|X) - h(X)$, suy ra

$$\mathbb{E}\Big[\big(Y - \mathbb{E}(Y|X)\big)\big(\mathbb{E}(Y|X) - h(X)\big)\Big] = \mathbb{E}\Big[\big(Y - \mathbb{E}(Y|X)\big)\,k(X)\Big]$$
$$= \mathbb{E}\big[k(X)\,Y\big] - \mathbb{E}\big[k(X)\,\mathbb{E}(Y|X)\big] = 0.$$

Do đó

$$\begin{aligned}\mathbb{E}\big(Y - h(X)\big) &= \mathbb{E}\big(Y - \mathbb{E}(Y|X)\big)^2 + \mathbb{E}\big(\mathbb{E}(Y|X) - h(X)\big)^2 \\ &\geqslant \mathbb{E}\big(Y - \mathbb{E}(Y|X)\big)^2.\end{aligned}$$

□

Như vậy $\mathbb{E}(Y|X)$ là hàm ước lượng Y có sai số bình phương trung bình nhỏ nhất. Phương trình $\varphi(X) = \mathbb{E}(Y|X)$ được gọi là *phương trình hồi quy* của Y theo X.

6.2.2 Ước lượng hệ số hồi quy tuyến tính thực nghiệm

Giả sử X là đại lượng ngẫu nhiên độc lập còn Y là đại lượng ngẫu nhiên phụ thuộc và giữa chúng có tương quan tuyến tính

$$\mathbb{E}(Y|X) = AX + B, \quad A \neq 0,$$

trong đó A, B chưa biết và được gọi là *hệ số hồi quy lý thuyết*.

Bài toán. Căn cứ vào n cặp quan sát $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ của (X, Y), ta cần đi tìm một phương trình $y = ax+b$ ước lượng cho phương trình hồi quy tuyến tính lý thuyết $\mathbb{E}(Y|X) = AX + B$.

Phương trình $y = ax + b$ được gọi là *phương trình hồi quy tuyến tính thực nghiệm*; a và b được gọi là *hệ số hồi quy tuyến tính thực nghiệm* của Y theo X. Chúng ta sử dụng phương pháp bình phương bé nhất để xác định giá trị của a và b.

Như vậy, giữa giá trị thực nghiệm và giá trị xác định từ phương trình hồi quy tuyến tính thực nghiệm tại x_i có sai số $|y_i-(ax_i+b)|$. Tiêu chuẩn để xác định phương trình hồi quy tuyến tính thực nghiệm $y = ax + b$ là đảm bảo được yêu cầu

$$F(a,b) = \sum_{i=1}^{n} \big(y_i - (ax_i + b)\big)^2 \Rightarrow \min .$$

Tìm cực tiểu của $F(a,b)$ dẫn đến hệ phương trình

$$\begin{cases} \dfrac{\partial F(a,\, b)}{\partial a} = -2\displaystyle\sum_{i=1}^{n}(y_i - ax_i - b)\, x_i = 0; \\ \dfrac{\partial F(a,\, b)}{\partial b} = -2\displaystyle\sum_{i=1}^{n}(y_i - ax_i - b) = 0, \end{cases}$$

tương đương với hệ

$$\begin{cases} \Big(\displaystyle\sum_{i=1}^{n} x_i^2\Big) a + \Big(\displaystyle\sum_{i=1}^{n} x_i\Big) b = \displaystyle\sum_{i=1}^{n} x_i\, y_i; \\ \Big(\displaystyle\sum_{i=1}^{n} x_i\Big) a + nb = \displaystyle\sum_{i=1}^{n} y_i. \end{cases}$$

Giải hệ phương trình bậc nhất đối với a và b, ta được

$$\begin{cases} a = \dfrac{n\displaystyle\sum_{i=1}^{n} x_i\, y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\displaystyle\sum_{i=1}^{n} x_i^2 - \Big(\sum_{i=1}^{n} x_i\Big)^2}; \\ b = \dfrac{\displaystyle\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i}{n}. \end{cases}$$

Ngoài ra, hệ số hồi quy tuyến tính thực nghiệm còn có thể xác định nhờ công thức tương đương

$$\begin{cases} a = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\hat{s}_X^2}; \\ b = \overline{y} - a\,\overline{x}\,. \end{cases}$$

Ví dụ 2. Với giả thiết như ở Ví dụ 1:

$$n = 5;\ \sum x_i = 40;\ \sum y_i = 350;\ \sum x_i^2 = 340;\ \sum x_i y_i = 2955.$$

a. Tìm phương trình hồi quy tuyến tính thực nghiệm của y theo x.

b. Nếu doanh thu của một tháng nào đó là 10 tỉ đồng, hãy dự đoán chi phí quảng cáo của công ty tháng đó là bao nhiêu.

Giải. a. Hệ số hồi quy tuyến tính thực nghiệm

$$a = \frac{5 \times 2955 - 40 \times 350}{5 \times 340 - (40)^2} = 7,75; \quad b = \frac{350 - 7,75 \times 40}{5} = 8.$$

Phương trình hồi quy tuyến tính thực nghiệm: $y = 7,75\,x + 8$.

b. $x = 10$ suy ra $y = 85{,}5$. Vậy chi phí quảng cáo của tháng đó khoảng 85,5 triệu đồng.

BÀI TẬP CHƯƠNG 6

1. Bảng số liệu sau đây là kết quả thống kê về tổng giá trị hàng nông sản (X) và tổng đầu tư xây dựng đường giao thông (Y) của một huyện trong 6 năm như sau: (đơn vị: tỉ đồng)

X	60	45	75	90	80	70
Y	7	5	8	11	9	10

   (a) Hãy xác định hệ số tương quan mẫu.

   (b) Tìm phương trình hồi quy tuyến tính thực nghiệm của y theo x.

   (c) Nếu tiền đầu tư xây dựng đường giao thông của một năm nào đó là 8,6 tỉ đồng, hãy dự đoán tổng giá trị hàng nông sản năm đó là bao nhiêu?

2. Bảng số liệu sau đây là kết quả thu được của một công ty về số tiền dành cho các hoạt động chăm sóc khách hàng (X) và doanh thu (Y) trong 6 tháng như sau:

X	8	9	7	10	9	11
Y	600	700	500	900	800	1100

(đơn vị: triệu đồng).

(a) Hãy xác định hệ số tương quan mẫu.

(b) Nếu chi phí dành cho các hoạt động chăm sóc khách hàng của một tháng nào đó là 10,5 triệu đồng, hãy dự đoán doanh thu của công ty tháng đó là bao nhiêu?

3. Thống kê ghi lại dân số của một tỉnh qua 8 năm từ năm 1985 đến 1992 được bảng số sau

Năm	1985	1986	1987	1988	1989	1990	1991	1992
Dân số (10000)	50	51	51	53	54	56	59	60

Để thuận tiện trong tính toán ta đặt $x =$ "năm" $- 1985$ và $y =$ "dân số" $- 50$ (đơn vị 10000 người). Hãy tìm phương trình hồi quy tuyến tính thực nghiệm của y theo x.

4. Tính hệ số tương quan mẫu và phương trình hồi quy tuyến tính thực nghiệm của y theo x dựa vào bảng tần số sau:

x_i	17	14	12	15	12	20
y_i	31	33	25	29	27	40
n_i	2	4	10	3	5	6

5. Bảng số liệu sau đây chỉ ra sự phụ thuộc của năng suất thu hoạch Y theo lượng phân bón X của một loại hoa màu trên 100 thửa ruộng.

Y	X			
	20	25	30	35
400	12	5	1	1
420	6	18	3	2
450	2		10	9
490		1	10	20

Tính hệ số tương quan mẫu và phương trình hồi quy tuyến tính thực nghiệm của năng suất thu hoạch theo lượng phân bón.

HƯỚNG DẪN GIẢI BÀI TẬP

Trong phần này, chúng tôi sẽ hướng dẫn giải một số bài tập tương đối mới, chưa được trình bày nhiều trong các tài liệu bằng tiếng Việt.

