

TRƯỜNG ĐẠI HỌC HÀNG HẢI VIỆT NAM
KHOA CƠ SỞ - CƠ BẢN
BỘ MÔN TOÁN
—ooOoo—

TÀI LIỆU HỌC TẬP
XÁC SUẤT THỐNG KÊ

TÊN HỌC PHẦN : XÁC SUẤT THỐNG KÊ
MÃ HỌC PHẦN : 18143
TRÌNH ĐỘ ĐÀO TẠO : ĐẠI HỌC CHÍNH QUY

Mục lục

Mục lục	3
1 Biến cố ngẫu nhiên và các công thức xác suất	1
1.1 Phép thử và biến cố	1
1.1.1 Định nghĩa	1
1.1.2 Phân loại biến cố	2
1.2 Định nghĩa xác suất	2
1.2.1 Định nghĩa cổ điển về xác suất	2
1.2.2 Định nghĩa hình học về xác suất	3
1.2.3 Định nghĩa thống kê về xác suất	4
1.3 Quan hệ giữa các biến cố	5
1.3.1 Tổng, tích và biến cố đối	5
1.3.2 Biến cố xung khắc	6
1.3.3 Nhóm biến cố đầy đủ	7
1.3.4 Biến cố độc lập	7
1.4 Công thức cộng và nhân xác suất	8
1.4.1 Công thức cộng	8
1.4.2 Công thức nhân	9
1.4.3 Xác suất có điều kiện và công thức nhân tổng quát	10
1.4.4 Công thức cộng tổng quát	11
1.4.5 Định lí liên hệ cộng và nhân xác suất	11
1.5 Công thức Bernoulli	12
1.5.1 Lược đồ Bernoulli	12
1.5.2 Công thức Bernoulli	13
1.5.3 Công thức Bernoulli mở rộng	14
1.6 Công thức đầy đủ và công thức Bayes	15
Bài tập chương 1	17
2 Đại lượng ngẫu nhiên và các quy luật phân phối xác suất	23
2.1 Đại lượng ngẫu nhiên	23

2.1.1	Các định nghĩa	23
2.1.2	Phân loại đại lượng ngẫu nhiên	23
2.2	Quy luật phân phối xác suất của đại lượng ngẫu nhiên	24
2.2.1	Bảng phân phối xác suất	24
2.2.2	Hàm phân phối xác suất	26
2.2.3	Hàm mật độ xác suất	29
2.3	Các tham số đặc trưng của đại lượng ngẫu nhiên	33
2.3.1	Kỳ vọng toán	33
2.3.2	Phương sai	35
2.3.3	Độ lệch tiêu chuẩn	37
2.3.4	Mốt	37
2.3.5	Phân vị	38
2.4	Một số quy luật phân phối xác suất thông dụng	39
2.4.1	Quy luật phân phối chuẩn hóa	39
2.4.2	Quy luật phân phối chuẩn	40
2.4.3	Quy luật không - một	43
2.4.4	Quy luật nhị thức	43
2.4.5	Quy luật Poisson	46
2.4.6	Quy luật siêu bội	48
2.4.7	Quy luật khi - bình phương	49
2.4.8	Quy luật Student	50
	Bài tập chương 2	52
3	Mẫu ngẫu nhiên và ước lượng tham số của đại lượng ngẫu nhiên	61
3.1	Tổng thể nghiên cứu	61
3.1.1	Định nghĩa	61
3.1.2	Các phương pháp mô tả tổng thể	62
3.1.3	Các tham số đặc trưng của tổng thể	63
3.2	Mẫu ngẫu nhiên	65
3.2.1	Định nghĩa	65
3.2.2	Các phương pháp mô tả mẫu ngẫu nhiên	65
3.2.3	Đồ thị của phân phối thực nghiệm	67
3.3	Thống kê	68
3.3.1	Định nghĩa	68
3.3.2	Trung bình mẫu	69
3.3.3	Phương sai mẫu	69
3.3.4	Độ lệch tiêu chuẩn mẫu	69
3.3.5	Tần suất mẫu	70

3.3.6	Tính giá trị cụ thể cho các thống kê	70
3.4	Mẫu ngẫu nhiên hai chiều	74
3.4.1	Khái niệm	74
3.4.2	Phương pháp mô tả mẫu ngẫu nhiên hai chiều	74
3.4.3	Một số thống kê đặc trưng của mẫu ngẫu nhiên hai chiều	74
3.5	Ước lượng tham số đại lượng ngẫu nhiên	77
3.5.1	Các khái niệm	77
3.5.2	Phân loại ước lượng	77
3.6	Bài toán ước lượng điểm	77
3.6.1	Ước lượng không chệch	78
3.6.2	Ước lượng hiệu quả	78
3.6.3	Ước lượng vững	79
3.6.4	Bài toán hợp lí cực đại	80
3.7	Bài toán ước lượng khoảng	81
3.7.1	Bài toán tổng quát	81
3.7.2	Khoảng tin cậy cho giá trị trung bình	82
3.7.3	Khoảng tin cậy cho tỉ lệ	88
	Bài tập chương 3	92
4	Kiểm định giả thuyết thống kê	101
4.1	Kiểm định giả thuyết	101
4.1.1	Giả thuyết thống kê	101
4.1.2	Cặp giả thuyết-đối thuyết	101
4.1.3	Quy tắc kiểm định giả thuyết thống kê	102
4.1.4	Các loại sai lầm	103
4.2	Kiểm định giá trị trung bình	103
4.2.1	Đặt vấn đề	103
4.2.2	Trường hợp đã biết phương sai $\sigma^2 = \sigma_0^2$	103
4.2.3	Trường hợp chưa biết phương sai σ^2 và kích thước mẫu $n < 30$	105
4.2.4	Trường hợp chưa biết phương sai σ^2 và kích thước mẫu $n \geq 30$	106
4.3	Kiểm định tỉ lệ	107
	Bài tập chương 4	109
	Tài liệu tham khảo	111
	Phụ lục	113
A	Đáp số bài tập chương 1	113
B	Đáp số bài tập chương 2	114
C	Đáp số bài tập chương 3	116

D	Đáp số bài tập chương 4	117
E	Bảng tra thống kê	118
E.1	Hàm mật độ của phân phối chuẩn hóa	118
E.2	Hàm tích phân Laplace	119
E.3	Phân vị chuẩn	120
E.4	Phân vị Student	121

Chương 1

Biến cố ngẫu nhiên và các công thức xác suất

1.1	Phép thử và biến cố	1
1.2	Định nghĩa xác suất	2
1.3	Quan hệ giữa các biến cố	5
1.4	Công thức cộng và nhân xác suất	8
1.5	Công thức Bernoulli	12
1.6	Công thức đầy đủ và công thức Bayes	15
	Bài tập chương 1	17

1.1 Phép thử và biến cố

1.1.1 Định nghĩa

Định nghĩa 1.1. ¹

- “Việc thực hiện một nhóm các điều kiện cơ bản để quan sát một hiện tượng nào đó có xảy ra hay không được gọi là thực hiện một phép thử.”
- “*Phép thử ngẫu nhiên*² là phép thử mà ta không đoán trước được kết quả của nó, mặc dù đã biết tập hợp Ω tất cả các kết quả có thể có của phép thử đó. Tập hợp Ω được gọi là không gian mẫu của phép thử.”
- “Hiện tượng có thể xảy ra trong kết quả của phép thử được gọi là *biến cố*. Biến cố là một tập con của không gian mẫu Ω .”

Ví dụ 1.1. Tung một con súc sắc xuống đất là một phép thử, còn việc lật lên một mặt nào đó là biến cố.

Ví dụ 1.2. Bắn ngẫu nhiên một phát súng vào bia. Việc bắn súng là phép thử, còn việc trúng vào một miền nào đó của bia là biến cố.

¹Các định nghĩa trong cuốn sách này được trích dẫn theo các giáo trình cơ bản về xác suất và thống kê trước đây, chẳng hạn ở [12, 9].

²Trong cuốn sách này phép thử được dùng để nói chung cho các phép thử ngẫu nhiên.

1.1.2 Phân loại biến cố

Một biến cố chỉ có thể xảy ra khi một phép thử gắn liền với nó được thực hiện. Trong thực tế có thể xảy ra các loại biến cố sau đây:

- Biến cố chắc chắn: là biến cố nhất định sẽ xảy ra khi thực hiện một phép thử. Ω là biến cố chắc chắn.
- Biến cố không thể có: là biến cố nhất định không xảy ra khi thực hiện một phép thử. Kí hiệu $\emptyset$ là biến cố không thể có.
- Biến cố ngẫu nhiên: là biến cố có thể xảy ra hoặc không xảy ra khi thực hiện một phép thử. Các biến cố ngẫu nhiên thường được kí hiệu là $A, B, C, \dots$ hoặc $A_1, A_2, \dots, A_n, B_1, B_2, \dots, B_n, \dots$.

Ví dụ 1.3. Tung một con súc sắc, xét các biến cố sau đây:

Ω = “Xuất hiện mặt có số chấm nhỏ hơn 7”. Ω là biến cố chắc chắn.

$\emptyset$ = “Xuất hiện mặt có 8 chấm”. $\emptyset$ là biến cố không thể có.

A = “Xuất hiện mặt có số chấm chẵn”. A là biến cố ngẫu nhiên.

A_i = “Xuất hiện mặt i chấm”, ($i = 1, 2, \dots, 6$). A_i là các biến cố ngẫu nhiên.

1.2 Định nghĩa xác suất

Quan sát các sự kiện ngẫu nhiên ta thấy khả năng xuất hiện của chúng nói chung là không đồng đều, một số sự kiện thường hay xảy ra, một số khác thường ít xảy ra. Từ đó nảy sinh vấn đề tìm cách đo lường “khả năng xảy ra của một sự kiện”. Muốn vậy người ta gán cho mỗi biến cố A một số không âm, gọi là xác suất của biến cố A .

Định nghĩa 1.2. “Xác suất của một biến cố là một con số đặc trưng cho khả năng khách quan xuất hiện biến cố đó khi thực hiện phép thử.

Kí hiệu xác suất của biến cố A là $\mathbb{P}(A)$.”

Để phù hợp với thước đo “độ chắc” của biến cố thì xác suất $\mathbb{P}(A)$ phải được xây dựng sao cho:

- $0 \leq \mathbb{P}(A) \leq 1$;
- Xác suất của biến cố chắc chắn bằng một: $\mathbb{P}(\Omega) = 1$;
- Xác suất của biến cố không thể có bằng không: $\mathbb{P}(\emptyset) = 0$.

Ta chú ý rằng, tính được xác suất của một biến cố chỉ giúp đánh giá việc biến cố đó là dễ hay khó xảy ra, chứ không giúp đoán trước được việc biến cố có xảy ra hay không.

1.2.1 Định nghĩa cổ điển về xác suất

Định nghĩa 1.3. “Xác suất xuất hiện biến cố A trong một phép thử là tỉ số giữa số kết cục thuận lợi cho A và tổng số các kết cục duy nhất đồng khả năng có thể xảy ra khi thực hiện phép thử đó.”

Nếu kí hiệu: m là số kết cục thuận lợi cho biến cố A ;
 n là số kết cục duy nhất đồng khả năng của phép thử,
 thì công thức tính xác suất của biến cố A như sau:

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

Ví dụ 1.4. Giả sử thực hiện phép thử tung một con súc sắc cân đối và đồng chất. Tính xác suất của biến cố $A = \text{“Xuất hiện mặt có số chấm chẵn”}$.

Lời giải: Khi tung một con súc sắc cân đối và đồng chất ta thấy có thể có 6 kết cục xảy ra là: xuất hiện các mặt 1 chấm, 2 chấm, ..., 6 chấm. Những kết cục này thoả mãn hai điều kiện: chúng duy nhất, tức là trong kết quả của phép thử xảy ra một và chỉ một kết cục trong số đó; hơn nữa chúng có khả năng xảy ra như nhau. Vậy số kết cục duy nhất đồng khả năng là $n = 6$.

Trong số 6 kết cục duy nhất đồng khả năng đó ta thấy chỉ có 3 kết cục mà nếu kết cục đó xảy ra thì biến cố A sẽ xảy ra, đó là những kết cục được mặt 2 chấm, 4 chấm, 6 chấm. Do đó, số kết cục thuận lợi cho biến cố A là $m = 3$.

Như vậy, xác suất của biến cố A bằng:

$$\mathbb{P}(A) = \frac{m}{n} = \frac{3}{6} = \frac{1}{2}.$$

Ví dụ 1.5. Một hộp có 10 sản phẩm, trong đó có 6 chính phẩm và 4 phế phẩm. Lấy ngẫu nhiên từ hộp đó 3 sản phẩm. Tìm xác suất để:

- Cả 3 sản phẩm lấy ra đều là chính phẩm;
- Trong 3 sản phẩm lấy ra có đúng 2 chính phẩm.

Lời giải: a) Gọi A là biến cố “Lấy được 3 chính phẩm”.

Số kết cục đồng khả năng trong phép thử bằng số cách chọn 3 sản phẩm (phân biệt và không kể thứ tự) từ 10 sản phẩm, như vậy $n = C_{10}^3 = 120$.

Số kết cục thuận lợi cho A xảy ra là số cách chọn được 3 sản phẩm từ 6 chính phẩm, vậy $m = C_6^3 = 20$.

Do đó, xác suất của biến cố A là: $\mathbb{P}(A) = \frac{m}{n} = \frac{20}{120} = \frac{1}{6}$.

b) Gọi B là biến cố “Trong 3 sản phẩm lấy ra có đúng 2 chính phẩm”.

Để biến cố B xảy ra ta phải thực hiện chọn theo 2 bước:

- Chọn 2 chính phẩm trong số 6 chính phẩm, số cách chọn là C_6^2 ;
- Chọn 1 phế phẩm trong số 4 phế phẩm, số cách chọn là C_4^1 .

Số kết cục thuận lợi cho biến cố B là số cách chọn cho biến cố B xảy ra:

$$m = C_6^2 \cdot C_4^1.$$

Vậy xác suất của biến cố B là:

$$\mathbb{P}(B) = \frac{m}{n} = \frac{C_6^2 \cdot C_4^1}{C_{10}^3} = \frac{1}{2}.$$

1.2.2 Định nghĩa hình học về xác suất

Định nghĩa 1.4. “Xét một phép thử có vô hạn kết cục đồng khả năng. Giả sử ta có thể biểu thị tập hợp mọi kết cục này bởi một miền hình học Ω nào đó và những kết cục thuận lợi cho

biến cố A được biểu thị bởi miền con $G \subset \Omega$. Khi đó, xác suất của biến cố A được tính bởi công thức:

$$\mathbb{P}(A) = \frac{\text{độ đo } (G)}{\text{độ đo } (\Omega)}.$$

Tùy theo Ω là tập con của đường thẳng, tập con của đường cong, miền phẳng, mảnh mặt cong hay khối không gian mà độ đo được hiểu là độ dài, diện tích hay thể tích.”

Ví dụ 1.6. Trong một trò chơi mô phỏng, khi người chơi ấn nút thì súng lazer sẽ bắn ngẫu nhiên vào một điểm trên bia là hình vuông MNPQ cạnh 40 cm. Tính xác suất của các biến cố:

- a) A = “Bắn trúng một điểm cách tâm O của hình vuông không quá 10 cm”;
b) B = “Bắn trúng một điểm trên đường chéo MP”.

Lời giải:

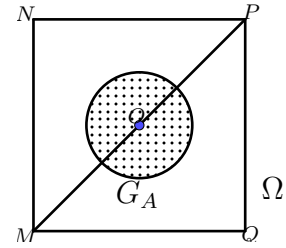
Khi thực hiện phép thử là bắn ngẫu nhiên vào hình vuông MNPQ thì mỗi kết cục đồng khả năng là một điểm của hình vuông đó. Vậy phép thử có vô hạn kết cục đồng khả năng, tập Ω tất cả các kết cục này chính là toàn bộ hình vuông MNPQ.

- a) Miền kết cục thuận lợi cho biến cố A chính là hình tròn tâm O bán kính 10 cm, kí hiệu miền này là G_A . Khi đó, xác suất của biến cố A theo định nghĩa hình học về xác suất là:

$$\mathbb{P}(A) = \frac{\text{diện tích } (G_A)}{\text{diện tích } (\Omega)} = \frac{\pi 10^2}{40^2} = \frac{\pi}{16} \approx 0,196.$$

- b) Miền kết cục thuận lợi cho biến cố B chính là đoạn thẳng MP , kí hiệu miền này là G_B . Khi đó, xác suất của biến cố B là:

$$\mathbb{P}(B) = \frac{\text{diện tích } (G_B)}{\text{diện tích } (\Omega)} = \frac{0}{40^2} = 0.$$



Hình 1.1 – Ví dụ 1.6

Nhận xét. Ta luôn có:

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

Điều ngược lại không đúng, tức là nếu $\mathbb{P}(A) = 0$ thì chưa thể kết luận A là biến cố không thể có. Tương tự như vậy, nếu $\mathbb{P}(A) = 1$ thì chưa chắc A đã là biến cố chắc chắn.

1.2.3 Định nghĩa thống kê về xác suất

Định nghĩa 1.5. “Tần suất xuất hiện biến cố trong các phép thử là tỉ số giữa số phép thử trong đó biến cố xuất hiện và tổng số phép thử được thực hiện.”

Như vậy, nếu kí hiệu số phép thử là n , số lần xuất hiện biến cố A là k , tần suất xuất hiện biến cố A là $f(A)$ thì:

$$f(A) = \frac{k}{n}.$$

Định nghĩa 1.6. “Xác suất xuất hiện biến cố A trong một phép thử là một số p không đổi mà tần suất f xuất hiện biến cố đó trong n phép thử sẽ hội tụ theo xác suất về p khi số phép thử tăng lên vô hạn.”

$$f \xrightarrow[n \rightarrow \infty]{\text{Hội tụ theo xác suất}} p \quad (\text{kí hiệu là } f \xrightarrow{\mathbb{P}} p).$$

Như vậy về mặt thực tế, với số phép thử n đủ lớn ta có thể lấy $\mathbb{P}(A) \approx f(A)$. Ta chú ý rằng ở đây tần suất f hội tụ theo xác suất về p chứ không phải hội tụ theo nghĩa thông thường của giải tích toán học. Điều đó có nghĩa là với mọi ε dương bé tùy ý ta luôn có:

$$\lim_{n \rightarrow \infty} \mathbb{P}(|f - p| < \varepsilon) = 1.$$

Ví dụ 1.7. Để nghiên cứu khả năng xuất hiện mặt sấp (biến cố A) khi tung một đồng xu, người ta tiến hành tung một đồng xu nhiều lần và thu được kết quả sau đây:

Người làm thí nghiệm	Số lần tung (n)	Số lần được mặt sấp (k)	Tần suất $f(A) = \frac{k}{n}$
Buffon	4040	2048	0,5069
Pearson	12000	6019	0,5016
Pearson	24000	12012	0,5005

Ta thấy khi số phép thử tăng lên thì tần suất xuất hiện mặt sấp sẽ dao động ngày càng ít hơn xung quanh giá trị không đổi là 0,5. Điều đó cho phép hy vọng rằng khi số phép thử tăng lên vô hạn, tần suất sẽ hội tụ về giá trị lí thuyết 0,5.

Ví dụ 1.8. Tỷ lệ phế phẩm trong số sản phẩm được sản xuất bởi một thiết bị tự động là 0,5%. Khi lấy ngẫu nhiên 1 sản phẩm trong số những sản phẩm do thiết bị đó sản xuất thì xác suất lấy được một phế phẩm bằng:

$$0,5\% = 0,005.$$

1.3 Quan hệ giữa các biến cố

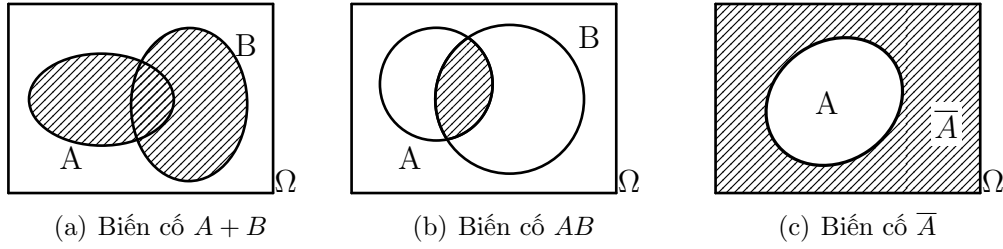
1.3.1 Tổng, tích và biến cố đối

Định nghĩa 1.7. Giả sử A và B là hai biến cố liên quan đến một phép thử. Khi đó,

- “Tổng của hai biến cố A và B là biến cố được kí hiệu $A + B$. Biến cố $A + B$ xảy ra khi và chỉ khi A xảy ra *hoặc* B xảy ra;”
- “Tích của hai biến cố A và B là biến cố được kí hiệu AB . Biến cố AB xảy ra khi và chỉ khi A xảy ra *và* B xảy ra;”
- “Biến cố đối của biến cố A được kí hiệu là $\overline{A}$. Biến cố $\overline{A}$ xảy ra khi và chỉ khi A *không* xảy ra.”

Định nghĩa 1.8. Giả sử $A_1, A_2, \dots, A_n$ là các biến cố liên quan đến một phép thử. Khi đó,

- “Tổng của n biến cố $A_1, A_2, \dots, A_n$ được kí hiệu $A_1 + A_2 + \dots + A_n$. Biến cố tổng này xảy ra khi và chỉ khi ít nhất một trong các biến cố $A_i, (i = \overline{1, n})$, xảy ra;”
- “Tích của n biến cố $A_1, A_2, \dots, A_n$ được kí hiệu là $A_1 A_2 \dots A_n$. Biến cố tích này xảy ra khi và chỉ khi tất cả các biến cố $A_i, (i = \overline{1, n})$, xảy ra.”



Hình 1.2 – Quan hệ tổng, tích và biến cố đối.

Công thức đối ngẫu De Morgan

$$\overline{A_1 + A_2 + \dots + A_n} = \bar{A}_1 \cdot \bar{A}_2 \dots \bar{A}_n$$

$$\overline{\bar{A}_1 \cdot \bar{A}_2 \dots \bar{A}_n} = \bar{A}_1 + \bar{A}_2 + \dots + \bar{A}_n$$

Ví dụ 1.9. Một con tàu được trang bị hai radar độc lập, kí hiệu các biến cố:

A = “Radar thứ nhất bị hỏng”;

B = “Radar thứ hai bị hỏng”.

Biểu diễn các biến cố sau đây theo biến cố A, B :

C = “Chỉ radar thứ nhất hỏng”;

D = “Chỉ có một radar hỏng”;

E = “Có radar hỏng”.

Lời giải: “Chỉ radar thứ nhất hỏng” tức là radar thứ nhất hỏng và radar thứ hai phải không hỏng. Do đó:

$$C = A\bar{B}.$$

Để xảy ra việc “chỉ có một radar hỏng” thì có hai khả năng: hoặc chỉ radar thứ nhất hỏng, hoặc chỉ radar thứ hai hỏng. Do đó:

$$D = A\bar{B} + \bar{A}B.$$

“Có radar hỏng” tức là radar thứ nhất hỏng hoặc radar thứ hai hỏng, vì vậy:

$$E = A + B.$$

1.3.2 Biến cố xung khắc

Định nghĩa 1.9. “Hai biến cố A và B được gọi là *xung khắc* với nhau (gọi tắt là *xung khắc*) nếu chúng không thể đồng thời xảy ra trong một phép thử.”

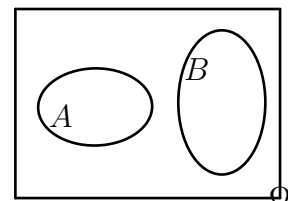
Nhóm n biến cố $A_1, A_2, \dots, A_n$ được gọi là *xung khắc từng đôi* nếu bất kì hai biến cố nào trong nhóm này cũng xung khắc với nhau.

Ví dụ 1.10. Một con tàu được trang bị hai radar độc lập. Các biến cố:

A = “Radar thứ nhất bị hỏng”;

và B = “Radar thứ hai bị hỏng”

không xung khắc vì chúng có thể cùng xảy ra trong một phép thử.



Hình 1.3 – Hai biến cố A, B xung khắc.

Còn các biến cố:

$C_1 = \text{“Chỉ radar thứ nhất hỏng”};$

và $C_2 = \text{“Chỉ radar thứ hai hỏng”}$

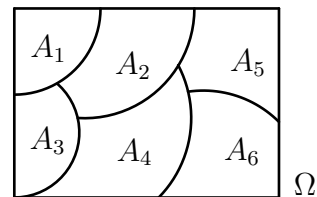
xung khắc vì chúng không thể cùng xảy ra trong một phép thử.

1.3.3 Nhóm biến cố đầy đủ

Định nghĩa 1.10. “Các biến cố $A_1, A_2, \dots, A_n$ được gọi là *nhóm biến cố đầy đủ* (hoặc *nhóm đầy đủ*) nếu trong kết quả của phép thử sẽ xảy ra một và chỉ một trong các biến cố đó.”

Nói cách khác, các biến cố $A_1, A_2, \dots, A_n$ sẽ tạo nên một nhóm đầy đủ nếu chúng xung khắc từng đôi với nhau và tổng của chúng là một biến cố chắc chắn.

Đặc biệt, hai biến cố $\{A, \bar{A}\}$ là một nhóm đầy đủ.



Ví dụ 1.11. Tung một súc sắc. Gọi biến cố $A_i = \text{“xuất hiện mặt } i \text{ chấm”}$. Khi đó $A_1, A_2, \dots, A_6$ tạo thành nhóm đầy đủ.

Hình 1.4 – Các biến cố $A_1, A_2, \dots, A_6$ lập thành một nhóm đầy đủ.

1.3.4 Biến cố độc lập

Định nghĩa 1.11. “Hai biến cố A, B được gọi là *độc lập với nhau* (gọi tắt là *độc lập*) nếu việc xảy ra hay không xảy ra biến cố này đều không ảnh hưởng đến xác suất để biến cố kia xảy ra.”

Nếu việc biến cố này xảy ra hay không xảy ra làm cho xác suất xảy ra của biến cố kia thay đổi thì hai biến cố đó được gọi là *phụ thuộc nhau*.

Định nghĩa 1.12. “Nhóm n biến cố được gọi là *độc lập toàn phần* (thường gọi tắt là *độc lập*) nếu mỗi biến cố độc lập với một tổ hợp bất kỳ các biến cố còn lại.”

Ví dụ 1.12. Trong bình có 3 quả cầu trắng và 2 quả cầu đen. Lấy ngẫu nhiên lần lượt 2 quả cầu theo phương thức có hoàn lại. Gọi các biến cố

$A = \text{“Lần thứ nhất lấy được quả cầu trắng”};$

$B = \text{“Lần thứ hai lấy được quả cầu trắng”}.$

Khi đó,

$$\mathbb{P}(A) = \frac{3}{5}; \quad \mathbb{P}(B) = \frac{3}{5}.$$

Ta thấy xác suất lấy được quả cầu trắng lần thứ hai (biến cố B) không phụ thuộc vào kết quả lấy của lần trước (biến cố A). Tương tự, xác suất lấy được quả cầu trắng lần thứ nhất (biến cố A) không phụ thuộc vào kết quả lấy của lần thứ hai (biến cố B). Vậy hai biến cố A và B độc lập với nhau.

Ví dụ 1.13. Trong bình có 3 quả cầu trắng và 2 quả cầu đen. Lấy ngẫu nhiên lần lượt 2 quả cầu theo phương thức không hoàn lại. Gọi các biến cố

$A = \text{“Lần thứ nhất lấy được quả cầu trắng”};$

$B = \text{“Lần thứ hai lấy được quả cầu trắng”}.$

Khi đó, xác suất lấy được quả cầu trắng lần thứ nhất là

$$\mathbb{P}(A) = \frac{3}{5}.$$

Tuy nhiên, xác suất lấy được quả cầu trắng lần thứ hai sẽ phụ thuộc vào kết quả lấy của lần trước:

- Nếu lần thứ nhất lấy được quả cầu trắng (A xảy ra) thì: $\mathbb{P}(B) = \frac{2}{4}$;
- Nếu lần thứ nhất lấy được quả cầu đen (A không xảy ra) thì: $\mathbb{P}(B) = \frac{3}{4}$;

Vậy A và B phụ thuộc nhau.

Nhận xét. Trong thực tế, việc nhận xét tính độc lập hay phụ thuộc của các biến cố chủ yếu dựa vào trực giác. Ta cũng nhận thấy rằng tính độc lập của các biến cố có tính tương hỗ. Nếu A và B độc lập với nhau thì A và $\overline{B}$, $\overline{A}$ và B , $\overline{A}$ và $\overline{B}$ cũng độc lập với nhau.

1.4 Công thức cộng và nhân xác suất

1.4.1 Công thức cộng

Định lý 1.1. Nếu A và B là hai biến cố xung khắc thì xác suất của tổng hai biến cố đó bằng tổng xác suất của hai biến cố A và B . Cụ thể,

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

Hệ quả 1.1.

- Nếu $A_1, A_2, \dots, A_n$ là các biến cố xung khắc từng đôi thì xác suất của tổng các biến cố đó bằng tổng các xác suất của từng biến cố thành phần. Cụ thể,

$$\mathbb{P}(A_1 + A_2 + \dots + A_n) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_n).$$

- Nếu $A_1, A_2, \dots, A_n$ là nhóm đầy đủ thì

$$\mathbb{P}(A_1) + \mathbb{P}(A_2) + \dots + \mathbb{P}(A_n) = 1.$$

- Với hai biến cố đối lập A và $\overline{A}$, ta luôn có

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

Ví dụ 1.14. Tung một đồng xu 6 lần. Tính xác suất để số lần được mặt sấp nhiều hơn số lần được mặt ngửa.

Lời giải: Khi tung đồng xu 6 lần, có các trường hợp sau đây có thể xảy ra

- A = “Số mặt sấp xuất hiện nhiều hơn số mặt ngửa”;
- B = “Số mặt ngửa xuất hiện nhiều hơn số mặt sấp”;
- C = “Số mặt sấp và số mặt ngửa xuất hiện bằng nhau”;

Các biến cố A, B và C tạo nên một nhóm đầy đủ. Vì vậy,

$$\mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) = 1.$$

Ta đi tìm xác suất của biến cố C . Khi tung đồng xu 6 lần số kết cục đồng khả năng là $n = 2^6$. Để có đúng 3 lần xuất hiện mặt sấp và 3 lần xuất hiện mặt ngửa thì số kết cục thuận lợi là $m = C_6^3$. Do đó,

$$\mathbb{P}(C) = \frac{m}{n} = \frac{C_6^3}{2^6} = \frac{20}{64} = \frac{5}{16}.$$

Vì mặt sấp và ngửa đối xứng nhau nên $\mathbb{P}(A) = \mathbb{P}(B)$. Từ đó, xác suất cần tìm là

$$\mathbb{P}(A) = \frac{1 - \mathbb{P}(C)}{2} = \frac{1 - \frac{5}{16}}{2} = \frac{11}{32}.$$

1.4.2 Công thức nhân

Định lý 1.2. Nếu A và B là hai biến cố độc lập thì xác suất của tích hai biến cố đó bằng tích các xác suất của A và B . Cụ thể,

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

Hệ quả 1.2. Nếu $A_1, A_2, \dots, A_n$ là các biến cố độc lập toàn phần thì xác suất của tích n biến cố đó bằng tích các xác suất của từng biến cố thành phần. Cụ thể,

$$\mathbb{P}(A_1 \cdot A_2 \cdot \dots \cdot A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \dots \cdot \mathbb{P}(A_n).$$

Ví dụ 1.15. Một xí nghiệp có 3 ô tô hoạt động độc lập. Xác suất để trong một ngày các ô tô bị hỏng tương ứng là 0, 1; 0, 2; 0, 15.

- Tính xác suất để cả 3 ô tô đều hỏng;
- Tính xác suất để trong một ngày có đúng 1 ô tô hỏng;
- Tính xác suất để trong một ngày có ít nhất một ô tô hỏng.

Lời giải: Gọi A_i là biến cố “ô tô thứ i bị hỏng trong ngày”, ($i = \overline{1, 3}$).

Theo giả thiết, $\mathbb{P}(A_1) = 0, 1$; $\mathbb{P}(A_2) = 0, 2$; $\mathbb{P}(A_3) = 0, 15$.

Do đó, $\mathbb{P}(\overline{A_1}) = 0, 9$; $\mathbb{P}(\overline{A_2}) = 0, 8$; $\mathbb{P}(\overline{A_3}) = 0, 85$.

a) Gọi A là biến cố “cả 3 ô tô đều hỏng”, ta có

$$A = A_1 \cdot A_2 \cdot A_3.$$

Vì các biến cố A_1, A_2, A_3 độc lập nên theo công thức nhân xác suất, ta được

$$\mathbb{P}(A) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) = 0, 1 \cdot 0, 2 \cdot 0, 15 = 0, 003.$$

b) Gọi B là biến cố “trong một ngày có đúng 1 ô tô hỏng”, ta có

$$B = A_1 \cdot \overline{A_2} \cdot \overline{A_3} + \overline{A_1} \cdot A_2 \cdot \overline{A_3} + \overline{A_1} \cdot \overline{A_2} \cdot A_3.$$

Vì các biến cố $(A_1 \cdot \overline{A_2} \cdot \overline{A_3})$, $(\overline{A_1} \cdot A_2 \cdot \overline{A_3})$ và $(\overline{A_1} \cdot \overline{A_2} \cdot A_3)$ là xung khắc từng đôi và mỗi biến cố đó là tích của các biến cố độc lập toàn phần với nhau. Do đó,

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(A_1 \cdot \overline{A_2} \cdot \overline{A_3} + \overline{A_1} \cdot A_2 \cdot \overline{A_3} + \overline{A_1} \cdot \overline{A_2} \cdot A_3) \\ &= \mathbb{P}(A_1) \cdot \mathbb{P}(\overline{A_2}) \cdot \mathbb{P}(\overline{A_3}) + \mathbb{P}(\overline{A_1}) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(\overline{A_3}) + \mathbb{P}(\overline{A_1}) \cdot \mathbb{P}(\overline{A_2}) \cdot \mathbb{P}(A_3) \\ &= 0, 1 \cdot 0, 8 \cdot 0, 85 + 0, 9 \cdot 0, 2 \cdot 0, 85 + 0, 9 \cdot 0, 8 \cdot 0, 15 = 0, 329. \end{aligned}$$

c) Gọi C là biến cố “trong một ngày có ít nhất một ô tô hỏng”, ta có

$$C = A_1 + A_2 + A_3.$$

Tuy nhiên, các biến cố A_1, A_2, A_3 không xung khắc từng đôi, vì vậy không thể áp dụng định lý cộng xác suất đã học. Ta xét biến cố đối lập của C là biến cố $\overline{C}$ = “tất cả các ô tô đều không hỏng” như sau

$$\overline{C} = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}.$$

Theo công thức nhân xác suất

$$\mathbb{P}(\overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}) = \mathbb{P}(\overline{A_1}) \cdot \mathbb{P}(\overline{A_2}) \cdot \mathbb{P}(\overline{A_3}) = 0,9 \cdot 0,8 \cdot 0,85 = 0,612.$$

Vậy xác suất cần tìm là $\mathbb{P}(C) = 1 - \mathbb{P}(\overline{C}) = 1 - 0,612 = 0,388$.

1.4.3 Xác suất có điều kiện và công thức nhân tổng quát

Định nghĩa 1.13. “Cho hai biến cố A và B . Xác suất của biến cố B được tính với điều kiện biến cố A đã xảy ra được gọi là *xác suất có điều kiện* của B và được kí hiệu là $\mathbb{P}(B|A)$.”

Định lý 1.3. *Xác suất của tích hai biến cố A và B bằng tích xác suất của một trong hai biến cố đó với xác suất điều kiện của biến cố còn lại. Cụ thể,*

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

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

Xác suất của tích n biến cố bằng tích xác suất của n biến cố đó, trong đó xác suất của mỗi biến cố tiếp theo đều được tính với điều kiện tất cả các biến cố trước đó đã xảy ra

$$\mathbb{P}(A_1 \cdot A_2 \cdot \dots \cdot A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2|A_1) \cdot \dots \cdot \mathbb{P}(A_n|A_1 \dots A_{n-1}).$$

Nhận xét. Nếu các biến cố A và B độc lập thì xác suất điều kiện trở thành xác suất thông thường. Khi đó ta nhận lại được các công thức nhân xác suất ở phần 1.4.2.

Ví dụ 1.16. Gieo đồng thời hai con súc sắc cân đối và đồng chất. Tính xác suất để tổng số chấm xuất hiện trên hai con súc sắc không nhỏ hơn 10 biết rằng trên ít nhất một con đã xuất hiện mặt 5 chấm.

Lời giải: Kí hiệu các biến cố:

A = “Trên ít nhất một súc sắc xuất hiện mặt 5 chấm”;

B = “Tổng số chấm xuất hiện trên hai con súc sắc không nhỏ hơn 10”.

Ta cần tính $\mathbb{P}(B|A)$.

Ta thấy

$$\mathbb{P}(A) = 1 - \mathbb{P}(\overline{A}) = 1 - \left(\frac{5}{6}\right)^2 = \frac{11}{36}.$$

Để tính $\mathbb{P}(AB)$, ta có không gian mẫu có 36 kết cục đồng khả năng trong đó có 3 kết cục thuận lợi cho AB là $(5, 6)$; $(6, 5)$ và $(5, 5)$. Do đó, $\mathbb{P}(AB) = \frac{3}{36}$.

Vậy xác suất cần tìm là:

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(AB)}{\mathbb{P}(A)} = \frac{3}{11}.$$

1.4.4 Công thức cộng tổng quát

Định lý 1.4. Với A, B là 2 biến cố bất kì, ta luôn có

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

Hệ quả 1.3.

- Nếu A, B là hai biến cố xung khắc thì

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

- Nếu hai biến cố A, B độc lập thì

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

- Xác suất của tổng n biến cố không xung khắc được xác định bằng công thức sau đây

$$\begin{aligned} \mathbb{P}\left(\sum_{i=1}^n A_i\right) &= \sum_i \mathbb{P}(A_i) - \sum_{i < j} \mathbb{P}(A_i A_j) + \sum_{i < j < k} \mathbb{P}(A_i A_j A_k) - \dots \\ &\quad \dots + (-1)^{n-1} \mathbb{P}(A_1 A_2 \dots A_n). \end{aligned}$$

Ví dụ 1.17. Trong một chương trình trò chơi điện tử, một máy bay chiến đấu với hai động cơ được thiết kế là mục tiêu cần tiêu diệt. Trong đó, xác suất để động cơ thứ nhất và thứ hai của máy bay trúng đạn lần lượt là 0,1 và 0,2; xác suất để phi công trúng đạn là 0,05; việc trúng đạn của các động cơ và phi công là độc lập với nhau. Tính xác suất máy bay rơi, biết máy bay sẽ rơi nếu cả hai động cơ trúng đạn hoặc phi công trúng đạn.

Lời giải: Gọi các biến cố:

A_1 = “động cơ thứ nhất của máy bay trúng đạn”;

A_2 = “động cơ thứ hai của máy bay trúng đạn”;

A_3 = “phi công trúng đạn”;

B = “máy bay rơi”.

Ta có

$$B = A_1 A_2 + A_3.$$

Các biến cố A_1, A_2, A_3 độc lập và không xung khắc, áp dụng định lý cộng xác suất cho hai biến cố $A_1 A_2$ và A_3 , sau đó áp dụng công thức nhân xác suất cho các biến cố độc lập, ta được

$$\begin{aligned} \mathbb{P}(B) &= \mathbb{P}(A_1 A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1 A_2 A_3) \\ &= \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \mathbb{P}(A_3) \\ &= 0,1 \cdot 0,2 + 0,05 - 0,1 \cdot 0,2 \cdot 0,05 = 0,069. \end{aligned}$$

1.4.5 Định lý liên hệ cộng và nhân xác suất

Định lý 1.5. Cho $A_1, A_2, \dots, A_n$ là các biến cố độc lập toàn phần. Khi đó,

$$\mathbb{P}\left(\sum_{i=1}^n A_i\right) = 1 - \prod_{i=1}^n \mathbb{P}(\overline{A_i}).$$

Hệ quả 1.4. Nếu $A_1, A_2, \dots, A_n$ là các biến cố độc lập toàn phần có cùng xác suất $\mathbb{P}(A_1) = \mathbb{P}(A_2) = \dots \mathbb{P}(A_n) = p$ thì

$$\mathbb{P}\left(\sum_{i=1}^n A_i\right) = 1 - (1 - p)^n.$$

Ví dụ 1.18. Phải tung một con súc sắc (cân đối và đồng chất) bao nhiêu lần để với xác suất không nhỏ hơn 0,5 có thể hy vọng rằng trong đó có ít nhất một lần được mặt 6 chấm?

Lời giải: Giả sử phải tung súc sắc n lần.

Gọi A_i là biến cố “tung lần thứ i được mặt 6 chấm”, ($i = \overline{1, n}$), gọi B là biến cố “trong n lần tung có ít nhất một lần được mặt 6 chấm”. Ta có

$$B = A_1 + A_2 + \dots + A_n = \sum_{i=1}^n A_i.$$

Các biến cố A_i là độc lập toàn phần và không xung khắc với nhau, và có

$$\mathbb{P}(A_1) = \mathbb{P}(A_2) = \dots = \mathbb{P}(A_n) = \frac{1}{6}.$$

Do đó,

$$\mathbb{P}(B) = 1 - \left(1 - \frac{1}{6}\right)^n = 1 - \left(\frac{5}{6}\right)^n.$$

Theo giả thiết xác suất của B không nhỏ hơn 0,5. Ta có bất phương trình sau

$$\begin{aligned} 1 - \left(\frac{5}{6}\right)^n \geq 0,5 &\Leftrightarrow \left(\frac{5}{6}\right)^n \leq 0,5 \\ &\Leftrightarrow n \cdot \ln \left(\frac{5}{6}\right) \leq \ln 0,5 \\ &\Leftrightarrow n \geq \frac{\ln 0,5}{\ln \left(\frac{5}{6}\right)} \approx 3,802. \end{aligned}$$

Như vậy, $n \geq 4$, tức là phải tung ít nhất 4 lần.

1.5 Công thức Bernoulli

1.5.1 Lược đồ Bernoulli

Bài toán được gọi là tuân theo *lược đồ Bernoulli* (n, A, p) nếu thỏa mãn các giả thiết sau

- Tiến hành n phép thử độc lập;
- Trong mỗi phép thử chỉ có hai trường hợp: hoặc biến cố A xảy ra, hoặc biến cố A không xảy ra;
- Trong mỗi phép thử xác suất xảy ra của biến cố A đều bằng p , ($0 \leq p \leq 1$), xác suất không xảy ra biến cố A đều bằng $q = 1 - p$.

1.5.2 Công thức Bernoulli

Định lý 1.6. Giả sử bài toán tuân theo lược đồ Bernoulli (n, A, p) . Khi đó, xác suất để trong n phép thử độc lập nói trên, biến cố A xuất hiện đúng k lần, kí hiệu là $P_n(k)$ hoặc p_k , được tính bằng công thức Bernoulli sau đây

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k} \quad (0 \leq k \leq n).$$

Hơn nữa, số lần xuất hiện chắc nhất của biến cố A là số $n_0 \in \mathbb{N}$ thỏa mãn

$$np - q \leq n_0 \leq np + p.$$

Ví dụ 1.19. Trong phân xưởng có 7 máy hoạt động độc lập, xác suất để trong ca mỗi máy bị hỏng đều bằng 0,15.

- Tìm xác suất để trong ca đó có đúng 2 máy hỏng;
- Nhiều nhất có thể có bao nhiêu máy hỏng?
- Số máy hỏng có khả năng xảy ra lớn nhất là bao nhiêu?

Lời giải: Bài toán tuân theo lược đồ Bernoulli với $n = 7$. Gọi A = “máy hỏng”, $p = 0,15$ và $q = 0,85$.

- Xác suất để trong ca đó có đúng 2 máy hỏng là

$$P_7(2) = C_7^2 \cdot 0,15^2 \cdot 0,85^5 \approx 0,20965.$$

- Vì có 7 máy hoạt động nên nhiều nhất có thể cả 7 máy hỏng.

- Số máy hỏng có khả năng xảy ra lớn nhất là n_0 thỏa mãn

$$\begin{aligned} np - q &\leq n_0 \leq np + p; \\ \Leftrightarrow 7 \cdot 0,15 - 0,85 &\leq n_0 \leq 7 \cdot 0,15 + 0,15; \\ \Leftrightarrow 0,2 &\leq n_0 \leq 1,2. \end{aligned}$$

Vậy $n_0 = 1$, khả năng lớn nhất là có 1 máy hỏng.

Ví dụ 1.20. Hai đấu thủ A và B thi đấu bóng bàn. Xác suất thắng của A trong mỗi ván là 0,6 (không có hòa). Trận đấu bao gồm 5 ván. Người nào thắng số ván lớn hơn là người thắng cuộc. Tính xác suất để B thắng cuộc.

Lời giải: Nếu coi mỗi ván thi đấu bóng bàn là một phép thử thì sẽ có $n = 5$ phép thử độc lập. Trong mỗi lần thử chỉ xảy ra hai khả năng: B thắng hoặc B thua. Xác suất thắng mỗi ván của B là 0,4. Vậy bài toán tuân theo lược đồ Bernoulli với $n = 5$ và $p = 0,4$, suy ra $q = 0,6$. Ta có

- Xác suất để B thắng 3 ván là $P_5(3) = C_5^3 \cdot (0,4)^3 \cdot (0,6)^2 = 0,2304$.
- Xác suất để B thắng 4 ván là $P_5(4) = C_5^4 \cdot (0,4)^4 \cdot (0,6)^1 = 0,0768$.
- Xác suất để B thắng 5 ván là $P_5(5) = (0,4)^5 = 0,0102$.

Vậy xác suất để B thắng cuộc là

$$P_5(3) + P_5(4) + P_5(5) = 0,3174.$$

1.5.3 Công thức Bernoulli mở rộng

Ta có thể tổng quát hoá lược đồ Bernoulli theo hướng như sau:

- Giả sử tiến hành n phép thử độc lập;
- Trong mỗi phép thử có m kết cục $A_1, A_2, \dots, A_m$ hợp thành nhóm đầy đủ;
- Trong mỗi phép thử, xác suất xuất hiện kết cục A_i bằng p_i .

Kí hiệu $P_n(k_1, k_2, \dots, k_m)$ là xác suất để trong n phép thử đó, biến cố A_i xảy ra k_i lần, ($i = \overline{1, m}$).

Ta có

$$\begin{aligned} P_n(k_1, k_2, \dots, k_m) &= C_n^{k_1} \cdot C_{n-k_1}^{k_2} \cdot \dots \cdot C_{n-k_1-\dots-k_{m-1}}^{k_m} \cdot p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_m^{k_m} \\ &= \frac{n!}{k_1! k_2! \dots k_m!} p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_m^{k_m}. \end{aligned}$$

Ví dụ 1.21. Tại một ga xếp có 10 khách lên tàu gồm 3 toa. Xác suất một khách bất kì lên toa I, II và III theo thứ tự là 0,2; 0,5 và 0,3. Tính xác suất

- Có đúng 3 khách lên toa I;
- Có 3 khách lên toa I, 5 khách lên toa II và 2 khách lên toa III;
- Có ít nhất 5 khách lên toa II và 4 khách lên toa III.

Lời giải: Coi sự kiện mỗi người khách lên tàu là 1 phép thử, như vậy ta có dãy 10 phép thử độc lập.

- Bài toán tuân theo lược đồ Bernoulli với $n = 10$; $p = 0,2$. Xác suất cần tìm là

$$p_3 = P_{10}(3) = C_{10}^3 \cdot (0,2)^3 \cdot (0,8)^7 \approx 0,201.$$

- Bài toán tuân theo lược đồ Bernoulli mở rộng với

$$n = 10 ; p_1 = 0,2 ; p_2 = 0,5 ; p_3 = 0,3.$$

Xác suất cần tìm là

$$P_{10}(3, 5, 2) = C_{10}^3 \cdot C_7^5 \cdot C_2^2 \cdot (0,2)^3 \cdot (0,5)^5 \cdot (0,3)^2 = 0,0567.$$

- Bài toán tuân theo lược đồ Bernoulli mở rộng tương tự phần trên. Có 3 trường hợp có thể xảy ra

- 6 khách lên toa II và 4 khách lên toa III;
- 5 khách lên toa II và 5 khách lên toa III;
- 1 khách lên toa I; 5 khách lên toa II và 4 khách lên toa III.