CHƯƠNG 1

5. a) Sử dụng đẳng thức $A \triangle B = (A \setminus B) \cup (B \setminus A)$.

b) Sử dụng công thức tính $\mathbb{P}(A \cup B \cup C)$.

c) Nhận xét $AB \cup AC \subset A$.

d) Sử dụng các kết quả a), b), c).

6. a) Chuyển vế, sử dụng công thức

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

Sau đó biến đổi tương đương.

b) Trước hết, nhận xét rằng, nếu $AB = \emptyset$ thì

$$\mathbb{P}(A)\mathbb{P}(B) \leqslant \Big(\frac{\mathbb{P}(A) + \mathbb{P}(B)}{2}\Big)^2 \leqslant \frac{1}{4}.$$

Xét trường hợp A, B bất kỳ. Khi đó

$$\begin{aligned}\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B) &\leqslant \mathbb{P}(AB) - \mathbb{P}(AB)\mathbb{P}(B)\\ &\leqslant \mathbb{P}(AB)\mathbb{P}(\bar{B}) \leqslant \frac{1}{4} \quad (1).\end{aligned}$$

$\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB) = [\mathbb{P}(A \setminus AB) + \mathbb{P}(AB)]\mathbb{P}(B) - \mathbb{P}(AB) = \mathbb{P}(A \setminus AB)\mathbb{P}(B) - \mathbb{P}(AB)(\mathbb{P}(B) - 1) \leqslant \mathbb{P}(A \setminus AB)\mathbb{P}(B) \leqslant \frac{1}{4}$ (2).

Từ (1) và (2) suy ra điều phải chứng minh.

7. Sử dụng định nghĩa xác suất có điều kiện, biến đổi vế phải để được điều phải chứng minh.

8.

$$\mathbb{P}(A|C) = \frac{\mathbb{P}(AC)}{\mathbb{P}(C)} = \frac{\sum_{k=1}^{n} \mathbb{P}(AB_kC)}{\mathbb{P}(C)} =$$

$$= \frac{\sum_{k=1}^{n} \mathbb{P}(A|B_kC)\mathbb{P}(B_kC)}{\mathbb{P}(C)} = \sum_{k=1}^{n} \mathbb{P}(A|B_kC)\mathbb{P}(B_k|C).$$

9. $1 = \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(AB) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(B) \Rightarrow 1 - \mathbb{P}(B) = \mathbb{P}(A)(1 - \mathbb{P}(B)) \Rightarrow 0 = (1 - \mathbb{P}(A))(1 - \mathbb{P}(B)) \Rightarrow \mathbb{P}(A) = 1$ hoặc $\mathbb{P}(B) = 1$.

10. $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(AB) = \mathbb{P}\big((A \cup B)AB\big) = \mathbb{P}(A \cup B)\mathbb{P}(AB) = (\mathbb{P}(A \cup B))\mathbb{P}(A)\mathbb{P}(B) \Rightarrow \mathbb{P}(A) = 0$ hoặc $\mathbb{P}(B) = 0$ hoặc $\mathbb{P}(A \cup B) = 1 \Rightarrow \mathbb{P}(A) = 0$ hoặc $\mathbb{P}(B) = 0$ hoặc $\mathbb{P}(A) = 1$ hoặc $\mathbb{P}(B) = 1$.

11. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) \Rightarrow \mathbb{P}(A) = 0$ hoặc $\mathbb{P}(B) = 0$. Kết hợp điều này với điều kiện $\mathbb{P}(A \triangle B) = p$, ta được $\mathbb{P}(A \setminus B) = p$ hoặc $\mathbb{P}(B \setminus A) = p$. Nhưng $\mathbb{P}(A \setminus B) < p$, nên $\mathbb{P}(B \setminus A) = p \Rightarrow \mathbb{P}(A \setminus B) = 0 \Rightarrow \mathbb{P}(B) = p; \mathbb{P}(A) = 0$.

12. $\mathbb{P}(ABBC) = \mathbb{P}(ABC) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) \neq \mathbb{P}(A)\mathbb{P}(B)^2\mathbb{P}(C) = \mathbb{P}(AB)\mathbb{P}(BC)$ suy ra AB và BC không độc lập. Tương tự, AB và AC, AC và BC cũng không độc lập.

13. Ta có

$$\mathbb{P}(ABC) = \mathbb{P}(B)\mathbb{P}(AC) \quad (1)$$

$$\mathbb{P}(ABC) = \mathbb{P}(C)\mathbb{P}(AB) \quad (2)$$

$$\mathbb{P}(ABC) = \mathbb{P}(A)\mathbb{P}(BC) \quad (3)$$

$\mathbb{P}(A)(\mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(BC)) = \mathbb{P}(A)\mathbb{P}(B \cup C) = \mathbb{P}(A(B \cup C)) = \mathbb{P}(AB \cup AC) = \mathbb{P}(AB) + \mathbb{P}(AC) - \mathbb{P}(ABC)$. Điều này kết hợp với (3) cho ta $\mathbb{P}(A)\mathbb{P}(B)+\mathbb{P}(A)\mathbb{P}(C)-\mathbb{P}(A)\mathbb{P}(BC)= \mathbb{P}(AB)+\mathbb{P}(AC)-\mathbb{P}(A)\mathbb{P}(BC)$ hay

$$\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AC) \quad (4).$$

Mặt khác, từ (1) và (2) suy ra

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

Thay hệ thức này vào (4), ta được

$$\mathbb{P}(B)\Big[\frac{\mathbb{P}(AC)}{\mathbb{P}(C)} - \mathbb{P}(A)\Big] = \mathbb{P}(A)\mathbb{P}(C) - \mathbb{P}(AC)$$

$$= \mathbb{P}(C)\Big[\mathbb{P}(A) - \frac{\mathbb{P}(AC)}{\mathbb{P}(C)}\Big].$$

Từ đó suy ra

$$\mathbb{P}(A) - \frac{\mathbb{P}(AC)}{\mathbb{P}(C)} = 0.$$

Hay

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

Điều này cùng với (1) cho ta

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

Từ hai hệ thức cuối cùng, kết hợp với (2) và (3) cho ta điều phải chứng minh.

14. a) Ta có

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

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

Do đó, nếu A và B ε-độc lập thì

$$|\mathbb{P}(\bar{A}B) - \mathbb{P}(\bar{A})\mathbb{P}(B)| = |\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)| \leqslant \varepsilon.$$

Tức là $\bar{A}$ và B ε-độc lập. Tương tự, ta cũng chứng minh được A và $\bar{B}$, $\bar{A}$ và $\bar{B}$ cũng ε-độc lập.

b) Giả sử $\mathbb{P}(A) \leqslant \varepsilon$. Khi đó

$$|\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(AB)| \leqslant \max(\mathbb{P}(A)\mathbb{P}(B), \mathbb{P}(AB)\} \leqslant \varepsilon.$$

Do đó A và B là ε-độc lập.

Nếu $\mathbb{P}(A) \geqslant 1 - \varepsilon$ thì $\mathbb{P}(\bar{A}) \leqslant \varepsilon$. Do đó $\bar{A}$ và B là ε-độc lập. Suy ra A và B bất kỳ cũng ε-độc lập.

c) **Cách 1.** Chỉ cần xét với $\varepsilon < 1/4$. Ta có

$$|\mathbb{P}(AA) - \mathbb{P}(A)\mathbb{P}(A)| \leqslant \varepsilon.$$

Do đó

$$\mathbb{P}(A)(1 - \mathbb{P}(A)) \leqslant \varepsilon.$$

Giải ra ta được

$\mathbb{P}(A) \leqslant \frac{1-\sqrt{1-4\varepsilon}}{2} \leqslant 2\varepsilon$ hoặc $\mathbb{P}(A) \geqslant \frac{1+\sqrt{1-4\varepsilon}}{2} \geqslant 1 - 2\varepsilon$. Vì $0 < 1 - 4\varepsilon < 1 \Rightarrow 1 - 4\varepsilon < \sqrt{1 - 4\varepsilon}$.

Cách 2. Xét hai trường hợp $\mathbb{P}(A) \geq 1/2$ và $\mathbb{P}(A) \leq 1/2$.

27. $\lim\limits_{n\to\infty} \mathbb{P}(A_n) = \lim\limits_{n\to\infty} \big(\mathbb{P}(A_n\bar{B}_n)+\mathbb{P}(A_nB_n)\big) \leqslant \lim\limits_{n\to\infty} \mathbb{P}(\bar{B}_n)+\lim\limits_{n\to\infty} \mathbb{P}(A_nB_n) = \lim\limits_{n\to\infty} \mathbb{P}(A_nB_n)$.

Bất đẳng thức ngược lại là hiển nhiên.

28. a) Chứng minh trực tiếp.

b) Dùng công thức đối ngẫu De-Morgan.

29. Dùng tính chất của các phép toán trên các tập hợp để biến đổi trực tiếp. Chẳng hạn:

$$\limsup(A_n \bigcup B_n) = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} (A_k \bigcup B_k) =$$

$$= \bigcap_{n=1}^{\infty} [(\bigcup_{k=n}^{\infty} A_k) \bigcup (\bigcup_{k=n}^{\infty} B_k)] =$$

$$= (\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k) \bigcup (\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} B_k) = \limsup A_n \bigcup \limsup B_n.$$

$$\limsup (A_n \bigcap B_n) = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \big(A_k \bigcap B_k\big) \subset$$

$$\subset \bigcap_{n=1}^{\infty} (\bigcup_{k=n}^{\infty} A_k) = \limsup A_n \quad (1).$$

Tương tự

$$\limsup (A_n \bigcup B_n) \subset \limsup B_n \quad (2).$$

Từ (1) và (2) suy ra

$$\limsup (A_n \cap B_n) \subset \limsup A_n \cap \limsup B_n.$$

c) $\limsup A_n = A \cap B$; $\liminf A_n = A \cup B$.

30. Đặt

$$B_1 = A_1, B_2 = \bar{A}_1 A_2, \ldots, B_n = \bar{A}_1 \bar{A}_2 \cdots \overline{A_{n-1}} A_n, \ldots.$$

Khi đó

$$B_i B_j = \emptyset, \quad \bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n.$$

Từ đó, suy ra điều phải chứng minh.

31. a)-b) Sử dụng tính chất liên tục của xác suất để chứng minh.

c) $\mathbb{P}(\liminf A_n) = \mathbb{P}\big(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\big) = \lim_{n\to\infty} \mathbb{P}\big(\bigcap_{k=n}^{\infty} A_n\big)$

$\leqslant \liminf \mathbb{P}(A_n) \leqslant \limsup \mathbb{P}(A_n) \leqslant \lim_{n\to\infty} \mathbb{P}\big(\bigcup_{k=n}^{\infty} A_n\big)$

$= \mathbb{P}\big(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\big) = \mathbb{P}(\limsup A_n)$.

(Vì $\bigcap_{k=n}^{\infty} A_k \subset A_l \subset \bigcup_{k=n}^{\infty} A_k$, với mọi $l \geqslant n$).

32. Trước hết chứng minh rằng với mọi họ hữu hạn các biến cố $A_{i_1}, A_{i_2}, \ldots, A_{i_n}$ của họ $(A_i)_{i \in I}$, ta đều có

$$\mathbb{P}(\overline{A_{i_1}} A_{i_2} \ldots A_{i_n}) = P(\overline{A_{i_1}}).\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_n}).$$

33.

$$\mathbb{P}(\bigcup_{k=1}^{n} A_k) = 1 - \mathbb{P}(\bigcap_{k=1}^{n} \bar{A}_k) = 1 - \prod_{k=1}^{n} \mathbb{P}(A_k).$$

34. Đặt

$$B_n = \bigcap_{k=1}^{n} A_k.$$

Khi đó, dãy $B_1, B_2, \ldots$ là dãy giảm và

$$\bigcap_{k=1}^{\infty} A_k = \bigcap_{k=1}^{\infty} B_k.$$

Suy ra

$$\mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = \mathbb{P}(\bigcap_{n=1}^{\infty} B_n) = \lim_{n \to \infty} \mathbb{P}(B_n)$$

$$= \lim_{n \to \infty} \mathbb{P}(\bigcap_{k=1}^{n} A_k) = \lim_{n \to \infty} \prod_{k=1}^{n} \mathbb{P}(A_k) = \prod_{n=1}^{\infty} \mathbb{P}(A_n).$$

35. Sử dụng Bài tập 33 và kết quả sau đây trong Giải tích: Tích vô hạn $\prod_{n=1}^{\infty}(1 + a_n)$ hội tụ tuyệt đối tới một giới hạn khác không khi và chỉ khi chuỗi $\sum_{n=1}^{\infty} a_n$ hội tụ tuyệt đối.

36. Giả sử $\mathcal{A}$ là một họ các biến cố độc lập đôi một. Khi đó, với mọi $n = 1, 2, \ldots$ và mọi $A_1, A_2 \ldots, A_n \in \mathcal{A}$, ta có $A_2 A_3 \cdots A_n$ độc lập với A_1 (vì họ các biến cố độc lập với A_1 lập thành một đại số). Vì vậy

$$\mathbb{P}(A_1 A_2 \cdots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2 \cdots A_n).$$

Lý luận tương tự đối với họ $A_2 A_3 \ldots A_n$, ta được

$$\mathbb{P}(A_2 A_3 \cdots A_n) = \mathbb{P}(A_2)\mathbb{P}(A_3 \cdots A_n).$$

....

Từ đó ta được

$$\mathbb{P}(A_1A_2\cdots A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2)\cdots\mathbb{P}(A_n).$$

Điều này chứng tỏ $\mathcal{A}$ là họ độc lập.

37. Nếu tất cả các biến cố của không gian xác suất đều có xác suất bằng 0 hoặc bằng 1 thì hiển nhiên chúng độc lập, do đó độc lập đôi một. Ngược lại, giả sử tập hợp tất cả các biến cố là tập độc lập đôi một. Khi đó, nếu tồn tại một biến cố A sao cho $0 < \mathbb{P}(A) < 1$ thì $0 < \mathbb{P}(\overline{A}) < 1$. Mặt khác $\mathbb{P}(A)\mathbb{P}(\overline{A}) = \mathbb{P}(A\overline{A}) = \mathbb{P}(\emptyset) = 0$, vô lý.

38. Không. Chẳng hạn, xét không gian xác suất mà trong đó tồn tại một biến cố có xác suất khác 0 và khác 1. Khi đó, theo bài trên, tồn tại hai biến cố không độc lập với nhau. Tuy nhiên, cả hai biến cố đó đều độc lập với biến cố $\emptyset$.

39. Không. Chẳng hạn, lấy $\Omega = (1,2,3,4\}; A = (1,2\}; B = (4\}; C = (2,3\};$

$\mathbb{P}(1) = \mathbb{P}(2) = \mathbb{P}(3) = \mathbb{P}(4) = 1/4$. Khi đó A, C không độc lập với B, nhưng C độc lập với A.

40. Điều kiện đủ là hiển nhiên. Ngược lại, giả sử quan hệ độc lập giữa các biến cố có tính bắc cầu. Khi đó, nếu tồn tại một biến cố A sao cho $0 < \mathbb{P}(A) < 1$ thì $0 < \mathbb{P}(\overline{A}) < 1$. Nhưng A và $\overline{A}$ đều độc lập với $\emptyset$, nên chúng độc lập với nhau. Suy ra $\mathbb{P}(A)\mathbb{P}(\overline{A}) = \mathbb{P}(A\overline{A}) = \mathbb{P}(\emptyset) = 0$, vô lý.