Vậy xác suất cần tìm là

$$\begin{aligned} &P_{10}(0, 6, 4) + P_{10}(0, 5, 5) + P_{10}(1, 5, 4) \\ &= C_{10}^6 \cdot (0,5)^6 \cdot (0,3)^4 + C_{10}^5 \cdot (0,5)^5 \cdot (0,3)^5 + C_{10}^1 \cdot C_9^5 \cdot (0,2) \cdot (0,5)^5 \cdot (0,3)^4 \\ &\approx 0,1095. \end{aligned}$$

1.6 Công thức đầy đủ và công thức Bayes

Định lý 1.7. Giả sử n biến cố $H_1, H_2, \dots, H_n$ lập thành nhóm đầy đủ; biến cố A có thể xảy ra đồng thời với một trong các biến cố $H_1, H_2, \dots, H_n$. Khi đó,

- **Công thức xác suất đầy đủ.** Xác suất của biến cố A được tính bằng công thức

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

- **Công thức Bayes.** Nếu biến cố A đã xảy ra thì xác suất điều kiện của biến cố H_k được tính bởi công thức sau

$$\mathbb{P}(H_k|A) = \frac{\mathbb{P}(H_k) \cdot \mathbb{P}(A|H_k)}{\sum_{i=1}^n \mathbb{P}(H_i) \cdot \mathbb{P}(A|H_i)}; \quad (k = \overline{1, n}).$$

Ví dụ 1.22. Có 2 hộp đựng sản phẩm. Hộp thứ nhất có 10 sản phẩm trong đó có 9 chính phẩm. Hộp thứ hai có 20 sản phẩm trong đó có 18 chính phẩm. Từ hộp thứ nhất lấy ngẫu nhiên một sản phẩm bỏ sang hộp thứ hai. Sau đó tiếp tục lấy ngẫu nhiên một sản phẩm từ hộp thứ hai ra ngoài.

- Tìm xác suất để sản phẩm lấy từ hộp thứ hai là chính phẩm;
- Giả sử sản phẩm lấy từ hộp thứ hai là chính phẩm, tính xác suất để sản phẩm lấy từ hộp một bỏ sang hộp hai cũng là một chính phẩm.

Lời giải: Gọi các biến cố:

H_1 = “sản phẩm bỏ từ hộp thứ nhất sang hộp thứ hai là chính phẩm”;

H_2 = “sản phẩm bỏ từ hộp thứ nhất sang hộp thứ hai là phế phẩm”;

A = “sản phẩm lấy từ hộp thứ hai là chính phẩm”.

Biến cố A có thể xảy ra đồng thời với một trong các biến cố H_1, H_2 và hai biến cố H_1, H_2 tạo nên một nhóm đầy đủ. Ta có

- Xác suất để lấy được một chính phẩm từ hộp một là $\mathbb{P}(H_1) = \frac{9}{10}$;
Xác suất để lấy được một phế phẩm từ hộp một là $\mathbb{P}(H_2) = \frac{1}{10}$.
- Nếu lấy một chính phẩm từ hộp một sang hộp hai thì xác suất lấy được một chính phẩm từ hộp hai là $\mathbb{P}(A|H_1) = \frac{19}{21}$;
Nếu lấy một phế phẩm từ hộp một sang hộp hai thì xác suất lấy được một chính phẩm từ hộp hai là $\mathbb{P}(A|H_2) = \frac{18}{21}$.

- Theo công thức đầy đủ, ta có

$$\begin{aligned} \mathbb{P}(A) &= \mathbb{P}(H_1)\mathbb{P}(A|H_1) + \mathbb{P}(H_2)\mathbb{P}(A|H_2) \\ &= \frac{9}{10} \cdot \frac{19}{21} + \frac{1}{10} \cdot \frac{18}{21} = \frac{9}{10}. \end{aligned}$$

b) Ta cần tính $\mathbb{P}(H_1|A)$. Theo công thức Bayes, ta có

$$\begin{aligned}\mathbb{P}(H_1|A) &= \frac{\mathbb{P}(H_1) \cdot \mathbb{P}(A|H_1)}{\mathbb{P}(H_1) \cdot \mathbb{P}(A|H_1) + \mathbb{P}(H_2) \cdot \mathbb{P}(A|H_2)} \\ &= \frac{\frac{9}{10} \cdot \frac{19}{21}}{\frac{9}{10} \cdot \frac{19}{21} + \frac{1}{10} \cdot \frac{18}{21}} = \frac{\frac{57}{70}}{\frac{57}{70} + \frac{18}{210}} = \frac{19}{21}.\end{aligned}$$

Ví dụ 1.23. Phương pháp xét nghiệm³ SARS-CoV-2 bằng RT-PCR có độ nhạy 0,87 và độ đặc hiệu 0,95. Tức là nếu một người thực sự nhiễm virus SARS-CoV-2 thì xác suất xét nghiệm có kết quả dương tính bằng 0,87, còn nếu thực sự không bị nhiễm thì xác suất xét nghiệm âm tính bằng 0,95.

Chọn ngẫu nhiên một người trong cộng đồng để xét nghiệm RT-PCR. Gọi các biến cố:

H_1 = “Người đó thực sự nhiễm virus SARS-CoV-2”;

H_2 = “Người đó thực sự *không* nhiễm virus SARS-CoV-2”.

A = “Người đó có kết quả xét nghiệm RT-PCR dương tính”;

B = “Người đó có kết quả xét nghiệm RT-PCR âm tính”.

Dễ thấy $\{H_1, H_2\}$ lập thành một nhóm đầy đủ và

$$\begin{aligned}\mathbb{P}(A|H_1) &= 0,87; & \mathbb{P}(A|H_2) &= 0,05; \\ \mathbb{P}(B|H_1) &= 0,13; & \mathbb{P}(B|H_2) &= 0,95.\end{aligned}$$

Xét một cộng đồng có nguy cơ rất cao, với tỉ lệ nhiễm virus trong cộng đồng là 10%. Tức là:

$$\mathbb{P}(H_1) = 0,1; \quad \mathbb{P}(H_2) = 0,9.$$

Khi đó nếu xét nghiệm toàn cộng đồng thì tỉ lệ kết quả xét nghiệm dương tính là:

$$\begin{aligned}\mathbb{P}(A) &= \mathbb{P}(H_1) \cdot \mathbb{P}(A|H_1) + \mathbb{P}(H_2) \cdot \mathbb{P}(A|H_2) \\ &= 0,1 \cdot 0,87 + 0,9 \cdot 0,05 = 0,132.\end{aligned}$$

Nếu một người có kết quả xét nghiệm dương tính thì khả năng người đó thực sự mắc bệnh là:

$$\mathbb{P}(H_1|A) = \frac{\mathbb{P}(H_1) \cdot \mathbb{P}(A|H_1)}{\mathbb{P}(A)} = \frac{0,1 \cdot 0,87}{0,132} = 0,659.$$

Suy ra tỉ lệ dương tính giả khá cao, bằng $\mathbb{P}(H_2|A) = 0,341$. Tức là trong số những người có kết quả xét nghiệm RT-PCR dương tính, chỉ có 65,9% người thực sự nhiễm virus, còn lại 34,1% là không bị nhiễm (dù có kết quả dương tính).

Tuy nhiên nếu một người có kết quả xét nghiệm âm tính thì khả năng người đó thực sự không bị nhiễm là:

$$\mathbb{P}(H_2|B) = \frac{\mathbb{P}(H_2) \cdot \mathbb{P}(B|H_2)}{\mathbb{P}(B)} = \frac{0,9 \cdot 0,95}{1 - 0,132} = 0,985.$$

Suy ra tỉ lệ âm tính giả chỉ bằng 1,5%. Tức là nếu một người có kết quả xét nghiệm RT-PCR âm tính thì khả năng cao (98,5%) là người đó không nhiễm SARS-CoV-2.

³Số liệu được tham khảo tại:

Arevalo-Rodriguez I & al. (2020) *False-negative results of initial RT-PCR assays for COVID-19: A systematic review*. PLOS ONE 15(12): e0242958. <https://doi.org/10.1371/journal.pone.0242958>

Bài tập chương 1

Định nghĩa cổ điển về xác suất

Bài 1.1. Gieo một con súc sắc cân đối đồng chất. Tính xác suất để

- a) Mặt xuất hiện có số chấm lớn hơn 2;
- b) Mặt xuất hiện có số chấm là một số chẵn.

Bài 1.2. Gieo đồng thời 2 con súc sắc cân đối đồng chất một cách ngẫu nhiên. Tính xác suất để được hai mặt có

- a) Tổng số chấm bằng 7;
- b) Tổng số chấm nhỏ hơn 8;
- c) Số chấm hơn kém nhau 2;
- d) Số chấm bằng nhau;
- e) Ít nhất một mặt 6 chấm.

Bài 1.3. Một công ty vận tải cần tuyển 2 nhân viên. Có 8 người nộp đơn ứng tuyển trong đó có 3 nam và 5 nữ, khả năng trúng tuyển của mỗi người là như nhau.

- a) Tính xác suất để 2 người trúng tuyển đều là nữ;
- b) Tính xác suất để có ít nhất 1 nữ trúng tuyển;
- c) Giả sử Mai là một trong 5 nữ ứng tuyển. Tính xác suất để Mai trúng tuyển.

Bài 1.4. Để kiểm tra chất lượng sản phẩm của một lô hàng có 50 sản phẩm, người ta lấy ngẫu nhiên ra 10 sản phẩm từ lô sản phẩm này. Nếu trong 10 sản phẩm lấy ra có không quá 1 phế phẩm thì lô hàng được xếp loại đạt chất lượng, ngược lại thì lô hàng không đạt chất lượng và bị loại. Biết rằng trong lô hàng có 5 phế phẩm, tính xác suất để lô hàng bị loại.

Bài 1.5. Nam có hai hộp bi, hộp 1 đựng 2 bi màu đỏ và 4 bi màu xanh, hộp 2 đựng 3 bi màu đỏ và 4 bi màu xanh. Kích thước các viên bi như nhau, Nam lấy ngẫu nhiên từ mỗi hộp ra 2 viên bi. Tính xác suất để

- a) Trong 4 viên lấy ra có 2 viên bi màu đỏ;
- b) Số viên bi màu đỏ lấy ra ở mỗi hộp bằng nhau;
- c) Số viên bi màu đỏ lấy ra ở hộp 2 ít hơn ở hộp 1.

Bài 1.6. Một lô hàng chi tiết máy có 30 sản phẩm cùng loại trong đó có 3 phế phẩm được chia ngẫu nhiên thành 3 hộp kín phân biệt với nhau, mỗi hộp 10 sản phẩm. Tính xác suất để có

- a) Mỗi hộp chứa 1 phế phẩm;
- b) Một hộp chứa 1 phế phẩm.

Bài 1.7. Về mặt hành chính, hiện nay ở thành phố Hải Phòng có 15 quận, huyện trực thuộc. Giả sử mỗi quận, huyện đề cử 2 thanh niên tiêu biểu là thủ lĩnh đoàn. Thành đoàn Hải Phòng cần chọn ngẫu nhiên 15 thủ lĩnh đoàn trong số 30 đề cử để thành lập một lớp tập huấn các hoạt động hè năm 2021. Tính xác suất để

- a) Mỗi quận, huyện đều có đúng một thủ lĩnh đoàn tham gia lớp tập huấn;
- b) Trong lớp tập huấn có ít nhất một thủ lĩnh đoàn của quận Ngô Quyền.

Bài 1.8. Trong buổi lễ Tuyên dương sinh viên tiêu biểu cuối năm học của trường Đại học Hàng hải Việt Nam có 4 sinh viên khoa Đóng tàu, 5 sinh viên khoa Máy tàu biển và 6 sinh viên khoa Kinh tế. Biết các sinh viên cùng lên nhận phần thưởng và xếp ngẫu nhiên thành một hàng ngang. Tính xác suất để

- Sinh viên mỗi khoa đều đứng cạnh nhau;
- Không có sinh viên khoa Đóng tàu nào đứng cạnh nhau.

Bài 1.9. Trong một thùng có 3 quả cầu màu đỏ và 5 quả cầu màu đen có cùng kích thước. Người ta lấy ngẫu nhiên mỗi lần một quả cầu không hoàn lại cho đến hết.

- Tính xác suất để lấy được quả cầu đầu tiên màu đỏ;
- Tính xác suất để lấy được quả cầu cuối cùng màu đỏ.

Bài 1.10. Một hộp kín có 100 tấm bìa hình vuông kích thước như nhau được đánh số từ 1 đến 100. Rút ngẫu nhiên 2 tấm bìa rồi đặt theo thứ tự từ trái sang phải.

- Tính xác suất để hai tấm bìa rút được theo thứ tự trên lập nên một số có 2 chữ số;
- Tính xác suất để hai tấm bìa rút được theo thứ tự trên lập nên một số chia hết cho 5.

Định nghĩa hình học và thống kê về xác suất

Bài 1.11. Hai tàu chở hàng sẽ cập vào một bến cảng độc lập nhau trong vòng một ngày đêm. Biết rằng thời gian đỗ lại bến để bốc xếp hàng của tàu thứ nhất là 1 giờ, của tàu thứ hai là 2 giờ. Tính xác suất để một trong hai tàu phải chờ để cập bến. Biết tại mỗi thời điểm, cảng chỉ cho phép một trong hai tàu trên cập bến.

Bài 1.12. Hai bạn Nam và Mai hẹn gặp nhau tại một quán cafe từ 8 đến 9 giờ với quy ước người nào đến trước chờ quá 20 phút thì sẽ về và hủy cuộc hẹn. Tìm xác suất để họ gặp nhau, biết rằng mỗi người có thể đến nơi hẹn vào thời điểm bất kì trong thời gian trên.

Bài 1.13. Trên mặt phẳng cho đường tròn bán kính R cố định và đường thẳng d (nằm ngoài đường tròn). Kẻ ngẫu nhiên các đường thẳng song song với d và cắt đường tròn đã cho. Tìm xác suất để cát tuyến nhận được có độ dài nhỏ hơn R .

Bài 1.14. Một thanh sắt thẳng được bẻ thành ba khúc một cách ngẫu nhiên. Tìm xác suất để ba khúc đó tạo thành một tam giác, biết rằng thanh sắt đó dài l (đơn vị dài).

Bài 1.15. Tỉ suất xuất hiện biến cố viên đạn trúng đích là 0,85. Tìm số viên đạn bắn trúng đích của một xạ thủ nếu người này bắn 200 viên đạn.

Bài 1.16. Có thể xem xác suất sinh được cháu trai là bao nhiêu nếu quan sát 88200 trường hợp sinh trong năm ở một vùng thấy có 45600 cháu trai.

Quan hệ giữa các biến cố

Bài 1.17. Ba xạ thủ cùng bắn mỗi người một viên đạn vào cùng một mục tiêu. Gọi A, B, C là các biến cố các xạ thủ tương ứng bắn trúng mục tiêu. Hãy viết bằng kí hiệu các biến cố sau đây

- Chỉ có 2 viên đạn trúng mục tiêu;

- b) Cả 3 viên đạn đều trúng mục tiêu;
- c) Không có viên đạn nào trúng mục tiêu;
- d) Chỉ có viên đạn của xạ thủ đầu tiên trúng mục tiêu;
- e) Chỉ có 1 viên đạn trúng mục tiêu;
- f) Có ít nhất 1 viên đạn trúng mục tiêu.

Bài 1.18. Hai cầu thủ bóng rổ Dũng và Nam mỗi người lần lượt ném hai quả bóng vào rổ. Gọi A_1, A_2 và B_1, B_2 tương ứng là biến cố Dũng và Nam ném trúng rổ quả thứ nhất và quả thứ hai. Biểu diễn các biến cố sau đây

- a) Số bóng trúng rổ của hai cầu thủ bằng nhau;
- b) Số bóng trúng rổ của Dũng ít hơn Nam;
- c) Tổng số bóng trúng rổ của cả hai cầu thủ bằng 3.

Công thức cộng và nhân xác suất

Bài 1.19. Cơ cấu chất lượng sản phẩm của một nhà máy như sau: sản phẩm loại I là 40%; sản phẩm loại II là 50%; còn lại là phế phẩm. Lấy ngẫu nhiên một sản phẩm của nhà máy. Tính xác suất để sản phẩm lấy ra thuộc loại I hoặc loại II.

Bài 1.20. Để được nhập kho, sản phẩm nhà máy phải qua 3 phòng kiểm tra chất lượng, xác suất phát hiện ra phế phẩm ở các phòng theo thứ tự là 0,8; 0,9 và 0,99. Biết các phòng kiểm tra hoạt động độc lập với nhau, tính xác suất phế phẩm được nhập kho.

Bài 1.21. Xác suất để khi đo một đại lượng vật lý phạm sai số vượt quá tiêu chuẩn cho phép là 0,4. Thực hiện 3 lần đo độc lập. Tìm xác suất sao cho có đúng một lần đo phạm sai số vượt quá tiêu chuẩn cho phép.

Bài 1.22. Xác suất thi đỗ lần 1 môn Xác suất thống kê của một sinh viên là 0,8; thi đỗ lần 2 là 0,7. Biết rằng khả năng đỗ mỗi lần thi của sinh viên là độc lập với nhau, sinh viên chỉ thi lần 2 khi thi trượt lần 1 và sinh viên chỉ được thi tối đa 2 lần. Tìm xác suất thi đỗ môn Xác suất thống kê của sinh viên đó?

Bài 1.23. Xác suất ném bóng trúng rổ của 3 cầu thủ tương ứng là 0,8; 0,7 và 0,6. Mỗi người được ném 1 quả bóng vào rổ, lần lượt người trước ném trượt thì người sau ném tiếp, còn nếu người trước ném trúng rổ thì các người sau dừng. Tính xác suất để có 1 người ném trúng rổ, biết khả năng ném trúng rổ của các cầu thủ độc lập với nhau?

Bài 1.24. Trong một hộp kín có 3 bi màu đỏ và 7 bi màu trắng, lấy ngẫu nhiên lần lượt từng viên bi không hoàn lại đến khi nào trong số bi lấy ra có 2 bi màu đỏ thì dừng lại. Tính xác suất để số bi lấy ra không vượt quá 4?

Bài 1.25. Ba xạ thủ cùng bắn một viên đạn vào một mục tiêu với xác suất trúng đích của từng người là 0,6; 0,7 và 0,8. Tính xác suất để

- a) Chỉ có người thứ hai bắn trúng mục tiêu;
- b) Có đúng 1 người bắn trúng mục tiêu;
- c) Có ít nhất 1 người bắn trúng mục tiêu;
- d) Cả 3 người đều bắn trúng mục tiêu;
- e) Có đúng 2 người bắn trúng mục tiêu;

- f) Có ít nhất 2 người bắn trúng mục tiêu;
- g) Có không quá 2 người bắn trúng mục tiêu.

Bài 1.26. Một tín hiệu thông tin được phát 3 lần với xác suất thu được của mỗi lần là 0,4.

- a) Tìm xác suất để nguồn thu nhận được thông tin đó;
- b) Muốn xác suất thu được thông tin lên đến 0,9 thì phải phát ít nhất bao nhiêu lần?

Bài 1.27. Một người bắn liên tiếp vào một mục tiêu cho đến khi viên đạn đầu tiên trúng mục tiêu thì dừng lại. Biết rằng xác suất trúng mục tiêu của mỗi viên đạn là 0,2 và các lần bắn độc lập với nhau. Tính xác suất để người đó phải bắn đến viên đạn thứ 6.

Bài 1.28. Một công ty sử dụng hai hình thức quảng cáo là vô tuyến truyền hình và đài phát thanh để quảng bá sản phẩm. Giả sử 25% khách hàng nắm được thông tin này qua vô tuyến truyền hình, 34% khách hàng nắm được thông tin qua đài phát thanh và 10% khách hàng nắm được thông tin qua cả hai hình thức quảng cáo. Tìm xác suất để chọn ngẫu nhiên một khách hàng của công ty thì người này không nắm được thông tin về sản phẩm qua cả hai hình thức quảng cáo trên?

Bài 1.29. Trong một vùng dân cư, tỉ lệ người mắc bệnh tim là 9%, mắc bệnh huyết áp là 12% và mắc cả hai bệnh là 7%. Chọn ngẫu nhiên một người trong vùng dân cư, tính xác suất để người đó

- a) Mắc bệnh tim hay bệnh huyết áp;
- b) Không mắc bệnh tim cũng không mắc bệnh huyết áp;
- c) Không mắc bệnh tim hay không mắc bệnh huyết áp;
- d) Bị bệnh tim nhưng không mắc bệnh huyết áp;
- e) Không mắc bệnh tim nhưng mắc bệnh huyết áp.

Bài 1.30. Bộ môn Ngoại ngữ của một trường đại học có 16 giảng viên, trong đó 9 người biết Tiếng Anh, 7 người biết tiếng Trung, 8 người biết tiếng Pháp, 2 người biết cả 3 thứ tiếng, 3 người biết tiếng Anh và tiếng Pháp, 4 người biết tiếng Anh và tiếng Trung, 4 người biết tiếng Pháp và tiếng Trung. Mời ngẫu nhiên một giảng viên, tìm xác suất để người đó biết ít nhất một thứ tiếng?

Bài 1.31. Điều tra mối liên hệ giữa người nghiện thuốc lá và bệnh ung thư phổi ở một vùng dân cư thu được kết quả: 15% số dân vừa nghiện thuốc lá vừa mắc bệnh ung thư phổi, 25% số dân nghiện thuốc lá mà không mắc bệnh ung thư phổi và 10% số dân không nghiện thuốc lá lại mắc bệnh ung thư phổi. Tính xác suất để một người nghiện thuốc lá mắc bệnh ung thư phổi và xác suất để một người không nghiện thuốc lá mắc bệnh ung thư phổi. Từ đó có nhận xét gì về mối quan hệ giữa thói nghiện thuốc lá và bệnh ung thư phổi?

Bài 1.32. Một lô hàng có 9 sản phẩm. Mỗi lần kiểm tra chất lượng lấy ngẫu nhiên ra 3 sản phẩm, sau khi kiểm tra xong lại trả lại vào lô hàng. Tính xác suất để sau 3 lần kiểm tra tất cả các sản phẩm của lô hàng đều được kiểm tra.

Bài 1.33. Một tủ kho có một chùm chìa khóa gồm 9 chiếc bề ngoài chúng giống hệt nhau nhưng chỉ có 1 chiếc mở được cửa kho. Anh ta thử ngẫu nhiên từng chìa (chìa nào không trúng thì bỏ ra). Tính xác suất để anh ta mở được cửa kho ở lần mở thứ 4.

Công thức Bernoulli

Bài 1.34. Một nữ công nhân quản lý 12 máy dệt, xác suất để mỗi máy trong khoảng thời gian t cần đến sự chăm sóc của nữ công nhân là $\frac{1}{3}$.

- Tìm xác suất để trong khoảng thời gian t có máy cần đến sự chăm sóc của nữ công nhân;
- Tìm xác suất để trong khoảng thời gian t có không nhiều hơn 2 máy cần đến sự chăm sóc của nữ công nhân;
- Tìm số máy cần đến sự chăm sóc của nữ công nhân có khả năng xảy ra nhiều nhất và xác suất để xảy ra điều đó.

Bài 1.35. Giả sử bạn đem giao một lô hàng có rất nhiều sản phẩm với tỉ lệ phế phẩm là 10%. Người nhận hàng đề nghị lấy ngẫu nhiên 6 sản phẩm để kiểm tra và chỉ nhận lô hàng nếu có không quá k phế phẩm. Bạn cần phải đề nghị k bằng bao nhiêu để vừa thuyết phục được người nhận, vừa hi vọng khả năng lô hàng không bị từ chối ít nhất là 95%?

Công thức đầy đủ và Bayes

Bài 1.36. Một nhà máy có 3 phân xưởng sản xuất cùng một loại sản phẩm. Trong đó phân xưởng I sản xuất 36%, phân xưởng II sản xuất 34%, phân xưởng III sản xuất 30% tổng sản lượng toàn nhà máy. Biết tỉ lệ phế phẩm tương ứng của các phân xưởng là 12%, 10% và 8%. Lấy ngẫu nhiên một sản phẩm của nhà máy.

- Tính tỉ lệ phế phẩm chung của nhà máy;
- Biết sản phẩm lấy ra là phế phẩm, tìm xác suất để sản phẩm đó thuộc phân xưởng II sản xuất;
- Giả sử sản phẩm lấy ra là phế phẩm, sản phẩm phế phẩm đó có khả năng thuộc phân xưởng nào nhiều nhất?

Bài 1.37. Một lớp học môn Xác suất thống kê của Trường Đại học Hàng hải Việt Nam có $\frac{2}{3}$ là sinh viên nam và $\frac{1}{3}$ là sinh viên nữ. Số sinh viên quê ở Hải Phòng chiếm tỉ lệ 40% trong nữ sinh viên và 60% trong nam sinh viên.

- Chọn ngẫu nhiên một sinh viên của lớp. Tính xác suất để chọn được một sinh viên quê ở Hải Phòng.
- Nếu biết rằng sinh viên vừa chọn quê ở Hải Phòng thì xác suất để sinh viên đó là nam bằng bao nhiêu?

Bài tập tổng hợp

Bài 1.38. Trong kì thi kết thúc học kỳ vừa qua ở trường Đại học Hàng hải Việt Nam, tỉ lệ sinh viên trượt môn Xác suất thống kê là 34%, trượt môn Vật lí là 20,5% và trong số các sinh viên trượt môn Xác suất thống kê có 50% sinh viên trượt môn Vật lí. Gặp ngẫu nhiên một sinh viên của trường đã thi xong hai môn trên.

- Tính xác suất để sinh viên đó trượt cả hai môn;

- b) Tính xác suất để sinh viên đó đỗ cả hai môn;
- c) Nếu biết sinh viên này trượt môn Vật lý thì xác suất để anh ta đỗ môn Xác suất thống kê là bao nhiêu?
- d) Chọn ngẫu nhiên 12 sinh viên của trường đã thi xong hai môn trên, nhiều khả năng nhất là có bao nhiêu sinh viên thi trượt cả hai môn ? Tính xác suất tương ứng;
- e) Phải chọn ngẫu nhiên bao nhiêu sinh viên của trường đã thi xong hai môn trên, sao cho với xác suất không bé hơn 0,99, trong số đó có ít nhất một sinh viên đỗ cả hai môn?

Bài 1.39. Có 30 tấm bìa được đánh số lần lượt từ 1 đến 30, trong đó có 15 tấm bìa chữ "SU", 11 tấm bìa chữ "ZU", 3 tấm bìa chữ "KI", 1 tấm bìa chữ "SUZUKI". Người chơi sẽ thắng cuộc nếu ghép được (hoặc chọn được) chữ SUZUKI. Người chơi được chọn ngẫu nhiên 2 tấm bìa và trả lời thêm 3 câu hỏi, trả lời đúng mỗi câu thì được chọn ngẫu nhiên thêm 1 tấm bìa. Biết xác suất trả lời đúng mỗi câu hỏi là 0,3, tính xác suất người chơi thắng cuộc.

Chương 2

Đại lượng ngẫu nhiên và các quy luật phân phối xác suất

2.1	Đại lượng ngẫu nhiên	23
2.2	Quy luật phân phối xác suất của đại lượng ngẫu nhiên	24
2.3	Các tham số đặc trưng của đại lượng ngẫu nhiên	33
2.4	Một số quy luật phân phối xác suất thông dụng	39
	Bài tập chương 2	52

2.1 Đại lượng ngẫu nhiên

2.1.1 Các định nghĩa

Định nghĩa 2.1. “*Đại lượng ngẫu nhiên* (hay còn gọi là *biến ngẫu nhiên*) là đại lượng mà trong kết quả của phép thử sẽ nhận một và chỉ một trong các giá trị có thể có của nó với một xác suất tương ứng xác định.”

Các đại lượng ngẫu nhiên được kí hiệu là $X, Y, \dots$, còn các giá trị có thể có của đại lượng ngẫu nhiên được kí hiệu là $x, x_1, x_2, \dots, x_n, y, y_1, y_2, \dots$

Đại lượng X được gọi là ngẫu nhiên vì trước khi tiến hành phép thử ta chưa thể nói chắc chắn nó sẽ nhận giá trị bằng bao nhiêu, mà chỉ có thể dự đoán điều đó với xác suất nhất định. Việc X nhận giá trị nào đó ($X = x$) hoặc ($a < X < b$) là các biến cố ngẫu nhiên.

2.1.2 Phân loại đại lượng ngẫu nhiên

Chúng ta quan tâm đến hai loại đại lượng ngẫu nhiên sau đây:

- Đại lượng ngẫu nhiên được gọi là *rời rạc* nếu các giá trị có thể có của nó là một tập hợp hữu hạn hoặc đếm được.
- Đại lượng ngẫu nhiên được gọi là *liên tục* nếu các giá trị có thể có của nó lấp đầy một khoảng trên trục số.

Ví dụ 2.1. Tung một con súc sắc. Gọi X là “Số chấm xuất hiện”. X là đại lượng ngẫu nhiên rời rạc vì trong kết quả của phép thử nó sẽ nhận một trong các giá trị có thể có là 1, 2, 3, 4, 5, 6 với xác suất tương ứng là $\mathbb{P}(X = k) = \frac{1}{6}$, $k = \overline{1, 6}$.

Ví dụ 2.2. Gọi X là “Số người vào mua hàng tại một siêu thị trong một ngày”. X là đại lượng ngẫu nhiên rời rạc với các giá trị có thể có của nó là một tập hợp đếm được $0, 1, 2, \dots$.

Ví dụ 2.3. Bắn ngẫu nhiên một phát súng vào bia hình tròn bán kính 40 cm, giả sử tất cả các viên đạn đều trúng bia. Gọi X là “Khoảng cách từ tâm bia đến điểm chạm của đạn vào bia”. X là đại lượng ngẫu nhiên liên tục vì ta không thể kể ra được tất cả các giá trị có thể có của nó, ta chỉ có thể nói rằng tập hợp đó là khoảng $[0; 40]$ cm.

Nhận xét. Nếu X là đại lượng ngẫu nhiên liên tục thì tại mỗi giá trị k cụ thể, ta luôn có: $\mathbb{P}(X = k) = 0$.

2.2 Quy luật phân phối xác suất của đại lượng ngẫu nhiên

Quy luật phân phối xác suất là một cách biểu diễn mối quan hệ giữa các giá trị có thể có của đại lượng ngẫu nhiên và các xác suất tương ứng của nó.

Ta thường sử dụng ba phương pháp để thiết lập quy luật phân phối xác suất của đại lượng ngẫu nhiên là bảng phân phối xác suất, hàm phân phối xác suất, hàm mật độ xác suất. Trong đó, bảng phân phối xác suất chỉ sử dụng cho đại lượng ngẫu nhiên rời rạc và hàm mật độ xác suất chỉ sử dụng cho đại lượng ngẫu nhiên liên tục.

2.2.1 Bảng phân phối xác suất

Giả sử đại lượng ngẫu nhiên rời rạc X có thể nhận một trong các giá trị có thể có là $x_1, x_2, \dots, x_n$ với các xác suất tương ứng là $p_1, p_2, \dots, p_n$. Bảng phân phối xác suất của đại lượng ngẫu nhiên rời rạc X có dạng

X	x_1	x_2	$\dots$	x_i	$\dots$	x_n
$\mathbb{P}$	p_1	p_2	$\dots$	p_i	$\dots$	p_n

Chú ý rằng, để tạo nên một quy luật phân phối xác suất thì các xác suất p_i phải thỏa mãn điều kiện sau

$$\begin{cases} 0 \leq p_i \leq 1, & \forall i = \overline{1, n}; \\ \sum_{i=1}^n p_i = 1. \end{cases}$$

Thông thường khi lập bảng phân phối xác suất người ta thường sắp xếp các giá trị có thể có theo thứ tự tăng dần: $x_1 < x_2 < \dots < x_n$.

Ví dụ 2.4. Tung một con súc sắc và xét đại lượng ngẫu nhiên $X =$ “Số chấm xuất hiện”.

Ta có X là đại lượng ngẫu nhiên rời rạc nhận các giá trị có thể có 1, 2, 3, 4, 5, 6 với các xác suất tương ứng đều bằng $\frac{1}{6}$. Vì vậy, bảng phân phối xác suất của X có dạng

X	1	2	3	4	5	6
$\mathbb{P}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Nhận xét thấy tất cả các xác suất đều nằm trong $[0; 1]$ và

$$\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1.$$

Ví dụ 2.5. Trong hộp có 10 sản phẩm trong đó có 6 chính phẩm. Lấy ngẫu nhiên 3 sản phẩm. Lập bảng phân phối xác suất của số chính phẩm được lấy ra.

Lời giải: Gọi X là số chính phẩm lấy ra trong 3 sản phẩm, X là đại lượng ngẫu nhiên rời rạc với các giá trị có thể có 0, 1, 2, 3.

• $X = 0$ là biến cố để cả 3 sản phẩm lấy ra đều là phế phẩm (không có chính phẩm), biến cố này có xác suất là

$$\mathbb{P}(X = 0) = \frac{C_4^3}{C_{10}^3} = \frac{4}{120} = \frac{1}{30}.$$

• $X = 1$ là biến cố để trong 3 sản phẩm lấy ra có đúng 1 chính phẩm, biến cố này có xác suất là

$$\mathbb{P}(X = 1) = \frac{C_6^1 \cdot C_4^2}{C_{10}^3} = \frac{36}{120} = \frac{3}{10}.$$

• $X = 2$ là biến cố để trong 3 sản phẩm lấy ra có đúng 2 chính phẩm, biến cố này có xác suất là

$$\mathbb{P}(X = 2) = \frac{C_6^2 \cdot C_4^1}{C_{10}^3} = \frac{60}{120} = \frac{1}{2}.$$

• $X = 3$ là biến cố để cả 3 sản phẩm đều là chính phẩm, biến cố này có xác suất là

$$\mathbb{P}(X = 3) = \frac{C_6^3}{C_{10}^3} = \frac{20}{120} = \frac{1}{6}.$$

Như vậy, bảng phân phối xác suất của X có dạng

X	0	1	2	3
$\mathbb{P}$	$\frac{1}{30}$	$\frac{3}{10}$	$\frac{1}{2}$	$\frac{1}{6}$

Ta nhận thấy tất cả các xác suất đều nằm trong $[0; 1]$ và

$$\frac{1}{30} + \frac{3}{10} + \frac{1}{2} + \frac{1}{6} = 1.$$

Ví dụ 2.6. Một tàu khu trục của công ty vận tải biển Isalco có 3 hệ thống máy trực vớt hoạt động độc lập. Khi ra biển hoạt động thì trong một ngày khả năng các hệ thống đó hỏng tương ứng là 0, 4; 0, 2 và 0, 3. Tìm quy luật phân phối xác suất của số hệ thống bị hỏng.

Lời giải: Gọi X là số hệ thống bị hỏng. X nhận các giá trị 0, 1, 2, 3.

$$\mathbb{P}(X = 0) = 0,6 \cdot 0,8 \cdot 0,7 = 0,336;$$

$$\mathbb{P}(X = 1) = 0,4 \cdot 0,8 \cdot 0,7 + 0,6 \cdot 0,2 \cdot 0,7 + 0,6 \cdot 0,8 \cdot 0,3 = 0,452;$$

$$\mathbb{P}(X = 2) = 0,4 \cdot 0,2 \cdot 0,7 + 0,4 \cdot 0,8 \cdot 0,3 + 0,6 \cdot 0,2 \cdot 0,3 = 0,188;$$

$$\mathbb{P}(X = 3) = 0,4 \cdot 0,2 \cdot 0,3 = 0,024.$$

Bảng phân phối xác suất của số hệ thống bị hỏng là

X	0	1	2	3
$\mathbb{P}$	0,336	0,452	0,188	0,024

Các xác suất đều nằm trong $[0; 1]$ và tổng các xác suất đó là

$$0,336 + 0,452 + 0,188 + 0,024 = 1.$$

Ví dụ 2.7. Xác suất để xạ thủ bắn trúng bia là 0,8. Xạ thủ được phát từng viên đạn để bắn cho đến khi trúng bia. Xây dựng quy luật phân phối của số đạn được phát.

Lời giải: Gọi X là “Số viên đạn xạ thủ được phát”. X là đại lượng ngẫu nhiên rời rạc với các giá trị có thể có $X = 1, 2, \dots, k, \dots$

Xác suất để xạ thủ được phát một viên đạn là xác suất để xạ thủ ấy bắn viên đạn đầu tiên trúng bia. Do đó,

$$\mathbb{P}(X = 1) = 0,8.$$

Xác suất để xạ thủ được phát 2 viên đạn là xác suất để xảy ra đồng thời hai biến cố, đó là phát thứ nhất bắn trượt và phát thứ hai bắn trúng. Theo định lí nhân xác suất, ta có

$$\mathbb{P}(X = 2) = 0,2 \cdot 0,8.$$

Biến cố $(X = k)$ là tích của k biến cố, đó là $k - 1$ phát đầu bắn trượt và phát thứ k bắn trúng. Theo định lí nhân xác suất, ta có

$$\mathbb{P}(X = k) = (0,2)^{k-1} \cdot 0,8.$$

Như vậy, bảng phân phối xác suất của X có dạng

X	1	2	3	$\dots$	k	$\dots$
$\mathbb{P}$	0,8	0,2.0,8	$(0,2)^2 \cdot 0,8$	$\dots$	$(0,2)^{k-1} \cdot 0,8$	$\dots$

Các xác suất đều nằm trong $[0; 1]$ và tổng các xác suất là tổng cấp số nhân lùi vô hạn, ta có

$$\sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} (0,2)^{k-1} \cdot 0,8 = \frac{0,8}{1 - 0,2} = 1.$$

2.2.2 Hàm phân phối xác suất

Định nghĩa 2.2. “Hàm phân phối xác suất của đại lượng ngẫu nhiên X , kí hiệu $F(x)$, là xác suất để đại lượng ngẫu nhiên X nhận giá trị nhỏ hơn x , với x là số thực bất kì.

$$F(x) = \mathbb{P}(X < x).”$$

Ý nghĩa của hàm phân phối xác suất

Hàm phân phối xác suất phản ánh mức độ tập trung xác suất về phía bên trái một số thực x nào đó. Giá trị của $F(x)$ cho biết có bao nhiêu phần của một đơn vị xác suất phân phối trong khoảng $(-\infty, x)$.

Nhận xét. Theo Định nghĩa 2.2, nếu X là đại lượng ngẫu nhiên rời rạc thì từ bảng phân phối xác suất của X , ta có thể xây dựng được hàm phân phối xác suất như sau

$$F(x) = \sum_{x_i < x} \mathbb{P}(X = x_i) = \sum_{x_i < x} p_i.$$

Ví dụ 2.8. Cho đại lượng ngẫu nhiên rời rạc X có bảng phân phối xác suất sau

X	1	3	4
P	0,1	0,5	0,4

Hãy xây dựng hàm phân phối xác suất của X và vẽ đồ thị.

Lời giải:

- Nếu $x \leq 1$, biến cố $(X < x)$ là biến cố không thể có, do đó

$$F(x) = 0.$$

- Nếu $1 < x \leq 3$, biến cố $(X < x)$ chỉ xảy ra khi $(X = 1)$, do đó

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

- Nếu $3 < x \leq 4$, biến cố $(X < x)$ sẽ xảy ra khi $(X = 1)$ hoặc khi $(X = 3)$, do đó

$$F(x) = \mathbb{P}(X = 1) + \mathbb{P}(X = 3) = 0,1 + 0,5 = 0,6.$$

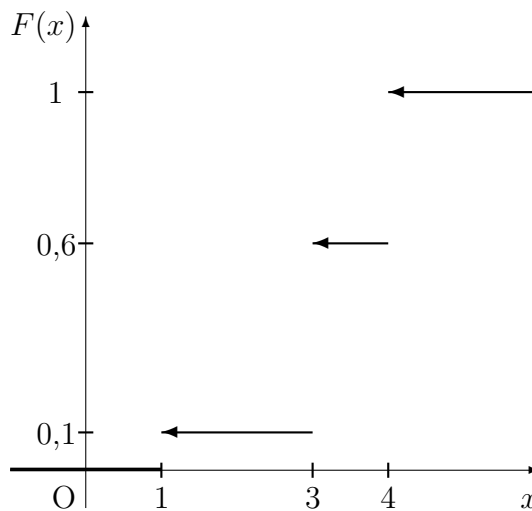
- Nếu $x > 4$, biến cố $(X < x)$ sẽ xảy ra khi $(X = 1)$, hoặc khi $(X = 3)$, hoặc khi $(X = 4)$, do đó

$$F(x) = \mathbb{P}(X = 1) + \mathbb{P}(X = 3) + \mathbb{P}(X = 4) = 0,1 + 0,5 + 0,4 = 1.$$

Vậy hàm phân phối xác suất của X là

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

Đồ thị hàm $F(x)$ như sau



Hình 2.1 – Đồ thị hàm $F(x)$.

Tính chất của hàm phân phối xác suất

1. Hàm phân phối xác suất luôn nhận giá trị trong đoạn $[0; 1]$:

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

2. Hàm phân phối xác suất của đại lượng ngẫu nhiên rời rạc là hàm bậc thang với số điểm gián đoạn bằng số giá trị có thể có của X .

Hàm phân phối xác suất của đại lượng ngẫu nhiên liên tục là hàm liên tục và khả vi hầu khắp nơi (số các điểm không khả vi là đếm được).

3. Hàm phân phối xác suất là hàm không giảm, tức là với $x_1 < x_2$ thì $F(x_1) \leq F(x_2)$.
4. Xác suất để đại lượng ngẫu nhiên X nhận giá trị trong khoảng $[a; b)$ bằng hiệu số giá trị của hàm phân phối xác suất ở hai đầu khoảng đó

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

Đặc biệt, nếu X là đại lượng ngẫu nhiên liên tục thì

$$\begin{aligned} \mathbb{P}(a < X < b) &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b) \\ &= F(b) - F(a). \end{aligned}$$

5. Nếu đại lượng ngẫu nhiên X chỉ nhận giá trị trong khoảng $(a; b)$ thì

$$\begin{aligned} F(x) &= 0 \quad \text{với } \forall x \leq a; \\ F(x) &= 1 \quad \text{với } \forall x > b. \end{aligned}$$

Hàm phân phối có giới hạn tại vô cực như sau

$$\lim_{x \rightarrow -\infty} F(x) = 0; \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

6. Điều kiện để hàm số $F(x)$ là hàm phân phối xác suất của đại lượng ngẫu nhiên liên tục là

$$\left\{ \begin{array}{l} \bullet F(x) \text{ là hàm liên tục;} \\ \bullet F(x) \text{ là hàm không giảm;} \\ \bullet \lim_{x \rightarrow -\infty} F(x) = 0; \quad \lim_{x \rightarrow +\infty} F(x) = 1. \end{array} \right.$$

Ví dụ 2.9. Đại lượng ngẫu nhiên liên tục X có hàm phân phối xác suất như sau

$$F(x) = \begin{cases} 0 & \text{với } x \leq -1; \\ ax + b & \text{với } -1 < x \leq \frac{1}{3}; \\ 1 & \text{với } x > \frac{1}{3}. \end{cases}$$

- a) Tìm các hệ số a, b ;
- b) Tìm xác suất để X nhận giá trị trong khoảng $[0, \frac{1}{3})$;

- c) Tìm xác suất để trong 3 phép thử độc lập, có đúng một lần X nhận giá trị trong khoảng $(0, 1)$.

Lời giải: a) Vì X là đại lượng ngẫu nhiên liên tục nên $F(x)$ là hàm liên tục, do đó nó phải liên tục tại -1 và $\frac{1}{3}$. Vì vậy,

$$\begin{cases} -a + b = 0 \\ \frac{1}{3}a + b = 1 \end{cases} \Leftrightarrow \begin{cases} a = \frac{3}{4}; \\ b = \frac{3}{4}. \end{cases}$$

Khi đó, $F(x)$ thỏa mãn các điều kiện của hàm phân phối xác suất.

- b) Theo tính chất của hàm phân phối, xác suất để X nhận giá trị trong khoảng $[0, \frac{1}{3})$ là

$$\mathbb{P}(0 \leq X < \frac{1}{3}) = F(\frac{1}{3}) - F(0) = \left(\frac{3}{4} \cdot \frac{1}{3} + \frac{3}{4}\right) - \left(\frac{3}{4} \cdot 0 + \frac{3}{4}\right) = \frac{1}{4}.$$

- c) Bài toán tuân theo lược đồ Bernoulli với $n = 3$ phép thử độc lập; xác suất để trong mỗi lần thử X nhận giá trị trong khoảng $(0, 1)$ là

$$\mathbb{P}(0 < X < 1) = F(1) - F(0) = 1 - \left(\frac{3}{4} \cdot 0 + \frac{3}{4}\right) = \frac{1}{4}.$$

Vậy xác suất cần tìm bằng

$$C_3^1 p^1 q^2 = 3 \cdot \frac{1}{4} \cdot \left(\frac{3}{4}\right)^2 = \frac{27}{64}.$$

2.2.3 Hàm mật độ xác suất

Định nghĩa 2.3. “Hàm mật độ xác suất của đại lượng ngẫu nhiên liên tục X , kí hiệu $f(x)$, là đạo hàm bậc nhất của hàm phân phối xác suất của đại lượng ngẫu nhiên đó:

$$f(x) = [F(x)]'.”$$

Ý nghĩa của hàm mật độ xác suất

Hàm mật độ xác suất của đại lượng ngẫu nhiên X tại điểm x cho biết mức độ tập trung xác suất tại điểm đó.

Các tính chất của hàm mật độ xác suất

1. Hàm mật độ xác suất luôn không âm:

$$f(x) \geq 0, \quad \text{với } \forall x.$$

Về mặt hình học, đồ thị của hàm số $f(x)$ không nằm thấp hơn trục Ox .

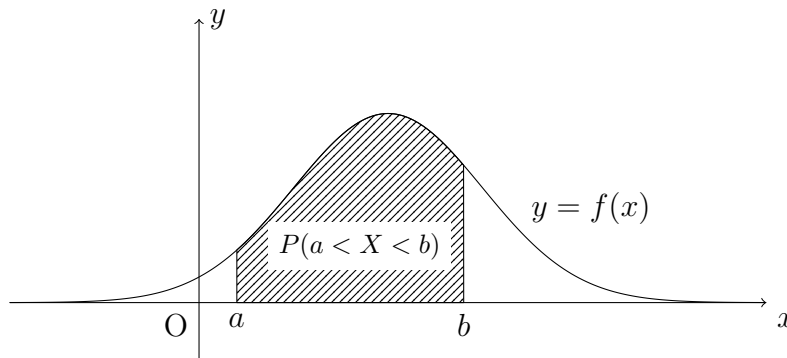
2. Xác suất để đại lượng ngẫu nhiên liên tục X nhận giá trị trong khoảng $[a, b)$ bằng

$$\mathbb{P}(a \leq X < b) = \int_a^b f(x) dx.$$

Vì X là đại lượng ngẫu nhiên liên tục nên từ tính chất trên và kết hợp với nhận xét cuối phần 2.1.2, ta có

$$\begin{aligned}\mathbb{P}(a < X < b) &= \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a \leq X \leq b) \\ &= \int_a^b f(x)dx.\end{aligned}$$

Về mặt hình học, xác suất để đại lượng ngẫu nhiên liên tục X nhận giá trị trong khoảng (a, b) bằng diện tích hình thang cong giới hạn bởi trục Ox , đường cong $f(x)$ và các đường thẳng $x = a$; $x = b$.

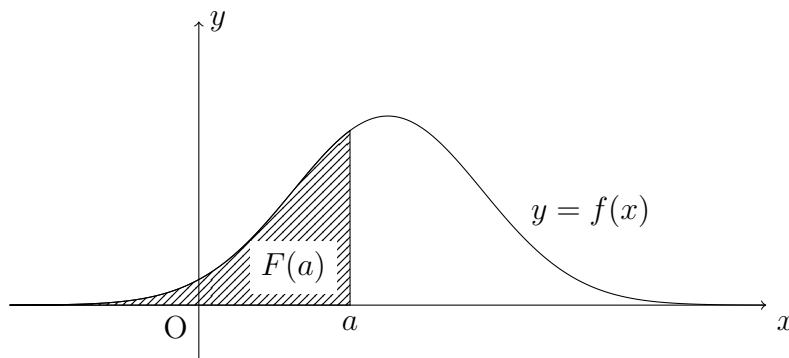


Hình 2.2 – Liên hệ giữa hàm mật độ và xác suất.