41. a) Hiển nhiên.

b)

$$(\bigcup_{n=1}^{\infty} A_n)\triangle(\bigcup_{n=1}^{\infty} B_n) = (\bigcup_{n=1}^{\infty} A_n)\overline{(\bigcup_{n=1}^{\infty} B_n)}\bigcup(\bigcup_{n=1}^{\infty} B_n)\overline{(\bigcup_{n=1}^{\infty} A_n)}$$

$$= (\bigcup_{n=1}^{\infty} A_n)(\bigcap_{n=1}^{\infty} \overline{B}_n)\bigcup(\bigcup_{n=1}^{\infty} B_n)(\bigcup_{n=1}^{\infty} \overline{A}_n)$$

$$\subset \bigcup_{n=1}^{\infty} \big[(A_n\overline{B}_n)\bigcup(B_n\overline{A}_n)\big] = \bigcup_{n=1}^{\infty}(A_n\triangle B_n).$$

c) Tương tự b).

42. Đặt

$$\mathcal{B} = (A: \ A \in \mathcal{F} \ \text{ và } \ \forall\varepsilon > 0, \exists B_\varepsilon \in \mathcal{A} \text{ sao cho } \mathbb{P}(A\triangle B_\varepsilon) < \varepsilon\}.$$

Khi đó $\mathcal{B} \supset \mathcal{A}$. Hơn nữa, $\mathcal{B}$ là σ- đại số. Thật vậy

$$\bar{A}\triangle\bar{B} = A\triangle B \Rightarrow \bar{A} \in \mathcal{B} \text{ nếu } A \in \mathcal{B}.$$

Ta lại có

$$\bigcup_{n=1}^{\infty} A_n\triangle\bigcup_{n=1}^{\infty} B_n \subset \bigcup_{n=1}^{\infty}(A_n\triangle B_n)$$

và

$$\lim_{n\to\infty} P\big[\bigcup_{j=1}^{\infty} A_j\triangle\bigcup_{j=1}^{n} B_j\big] = P\big(\bigcup_{j=1}^{\infty} A_j\triangle\bigcup_{j=1}^{\infty} B_j\big).$$

Suy ra $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$ nếu $A_n \in \mathcal{B}$ $(\forall n \geqslant 1)$.

Vậy $\mathcal{B}$ là σ- đại số. Do đó $\mathcal{F} = \sigma(\mathcal{A}) \subset B$.

43. Giả sử $\mathcal{A}$ và $\mathcal{B}$ là hai σ-đại số độc lập đã cho. Khi đó, tồn tại hai biến cố A và B sao cho $A \in \mathcal{A}, B \in \mathcal{B}$, $0 < \mathbb{P}(A) < 1$,$0 < \mathbb{P}(B) < 1$, $0 < \mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(AB) < 1$. Nếu $\mathcal{A} \cup \mathcal{B}$ là đại số thì $AB \in \mathcal{A} \cup \mathcal{B}$. Do đó $AB \in \mathcal{A}$ hoặc $AB \in \mathcal{B}$. Nếu $AB \in \mathcal{A}$ thì AB độc lập với B, do đó $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(AB) = \mathbb{P}(ABB) = \mathbb{P}(A)\mathbb{P}(B)^2$. Điều này kéo theo $\mathbb{P}(B) = 1$, trái với giả thiết. Nếu $AB \in \mathcal{B}$ thì AB độc lập với A, do đó $\mathbb{P}(A)\mathbb{P}(B) = \mathbb{P}(AB) = \mathbb{P}(AAB) = \mathbb{P}(A)^2\mathbb{P}(B)$. Điều này kéo theo $\mathbb{P}(A) = 1$, cũng trái với giả thiết.

44.

$$\frac{\mathbb{P}(A_n)}{\mathbb{P}(A_nB_n)} = \frac{\mathbb{P}(A_n\overline{B_n}) + \mathbb{P}(A_nB_n)}{\mathbb{P}(A_nB_n)} = 1 + \frac{\mathbb{P}(A_n\overline{B_n})}{\mathbb{P}(A_nB_n)}.$$

45. Ta có $\mathbb{P}(\limsup A_n) = \lim\limits_{n\to\infty} \mathbb{P}\big(\bigcup_{k=n}^{\infty} A_k\big)$.
Mặt khác, $\mathbb{P}\big(\bigcup_{k=n}^{\infty} A_n\big) = \lim\limits_{m\to\infty} \mathbb{P}\big(\bigcup_{k=n}^{n+m} A_k\big)$
$= \lim\limits_{m\to\infty} \mathbb{P}\Big(\big(\bigcup_{k=n}^{n+m-1} A_k\bar{A}_{k+1}\big) \bigcup A_{n+m}\Big)$
$\leqslant \lim\limits_{m\to\infty} \Big(\sum_{k=n}^{n+m-1} \mathbb{P}(A_k\bar{A}_{k+1}) + \mathbb{P}(A_{n+m})\Big)$
$\leqslant \sum_{k=n}^{n+m-1} \mathbb{P}(A_k\bar{A}_{k+1}) + \lim\limits_{m\to\infty} \mathbb{P}(A_{n+m})$
$= \sum_{k=n}^{n+m-1} \mathbb{P}(A_k\bar{A}_{k+1}) \to 0$ khi $n \to \infty$.

CHƯƠNG 2

28. Đặt $Z = \max(X, Y\}$, $T = \min(X, Y\}$, $U = \max(2X, Y\}$, $V = \min(X^2, Y\}$. Ta có

$$F_Z(x) = \mathbb{P}(\max(X, Y\} < x) = \mathbb{P}(X < x; Y < x)$$

$$= \mathbb{P}(X < x)\mathbb{P}(Y < x) = F(x)G(x).$$

$$F_T(x) = \mathbb{P}(\min(X, Y\} < x) = 1 - \mathbb{P}(\min(X; Y\} \geqslant x)$$

$$= 1 - \mathbb{P}(X \geqslant x; Y \geqslant x) = 1 - \mathbb{P}(X \geqslant x)\mathbb{P}(Y \geqslant x)$$

$$= 1 - (1 - F(x))(1 - G(x)).$$

$$F_U(x) = \mathbb{P}(\max(2X, Y\} < x) = \mathbb{P}(2X < x; Y < x) =$$

$$\mathbb{P}(X < x/2)\mathbb{P}(Y < x) = F(x/2)G(x).$$

$$F_V(x) = \mathbb{P}(\min(X, Y^3\} < x) = 1 - \mathbb{P}(\min(X; Y^3\} \geqslant x)$$

$$= 1 - \mathbb{P}(X \geqslant x; Y^3 \geqslant x) = 1 - \mathbb{P}(X \geqslant x)\mathbb{P}(Y \geqslant \sqrt[3]{x})$$

$$= 1 - (1 - F(x))(1 - G(\sqrt[3]{x})).$$

29. Đặt

$$Z = X^3,\ T = \max\{X, Y^3\},\ U = \min\{X, Y^3\}.$$

Lập luận như bài trên, ta được

$$F_Z(x) = \begin{cases} 1 - e^{-\lambda\sqrt[3]{x}} & \text{nếu } x \geqslant 0 \\ 0 & \text{nếu } x < 0 \end{cases}.$$

$$p_Z(x) = \begin{cases} \frac{\lambda}{3}x^{-\frac{2}{3}}e^{-\lambda\sqrt[3]{x}} & \text{nếu } x \geqslant 0 \\ 0 & \text{nếu } x < 0 \end{cases}.$$

$$F_T(x) = \begin{cases} (1 - e^{-\lambda x})(1 - e^{-\lambda\sqrt[3]{x})} & \text{nếu } x \geqslant 0 \\ 0 & \text{nếu } x < 0 \end{cases}.$$

$$p_T(x) = \begin{cases} \lambda e^{-\lambda x}(1 - e^{-\lambda\sqrt[3]{x}}) + \frac{\lambda}{3}x^{-\frac{2}{3}}e^{-\lambda\sqrt[3]{x}}(1 - e^{-\lambda x}) \\ \qquad \text{nếu } x \geqslant 0 \\ 0 \qquad \text{nếu } x < 0 \end{cases}.$$

$$F_U(x) = \begin{cases} 1 - e^{-\lambda(x+\sqrt[3]{x})} & \text{nếu } x \geqslant 0 \\ 0 & \text{nếu } x < 0 \end{cases}.$$

$$p_U(x) = \begin{cases} \lambda\Big(1 + \frac{1}{3}x^{-\frac{2}{3}}\Big)e^{-\lambda(x+\sqrt[3]{x})} & \text{nếu } x \geqslant 0 \\ 0 & \text{nếu } x < 0 \end{cases}.$$

30. Đặt

$$Z = X^3,\ T = \max\{X, Y^3\},\ U = \min\{X, Y^3\}.$$

Lập luận như bài trên, ta được

$$F_Z(x) = \begin{cases} 0 & \text{nếu } x < -1 \\ \frac{1}{4}(x+1)(\sqrt[3]{x}+1) & \text{nếu } |x| \leqslant 1 \\ 1 & \text{nếu } x > 1 \end{cases}.$$

$$p_Z(x) = \begin{cases} \frac{1}{6}x^{-2/3} & \text{nếu } |x| \leqslant 1 \\ 0 & \text{nếu } |x| > 1 \end{cases}.$$

$$F_T(x) = \begin{cases} 0 & \text{nếu } x < -1 \\ \frac{1}{2}\sqrt[3]{x} + \frac{1}{2} & \text{nếu } |x| \leqslant 1 \\ 1 & \text{nếu } x > 1 \end{cases}.$$

$$p_T(x) = \begin{cases} \frac{1}{4}\Big(\frac{4}{3}x^{1/3} + \frac{1}{3}x^{-2/3} + 1\Big) & \text{nếu } |x| \leqslant 1 \\ 0 & \text{nếu } |x| > 1 \end{cases}.$$

$$F_U(x) = \begin{cases} 0 \quad \text{nếu } x < -1 \\ 1 - \Big(1 - \frac{1}{2}(x+1)\Big)\Big(1 - \frac{1}{2}(x^{1/3}+1)\Big) \\ \quad \text{nếu } |x| \leqslant 1 \\ 1 \quad \text{nếu } x > 1 \end{cases}.$$

$$p_Z(x) = \begin{cases} -\frac{1}{3}x^{1/3} + \frac{1}{12}x^{-2/3} + \frac{1}{2} & \text{nếu } |x| \leqslant 1 \\ 0 & \text{nếu } |x| > 1 \end{cases}.$$

31. Đặt $Y = f(X)$. Khi đó

$$F_Y(x) = \mathbb{P}\big(f(X) < x\big) = F_X\big(f^{-1}(x)\big).$$

Do đó

$$p_Y(x) = \Big[F_X\big(f^{-1}(x)\big)\Big] = \begin{cases} \frac{1}{f'(x)}e^{-\lambda f^{-1}(x)} & \text{nếu } x \geqslant 0 \\ 0 & \text{nếu } x < 0 \end{cases}.$$

32. ứng với mỗi $k = 1, 2, \ldots$, chỉ tồn tại hữu hạn điểm x mà xác suất tại đó là $p = F(x+0) - F(x)$ thoả mãn hệ thức $\frac{1}{2^{k+1}} < p \leqslant \frac{1}{2^k}$.

33. Giả sử $\varepsilon > 0$ chọn tuỳ ý. Khi đó, tìm được số $A > 0$ đủ lớn sao cho $1 - F(A) \leqslant \varepsilon$ và $F(-A) \leqslant \varepsilon$. Trên đoạn $[-A, A]$, $F(x)$ liên tục đều nên tìm được $\delta > 0$ sao cho với mọi $x_1, x_2 \in [-A, A]$, $|x_1 - x_2| < \delta$ thì $|F(x_1) - F(x_2)| < \varepsilon/2$. Khi đó $|F(x_1) - F(x_2)| < \varepsilon$ với mọi $x_1, x_2 \in R$, $|x_1 - x_2| < \delta$.

34. Đặt

$$Y = \max(X_1, \dots, X_n\}; \quad Z = \min(X_1, \dots, X_n\}.$$

Khi đó

$$Y \geqslant X_i \geqslant Z, \ \forall i = 1, \dots, n.$$

Từ đó suy ra các bất đẳng thức cần chứng minh.

35. Với mọi số thực $a \in R$, ta có

$$\mathbb{E}(X-a)^2 - DX = (a - \mathbb{E}X)^2 \geqslant 0.$$

Từ đó suy ra điều phải chứng minh.

36. $\mathbb{P}(|X-Y| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/2) + \mathbb{P}(|Y_n - X| > \varepsilon/2) \to 0$ khi $n \to \infty$ $(\forall \varepsilon > 0)$. Suy ra $\mathbb{P}(|X-Y| > \varepsilon) = 0$ $(\forall \varepsilon > 0)$.

37. Dùng định nghĩa.

38. $\mathbb{P}(|Y_n - X| > \varepsilon) \leq \mathbb{P}[(|Y_n - X| > \varepsilon)(X_n = Y_n)] + \mathbb{P}[(|Y_n - X| > \varepsilon)(X_n \neq Y_n)] \leq \mathbb{P}(|X_n - X| > \varepsilon) + P((X_n \neq Y_n)$.

39. $\mathbb{P}(|X_n - Y_n| > \varepsilon) \leq \mathbb{P}(|X_n - X| > \varepsilon/3) + \mathbb{P}(|X-Y| > \varepsilon/3) + \mathbb{P}(|Y_n - Y| > \varepsilon/3)$.

40. Sử dụng tính chất: Từ một dãy hội tụ theo xác suất có thể trích ra được một dãy con hội tụ hầu chắc chắn.

41. Sử dung câu 40).

42. Sử dụng bổ đề Borel-Cantelli chứng minh rằng với xác suất 1, chỉ có hữu hạn X_n khác Y_n.

43. $\mathbb{P}(X_n + Y_n < x) \leq P(X_n < x + \varepsilon) + \mathbb{P}(|Y_n| > \varepsilon)$ $(\forall \varepsilon > 0)$.

$\mathbb{P}(|X_n Y_n| \leq \varepsilon) \geq \mathbb{P}(|X_n Y_n| \leq \varepsilon, |X_n| \leq A) \geq \mathbb{P}(|Y_n| \geq \varepsilon/A),\ |X_n| \geq A) \geq 1 - \mathbb{P}(|Y_n| \geq \varepsilon/A) - \mathbb{P}(|X_n| \geq A)$.

44-46. Sử dụng kết quả bài 40.

54. Giả sử $\varepsilon < 1$. Khi đó

$\mathbb{P}\big(\big|\frac{X_1 + \cdots + X_n}{n}\big| > \varepsilon\big) \geq \mathbb{P}(X_n = 2^n, X_{n-1} = 2^{n-1})$

$\times\, \mathbb{P}\big(\big|\frac{X_1+\dots+X_{n-2}+2^{n-1}+2^n}{n}\big| > \varepsilon\big) \quad = \mathbb{P}(X_n = 2^n, X^{n-1}) = 1/4.$

56. a. Sử dụng định lý giới hạn trung tâm

$$\mathbb{P}\big(a \le S_n \le b\big) = \mathbb{P}\big(\frac{a - mn}{\sigma\sqrt{n}} \le \frac{S_n - mn}{\sigma\sqrt{n}} \le \frac{b - mn}{\sigma\sqrt{n}}\big) \to 0.$$

b. Tương tự câu a). Giới hạn bằng 0 nếu $\mathbb{E}X_1 > 0$, bằng 1 nếu $\mathbb{E}X_1 < 0$, bằng 1/2 nếu $\mathbb{E}X_1 = 0$.

CHƯƠNG 3

??. Bảng tần số và bảng tần suất

x_i	12	13	14	15	16
n_i	3	7	4	10	6

x_i	12	13	14	15	16
f_i	3/30	7/30	4/30	10/30	6/30

5. $\overline{x} = 20$; $\hat{s}^2 = 191,67$; $s^2 = 195,74$.