3. Cho trước hàm mật độ thì hàm phân phối xác suất của đại lượng ngẫu nhiên liên tục được tính bởi công thức sau

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

Về mặt hình học, giá trị của hàm phân phối xác suất $F(x)$ tại điểm a bằng diện tích của hình thang cong suy rộng giới hạn bởi trục Ox , đường cong $f(x)$ và phía bên trái đường thẳng $x = a$.

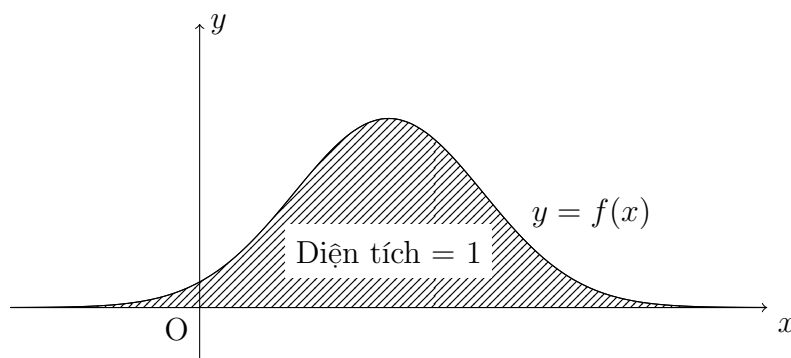


Hình 2.3 – Liên hệ giữa hàm mật độ và hàm phân phối.

4. Tích phân suy rộng trong khoảng $(-\infty, +\infty)$ của hàm mật độ xác suất luôn bằng 1

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

Về mặt hình học, điều đó có nghĩa là diện tích hình giới hạn bởi đường cong $f(x)$ và trục Ox bằng 1.



Hình 2.4 – Hàm mật độ và 1 đơn vị xác suất.

5. Điều kiện để một hàm số $f(x)$ có thể là hàm mật độ xác suất của một đại lượng ngẫu nhiên liên tục:

$$\begin{cases} f(x) \geq 0 & \forall x; \\ \int_{-\infty}^{+\infty} f(x)dx = 1. \end{cases}$$

Ví dụ 2.10. Thời gian xếp hàng chờ mua hàng của khách là đại lượng ngẫu nhiên liên tục với hàm phân phối xác suất như sau (đơn vị: phút)

$$F(x) = \begin{cases} 0 & \text{với } x \leq 0; \\ 2x^3 - 3x^2 + 2x & \text{với } 0 < x \leq 1; \\ 1 & \text{với } x > 1. \end{cases}$$

Hàm mật độ xác suất của đại lượng ngẫu nhiên này là⁴

$$f(x) = [F(x)]' = \begin{cases} 6x^2 - 6x + 2 & \text{với } x \in (0; 1); \\ 0 & \text{với } x \notin (0; 1). \end{cases}$$

Ví dụ 2.11. Đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} a \cos x & \text{với } x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]; \\ 0 & \text{với } x \notin \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]. \end{cases}$$

- Tìm hệ số a ;
- Tìm xác suất để đại lượng ngẫu nhiên X nhận giá trị trong khoảng $(0, \frac{\pi}{4})$;
- Tính xác suất trong 3 lần thử độc lập, có đúng 1 lần X nhận giá trị lớn hơn $\frac{\pi}{6}$;
- Tìm hàm phân phối $F(x)$.

⁴Theo định nghĩa đạo hàm thì không tồn tại $F'(0)$ và $F'(1)$. Tuy nhiên, theo ý nghĩa của hàm mật độ thì có thể coi mật độ xác suất tại $x = 0$ và $x = 1$ bằng 0, tức là $f(0) = f(1) = 0$.

Lời giải: a) Điều kiện để $f(x)$ là hàm mật độ là

$$\begin{cases} f(x) \geq 0 & \forall x; \\ \int_{-\infty}^{+\infty} f(x)dx = 1. \end{cases}$$

Ta có

$$1 = \int_{-\infty}^{+\infty} f(x)dx = \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} a \cos x dx = a \cdot \sin x \Big|_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} = 2a.$$

Từ đó, $a = \frac{1}{2}$ thỏa mãn các điều kiện trên.

b) Theo tính chất của hàm mật độ xác suất

$$\mathbb{P}(0 < X < \frac{\pi}{4}) = \int_0^{\frac{\pi}{4}} f(x)dx = \int_0^{\frac{\pi}{4}} a \cos x dx = \frac{1}{2} \sin x \Big|_0^{\frac{\pi}{4}} = \frac{\sqrt{2}}{4}.$$

c) Xác suất để X nhận giá trị lớn hơn $\frac{\pi}{6}$ là

$$\mathbb{P}(X > \frac{\pi}{6}) = \int_{\frac{\pi}{6}}^{+\infty} f(x)dx = \int_{\frac{\pi}{6}}^{\frac{\pi}{2}} a \cos x dx = \frac{1}{2} \sin x \Big|_{\frac{\pi}{6}}^{\frac{\pi}{2}} = \frac{1}{4}.$$

Khi đó, theo công thức Bernoulli, xác suất để trong 3 phép thử độc lập có 1 lần X nhận giá trị lớn hơn $\frac{\pi}{6}$ là

$$P_3(1) = C_3^1 \left(\frac{1}{4}\right)^1 \left(\frac{3}{4}\right)^2 = \frac{27}{64}.$$

d) Để tìm hàm phân phối xác suất ta sử dụng tính chất $F(x) = \int_{-\infty}^x f(t)dt$.

$$\bullet x < -\frac{\pi}{2}: F(x) = \int_{-\infty}^x 0dt = 0.$$

$$\bullet -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}: F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^{-\frac{\pi}{2}} 0dt + \int_{-\frac{\pi}{2}}^x \frac{1}{2} \cos t dt = \frac{1}{2}(\sin x + 1).$$

$$\bullet x > \frac{\pi}{2}: F(x) = \int_{-\infty}^x f(t)dt = \int_{-\infty}^{-\frac{\pi}{2}} 0dt + \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2} \cos t dt + \int_{\frac{\pi}{2}}^x 0dt = 1.$$

Vậy hàm phân phối xác suất của X là

$$F(x) = \begin{cases} 0 & \text{với } x < -\frac{\pi}{2}; \\ \frac{1}{2}(\sin x + 1) & \text{với } -\frac{\pi}{2} \leq x \leq \frac{\pi}{2}; \\ 1 & \text{với } x > \frac{\pi}{2}. \end{cases}$$

2.3 Các tham số đặc trưng của đại lượng ngẫu nhiên

Trong thực tế, các nghiên cứu đã chỉ ra rằng, việc khảo sát các giá trị của đại lượng ngẫu nhiên dựa vào các phương pháp được nêu trong mục 2.2 không hiệu quả bằng việc khảo sát các tham số đặc trưng của đại lượng ngẫu nhiên đó. Các tham số đặc trưng của đại lượng ngẫu nhiên có một ý nghĩa rất lớn trong đời sống thực tế. Mục này sẽ trình bày định nghĩa, tính chất, ý nghĩa và cách tính cụ thể giá trị của các tham số đặc trưng của đại lượng ngẫu nhiên rời rạc và liên tục.

2.3.1 Kỳ vọng toán

Định nghĩa 2.4. “Kỳ vọng toán của đại lượng ngẫu nhiên X , kí hiệu⁵ $\mathbb{E}(X)$, là một số được xác định như sau

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

X	x_1	x_2	$\dots$	x_i	$\dots$	x_n
$\mathbb{P}$	p_1	p_2	$\dots$	p_i	$\dots$	p_n

thì kỳ vọng toán $\mathbb{E}(X)$ được xác định bằng biểu thức

$$\mathbb{E}(X) = \sum_{i=1}^n x_i p_i.$$

- Nếu X là đại lượng ngẫu nhiên liên tục với hàm mật độ xác suất $f(x)$ thì kỳ vọng toán $\mathbb{E}(X)$ được xác định bằng biểu thức

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} x f(x) dx."$$

Nhận xét. Nếu X là đại lượng ngẫu nhiên rời rạc có tập hợp các giá trị có thể có là tập hợp vô hạn đếm được $\{x_1, x_2, \dots, x_n, \dots\}$ thì kỳ vọng của X được xác định bởi

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i p_i.$$

Ví dụ 2.12. Cho X là đại lượng ngẫu nhiên rời rạc có bảng phân phối xác suất

X	1	3	4
$\mathbb{P}$	0,1	0,5	0,4

Khi đó, kỳ vọng toán của đại lượng ngẫu nhiên X bằng

$$\mathbb{E}(X) = 1 \cdot 0,1 + 3 \cdot 0,5 + 4 \cdot 0,4 = 3,2.$$

⁵Kỳ vọng toán của đại lượng ngẫu nhiên X còn được kí hiệu bởi $M(X)$.

Ví dụ 2.13. Cho X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} \frac{3}{4}(x^2 + 2x) & \text{với } x \in (0, 1); \\ 0 & \text{với } x \notin (0, 1). \end{cases}$$

Khi đó, kỳ vọng toán của đại lượng ngẫu nhiên X bằng

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^0 x \cdot 0dx + \int_0^1 \frac{3}{4}x(x^2 + 2x)dx + \int_1^{+\infty} x \cdot 0dx \\ &= \frac{3}{4} \int_0^1 (x^3 + 2x^2)dx = \frac{3}{4} \left(\frac{x^4}{4} + \frac{2x^3}{3} \right) \Big|_0^1 = \frac{11}{16}. \end{aligned}$$

Ý nghĩa của kỳ vọng toán

Kỳ vọng toán $\mathbb{E}(X)$ mang ý nghĩa là trung bình số học của các giá trị quan sát của đại lượng ngẫu nhiên X . Nó phản ánh giá trị trung tâm của phân phối xác suất. Khái niệm kỳ vọng được áp dụng rộng rãi trong nhiều lĩnh vực. Chẳng hạn, trong kinh doanh và quản lí, kỳ vọng toán được ứng dụng dưới dạng lợi nhuận kỳ vọng hay doanh số kỳ vọng.

Tính chất của kỳ vọng toán

1. Kỳ vọng toán của hằng số bằng chính hằng số đó. Như vậy, nếu C là hằng số thì

$$\mathbb{E}(C) = C.$$

2. Kỳ vọng toán của tích giữa một hằng số và đại lượng ngẫu nhiên bằng tích giữa hằng số đó và kỳ vọng toán của đại lượng ngẫu nhiên ấy

$$\mathbb{E}(CX) = C\mathbb{E}(X).$$

3. Kỳ vọng toán của tổng các đại lượng ngẫu nhiên bằng tổng các kỳ vọng toán thành phần

$$\mathbb{E}(X_1 + X_2 + \dots + X_n) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \dots + \mathbb{E}(X_n).$$

4. Kỳ vọng toán của tích các đại lượng ngẫu nhiên độc lập bằng tích các kỳ vọng toán thành phần

$$\mathbb{E}(X_1 \cdot X_2 \cdot \dots \cdot X_n) = \mathbb{E}(X_1) \cdot \mathbb{E}(X_2) \cdot \dots \cdot \mathbb{E}(X_n).$$

5. Cho X là đại lượng ngẫu nhiên và $g(t)$ là hàm số liên tục có tập xác định chứa tất cả các giá trị có thể có của X . Khi đó, đại lượng ngẫu nhiên $g(X)$ có kỳ vọng toán $\mathbb{E}[g(X)]$ được xác định như sau

- Nếu X là đại lượng ngẫu nhiên rời rạc có phân phối xác suất $p_i = \mathbb{P}(X = x_i), i = \overline{1, n}$ thì

$$\mathbb{E}[g(X)] = \sum_{i=1}^n g(x_i)p_i.$$

- Nếu X là đại lượng ngẫu nhiên liên tục có hàm mật độ $f(x)$ thì

$$\mathbb{E}[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x)dx.$$

2.3.2 Phương sai

Ngoài việc khảo sát kỳ vọng toán để biết giá trị trung bình hoặc giá trị kỳ vọng xảy ra của một đại lượng ngẫu nhiên, ta còn cần xét thêm mức độ sai lệch của các giá trị có thể có của đại lượng ngẫu nhiên đó so với giá trị trung bình của nó. Vì vậy, ta cần khảo sát thêm một tham số đặc trưng nữa của đại lượng ngẫu nhiên đó là phương sai.

Định nghĩa 2.5. “*Phương sai* của đại lượng ngẫu nhiên X , kí hiệu⁶ $\text{var}(X)$, là kỳ vọng toán của bình phương sai lệch của đại lượng ngẫu nhiên so với kỳ vọng toán của nó.

$$\text{var}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2.”$$

Cách tính phương sai

Theo định nghĩa phương sai và các tính chất của kỳ vọng toán, ta có

$$\begin{aligned}\text{var}(X) &= \mathbb{E}[X - \mathbb{E}(X)]^2 = \mathbb{E}\{X^2 - 2X.\mathbb{E}(X) + [\mathbb{E}(X)]^2\} \\ &= \mathbb{E}(X^2) + \mathbb{E}[-2X.\mathbb{E}(X)] + \mathbb{E}[\mathbb{E}(X)]^2 \\ &= \mathbb{E}(X^2) - 2\mathbb{E}(X).\mathbb{E}(X) + [\mathbb{E}(X)]^2 \\ &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.\end{aligned}$$

Ví dụ 2.14. Cho X là đại lượng ngẫu nhiên rời rạc có bảng phân phối xác suất

X	1	3	4
$\mathbb{P}$	0,1	0,5	0,4

Tìm phương sai $\text{var}(X)$.

Lời giải: Kỳ vọng toán của đại lượng ngẫu nhiên X bằng

$$\mathbb{E}(X) = 1. 0,1 + 3. 0,5 + 4. 0,4 = 3,2.$$

Kỳ vọng toán $\mathbb{E}(X^2)$ bằng

$$\mathbb{E}(X^2) = 1^2. 0,1 + 3^2. 0,5 + 4^2. 0,4 = 11.$$

Phương sai của đại lượng ngẫu nhiên X bằng

$$\text{var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 11 - (3,2)^2 = 0,76.$$

Ví dụ 2.15. Cho X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} \frac{3}{4}(x^2 + 2x) & \text{với } x \in (0, 1); \\ 0 & \text{với } x \notin (0, 1). \end{cases}$$

Tìm phương sai $\text{var}(X)$.

⁶Phương sai của đại lượng ngẫu nhiên X còn được kí hiệu bởi $D(X)$ hoặc σ_X^2 .

Lời giải: Kỳ vọng toán của đại lượng ngẫu nhiên X bằng

$$\begin{aligned}\mathbb{E}(X) &= \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^0 x \cdot 0 dx + \int_0^1 \frac{3}{4} x(x^2 + 2x) dx + \int_1^{+\infty} x \cdot 0 dx \\ &= \frac{3}{4} \int_0^1 (x^3 + 2x^2) dx = \frac{3}{4} \left(\frac{x^4}{4} + \frac{2x^3}{3} \right) \Big|_0^1 = \frac{11}{16}.\end{aligned}$$

Kỳ vọng toán $\mathbb{E}(X^2)$ bằng

$$\begin{aligned}\mathbb{E}(X^2) &= \int_{-\infty}^{+\infty} x^2 f(x) dx = \int_{-\infty}^0 x^2 \cdot 0 dx + \int_0^1 \frac{3}{4} x^2(x^2 + 2x) dx + \int_1^{+\infty} x^2 \cdot 0 dx \\ &= \frac{3}{4} \int_0^1 (x^4 + 2x^3) dx = \frac{3}{4} \left(\frac{x^5}{5} + \frac{2x^4}{4} \right) \Big|_0^1 = \frac{21}{40}.\end{aligned}$$

Phương sai của đại lượng ngẫu nhiên X bằng

$$\text{var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{21}{40} - \left(\frac{11}{16} \right)^2 = \frac{67}{1280} \approx 0,052.$$

Ý nghĩa của phương sai

$\text{Var}(X)$ chính là trung bình của các bình phương sai lệch giữa giá trị có thể có của X và $\mathbb{E}(X)$. Do đó phương sai phản ánh mức độ phân tán của X xung quanh giá trị trung bình $\mathbb{E}(X)$. Trong kĩ thuật, phương sai được sử dụng để xác định mức độ phân tán của kích thước các chi tiết gia công hay sai số của thiết bị. Trong quản lí kinh doanh phương sai được sử dụng để đánh giá mức độ rủi ro của các quyết định.

Tính chất của phương sai

1. Phương sai của đại lượng ngẫu nhiên là một giá trị xác định không âm; phương sai của một hằng số bằng không

$$\begin{aligned}\text{var}(X) &\geq 0, \quad \forall X, \\ \text{var}(C) &= 0, \quad C \text{ là hằng số.}\end{aligned}$$

2. Phương sai của tích giữa một hằng số và một đại lượng ngẫu nhiên bằng tích giữa bình phương hằng số đó và phương sai của đại lượng ngẫu nhiên ấy

$$\text{var}(CX) = C^2 \text{var}(X).$$

3. Phương sai của tổng hai đại lượng ngẫu nhiên độc lập bằng tổng các phương sai thành phần

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

Từ tính chất trên, suy ra

$$\begin{aligned}\text{var}(X + C) &= \text{var}(X), \quad C \text{ là hằng số;} \\ \text{var}(X - Y) &= \text{var}(X) + \text{var}(Y), \quad X, Y \text{ độc lập.}\end{aligned}$$

Nếu $X_1, X_2, \dots, X_n$ là các đại lượng ngẫu nhiên độc lập thì

$$\text{var}(X_1 + X_2 + \dots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \dots + \text{var}(X_n).$$

2.3.3 Độ lệch tiêu chuẩn

Định nghĩa 2.6. “Độ lệch tiêu chuẩn (hoặc độ lệch chuẩn) của đại lượng ngẫu nhiên X , kí hiệu σ_X , là căn bậc hai của phương sai

$$\sigma_X = \sqrt{\text{var}(X)}.”$$

Ý nghĩa của độ lệch tiêu chuẩn

Đơn vị đo của độ lệch tiêu chuẩn trùng với đơn vị đo của đại lượng ngẫu nhiên. Vì vậy độ lệch chuẩn được sử dụng để đánh giá mức độ phân tán của đại lượng ngẫu nhiên theo đơn vị đo của nó.

2.3.4 Mốt

Định nghĩa 2.7. “Mốt X , kí hiệu $\text{mod}(X)$, là giá trị của đại lượng ngẫu nhiên X mà tại đó hàm mật độ $f(x)$ của X đạt giá trị cực đại (nếu X là đại lượng ngẫu nhiên liên tục). Đối với đại lượng ngẫu nhiên rời rạc, $\text{mod}(X)$ là giá trị có thể có của X tương ứng với xác suất lớn nhất.”

Ví dụ 2.16. Đại lượng ngẫu nhiên X có bảng phân phối xác suất

X	-2	-1	0	3
$\mathbb{P}$	0,1	0,7	0,05	0,15

có $\text{mod}(X) = -1$.

Ví dụ 2.17. Đại lượng ngẫu nhiên X có hàm mật độ xác suất

$$f(x) = \begin{cases} 0; & x \leq 0; \\ \frac{x}{2}e^{-\frac{x^2}{4}}; & x > 0. \end{cases}$$

Xác định $\text{mod}(X)$.

Lời giải:

Với $x > 0$, ta có

$$f'(x) = \frac{1}{2}e^{-\frac{x^2}{4}} - \frac{x^2}{4}e^{-\frac{x^2}{4}}.$$

$$f'(x) = 0 \Leftrightarrow \frac{1}{4}e^{-\frac{x^2}{4}}(2 - x^2) = 0 \Leftrightarrow x = \pm\sqrt{2}.$$

Kết hợp với điều kiện $x > 0$, suy ra $x = \sqrt{2}$.

Bảng biến thiên của $f(x)$ là

x	$-\infty$	$-\sqrt{2}$	0	$\sqrt{2}$	$+\infty$
$f'(x)$				$+$	0
				$-$	
$f(x)$	CD				
	0	$\longrightarrow$	0	$\nearrow$	0
				$\searrow$	0

Vậy, $\text{mod}(X) = \sqrt{2}$.

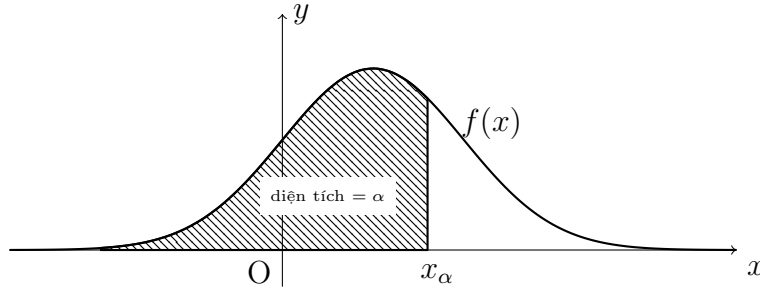
2.3.5 Phân vị

Định nghĩa 2.8. “Phân vị mức α của đại lượng ngẫu nhiên X là giá trị x_α thoả mãn

$$\mathbb{P}(X < x_\alpha) = \alpha.$$

Đặc biệt, với $\alpha = 0,5$ thì phân vị $x_{0,5}$ còn được gọi là *trung vị* của đại lượng ngẫu nhiên và được kí hiệu là $\text{med}(X)$.”

Như vậy, nếu đại lượng ngẫu nhiên X có hàm phân phối xác suất $F(x)$ thì phân vị mức α là nghiệm x của phương trình $F(x) = \alpha$.



Hình 2.5 – Mô tả hình học về phân vị mức α

Ví dụ 2.18. Đại lượng ngẫu nhiên X có hàm phân phối xác suất

$$F(x) = \begin{cases} 0 & \text{với } x \leq 0; \\ x^2 & \text{với } 0 < x \leq 1; \\ 1 & \text{với } x > 1. \end{cases}$$

Tìm $\text{med}(X)$.

Lời giải: Xét phương trình $F(x) = \frac{1}{2}$.

$$F(x) = \frac{1}{2} \Leftrightarrow \begin{cases} 0 < x \leq 1 \\ x^2 = \frac{1}{2} \end{cases} \Leftrightarrow \begin{cases} 0 < x \leq 1 \\ x = \pm \frac{1}{\sqrt{2}} \end{cases} \Leftrightarrow x = \frac{1}{\sqrt{2}}.$$

$$\text{Vậy, } \text{med}(X) = \frac{1}{\sqrt{2}}.$$

Ví dụ 2.19. Đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất

$$f(x) = \begin{cases} e^{-x} & \text{với } x > 0; \\ 0 & \text{với } x \leq 0. \end{cases}$$

Tính các giá trị phân vị $x_{0,95}, x_{0,5}, x_{0,05}$.

Lời giải: Từ hàm mật độ, ta suy ra hàm phân phối xác suất của X như sau

$$F(x) = \begin{cases} 1 - e^{-x} & \text{với } x > 0; \\ 0 & \text{với } x \leq 0. \end{cases}$$

Với $\alpha \in (0; 1)$ xét phương trình

$$F(x) = \alpha \Leftrightarrow 1 - e^{-x} = \alpha \Leftrightarrow e^{-x} = 1 - \alpha \Leftrightarrow x = -\ln(1 - \alpha).$$

Do đó, ta có các phân vị

- $x_{0,95} = -\ln 0,05 \approx 2,996$.
- $x_{0,5} = -\ln 0,5 \approx 0,693$. Lưu ý rằng đây chính là $\text{med}(X)$.
- $x_{0,05} = -\ln 0,95 \approx 0,051$.

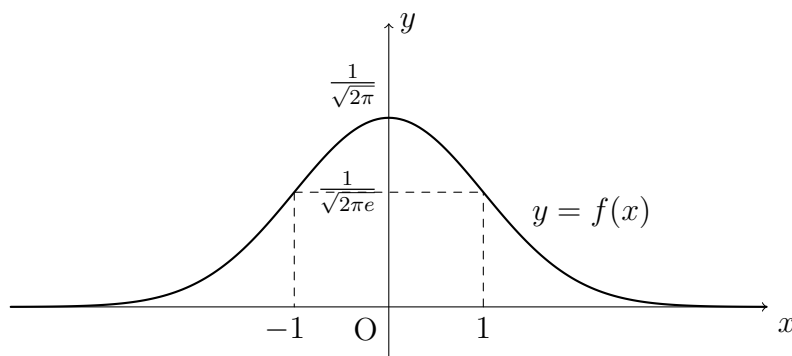
2.4 Một số quy luật phân phối xác suất thông dụng

2.4.1 Quy luật phân phối chuẩn hóa

Định nghĩa 2.9. “Đại lượng ngẫu nhiên liên tục U nhận các giá trị trong khoảng $(-\infty, +\infty)$ được gọi là tuân theo quy luật phân phối chuẩn hoá $\mathcal{N}(0, 1)$ nếu U có hàm mật độ xác suất là

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.”$$

a. Hàm mật độ chuẩn hóa



Hình 2.6 – Hàm mật độ của quy luật phân phối chuẩn hóa.

Hàm mật độ $f(x)$ của quy luật phân phối chuẩn hoá có đồ thị đối xứng qua trục tung Oy và nhận trục hoành Ox làm tiệm cận ngang.

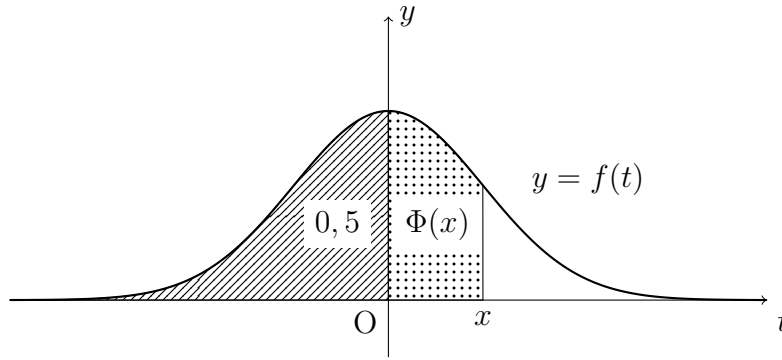
Dễ thấy: • Hàm số $f(x)$ là hàm số chẵn: $f(-x) = f(x)$;
• $\lim_{x \rightarrow \pm\infty} f(x) = 0$.

Các giá trị của hàm $f(x)$ được tính sẵn trong phần phụ lục E.1.

b. Hàm phân phối chuẩn hóa và hàm tích phân Laplace

Từ hàm mật độ $f(x)$ ta có hàm phân phối xác suất $F(x)$ của đại lượng ngẫu nhiên phân phối chuẩn hoá là $F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.

Xét hàm số (tích phân Laplace) $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$.



Hình 2.7 – Hàm tích phân Laplace.

- Dễ thấy:
- $F(x) = 0,5 + \Phi(x)$;
 - Hàm số $\Phi(x)$ là hàm số lẻ: $\Phi(-x) = -\Phi(x)$;
 - $\lim_{x \rightarrow +\infty} \Phi(x) = 0,5$.

Bảng tính sẵn giá trị hàm Laplace $\Phi(x)$ được trình bày trong phần phụ lục E.2.

c. Các tham số đặc trưng của quy luật phân phối chuẩn hóa

Cho U là đại lượng ngẫu nhiên phân phối theo quy luật chuẩn hóa $\mathcal{N}(0, 1)$. Khi đó

1. Kỳ vọng toán: $\mathbb{E}(U) = 0$.
2. Phương sai: $\text{var}(U) = 1$.
Độ lệch tiêu chuẩn: $\sigma_U = 1$.
3. Mốt của U : $\text{mod}(U) = 0$.

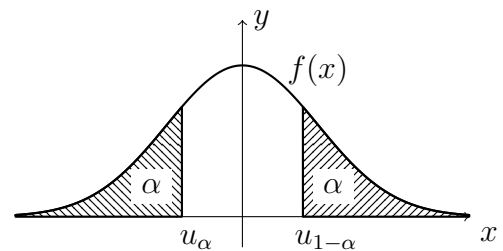
d. Phân vị chuẩn

Phân vị chuẩn mức α ($0 \leq \alpha \leq 1$), kí hiệu u_α , là phân vị mức α của đại lượng ngẫu nhiên U có phân phối chuẩn hoá. Như vậy, u_α thoả mãn điều kiện $\mathbb{P}(U < u_\alpha) = \alpha$.

Bảng tính sẵn giá trị u_α tương ứng với mức α cho trước được trình bày trong phần phụ lục E.3.

Do tính chất đối xứng của hàm mật độ phân phối chuẩn hoá nên phân vị chuẩn có tính chất

$$u_\alpha = -u_{1-\alpha}.$$

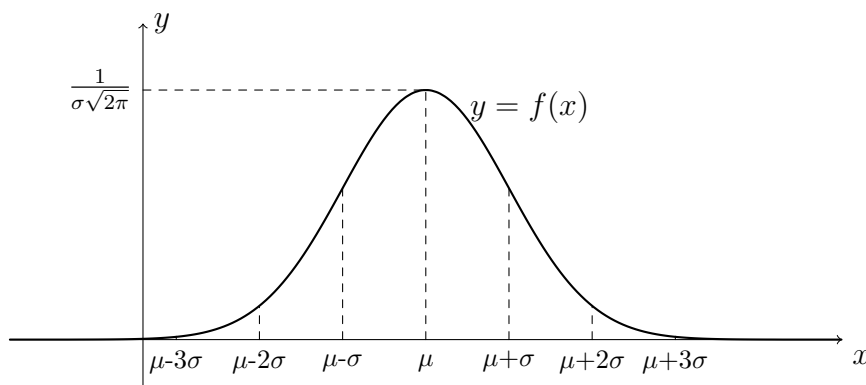


Hình 2.8 – Phân vị chuẩn

2.4.2 Quy luật phân phối chuẩn

Định nghĩa 2.10. “Đại lượng ngẫu nhiên liên tục X nhận giá trị trong khoảng $(-\infty, +\infty)$ được gọi là phân phối theo quy luật chuẩn với hai tham số μ và σ , kí hiệu là $\mathcal{N}(\mu, \sigma^2)$, nếu hàm mật độ xác suất của nó có dạng

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

**Hình 2.9** – Hàm mật độ của quy luật phân phối chuẩn

Nhận xét. Quy luật phân phối chuẩn hóa là trường hợp đặc biệt của quy luật phân phối chuẩn với hai tham số $\mu = 0$ và $\sigma = 1$.

Các tính chất của quy luật phân phối chuẩn

Cho X là đại lượng ngẫu nhiên phân phối theo quy luật chuẩn $\mathcal{N}(\mu, \sigma^2)$. Khi đó

1. Kỳ vọng toán: $\mathbb{E}(X) = \mu$.
2. Phương sai: $\text{var}(X) = \sigma^2$.
Độ lệch tiêu chuẩn: $\sigma_X = \sigma$.
3. Mốt của X : $\text{mod}(X) = \mu$.
4. Đặt $U = \frac{X - \mu}{\sigma}$ thì đại lượng ngẫu nhiên U phân phối theo quy luật chuẩn hóa $\mathcal{N}(0, 1)$.

Các công thức xác suất liên quan đến quy luật phân phối chuẩn

Cho đại lượng ngẫu nhiên liên tục X phân phối theo quy luật chuẩn $\mathcal{N}(\mu, \sigma^2)$. Khi đó

1. Xác suất để đại lượng ngẫu nhiên X nhận giá trị trong khoảng (a, b)

$$\mathbb{P}(a < X < b) = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right),$$

trong đó $\Phi(x)$ là hàm tích phân Laplace có giá trị cho ở phụ lục E.2.

2. Xác suất để đại lượng ngẫu nhiên X sai lệch so với kỳ vọng toán của nó một số nhỏ hơn ε là

$$\mathbb{P}(|X - \mu| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

3. Quy tắc "3 - σ " như sau

Nếu đại lượng ngẫu nhiên X phân phối chuẩn thì hầu như chắc chắn nó sẽ nhận giá trị trong khoảng $(\mu - 3\sigma; \mu + 3\sigma)$.

Mặt khác, nếu quy luật phân phối xác suất của đại lượng ngẫu nhiên được nghiên cứu chưa biết, song nó thỏa mãn điều kiện của quy tắc "3 - σ " thì có thể xem nó là đại lượng ngẫu nhiên phân phối chuẩn.

Ví dụ 2.20. Kích thước của các chi tiết do một máy sản xuất ra là đại lượng ngẫu nhiên phân phối chuẩn với kích thước trung bình $\mu = 5$ cm và độ lệch tiêu chuẩn $\sigma = 0,9$ cm.

- Tính tỉ lệ chi tiết có kích thước nằm trong khoảng từ 4 cm đến 7 cm.
- Tìm xác suất để lấy ngẫu nhiên 3 chi tiết có đúng 1 chi tiết có kích thước sai lệch so với trung bình không quá 1 cm.
- Hầu hết các chi tiết có kích thước như thế nào?

Lời giải: Gọi X cm là kích thước chi tiết do máy sản xuất. Theo giả thiết X phân phối chuẩn với hai tham số $\mu = 5$ cm và $\sigma = 0,9$ cm.

a) Ta có

$$\begin{aligned}\mathbb{P}(4 < X < 7) &= \Phi\left(\frac{7-5}{0,9}\right) - \Phi\left(\frac{4-5}{0,9}\right) \\ &= \Phi(2,22) - \Phi(-1,11) = \Phi(2,22) + \Phi(1,11) \\ &= 0,48679 + 0,36650 = 0,85329.\end{aligned}$$

Vậy tỉ lệ chi tiết có kích thước nằm trong khoảng từ 4 cm đến 7 cm là 85,329%.

b) Xác suất để lấy ngẫu nhiên 1 chi tiết thì thấy nó có kích thước sai lệch so với trung bình không quá 1 cm là

$$\begin{aligned}\mathbb{P}(|X - \mu| \leq 1) &= 2\Phi\left(\frac{1}{0,9}\right) = 2\Phi(1,11) \\ &= 2 \cdot 0,36650 = 0,733.\end{aligned}$$

Theo công thức Bernoulli, xác suất để lấy ngẫu nhiên 3 chi tiết có đúng 1 chi tiết có kích thước sai lệch so với trung bình không quá 1 cm bằng

$$P_3(1) = C_3^1(0,733)^1 \cdot (1 - 0,733)^2 \approx 0,157.$$

c) Theo quy tắc "3 - σ ", hầu hết chi tiết có kích thước trong khoảng

$$(\mu - 3\sigma; \mu + 3\sigma) = (5 - 3 \cdot 0,9; 5 + 3 \cdot 0,9) = (2,3; 7,7) \text{ cm}.$$

Ứng dụng của quy luật chuẩn

Quy luật phân phối chuẩn là quy luật phân phối xác suất được áp dụng rất rộng rãi trong thực tế. Nhà toán học Nga Lyapunov đã chứng minh được rằng, nếu đại lượng ngẫu nhiên X là tổng của một số lớn các đại lượng ngẫu nhiên độc lập và giá trị của mỗi đại lượng chỉ chiếm vị trí rất nhỏ trong tổng đó thì X sẽ có phân phối xấp xỉ chuẩn.

Trong thực tế, ta gặp chính các đại lượng ngẫu nhiên như vậy. Chẳng hạn, kích thước của các chi tiết do các nhà máy sản xuất ra sẽ phân phối chuẩn nếu quá trình sản xuất diễn ra bình thường; năng suất của cùng một loại cây trồng tại các thửa ruộng khác nhau cũng phân phối chuẩn; năng suất lao động của các công nhân có cùng tay nghề và làm cùng một công việc như nhau cũng phân phối chuẩn; một số chỉ tiêu về sinh lí của những người cùng giới (chiều cao, cân nặng, vòng ngực, chiều dài cánh tay ...) cũng phân phối theo quy luật chuẩn. Sự nhận biết này cho phép lập kế hoạch sản xuất phù hợp năng suất lao động và nhu cầu người tiêu dùng...

2.4.3 Quy luật không - một

Định nghĩa 2.11. “Đại lượng ngẫu nhiên rời rạc X nhận một trong hai giá trị có thể có $X = 0; 1$ được gọi là phân phối theo quy luật không - một với tham số p ($0 \leq p \leq 1$) nếu các xác suất được tính bằng công thức sau

$$\mathbb{P}(X = x) = p^x \cdot q^{1-x} \quad \text{với } x = 0; 1, \quad \text{trong đó } q = 1 - p.$$

Quy luật không - một với tham số p được kí hiệu là $\mathcal{A}(p)$.”

Ví dụ 2.21. Trong kho đang lưu giữ một số lượng lớn sản phẩm, trong đó có 60% sản phẩm loại I. Lấy ngẫu nhiên 1 sản phẩm và gọi $X =$ “Số sản phẩm loại I được lấy ra”. Khi đó X là đại lượng ngẫu nhiên tuân theo quy luật phân phối không - một $\mathcal{A}(p = 0,6)$.

Các tham số đặc trưng của quy luật không - một

Gọi X là đại lượng ngẫu nhiên tuân theo quy luật không - một $\mathcal{A}(p)$.

1. Bảng phân phối xác suất của X là

X	0	1
$\mathbb{P}$	q	p
2. Kỳ vọng toán: $\mathbb{E}(X) = p$.
3. Phương sai: $\text{var}(X) = pq$;
Độ lệch tiêu chuẩn: $\sigma_x = \sqrt{pq}$.

2.4.4 Quy luật nhị thức

Định nghĩa 2.12. “Đại lượng ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có $X = 0, 1, 2, \dots, n$ được gọi là phân phối theo quy luật nhị thức với các tham số là n và p , ($0 \leq p \leq 1$) nếu các xác suất tương ứng của nó được tính bởi công thức

$$\mathbb{P}(X = k) = C_n^k p^k q^{n-k}, \quad k = \overline{0, n}, \quad q = 1 - p.$$

Quy luật nhị thức được kí hiệu là $\mathcal{B}(n, p)$.”

Nhận xét. Giả sử bài toán tuân theo lược đồ Bernoulli

- Thực hiện dãy n phép thử độc lập;
- Trong mỗi phép thử chỉ có hai khả năng, đó là biến cố A xảy ra hoặc biến cố A không xảy ra;
- Xác suất xảy ra biến cố A trong mỗi phép thử đều bằng p ; xác suất xảy ra biến cố $\bar{A}$ trong mỗi phép thử là $q = 1 - p$.

Khi đó, nếu gọi X là số lần xuất hiện biến cố A trong dãy n phép thử trên thì X là đại lượng ngẫu nhiên phân phối theo quy luật nhị thức $\mathcal{B}(n, p)$.

Các tính chất của quy luật nhị thức

Giả sử X là đại lượng ngẫu nhiên phân phối theo quy luật nhị thức $\mathcal{B}(n, p)$.

1. Bảng phân phối xác suất của X như sau

X	0	1	2	...	k	...	n
$\mathbb{P}$	q^n	$C_n^1 p q^{n-1}$	$C_n^2 p^2 q^{n-2}$	...	$C_n^k p^k q^{n-k}$	...	p^n

2. Kỳ vọng toán: $\mathbb{E}(X) = np$.

3. Phương sai: $\text{var}(X) = npq$.

Độ lệch tiêu chuẩn: $\sigma_X = \sqrt{npq}$.

4. Giá trị của X tương ứng với xác suất lớn nhất là $\text{mod}(X)$ sao cho

$$np - q \leq \text{mod}(X) \leq np + p.$$

Liên hệ giữa quy luật nhị thức và quy luật chuẩn

Giả sử X là đại lượng ngẫu nhiên phân phối nhị thức $\mathcal{B}(n, p)$. Trong thực tế, nếu n khá lớn thì việc tính toán theo công thức Bernoulli sẽ gặp khó khăn, lúc đó có thể dùng quy luật chuẩn để thay thế cho quy luật nhị thức nếu thoả mãn một trong hai điều kiện sau⁷

$$\begin{cases} np \geq 5 \\ nq \geq 5 \end{cases} \quad \text{hoặc} \quad \begin{cases} n > 5 \\ \frac{|p - q|}{\sqrt{npq}} < 0,3 \end{cases} \quad (2.1)$$

Lúc đó, đại lượng ngẫu nhiên X phân phối nhị thức $\mathcal{B}(n, p)$ có thể coi như phân phối xấp xỉ chuẩn $\mathcal{N}(\mu, \sigma^2)$ với kỳ vọng toán $\mu = np$ và phương sai $\sigma^2 = npq$.

Công thức xấp xỉ Laplace

Cho X là đại lượng ngẫu nhiên phân phối nhị thức $\mathcal{B}(n, p)$ với các tham số n và p thoả mãn điều kiện (2.1). Khi đó, các công thức xấp xỉ Laplace như sau

1. Xác suất tại một điểm

$$\mathbb{P}(X = k) = C_n^k p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} f\left(\frac{k - np}{\sqrt{npq}}\right),$$

trong đó $f(x)$ là hàm mật độ của quy luật phân phối chuẩn hoá, có giá trị cho ở phụ lục E.1.

2. Xác suất để đại lượng ngẫu nhiên X nhận giá trị trong nửa khoảng $[a; b)$ là⁸

$$\mathbb{P}(a \leq X < b) \approx \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right),$$

trong đó $\Phi(x)$ là hàm tích phân Laplace, có giá trị cho ở phụ lục E.2.

⁷Cơ sở của các điều kiện này có thể được tìm thấy ở [6, 11]

⁸Khi điều kiện (2.1) được thoả mãn thì có thể coi X là đại lượng ngẫu nhiên liên tục có phân phối chuẩn. Kết hợp với nhận xét cuối mục 2.1.2 thì công thức này cũng được dùng để tính gần đúng xác suất để X nhận giá trị trong $(a; b)$; $(a; b]$ hay $[a; b]$.

3. Xác suất để đại lượng ngẫu nhiên X sai lệch so với kỳ vọng toán của nó một số nhỏ hơn ε là

$$\mathbb{P}(|X - np| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right).$$

Ví dụ 2.22. Xác suất để thắng trong một trò chơi trên máy tự động là 0,2.

- Giả sử một người được chơi 12 lượt, tính số lần thắng cuộc có khả năng xảy ra nhiều nhất và xác suất để xảy ra điều đó.
- Muốn số lần thắng cuộc trung bình không nhỏ hơn 12 thì người chơi cần phải chơi ít nhất bao nhiêu lần?

Lời giải: Gọi X là số lần thắng cuộc của người chơi.

a) Người đó chơi $n = 12$ lần, khi đó X tuân theo quy luật nhị thức $\mathcal{B}(n = 12; p = 0,2)$. Số lần thắng cuộc có khả năng xảy ra nhiều nhất là $\text{mod}(X)$ được xác định như sau

$$\begin{aligned} np - q &\leq \text{mod}(X) \leq np + p \\ \Leftrightarrow 12 \cdot 0,2 - 0,8 &\leq \text{mod}(X) \leq 12 \cdot 0,2 + 0,2 \\ \Leftrightarrow 1,6 &\leq \text{mod}(X) \leq 2,6 \quad \Rightarrow \text{mod}(X) = 2. \end{aligned}$$

Xác suất để người đó thắng cuộc 2 lần là

$$P(X = 2) = C_{12}^2(0,2)^2 \cdot (0,8)^{10} \approx 0,2835.$$

b) Giả sử người đó được chơi n lượt, khi đó X tuân theo quy luật nhị thức $\mathcal{B}(n; p = 0,2)$. Theo giả thiết, ta có

$$\mathbb{E}(X) \geq 12 \Leftrightarrow np \geq 12 \Leftrightarrow n \geq \frac{12}{0,2} = 60.$$

Vậy muốn số lần thắng cuộc trung bình không nhỏ hơn 12, người đó phải chơi ít nhất 60 lần.

Ví dụ 2.23. Xác suất để sản phẩm sau khi sản xuất không được kiểm tra chất lượng bằng 0,2. Giả sử trong kho của nhà máy có 400 sản phẩm vừa sản xuất ra.

- Tính xác suất để có 80 sản phẩm không được kiểm tra chất lượng.
- Tính xác suất để có từ 70 đến 100 sản phẩm không được kiểm tra chất lượng.
- Với xác suất 0,9545 hãy tính xem số lượng sản phẩm không được kiểm tra chất lượng nằm trong khoảng đối xứng nào xung quanh giá trị trung bình?

Lời giải: Gọi X là số sản phẩm không được kiểm tra chất lượng trong số 400 sản phẩm. Khi đó, X tuân theo quy luật phân phối nhị thức $\mathcal{B}(n = 400; p = 0,2)$. Tuy nhiên, vì n khá lớn $n = 400 > 5$ và

$$\frac{|p - q|}{\sqrt{npq}} = \frac{|0,2 - 0,8|}{\sqrt{400 \cdot 0,2 \cdot 0,8}} = 0,075 < 0,3$$

nên có thể coi X tuân theo quy luật phân phối chuẩn $\mathcal{N}(\mu = np = 80; \sigma^2 = npq = 64)$.

a) Xác suất để có 80 sản phẩm không được kiểm tra chất lượng là

$$\mathbb{P}(X = 80) \approx \frac{1}{\sqrt{64}} \cdot f\left(\frac{80 - 80}{\sqrt{64}}\right) = \frac{1}{8} f(0) = \frac{1}{8} 0,39894 \approx 0,04987.$$

b) Xác suất để có từ 70 đến 100 sản phẩm không được kiểm tra chất lượng là

$$\begin{aligned}\mathbb{P}(70 \leq X \leq 100) &\approx \Phi\left(\frac{100-80}{\sqrt{64}}\right) - \Phi\left(\frac{70-80}{\sqrt{64}}\right) = \Phi(2,5) - \Phi(-1,25) \\ &= \Phi(2,5) + \Phi(1,25) = 0,49379 + 0,39435 = 0,88814.\end{aligned}$$

c) Giả sử khoảng đối xứng cần tìm có bán kính là ε . Ta phải tìm ε sao cho $\mathbb{P}(|X - np| < \varepsilon) = 0,9545$. Ta có

$$\mathbb{P}(|X - np| < \varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right) = 2\Phi\left(\frac{\varepsilon}{8}\right).$$

Do đó

$$2\Phi\left(\frac{\varepsilon}{8}\right) = 0,9545 \Leftrightarrow \Phi\left(\frac{\varepsilon}{8}\right) = 0,47725 \Leftrightarrow \frac{\varepsilon}{8} = 2,0 \Leftrightarrow \varepsilon = 16.$$

Vậy số lượng sản phẩm không được kiểm tra chất lượng nằm trong khoảng

$$(np - \varepsilon ; np + \varepsilon) = (80 - 16 ; 80 + 16) = (64 ; 96).$$

2.4.5 Quy luật Poisson

Định nghĩa 2.13. “Đại lượng ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có $X = 0, 1, \dots$ được gọi là phân phối theo quy luật Poisson với tham số λ ($\lambda > 0$) nếu các xác suất tương ứng của nó được tính bởi công thức sau

$$\mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots$$

Quy luật Poisson được kí hiệu là $\mathcal{P}(\lambda)$.”

Các tính chất của quy luật Poisson

Giả sử X là đại lượng ngẫu nhiên phân phối theo quy luật Poisson $\mathcal{P}(\lambda)$. Khi đó

1. Bảng phân phối xác suất của X có dạng

X	0	1	2	...	k	...
$\mathbb{P}$	$e^{-\lambda}$	$\frac{\lambda^1}{1!} e^{-\lambda}$	$\frac{\lambda^2}{2!} e^{-\lambda}$	...	$\frac{\lambda^k}{k!} e^{-\lambda}$	...

2. Kỳ vọng toán: $\mathbb{E}(X) = \lambda$.

3. Phương sai: $\text{var}(X) = \lambda$.

Độ lệch tiêu chuẩn: $\sigma_X = \sqrt{\lambda}$.

4. Giá trị của X tương ứng với xác suất lớn nhất là $\text{mod}(X)$ sao cho:

$$\lambda - 1 \leq \text{mod}(X) \leq \lambda.$$

Trong tình huống nào chúng ta gặp phân phối Poisson?

Xét một biến cố E xuất hiện ở những thời điểm ngẫu nhiên. Giả thiết rằng số lần xuất hiện của E trong khoảng thời gian $(t_1; t_2)$ không ảnh hưởng đến xác suất xuất hiện của E trong những khoảng thời gian kế tiếp. Thêm vào đó “cường độ” xuất hiện của E là không đổi, nghĩa là số lần E xuất hiện trung bình trong khoảng thời gian $(t_1; t_2)$ tỉ lệ với độ dài khoảng đó. Gọi X là số lần xuất hiện của E trong khoảng thời gian $(t_1; t_2)$. Người ta chứng minh được rằng X có phân phối Poisson với tham số $\lambda = c(t_2 - t_1)$, trong đó c là hằng số và là số lần xuất hiện trung bình của E trong một đơn vị thời gian. Hằng số c được gọi là cường độ xuất hiện của E .