Lớp	x_i	n_i	$n_i x_i$	$n_i x_i^2$
0-10	5	12	60	300
10-20	15	18	270	4050
20-30	25	8	200	5000
30-40	35	4	140	4900
40-50	45	4	180	8100
50-60	55	2	110	6050
$\sum$		48	960	28400

6. a

x_i	64	65	66
n_i	9	17	9

x_i	64	65	66
f_i	9/35	17/35	9/35

c. $\overline{y}/[x > 10] = 65$

7. $\bar{x} = 20{,}45; \quad s \approx 1{,}08$

8. a. $\bar{x} = 20; \quad s^2 \approx 0{,}373; \quad$ b. $\bar{x} = 484; \quad s^2 = 212{,}5.$

CHƯƠNG 4

??. Vì các đại lượng ngẫu nhiên X_i độc lập với nhau và cùng phân phối với X nên

$$\begin{aligned}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\right) &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i-\mathbb{E}X_i)^2 \\ &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{D}X_i = \sigma^2.\end{aligned}$$

2. Gợi ý: Chứng minh $\overline{X}$ là ước lượng hiệu quả của θ.

   + Chứng minh $\overline{X}$ là ước lượng không chệch của θ: $\mathbb{E}\overline{X} = \mathbb{E}X = \theta$.

   + Phương sai nhỏ nhất: sử dụng bất đẳng thức Cramer-Rao, trong đó $\mathbb{D}\overline{X} = \dfrac{\mathbb{D}X}{n} = \dfrac{\theta^2}{n}$.

3. Chứng minh $\overline{X}$ là ước lượng hợp lý tối đa của λ. Vì

$$f(x,\lambda) = \frac{\lambda^x e^{-\lambda}}{x}\,; \quad x = 0, 1, ...$$

Do đó

$$\ln L(x,\lambda) = \ln\prod_{i=1}^{n}\frac{\lambda^{X_i}e^{-\lambda}}{X_i} = \sum_{i=1}^{n}(X_i\ln\lambda - \lambda - \ln X_i).$$

Suy ra $\dfrac{\partial \ln L(x,\lambda)}{\partial\lambda} = \displaystyle\sum_{i=1}^{n}\left(\frac{X_i}{\lambda}-1\right) = 0$ nên $\hat{\lambda} = \overline{X}$.

Kết hợp với $\dfrac{\partial^2 \ln L(x,\lambda)}{\partial\lambda^2} = -\dfrac{1}{\lambda^2}\sum_{i=1}^{n} X_i < 0$. Ta có điều phải chứng minh.

4. Gợi ý: sử dụng Bài tập 1.

5. a. $\bar{x} = 40{,}5$; $\overline{x^2} = 1641{,}5$; $s^2 = \dfrac{40}{39}(1641{,}5 - 40{,}5^2) \approx 1{,}28 \approx 1{,}13^2$.

   b. $t_{\alpha/2} = 1{,}96$; $\varepsilon = 1{,}96\,\dfrac{1{,}13}{\sqrt{40}} = 0,35$; $(40{,}15\,;\ 40{,}85)$.

6. a. $\bar{x} = 150,2$; $\overline{x^2} = 22562,2$; $s^2 = \dfrac{15}{14}(22562,2 - 150,2^2) \approx 2,314$.

   b. $\varepsilon = t_{(n-1;\,\alpha/2)}\,\dfrac{s}{\sqrt{n}} = 1,761\,\dfrac{1{,}52}{\sqrt{15}} = 0,69$; $(149,53\,;\ 150,87)$.

7. a. $\bar{x} = 49{,}8$; $s^2 \approx 0{,}158 \approx 0{,}4^2$.

   b. $t_{(n-1;\,\alpha/2)} = 2{,}539$; $\varepsilon = 0{,}227$; $(49{,}537\,;\ 50{,}027)$.

8. a. $t_{\alpha/2} = 2{,}58$; $\varepsilon = 2{,}92$; $(94{,}08; 99{,}92)$.

   b. $t_\alpha = 2{,}33$; $\varepsilon = 2{,}64$; $\mu > \overline{x} - \varepsilon = 94{,}36$.

   c. $n = \left[\dfrac{t^2_{\alpha/2}\,\sigma^2}{\varepsilon_0^2}\right] + 1 = 107$. Điều tra thêm 107-50=57.

9. $t_\alpha = 2,33$; $\varepsilon = 1,165$.

   Doanh thu tối thiểu của công ty: $N(\overline{x} - \varepsilon) = 41.667,5$ triệu;

   Doanh thu tối đa của công ty: $N(\overline{x} + \varepsilon) = 42.832,5$ triệu.

11. $t_{\alpha/2} = \dfrac{\varepsilon\sqrt{n}}{s} \approx 2,66$; $1 - \alpha = 2\varphi(t_{\alpha/2}) = 0,992$.

13. $n \geqslant 3458$.

14. $t_{\alpha/2} = 1,96$; $\varepsilon \approx 0,0849$; $(f - \varepsilon; f + \varepsilon) = (0,1651; 0,3349)$.

    $(N(f - \varepsilon); N(f + \varepsilon)) = (1651; 3349)$.

15. Gợi ý: tương tự Bài tập 14, dùng ước lượng tỉ lệ tối thiểu.

16. a. $f = 0,15$; $t_{\alpha/2} = 1,96$; $\varepsilon \approx 0,0495$

Trữ lượng cá: $\big([N/(f+\varepsilon)]; [N/(f-\varepsilon)]\big) = (1819; 2219)$.

b. $\varepsilon_0 = \varepsilon/2 \approx 0,02475$; $n = \left[\dfrac{f(1-f)}{\varepsilon_0^2}\, t_{\alpha/2}^2\right] + 1 = 800$.

CHƯƠNG 5

1. $\mu_0 = 300$; $\sigma = 3$; $\alpha = 0,05$; $t_{\alpha/2} = 1{,}96$; $n = 40$; $\overline{x} = 301{,}2$.

$$H:\ \mu = \mu_0\ ;\quad K:\ \mu \neq \mu_0.$$

$$t_{tn} = \frac{301{,}2 - 300}{3}\sqrt{40} \approx 2{,}53,$$

Vì $|t_{tn}| > t_{\alpha/2}$ nên ta bác bỏ H, chấp nhận K. Nghĩa là chấp nhận khẳng định rằng lô hàng trên không được chấp nhận.

2. $\mu_0 = 19{,}4$; $s = 3{,}24$; $\alpha = 0,08$; $t_{\alpha/2} = 1{,}75$; $n = 49$; $\overline{x} = 18{,}9$.

$$H:\ \mu = \mu_0\ ;\quad K:\ \mu \neq \mu_0.$$

$$t_{tn} = \frac{18{,}9 - 19{,}4}{3{,}24}\sqrt{49} \approx -1{,}08.$$

Vì $|t_{tn}| \leqslant t_{\alpha/2}$ nên ta chấp nhận H.

3. $H: \ \mu = \mu_0 = 50$; $\quad K: \ \mu < \mu_0$.

   $t_{(n-1,\alpha)} = 2{,}473$; $t_{tn} = \dfrac{49{,}8 - 50}{0{,}6}\sqrt{28} \approx -1{,}76$.

   Vì $t_{tn} \geqslant -t_{(n-1,\alpha)}$ nên ta chấp nhận H. Do đó nghi ngờ trên là chưa có cơ sở.

4. Gợi ý: $H: \ \mu = \mu_0 = 50$; $\quad K: \ \mu > \mu_0$.

5. $H: \ p = p_0 = 0{,}3$; $\quad K: \ p \neq p_0$; $t_{\alpha/2} = 1{,}75$; $t_{tn} = -3{,}7$.

   Vì $|t_{tn}| > t_{\alpha/2}$ nên ta bác bỏ H, chấp nhận K. Do đó tuyên bố trên không đúng với thực tế.

6. Gợi ý: Kiểm định một phía: $H: \ p = p_0$; $\quad K: \ p < p_0$.

7. Gợi ý: Kiểm định một phía: $H: \ p = p_0$; $\quad K: \ p > p_0$.

9. $H: \ \mu_1 = \mu_2$; $\ K: \ \mu_1 > \mu_2$.

   $$s^2 = \frac{19 \times 12^2 + 19 \times 10^2}{38} \approx 122; \ t_{tn} = \frac{140 - 120}{\sqrt{122\left(\frac{1}{20} + \frac{1}{20}\right)}} \approx 5,73.$$

   Vì $t_{tn} > t_{(28,0,05)} = 1,701$ nên ta bác bỏ H, chấp nhận K. Do đó việc khuyến mãi đã mang lại hiệu quả.

10. $H: \ \mu_1 = \mu_2$; $\ K: \ \mu_1 < \mu_2$.

    $$t_\alpha = 2{,}32; \ t_{tn} = (560 - 580)\Bigg/\sqrt{\frac{20^2}{35} + \frac{30^2}{35}} \approx -3,28.$$

    Việc thay đổi kiểu đóng gói đã mang lại hiệu quả.

11. $H: \ p_1 = p_2$; $\ K: \ p_1 \neq p_2$.

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{(0,75 - 0,7)}{\sqrt{0,72 \times 0,28\big(1/200 + 1/300\big)}} \approx 1,22.$$

$|t_{tn}| < t_{\alpha/2} = 1{,}96$: chấp nhận H.

12. $H: \ p_1 = p_2; \quad K: \ p_1 > p_2$.

Giá trị kiểm định thực nghiệm

$$t_{tn} = \frac{0,846 - 0,810}{\sqrt{0,834 \times 0,166\left(\frac{1}{2000} + \frac{1}{1000}\right)}} \approx 2,5.$$

$t_{tn} > t_\alpha = 1{,}64$: bác bỏ H, chấp nhận K.

13. Kết quả: $K: \ p_1 > p_2$; $t_{tn} \approx 0,212$; $t_\alpha = 1{,}4$; chấp nhận H.

15. H : chất lượng sản phẩm không phụ thuộc vào ca làm việc; K : chất lượng sản phẩm phụ thuộc vào ca làm việc.

$t_{tn} \approx 7{,}38 > \chi^2(2; 0{,}05) = 5{,}991$: bác bỏ H, chấp nhận K.

16. $t_{tn} \approx 18{,}197 > \chi^2(6; 0{,}01) = 16{,}812$: bác bỏ H, chấp nhận K.

CHƯƠNG 6

1. $n = 6$, $\sum x_i = 420$; $\sum y_i = 50$; $\sum x_i^2 = 30650$; $\sum x_i y_i = 3655$.

a. Hệ số tương quan mẫu

$$r = \frac{6 \cdot 3655 - 420 \cdot 50}{\sqrt{6 \cdot 30650 - 420^2}\ \sqrt{6 \cdot 440 - 50^2}} \approx 0{,}908.$$

b. Hệ số hồi quy tuyến tính thực nghiệm

$$a = \frac{6 \times 3655 - 420 \times 50}{6 \times 30650 - 420^2} = \frac{930}{7500} = 0,124\,;$$

$$b = \frac{50 - 0,124 \times 420}{6} \approx -0,3467.$$

PT HQTT: $y = 0,124x - 0,3467.$

c. $y = 8,6 \quad \Rightarrow \quad x = 72,15$ (tỉ).

2. a. Hệ số tương quan mẫu: $r \approx 0{,}982$.

   b. PT HQTT: $y = 150x - 583,33$; $x = 10,5 \quad \Rightarrow \quad y = 991,7$ (triệu).

3. Bảng số liệu chuyển đổi

x	0	1	2	3	4	5	6	7
y	0	1	1	3	4	6	9	10

Vậy $a = \dfrac{8 \cdot 182 - 28 \cdot 34}{8 \cdot 140 - 28^2} = 1,5$ và $b = 34/8 - 1,5 \cdot 28/8 = -1.$

Do đó $y = 1,5x - 1.$

4. Gợi ý: Sử dụng các công thức

$$r = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\hat{s}_x\,\hat{s}_y}\,;\;\; a = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\hat{s}_x^2}\,;\;\; b = \overline{y} - a\,\overline{x}\,.$$

5. Tương tự Bài tập 4.

TÀI LIỆU THAM KHẢO

1. Nguyễn Quang Báu, *Lý thuyết xác suất và thống kê toán học*, Đại học quốc gia Hà Nội, 2004.

2. Lê Sĩ Đồng, *Xác suất thống kê và ứng dụng*, NXB Giáo dục, 2004.

3. Đặng Hấn, *Xác suất thống kê*, NXB Thống kê, 1996.

4. Đào Hữu Hồ, *Xác suất thống kê*, NXB Đại học quốc gia Hà Nội, 2006.

5. Đào Hữu Hồ, Nguyễn Văn Hữu, Hoàng Hữu Như, *Thống kê toán học*, NXB Đại học quốc gia Hà Nội, 2004.

6. Nguyễn Văn Quảng, *Giáo trình xác suất*, NXB Đại học quốc gia Hà Nội, 2007.

7. Đặng Hùng Thắng, *Mở đầu lý thuyết xác suất*, NXB Giáo dục, 2000.

8. Nguyễn Duy Tiến - Vũ Việt Yên, *Lý thuyết xác suất*, NXB Giáo dục, 2000.

9. Y.S. Chow and H. Teicher; *Probabylity Theory: Independence, Interchangeability, martingales,* Springer-Verlag, Berlin and New York, 1988.

CÁC BẢNG SỐ THÔNG DỤNG

Bảng 1: Giá trị của hàm: $P_\lambda(k) = e^{-\lambda}\,\dfrac{\lambda^k}{k!}$

	λ									
k	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,0905	0,1637	0,2222	0,2681	0,3033	0,3293	0,3476	0,3595	0,3659	0,3679
2	0,0045	0,0164	0,0333	0,0536	0,0758	0,0988	0,1217	0,1438	0,1647	0,1839
3	0,0002	0,0011	0,0033	0,0072	0,0126	0,0198	0,0284	0,0383	0,0494	0,0613
4	0,0000	0,0001	0,0003	0,0007	0,0016	0,0030	0,0050	0,0077	0,0111	0,0153
5	0,0000	0,0000	0,0000	0,0001	0,0002	0,0004	0,0007	0,0012	0,0020	0,0031
6	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0002	0,0003	0,0005
7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001

k	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,3662	0,3614	0,3543	0,3452	0,3347	0,3230	0,3106	0,2975	0,2842	0,2707
2	0,2014	0,2169	0,2303	0,2417	0,2510	0,2584	0,2640	0,2678	0,2700	0,2707
3	0,0738	0,0867	0,0998	0,1128	0,1255	0,1378	0,1496	0,1607	0,1710	0,1804
4	0,0203	0,0260	0,0324	0,0395	0,0471	0,0551	0,0636	0,0723	0,0812	0,0902
5	0,0045	0,0062	0,0084	0,0111	0,0141	0,0176	0,0216	0,0260	0,0309	0,0361
6	0,0008	0,0012	0,0018	0,0026	0,0035	0,0047	0,0061	0,0078	0,0098	0,0120
7	0,0001	0,0002	0,0003	0,0005	0,0008	0,0011	0,0015	0,0020	0,0027	0,0034
8	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002	0,0003	0,0005	0,0006	0,0009
9	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0002

k	3	4	5	6	7	8	9	10	11	12
0	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0000	0,0000	0,0000
1	0,1494	0,0733	0,0337	0,0149	0,0064	0,0027	0,0011	0,0005	0,0002	0,0001
2	0,2240	0,1465	0,0842	0,0446	0,0223	0,0107	0,0050	0,0023	0,0010	0,0004
3	0,2240	0,1954	0,1404	0,0892	0,0521	0,0286	0,0150	0,0076	0,0037	0,0018
4	0,1680	0,1954	0,1755	0,1339	0,0912	0,0573	0,0337	0,0189	0,0102	0,0053
5	0,1008	0,1563	0,1755	0,1606	0,1277	0,0916	0,0607	0,0378	0,0224	0,0127
6	0,0504	0,1042	0,1462	0,1606	0,1490	0,1221	0,0911	0,0631	0,0411	0,0255
7	0,0216	0,0595	0,1044	0,1377	0,1490	0,1396	0,1171	0,0901	0,0646	0,0437