Ví dụ 2.24. Ở một tổng đài bưu điện, các cuộc điện thoại gọi đến xuất hiện ngẫu nhiên, độc lập với nhau và tốc độ trung bình 2 cuộc gọi đến trên phút. Tìm xác suất để

- Có 5 cuộc gọi đến trong 2 phút;
- Có ít nhất 1 cuộc điện thoại trong 10 giây.

Lời giải: a) Số cuộc điện thoại gọi đến trong khoảng thời gian 2 phút là đại lượng ngẫu nhiên X có phân phối Poisson với tham số $\lambda = 4$. Vậy

$$\mathbb{P}(X = 5) = e^{-4} \frac{4^5}{5!} \approx 0,156.$$

b) Số cuộc điện thoại gọi đến trong khoảng thời gian 10 giây là đại lượng ngẫu nhiên $Y \sim \mathcal{P}(\frac{1}{3})$. Vậy

$$\mathbb{P}(Y \geq 1) = 1 - \mathbb{P}(Y = 0) = 1 - e^{-\frac{1}{3}} \approx 0,2835.$$

Liên hệ giữa quy luật Poisson với quy luật nhị thức và quy luật chuẩn

Giả sử X là đại lượng ngẫu nhiên phân phối nhị thức $\mathcal{B}(n, p)$. Trong trường hợp n khá lớn và p khá nhỏ ($p < 0,1$), phân phối nhị thức được xấp xỉ bởi phân phối Poisson. Nói cách khác, trong trường hợp số phép thử n rất lớn và xác suất p rất nhỏ, các xác suất p_k của công thức Bernoulli có thể thay thế bằng công thức xấp xỉ Poisson với $\lambda = np$.

Mặt khác, đối với quy luật Poisson thì quá trình hội tụ của nó về quy luật chuẩn sẽ diễn ra khi λ lớn hơn 20. Vì vậy, nếu đại lượng ngẫu nhiên X phân phối theo quy luật Poisson song $\lambda > 20$ thì có thể xem X phân phối xấp xỉ chuẩn với $\mu = \lambda$ và $\sigma^2 = \lambda$.

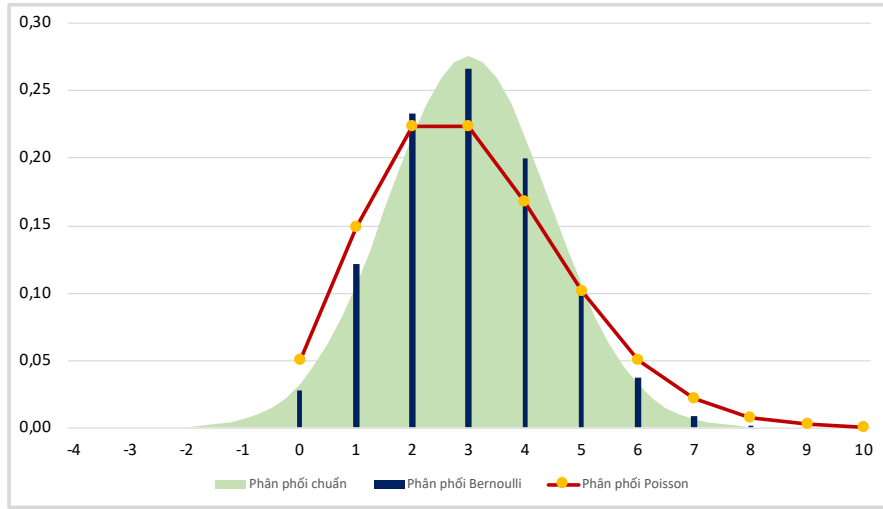
Ví dụ 2.25. Một máy dệt có 5000 ống sợi, xác suất để trong một phút một ống sợi bị đứt bằng 0,0002. Tìm xác suất để trong một phút có không quá 2 ống sợi bị đứt.

Lời giải: Gọi X là số ống sợi bị đứt trong một phút. Dễ thấy X là đại lượng ngẫu nhiên phân phối nhị thức $\mathcal{B}(n = 5000, p = 0,0002)$. Tuy nhiên, vì $n = 5000$ rất lớn, $p = 0,0002$ khá nhỏ và tích $np = 5000 \cdot 0,0002 = 1$ nên có thể coi X phân phối theo quy luật Poisson $\mathcal{P}(\lambda = 1)$.

Xác suất để trong một phút có không quá 2 ống sợi bị đứt là

$$\begin{aligned} \mathbb{P}(X \leq 2) &= p_0 + p_1 + p_2 = e^{-\lambda} \frac{\lambda^0}{0!} + e^{-\lambda} \frac{\lambda^1}{1!} + e^{-\lambda} \frac{\lambda^2}{2!} \\ &= e^{-1} \left(1 + \frac{1^1}{1!} + \frac{1^2}{2!} \right) = \frac{5}{2e} \approx 0,9197. \end{aligned}$$

Ví dụ 2.26. Biểu đồ sau đây thể hiện phân phối xác suất của quy luật phân phối nhị thức $\mathcal{B}(n = 10, p = 0,3)$; quy luật phân phối Poisson $\mathcal{P}(\lambda = np = 3)$ và quy luật phân phối chuẩn $\mathcal{N}(\mu = np = 3; \sigma^2 = npq = 2,1)$.



Hình 2.10 – Liên hệ giữa các quy luật nhị thức, Poisson và quy luật chuẩn

2.4.6 Quy luật siêu bội

Định nghĩa 2.14. “Đại lượng ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có $X = 0, 1, 2, \dots$ được gọi là phân phối theo quy luật siêu bội với các tham số N, M, n (N, M, n là các số tự nhiên thỏa mãn $n \leq N; M \leq N$) nếu với mỗi $k \in \mathbb{N}$ thỏa mãn

$$\max(0; n - N + M) \leq k \leq \min(M; n),$$

xác suất tương ứng được tính bởi công thức

$$\mathbb{P}(X = k) = p_k = \frac{C_M^k \cdot C_{N-M}^{n-k}}{C_N^n}.$$

Quy luật siêu bội được kí hiệu là $\mathcal{M}(N, M, n)$.”

Ví dụ 2.27. Trong 10 sản phẩm có 6 sản phẩm loại A. Lấy ngẫu nhiên 4 sản phẩm. Gọi X = “Số sản phẩm loại A được lấy ra”. Khi đó X là đại lượng ngẫu nhiên rời rạc tuân theo quy luật phân phối siêu bội $\mathcal{M}(N = 10; M = 6; n = 4)$.

Các tính chất của quy luật siêu bội

Giả sử X là đại lượng ngẫu nhiên phân phối theo quy luật siêu bội $\mathcal{M}(N, M, n)$. Khi đó, với $N \geq 2$ ta có

1. Kỳ vọng toán: $\mathbb{E}(X) = n \cdot \frac{M}{N}$.
2. Phương sai: $\text{var}(X) = n \cdot \frac{M}{N} \cdot \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}$.
3. Khi số phép thử n là rất bé so với N thì phân phối siêu bội không khác nhiều so với phân phối nhị thức với

$$p_k = C_n^k \left(\frac{M}{N}\right)^k \left(1 - \frac{M}{N}\right)^{n-k}.$$

2.4.7 Quy luật khi - bình phương

Định nghĩa 2.15. “Đại lượng ngẫu nhiên liên tục X được gọi là tuân theo quy luật phân phối khi - bình phương với n bậc tự do, kí hiệu $\chi^2(n)$, nếu hàm mật độ xác suất của nó có dạng sau

$$f(x) = \begin{cases} 0 & \text{với } x \leq 0; \\ \frac{1}{2^{\frac{n}{2}} \cdot \Gamma(\frac{n}{2})} e^{-\frac{x}{2}} \cdot x^{\frac{n}{2}-1} & \text{với } x > 0. \end{cases}$$

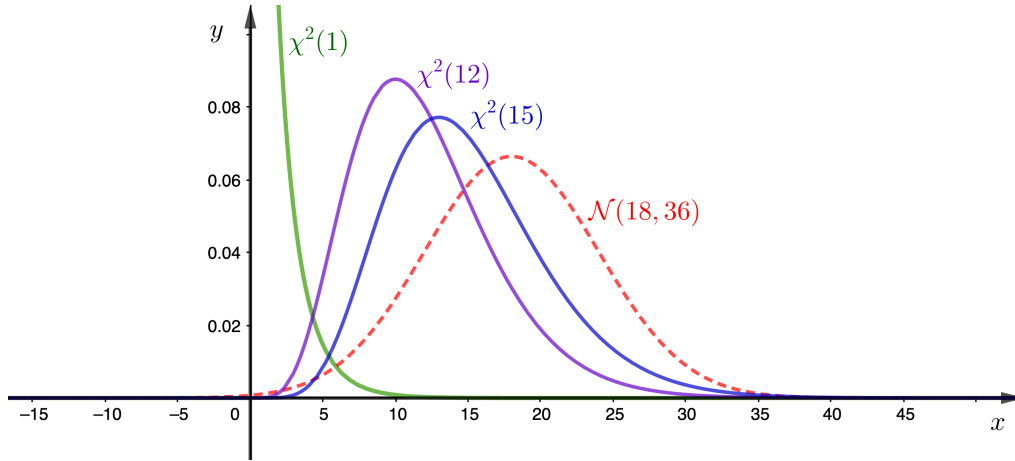
Trong đó

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt \text{ là hàm Gamma;}$$

$$\Gamma(1) = 1; \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi};$$

$$\Gamma(x) = (x-1)\Gamma(x-1). \text{ Đặc biệt với } n \in \mathbb{N}^* \text{ thì } \Gamma(n) = (n-1)!.$$

Sau đây là một số đồ thị hàm mật độ của quy luật khi - bình phương.



Hình 2.11 – Hàm mật độ của quy luật khi - bình phương

Các tính chất của quy luật khi - bình phương

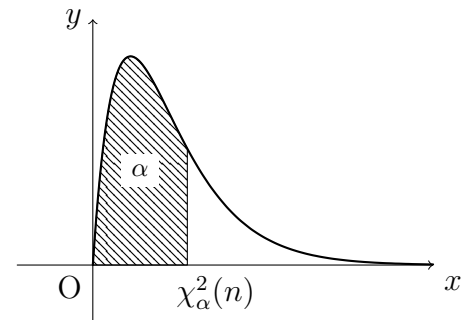
Giả sử X là đại lượng ngẫu nhiên phân phối theo quy luật khi - bình phương $\chi^2(n)$. Khi đó

1. Kỳ vọng toán: $\mathbb{E}(X) = n$.
2. Phương sai: $\text{var}(X) = 2n$.

Phân vị khi - bình phương

Phân vị khi - bình phương mức α , kí hiệu là $\chi_{\alpha}^2(n)$, là giá trị của đại lượng ngẫu nhiên X tuân theo quy luật phân phối khi - bình phương với n bậc tự do thoả mãn điều kiện

$$\mathbb{P}[X < \chi_{\alpha}^2(n)] = \alpha.$$



Hình 2.12 – Phân vị khi - bình phương

Liên hệ giữa quy luật khi - bình phương với quy luật chuẩn

Nếu có n đại lượng ngẫu nhiên độc lập X_i , ($i = \overline{1, n}$), cùng phân phối theo quy luật chuẩn hóa $\mathcal{N}(0, 1)$ thì tổng bình phương các đại lượng đó

$$X = \sum_{i=1}^n X_i^2$$

là đại lượng ngẫu nhiên phân phối theo quy luật khi - bình phương với n bậc tự do; còn nếu n đại lượng đó liên hệ với nhau bằng một hệ thức tuyến tính thì tổng bình phương X tuân theo quy luật khi - bình phương với $n - 1$ bậc tự do.

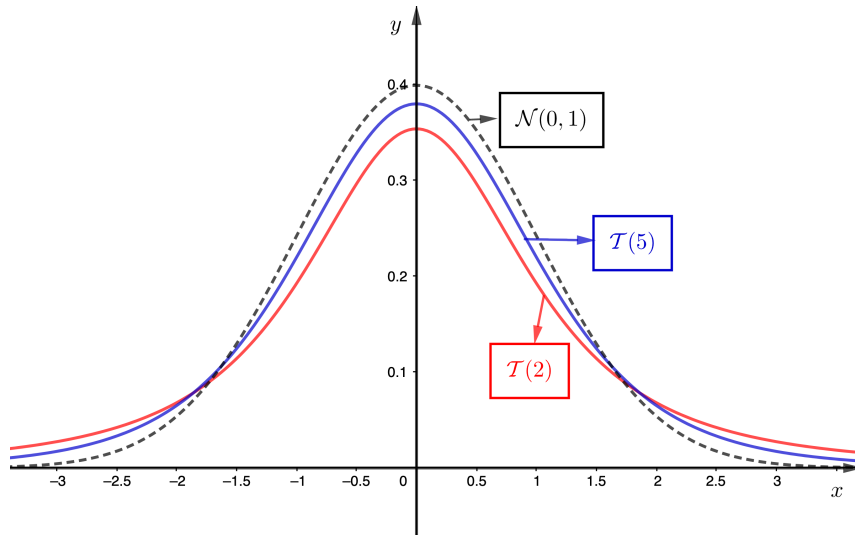
Khi số bậc tự do n tăng lên, quy luật khi - bình phương xấp xỉ với quy luật chuẩn.

2.4.8 Quy luật Student

Định nghĩa 2.16. “Đại lượng ngẫu nhiên liên tục X được gọi là tuân theo quy luật phân phối Student với n bậc tự do với $n \geq 2$, kí hiệu $\mathcal{T}(n)$, nếu hàm mật độ xác suất của nó có dạng sau

$$f(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{\pi(n-1)} \cdot \Gamma\left(\frac{n-1}{2}\right)} \left[1 + \frac{x^2}{n-1}\right]^{-\frac{n}{2}},$$

trong đó $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ là hàm Gamma.”



Hình 2.13 – Hàm mật độ của quy luật Student

Các tính chất của quy luật Student

Giả sử X là đại lượng ngẫu nhiên phân phối theo quy luật Student $\mathcal{T}(n)$. Khi đó

1. Kỳ vọng toán: $\mathbb{E}(X) = 0$.

2. Phương sai: $\text{var}(X) = \frac{n}{n-2}, \quad n > 2.$

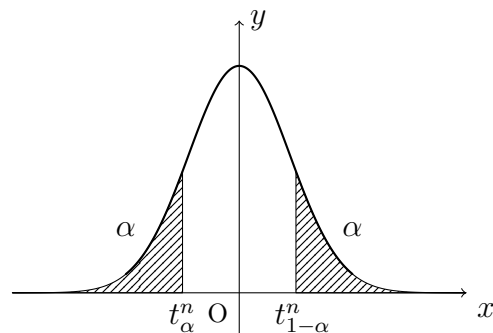
Phân vị Student

Phân vị Student mức α , kí hiệu là t_α^n , là giá trị của đại lượng ngẫu nhiên X phân phối theo quy luật Student với n bậc tự do thoả mãn điều kiện

$$\mathbb{P}(X < t_\alpha^n) = \alpha.$$

Giá trị của phân vị Student được cho trong phần phụ lục E.4. Lưu ý rằng,

$$t_\alpha^n = -t_{1-\alpha}^n.$$



Hình 2.14 – Phân vị Student

Liên hệ giữa quy luật Student, quy luật khi - bình phương và quy luật chuẩn

Giả sử U là đại lượng ngẫu nhiên phân phối chuẩn hoá $\mathcal{N}(0, 1)$ và V là đại lượng ngẫu nhiên độc lập với U , phân phối theo quy luật khi - bình phương với n bậc tự do $\chi^2(n)$. Khi đó, đại lượng ngẫu nhiên

$$X = \frac{U}{\sqrt{\frac{V}{n}}},$$

sẽ phân phối theo quy luật Student với n bậc tự do $\mathcal{T}(n)$.

Khi số bậc tự do n tăng lên, phân phối Student sẽ hội tụ nhanh về phân phối chuẩn hoá $\mathcal{N}(0, 1)$. Do đó, nếu n khá lớn ($n \geq 30$), có thể dùng phân phối chuẩn hoá thay thế cho phân phối Student.

Bài tập chương 2

Đại lượng ngẫu nhiên rời rạc

Bài 2.1. Một tàu khu trục của công ty vận tải biển Isalco có 3 hệ thống máy trục vớt hoạt động độc lập. Khi ra biển hoạt động thì trong một ngày khả năng các hệ thống đó hỏng tương ứng là 0, 4; 0, 2 và 0, 3.

- Tìm quy luật phân phối xác suất của số hệ thống bị hỏng X ;
- Thiết lập hàm phân phối xác suất của X ;
- Tính xác suất trong một ngày có không quá 2 hệ thống hỏng;
- Tìm $\text{mod}(X)$.

Bài 2.2. Một xí nghiệp có hai ô tô vận tải hoạt động độc lập. Xác suất trong ngày làm việc các ô tô bị hỏng tương ứng bằng 0, 1 và 0, 2. Gọi X là số ô tô bị hỏng trong ngày làm việc.

- Tìm quy luật phân phối xác suất của X ;
- Thiết lập hàm phân phối xác suất của X và vẽ đồ thị của nó;
- Tính $\mathbb{E}(X)$, $\text{var}(X)$ và $\text{mod}(X)$.

Bài 2.3. Bảng phân phối xác suất của một biến ngẫu nhiên X như sau

X	-5	2	3	4
$\mathbb{P}$	0, 4	0, 3	0, 1	0, 2

- Tính $\mathbb{E}(X)$, $\text{var}(X)$ và σ_X ;
- Tìm $\text{mod}(X)$.

Bài 2.4. Cho X_1 và X_2 là hai đại lượng ngẫu nhiên độc lập có bảng phân phối xác suất như sau

X_1	2	3	5
$\mathbb{P}$	0, 3	0, 5	0, 2

X_2	1	4
$\mathbb{P}$	0, 2	0, 8

- Tính $\mathbb{E}(X_1)$, $\mathbb{E}(X_2)$, $\text{var}(X_1)$, $\text{var}(X_2)$;
- Tính $\mathbb{E}(X_1 + X_2)$, $\text{var}(X_1 + X_2)$;
- Tính $\mathbb{E}(2X_1 - 3X_2)$, $\text{var}(2X_1 - 3X_2)$, $E(X_1^2 - X_2^2)$.

Bài 2.5. Một nhà đầu tư đang cân nhắc giữa việc đầu tư vào hai dự án A và B . Khả năng thu hồi vốn sau 2 năm (tính bằng %) của hai dự án là các biến ngẫu nhiên có bảng phân phối xác suất như sau

X_A	65	67	68	69	70	71	73
$\mathbb{P}$	0, 04	0, 12	0, 16	0, 28	0, 24	0, 08	0, 08

X_B	66	68	69	70	71
$\mathbb{P}$	0, 12	0, 28	0, 32	0, 2	0, 08

Hãy so sánh khả năng và mức độ rủi ro thu hồi vốn của 2 dự án đầu tư trên.

Bài 2.6. Theo thống kê, xác suất để một người Việt Nam 25 tuổi sẽ sống thêm trên một năm là 0,992 và xác suất để người đó chết trong vòng một năm tới là 0,008. Một chương trình bảo hiểm đề nghị người đó bảo hiểm sinh mạng cho một năm với số tiền chi trả là 1000 đô la Mỹ và tiền đóng bảo hiểm là 10 đô la Mỹ. Hỏi lợi nhuận trung bình của công ty bảo hiểm là bao nhiêu? Có nhận xét gì về độ rủi ro của hình thức kinh doanh trên của công ty bảo hiểm?

Bài 2.7. Một người đi từ nhà đến cơ quan phải qua 3 ngã tư, xác suất để người đó gặp đèn đỏ ở các ngã tư tương ứng là 0,2; 0,4; 0,5. Hỏi thời gian trung bình phải chờ trên đường là bao nhiêu, biết rằng mỗi khi gặp đèn đỏ người ấy phải đợi khoảng 2 phút?

Bài 2.8. Cho X, Y là hai đại lượng ngẫu nhiên độc lập có bảng phân phối xác suất

$$\begin{array}{c|ccc} X & 1 & 0 & 2 \\ \hline \mathbb{P} & x & 0,2 & y \end{array}, \quad \begin{array}{c|cc} Y & -1 & 1 \\ \hline \mathbb{P} & z & t \end{array}.$$

- Tìm x, y, z, t biết $\mathbb{E}(X + Y) = 1,3; \mathbb{E}(X^2 + Y^2) = 3,3$;
- Lập bảng phân phối xác suất của $X^2 - 2X, XY, |X - Y|$;
- Tìm số $p \in (0; 1)$ sao cho đại lượng ngẫu nhiên $pX^2 + (1 - p)Y^3$ có độ lệch tiêu chuẩn nhỏ nhất. Tìm độ lệch tiêu chuẩn nhỏ nhất đó.

Đại lượng ngẫu nhiên liên tục

Bài 2.9. Thời gian xếp hàng chờ mua hàng của khách là đại lượng ngẫu nhiên liên tục với hàm phân phối xác suất như sau (đơn vị là phút)

$$F(x) = \begin{cases} 0 & \text{với } x \leq 0; \\ ax^3 - 3x^2 + 2x & \text{với } 0 < x \leq 1; \\ 1 & \text{với } x > 1. \end{cases}$$

- Tìm hệ số a ;
- Tìm thời gian xếp hàng trung bình;
- Tìm xác suất để trong 3 người xếp hàng thì có không quá 2 người phải chờ quá 0,5 phút.

Bài 2.10. Cho đại lượng ngẫu nhiên liên tục X có hàm phân phối xác suất như sau

$$F(x) = \begin{cases} \frac{1}{2}e^{2x} & \text{với } x < 0; \\ ax^2 + x + b & \text{với } 0 \leq x \leq 1; \\ 1 & \text{với } x > 1. \end{cases}$$

- Tìm hệ số a, b ;
- Tính xác suất để trong 1 phép thử, X nhận giá trị trong khoảng $\left(\frac{1}{3}, \frac{1}{2}\right)$;
- Tính xác suất để trong 4 phép thử có đúng 1 lần X nhận giá trị trong khoảng $\left(\frac{1}{2}, 3\right)$;
- Tìm $\text{med}(X)$, phân vị mức 0,95 của X ;
- Tính kỳ vọng, phương sai và độ lệch chuẩn của X .

Bài 2.11. Cho X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} k(x^2 + 2x) & \text{với } x \in (0, 1); \\ 0 & \text{với } x \notin (0, 1). \end{cases}$$

- Tìm hệ số k ;
- Tính $\mathbb{E}(X)$, $\text{var}(X)$, σ_X ;
- Tính $\mathbb{P}\left(\frac{1}{3} < X < \frac{1}{2}\right)$;
- Tính xác suất để trong 5 lần thử thì có 3 lần giá trị của X thỏa mãn: $|X - \mathbb{E}(X)| \leq 2\sigma_X$.

Bài 2.12. Tuổi thọ (tính theo giờ) của một trò chơi điện tử bấm tay là đại lượng ngẫu nhiên có hàm mật độ xác suất như sau (k là hằng số)

$$f(x) = \begin{cases} 0 & \text{với } x < 0; \\ ke^{-\frac{x}{100}} & \text{với } x \geq 0. \end{cases}$$

- Tính xác suất để tuổi thọ của trò chơi này nằm trong khoảng từ 50 đến 150 giờ;
- Tính xác suất để tuổi thọ của trò chơi này ít hơn 100 giờ.

Bài 2.13. Tuổi thọ (tính theo giờ) của một loại van điện lắp trong một thiết bị là đại lượng ngẫu nhiên có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} 0 & \text{với } x \leq 100; \\ \frac{100}{x^2} & \text{với } x > 100. \end{cases}$$

Tìm xác suất để có 2 trong 5 van điện này bị thay thế trong 150 giờ hoạt động đầu tiên, biết rằng việc hỏng của chúng độc lập nhau.

Bài 2.14. Tuổi thọ của một loại máy là biến ngẫu nhiên liên tục X (năm) có hàm mật độ

$$f(x) = \begin{cases} \frac{k}{x^4} & \text{nếu } x \geq 2; \\ 0 & \text{nếu } x < 2. \end{cases}$$

Một cửa hàng kinh doanh loại máy trên với thời hạn bảo hành 3 năm và chi phí bảo hành 900 nghìn đồng một máy. Vậy cửa hàng phải bán với giá bao nhiêu để chi phí bảo hành trung bình bằng $\frac{1}{3}$ giá bán?