Bảng 2: Hàm phân phối Poisson:

$$F_\lambda(x) = \sum_{k=0}^{x} \frac{e^{-\lambda}\lambda^k}{k!}$$

λ										
x	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
0	0,9048	0,8187	0,7408	0,6703	0,6065	0,5488	0,4966	0,4493	0,4066	0,3679
1	0,9953	0,9825	0,9631	0,9384	0,9098	0,8781	0,8442	0,8088	0,7725	0,7358
2	0,9998	0,9989	0,9964	0,9921	0,9856	0,9769	0,9659	0,9526	0,9371	0,9197
3	1,0000	0,9999	0,9997	0,9992	0,9982	0,9966	0,9942	0,9909	0,9865	0,9810
4	1,0000	1,0000	1,0000	0,9999	0,9998	0,9996	0,9992	0,9986	0,9977	0,9963
5	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9998	0,9997	0,9994
6	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999
x	1,1	1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0
0	0,3329	0,3012	0,2725	0,2466	0,2231	0,2019	0,1827	0,1653	0,1496	0,1353
1	0,6990	0,6626	0,6268	0,5918	0,5578	0,5249	0,4932	0,4628	0,4337	0,4060
2	0,9004	0,8795	0,8571	0,8335	0,8088	0,7834	0,7572	0,7306	0,7037	0,6767
3	0,9743	0,9662	0,9569	0,9463	0,9344	0,9212	0,9068	0,8913	0,8747	0,8571
4	0,9946	0,9923	0,9893	0,9857	0,9814	0,9763	0,9704	0,9636	0,9559	0,9473
5	0,9990	0,9985	0,9978	0,9968	0,9955	0,9940	0,9920	0,9896	0,9868	0,9834
6	0,9999	0,9997	0,9996	0,9994	0,9991	0,9987	0,9981	0,9974	0,9966	0,9955
7	1,0000	1,0000	0,9999	0,9999	0,9998	0,9997	0,9996	0,9994	0,9992	0,9989
8	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9999	0,9999	0,9998	0,9998
x	3	4	5	6	7	8	9	10	11	12
0	0,0498	0,0183	0,0067	0,0025	0,0009	0,0003	0,0001	0,0000	0,0000	0,0000
1	0,1991	0,0916	0,0404	0,0174	0,0073	0,0030	0,0012	0,0005	0,0002	0,0001
2	0,4232	0,2381	0,1247	0,0620	0,0296	0,0138	0,0062	0,0028	0,0012	0,0005
3	0,6472	0,4335	0,2650	0,1512	0,0818	0,0424	0,0212	0,0103	0,0049	0,0023
4	0,8153	0,6288	0,4405	0,2851	0,1730	0,0996	0,0550	0,0293	0,0151	0,0076
5	0,9161	0,7851	0,6160	0,4457	0,3007	0,1912	0,1157	0,0671	0,0375	0,0203
6	0,9665	0,8893	0,7622	0,6063	0,4497	0,3134	0,2068	0,1301	0,0786	0,0458
7	0,9881	0,9489	0,8666	0,7440	0,5987	0,4530	0,3239	0,2202	0,1432	0,0895
8	0,9962	0,9786	0,9319	0,8472	0,7291	0,5925	0,4557	0,3328	0,2320	0,1550
9	0,9989	0,9919	0,9682	0,9161	0,8305	0,7166	0,5874	0,4579	0,3405	0,2424

Bảng 3: Giá trị hàm Gauss: $\varphi(x) = \dfrac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	0,3989	0,3989	0,3988	0,3986	0,3984	0,3982	0,3980	0,3977	0,3973
0,1	0,3970	0,3965	0,3961	0,3956	0,3951	0,3945	0,3939	0,3932	0,3925	0,3918
0,2	0,3910	0,3902	0,3894	0,3885	0,3876	0,3867	0,3857	0,3847	0,3836	0,3825
0,3	0,3814	0,3802	0,3790	0,3778	0,3765	0,3752	0,3739	0,3725	0,3712	0,3697
0,4	0,3683	0,3668	0,3653	0,3637	0,3621	0,3605	0,3589	0,3572	0,3555	0,3538
0,5	0,3521	0,3503	0,3485	0,3467	0,3448	0,3429	0,3410	0,3391	0,3372	0,3352
0,6	0,3332	0,3312	0,3292	0,3271	0,3251	0,3230	0,3209	0,3187	0,3166	0,3144
0,7	0,3123	0,3101	0,3079	0,3056	0,3034	0,3011	0,2989	0,2966	0,2943	0,2920
0,8	0,2897	0,2874	0,2850	0,2827	0,2803	0,2780	0,2756	0,2732	0,2709	0,2685
0,9	0,2661	0,2637	0,2613	0,2589	0,2565	0,2541	0,2516	0,2492	0,2468	0,2444
1,0	0,2420	0,2396	0,2371	0,2347	0,2323	0,2299	0,2275	0,2251	0,2227	0,2203
1,1	0,2179	0,2155	0,2131	0,2107	0,2083	0,2059	0,2036	0,2012	0,1989	0,1965
1,2	0,1942	0,1919	0,1895	0,1872	0,1849	0,1826	0,1804	0,1781	0,1758	0,1736
1,3	0,1714	0,1691	0,1669	0,1647	0,1626	0,1604	0,1582	0,1561	0,1539	0,1518
1,4	0,1497	0,1476	0,1456	0,1435	0,1415	0,1394	0,1374	0,1354	0,1334	0,1315
1,5	0,1295	0,1276	0,1257	0,1238	0,1219	0,1200	0,1182	0,1163	0,1145	0,1127
1,6	0,1109	0,1092	0,1074	0,1057	0,1040	0,1023	0,1006	0,0989	0,0973	0,0957
1,7	0,0940	0,0925	0,0909	0,0893	0,0878	0,0863	0,0848	0,0833	0,0818	0,0804
1,8	0,0790	0,0775	0,0761	0,0748	0,0734	0,0721	0,0707	0,0694	0,0681	0,0669
1,9	0,0656	0,0644	0,0632	0,0620	0,0608	0,0596	0,0584	0,0573	0,0562	0,0551
2,0	0,0540	0,0529	0,0519	0,0508	0,0498	0,0488	0,0478	0,0468	0,0459	0,0449
2,1	0,0440	0,0431	0,0422	0,0413	0,0404	0,0396	0,0387	0,0379	0,0371	0,0363
2,2	0,0355	0,0347	0,0339	0,0332	0,0325	0,0317	0,0310	0,0303	0,0297	0,0290
2,3	0,0283	0,0277	0,0270	0,0264	0,0258	0,0252	0,0246	0,0241	0,0235	0,0229
2,4	0,0224	0,0219	0,0213	0,0208	0,0203	0,0198	0,0194	0,0189	0,0184	0,0180
2,5	0,0175	0,0171	0,0167	0,0163	0,0158	0,0154	0,0151	0,0147	0,0143	0,0139
2,6	0,0136	0,0132	0,0129	0,0126	0,0122	0,0119	0,0116	0,0113	0,0110	0,0107
2,7	0,0104	0,0101	0,0099	0,0096	0,0093	0,0091	0,0088	0,0086	0,0084	0,0081
2,8	0,0079	0,0077	0,0075	0,0073	0,0071	0,0069	0,0067	0,0065	0,0063	0,0061

Bảng 4: Giá trị hàm phân phối chuẩn $N(0,1)$:

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

x	0	1	2	3	4	5	6	7	8	9
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981