Bài 2.15. Cho X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất như sau

$$f(x) = \begin{cases} ax^2 + bx & \text{với } x \in [1, 2]; \\ 0 & \text{với } x \notin [1, 2]. \end{cases}$$

Biết kỳ vọng của X bằng $\mathbb{E}(X) = 1,375$.

- Tìm giá trị a, b ;
- Tính xác suất để trong 1 phép thử, X nhận giá trị thuộc khoảng $(1, 5; 5, 0)$.

Bài 2.16. Lãi suất cổ phiếu của 2 công ty M và N hoạt động trong 2 lĩnh vực độc lập nhau là các đại lượng ngẫu nhiên liên tục với hàm mật độ xác suất lần lượt là

$$f_M(x) = \begin{cases} \frac{3}{4}x(2-x) & \text{với } x \in [0, 2]; \\ 0 & \text{với } x \notin [0, 2]. \end{cases}$$

$$f_N(x) = \begin{cases} \frac{2}{9}x(3-x) & \text{với } x \in [0, 3]; \\ 0 & \text{với } x \notin [0, 3]. \end{cases}$$

- Anh Chí mua cổ phiếu của 2 công ty trên với tỉ lệ 50% công ty M : 50% công ty N . Tính lãi suất trung bình anh Chí thu được;
- Nếu là nhà đầu tư thì theo bạn nên mua cổ phiếu của cả 2 công ty trên với tỉ lệ bao nhiêu để rủi ro là ít nhất?

Quy luật phân phối nhị thức

Bài 2.17. Một nữ công nhân quản lý 12 máy dệt. Xác suất để mỗi máy trong khoảng thời gian t cần đến sự chăm sóc của nữ công nhân là 0,3.

- Tính xác suất để trong khoảng thời gian t có 4 máy cần đến sự chăm sóc của nữ công nhân;
- Tính xác suất để trong khoảng thời gian t có từ 3 đến 6 máy cần đến sự chăm sóc của nữ công nhân;
- Tìm số máy có khả năng cần đến sự chăm sóc của nữ công nhân nhiều nhất.

Bài 2.18. Việc sản xuất ra các sản phẩm được tiến hành độc lập. Hỏi phải sản xuất mỗi đợt bao nhiêu sản phẩm để trung bình có được 10 sản phẩm đạt tiêu chuẩn biết rằng xác suất được sản phẩm đạt tiêu chuẩn bằng 0,8?

Bài 2.19. Đại lượng ngẫu nhiên rời rạc X tuân theo quy luật phân phối nhị thức với $\mathbb{E}(X) = 3,6$ và $\text{var}(X) = 2,52$.

- Lập bảng phân phối xác suất của X ;
- Tính $\text{mod}(X)$;
- Tính xác suất $\mathbb{P}(6 < X < 9)$;
- Tính xác suất để trong 5 phép thử có ít nhất 1 lần $|X - \mathbb{E}(X)| < \sigma_X$.

Bài 2.20. Tiến hành kiểm tra chất lượng 900 chi tiết, xác suất để một chi tiết đạt tiêu chuẩn là 0,9.

- Tính xác suất để có 810 chi tiết đạt chuẩn;
- Tính xác suất để có từ 800 đến 825 chi tiết đạt chuẩn;
- Hãy tìm với xác suất 0,9544 xem số sản phẩm đạt tiêu chuẩn nằm trong khoảng đối xứng nào xung quanh giá trị trung bình?

Quy luật phân phối chuẩn

Bài 2.21. Một công ty kinh doanh mặt hàng A dự định sẽ áp dụng một trong hai phương án kinh doanh. Giả thiết rằng lợi nhuận thu được từ mỗi phương án đều là biến ngẫu nhiên tuân theo quy luật chuẩn.

- Theo phương án thứ 1, lợi nhuận trung bình thu được là 140 triệu đồng/tháng và độ lệch tiêu chuẩn là 50 triệu đồng/tháng.
- Theo phương án thứ 2, lợi nhuận trung bình thu được là 200 triệu đồng/tháng và độ lệch tiêu chuẩn là 60 triệu đồng/tháng.

Để công ty tồn tại và phát triển thì lợi nhuận thu được từ kinh doanh mặt hàng A phải đạt ít nhất 80 triệu đồng/tháng. Hãy cho biết công ty nên áp dụng phương án nào để kinh doanh mặt hàng A ? Vì sao?

Bài 2.22. Trong hệ thống tỉ giá hối đoái thả nổi, sự biến động của tỉ giá hối đoái chịu sự tác động của rất nhiều nhân tố và có thể xem như đại lượng ngẫu nhiên phân phối chuẩn. Giả sử ở một giai đoạn nào đó tỉ giá của USD với VNĐ có trung bình là 15000 đồng và độ lệch chuẩn là 500 đồng. Tìm xác suất trong một ngày nào đó trong giai đoạn nói trên có

- Tỷ giá sẽ cao hơn 16000 đồng;
- Tỷ giá sẽ thấp hơn 16000 đồng;
- Tỷ giá sẽ nằm trong khoảng từ 14500 đồng đến 16500 đồng.

Bài 2.23. Doanh số hàng tháng của công ty là đại lượng ngẫu nhiên có phân phối chuẩn. Xác suất để đạt được doanh số lớn hơn 200 triệu là 0,9; xác suất để đạt được doanh số nhỏ hơn 500 triệu là 0,8. Tính doanh số trung bình hàng tháng của công ty đó.

Bài 2.24. Lãi suất đầu tư vào một công ty là đại lượng ngẫu nhiên phân phối chuẩn. Biết xác suất để đạt được lãi suất trên 20% một năm là 0,2 và dưới 10% một năm là 0,1. Tìm xác suất để khi đầu tư vào công ty đó sẽ được lãi suất thấp nhất là 14% một năm.

Bài 2.25. Lãi suất (%) đầu tư vào một dự án năm 2021 được coi là đại lượng ngẫu nhiên phân phối chuẩn. Theo đánh giá của ủy ban đầu tư thì lãi suất cao hơn 20% có xác suất là 0,1587 và lãi suất cao hơn 25% có xác suất là 0,0228. Vậy khả năng đầu tư vào dự án mà không bị thua lỗ là bao nhiêu?

Bài 2.26. Lượng điện tiêu thụ hàng tháng của các hộ gia đình ở Hải Phòng là đại lượng ngẫu nhiên phân phối chuẩn với trung bình là 200 kWh và độ lệch chuẩn 40 kWh. Tìm xác suất để chọn ngẫu nhiên một hộ gia đình thì hộ đó

- Có mức tiêu dùng điện hàng tháng trên 250 kWh;
- Có mức tiêu dùng điện hàng tháng dưới 180 kWh.

Bài 2.27. Thời gian hoạt động tốt (không phải sửa chữa) của một loại tivi là đại lượng ngẫu nhiên phân phối chuẩn $\mu = 4300$ giờ và $\sigma = 250$ giờ. Giả thiết mỗi ngày người ta dùng trung bình là 10 giờ và thời gian bảo hành miễn phí là 1 năm (365 ngày).

- Tính tỉ lệ sản phẩm bảo hành;
- Nếu muốn tỉ lệ bảo hành vẫn như trên và có thể nâng thời gian bảo hành thành 2 năm thì phải cải thiện thời gian hoạt động tốt trung bình của sản phẩm là bao nhiêu để nâng cao chất lượng sản phẩm ?

Bài 2.28. Thời gian đi từ nhà đến trường của sinh viên Đông là một đại lượng ngẫu nhiên (đơn vị là phút) có phân phối chuẩn. Biết rằng 65% số ngày Đông đến trường mất hơn 20 phút và 8% số ngày mất hơn 30 phút.

- Tính thời gian đến trường trung bình của Đông;

- b) Giả sử Đông xuất phát từ nhà trước giờ vào học 25 phút, tính xác suất để Đông bị muộn học?

Bài 2.29. Một người cân nhắc việc mua cổ phiếu của 2 công ty A và B hoạt động trong 2 lĩnh vực độc lập nhau. Biết lãi suất cổ phiếu của 2 công ty là đại lượng ngẫu nhiên phân phối chuẩn với các tham số đặc trưng như sau

	Kỳ vọng toán (%)	Độ lệch chuẩn (%)
Công ty A	11	4
Công ty B	10,4	2,6

- a) Nếu người đó muốn đạt được lãi suất tối thiểu là 10% thì nên mua cổ phiếu của công ty nào?
- b) Nếu người đó muốn hạn chế rủi ro bằng cách mua cổ phiếu của cả hai công ty thì nên mua theo tỉ lệ bao nhiêu để mức độ rủi ro về lãi suất là nhỏ nhất?

Bài tập tổng hợp

Bài 2.30. Nhu cầu về một loại thực phẩm ở một khu dân cư có bảng phân phối xác suất như sau

Nhu cầu (kg)	50	60	70	80	90	100
Xác suất	0,15	0,25	0,35	0,12	0,07	0,06

- a) Tính nhu cầu trung bình về loại thực phẩm đó ở khu dân cư nói trên;
- b) Mỗi kg thực phẩm cửa hàng mua vào 45000 đồng, bán ra 60000 đồng. Giả sử cửa hàng nhập 70 kg thực phẩm, hãy lập bảng phân phối xác suất của lượng thực phẩm bán ra. Tính tiền lãi trung bình khi đó;
- c) Nếu bị ế thì cửa hàng giảm giá bán ra chỉ còn 30000 đồng thì sẽ hết hàng. Hỏi cửa hàng nên nhập 70 kg hay 80 kg thì tốt hơn?

Bài 2.31. Một chi tiết do máy tiện sản xuất ra là một đại lượng ngẫu nhiên có phân phối chuẩn $\mathcal{N}(20\text{mm}; 4\text{mm}^2)$. Chi tiết được gọi là hợp quy cách nếu X dao động từ 18 mm đến 22 mm. Cho máy đó sản xuất 100 chi tiết, tính xác suất để có

- a) 50 chi tiết máy hợp quy cách;
- b) không quá 80 chi tiết máy hợp quy cách.

Bài 2.32. Thời gian bảo hành một sản phẩm được quy định là 3 năm. Nếu bán được 1 sản phẩm thì cửa hàng có lãi 150 ngàn song nếu sản phẩm bị hỏng trong thời gian bảo hành thì cửa hàng phải chi phí 500 ngàn cho việc bảo hành. Biết rằng tuổi thọ sản phẩm là đại lượng ngẫu nhiên phân phối chuẩn với tuổi thọ trung bình là 4,2 năm và độ lệch chuẩn là 1,8 năm.

- a) Tìm số tiền lãi mà cửa hàng hy vọng thu được khi bán mỗi sản phẩm;
- b) Nếu muốn số tiền lãi cho mỗi sản phẩm bán ra là 50 ngàn thì phải quy định thời gian bảo hành là bao nhiêu?
- c) Nếu muốn tỉ lệ sản phẩm phải bảo hành không quá 10% thì phải quy định thời gian bảo hành tối đa là bao nhiêu?

Bài 2.33. Gọi X là lượng điện (tính bằng kWh) mà mỗi hộ tiêu thụ hàng tháng. Giả sử X là đại lượng ngẫu nhiên tuân theo quy luật chuẩn với lượng điện tiêu thụ trung bình của các hộ dân là 60 kWh và độ lệch chuẩn là 40 kWh. Định mức điện năng tiêu thụ cho một hộ dân là 70 kWh/tháng. Số tiền điện một hộ dân phải trả trong một tháng là 1000 đồng cho 1 kWh điện tiêu thụ trong định mức, 3000 đồng cho 1 kWh điện tiêu thụ vượt định mức. Gọi Y là số tiền điện (nghìn đồng) phải trả hàng tháng của một hộ dân. Tính xác suất để

- Số tiền điện phải trả hàng tháng của một hộ dân từ 100 đến 130 nghìn đồng;
- Số tiền điện phải trả hàng tháng của một hộ dân trên 70 nghìn đồng;
- Số tiền điện phải trả hàng tháng của một hộ dân từ 40 đến 130 nghìn đồng;
- Nếu thành phố có 300.000 hộ thì ước tính trung bình có bao nhiêu hộ dùng điện vượt định mức?

Bài 2.34. Nhu cầu hàng năm về loại hàng A là đại lượng ngẫu nhiên X có hàm mật độ xác suất như sau (đơn vị: ngàn sản phẩm):

$$f(x) = \begin{cases} k(30 - x) & \text{với } x \in (0; 30); \\ 0 & \text{với } x \notin (0; 30). \end{cases}$$

- Tìm hệ số k ;
- Tính xác suất để nhu cầu về loại hàng đó không vượt quá 12 ngàn sản phẩm trong một năm;
- Tính $\mathbb{E}(X)$, $\text{var}(X)$, σ_X .
- Tính xác suất $\mathbb{P}(|X - \mathbb{E}(X)| \leq 12)$.
- Tính xác suất để trong 5 năm liên tiếp, nhu cầu về loại hàng đó đều không vượt quá 12 ngàn sản phẩm trong một năm.

Bài 2.35. Cho X là đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất như sau:

$$f(x) = \begin{cases} ax^2 + bx & \text{với } x \in (0, 1); \\ 0 & \text{với } x \notin (0, 1). \end{cases}$$

- Biết $\mathbb{E}(X) = \frac{11}{16}$. Tìm các hệ số a, b .
- Tính $\mathbb{P}[(X - \mathbb{E}(X))^2 < \text{var}(X)]$.
- Đặt $Y = X^2 - 2X$. Tính $\mathbb{E}(Y)$, σ_Y .
- Biết khả năng để $X > \varepsilon$ bằng 0,78125. Tính ε .
- Biết khi thử 3 lần thì khả năng để có lần xảy ra $y < X < 1,5$ bằng 0,657. Tìm y .

Bài 2.36. Tuổi thọ một loại sản phẩm là đại lượng ngẫu nhiên liên tục X (năm) có hàm mật độ

$$f(x) = \begin{cases} \frac{a}{x^4} + \frac{b}{x^3} & \text{với } x > 1; \\ 0 & \text{với } x \leq 1. \end{cases}$$

Biết khi chọn ngẫu nhiên 3 sản phẩm thì xác suất để có ít nhất 1 sản phẩm có tuổi thọ nhỏ hơn 2 năm bằng $\frac{4069}{4096}$.

- Tính a, b .

- b) Nếu quy định thời gian bảo hành là 2 năm thì tỉ lệ sản phẩm phải bảo hành bằng bao nhiêu?
- c) Giả sử thời hạn bảo hành sản phẩm này là 18 tháng và chi phí bảo hành một sản phẩm là 1,5 triệu đồng. Khi đó chi phí bảo hành trung bình bằng bao nhiêu? Muốn chi phí bảo hành trung bình chỉ chiếm 25% giá bán thì cần quy định giá bán sản phẩm bằng bao nhiêu?
- d) Giả sử thời hạn bảo hành sản phẩm này là 15 tháng. Khi bán mỗi sản phẩm được lãi 1 triệu đồng, tuy nhiên nếu sản phẩm bị hỏng trong thời gian bảo hành thì cần chi phí 1,5 triệu đồng cho việc sửa chữa. Vậy tiền lãi trung bình khi bán mỗi sản phẩm bằng bao nhiêu?
- e) Giả sử thời hạn bảo hành sản phẩm là 2 năm. Tính xác suất để khi bán 500 sản phẩm thì có ít hơn 400 sản phẩm cần phải bảo hành.

Chương 3

Mẫu ngẫu nhiên và ước lượng tham số của đại lượng ngẫu nhiên

3.1	Tổng thể nghiên cứu	61
3.2	Mẫu ngẫu nhiên	65
3.3	Thống kê	68
3.4	Mẫu ngẫu nhiên hai chiều	74
3.5	Ước lượng tham số đại lượng ngẫu nhiên	77
3.6	Bài toán ước lượng điểm	77
3.7	Bài toán ước lượng khoảng	81
	Bài tập chương 3	92

3.1 Tổng thể nghiên cứu

3.1.1 Định nghĩa

Định nghĩa 3.1. “Tập hợp toàn bộ các phần tử mang cùng một dấu hiệu nghiên cứu định tính hoặc định lượng nào đó được gọi là tổng thể nghiên cứu hay tổng thể.”

Số lượng các phần tử của tổng thể được gọi là kích thước tổng thể, kí hiệu là N . Với mỗi tổng thể ta không nghiên cứu trực tiếp tổng thể mà thông qua một hay nhiều dấu hiệu đặc trưng cho tổng thể đó. Chúng được gọi là dấu hiệu nghiên cứu, kí hiệu là χ , các dấu hiệu này có thể là định tính hoặc định lượng.

Ví dụ 3.1. Để nghiên cứu chiều cao nam thanh niên Việt Nam tuổi từ 18 đến 22, có thể tiến hành tổng điều tra nam thanh niên và đo chiều cao từng người. Khi đó tổng thể nghiên cứu là tập hợp tất cả nam thanh niên Việt Nam, dấu hiệu nghiên cứu χ là chiều cao từng người. Kích thước tổng thể N là số lượng nam thanh niên cả nước.

3.1.2 Các phương pháp mô tả tổng thể

Xét tổng thể kích thước N , dấu hiệu nghiên cứu định lượng χ nhận k giá trị $x_1, x_2, \dots, x_k$. Trong đó có N_i phần tử mà dấu hiệu nghiên cứu χ nhận giá trị bằng x_i , $i = \overline{1, k}$.

a. Bảng phân phối tần số

Tổng thể có thể được mô tả bằng bảng phân phối tần số như sau

Giá trị của χ	x_1	x_2	$\dots$	x_i	$\dots$	x_k
Tần số	N_1	N_2	$\dots$	N_i	$\dots$	N_k

trong đó

$$\begin{cases} 0 \leq N_i \leq N, & \forall i = \overline{1, k}, \\ \sum_{i=1}^k N_i = N \end{cases}.$$

b. Bảng phân phối tần suất

Kí hiệu p_i , ($i = \overline{1, k}$) là tần suất của x_i cho bởi

$$p_i = \frac{N_i}{N}, \quad i = \overline{1, k}.$$

Khi đó, tổng thể còn được mô tả bởi bảng phân phối tần suất như sau

Giá trị của χ	x_1	x_2	$\dots$	x_i	$\dots$	x_k
Tần suất	p_1	p_2	$\dots$	p_i	$\dots$	p_k

Dễ thấy

$$\begin{cases} 0 \leq p_i \leq 1 & \forall i = \overline{1, k}, \\ \sum_{i=1}^k p_i = 1 \end{cases}.$$

c. Tần số tích lũy - Tần suất tích lũy

Kí hiệu $w(x)$ là tần số tích lũy của x , tức là tổng số phần tử có giá trị nhỏ hơn x cho bởi

$$w(x) = \sum_{x_j < x} N_j$$

và $F(x)$ là tần suất tích lũy của x

$$F(x) = \frac{w(x)}{N} = \sum_{x_j < x} p_j.$$

Việc mô tả tổng thể theo bảng phân phối tần suất và tần suất tích lũy cho thấy dấu hiệu định lượng χ có thể mô hình hoá bằng một đại lượng ngẫu nhiên rời rạc X . Điều này cũng đúng đối với các tổng thể mang dấu hiệu nghiên cứu χ phân phối liên tục. Đại lượng ngẫu nhiên X dùng để mô hình hoá dấu hiệu nghiên cứu χ được gọi là đại lượng ngẫu nhiên gốc, quy luật phân phối xác suất của nó là quy luật phân phối gốc.

3.1.3 Các tham số đặc trưng của tổng thể

a. Trung bình tổng thể

Giả sử trong tổng thể kích thước N , dấu hiệu định lượng χ nhận các giá trị $x_1, x_2, \dots, x_N$. Trung bình tổng thể, kí hiệu là m , là trung bình số học của các giá trị của χ trong tổng thể cho bởi

$$m = \frac{1}{N} \sum_{i=1}^N x_i.$$

Nếu tổng thể được mô tả bởi bảng phân phối tần số ở phần (3.1.2a) thì trung bình tổng thể còn được tính bởi

$$m = \frac{1}{N} \sum_{i=1}^k x_i N_i.$$

Nếu xem dấu hiệu nghiên cứu χ như đại lượng ngẫu nhiên thì trung bình tổng thể chính là kỳ vọng toán của đại lượng ngẫu nhiên đó.

b. Phương sai tổng thể

Xét tổng thể kích thước N với dấu hiệu định lượng χ nhận các giá trị $x_1, x_2, \dots, x_N$. Phương sai tổng thể, kí hiệu là σ^2 , là trung bình số học của bình phương các sai lệch giữa các giá trị của dấu hiệu χ trong tổng thể và trung bình tổng thể

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - m)^2.$$

Nếu tổng thể được mô tả bởi bảng phân phối tần số ở phần (3.1.2a) thì phương sai tổng thể còn được tính bởi

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^k (x_i - m)^2 N_i.$$

Độ lệch tiêu chuẩn tổng thể là căn bậc hai của phương sai tổng thể

$$\sigma = \sqrt{\sigma^2}.$$

Thực chất phương sai tổng thể chính là phương sai của dấu hiệu nghiên cứu trong tổng thể. Tương tự với công thức tính phương sai của đại lượng ngẫu nhiên, ta còn có thể tính phương sai tổng thể bằng công thức sau

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N x_i^2 - \left(\frac{1}{N} \sum_{i=1}^N x_i \right)^2.$$

Nếu tổng thể được mô tả bởi bảng phân phối tần số thì phương sai tổng thể còn được tính bởi công thức

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^k x_i^2 N_i - \left(\frac{1}{N} \sum_{i=1}^k x_i N_i \right)^2.$$

c. Tần suất tổng thể

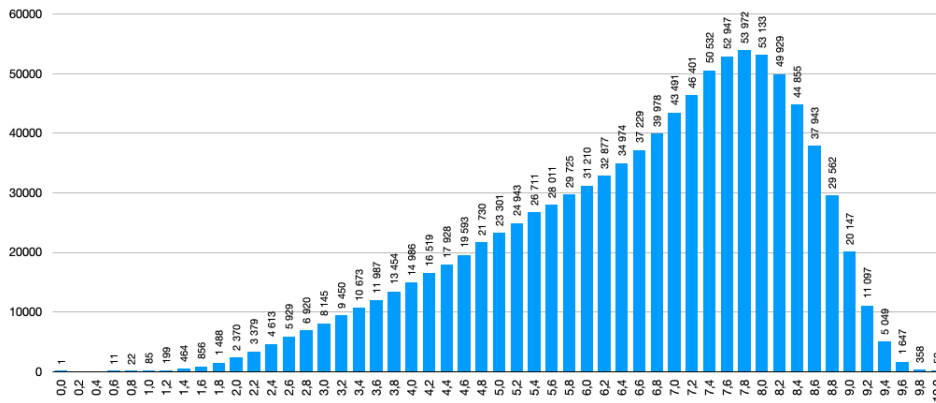
Giả sử trong tổng thể kích thước N có M phần tử mang dấu hiệu nghiên cứu, còn $N - M$ phần tử không mang dấu hiệu đó. Khi đó, tần suất của tổng thể, kí hiệu p , là tỉ lệ phần tử mang dấu hiệu nghiên cứu trong tổng thể đó

$$p = \frac{M}{N}.$$

Nếu xem dấu hiệu nghiên cứu χ là đại lượng ngẫu nhiên tuân theo quy luật không - một thì tần suất tổng thể p chính là kỳ vọng toán của đại lượng đó.

Ví dụ 3.2. Số liệu về điểm thi tốt nghiệp THPT môn Toán của cả nước năm 2021⁹

Điểm	0,0	0,2	0,4	0,6	0,8	1,0	1,2	1,4	1,6	1,8	
Số lượng	1	0	0	11	22	85	199	464	856	1488	
Điểm	2,0	2,2	2,4	2,6	2,8	3,0	3,2	3,4	3,6	3,8	
Số lượng	2370	3379	4613	5929	6920	8145	9450	10673	11987	13454	
Điểm	4,0	4,2	4,4	4,6	4,8	5,0	5,2	5,4	5,6	5,8	
Số lượng	14986	16519	17928	19593	21730	23301	24943	26711	28011	29725	
Điểm	6,0	6,2	6,4	6,6	6,8	7,0	7,2	7,4	7,6	7,8	
Số lượng	31210	32877	34974	37229	39978	43491	46401	50532	52947	53972	
Điểm	8,0	8,2	8,4	8,6	8,8	9,0	9,2	9,4	9,6	9,8	10,0
Số lượng	53133	49929	44855	37943	29562	20147	11097	5049	1647	358	52



Hình 3.1 – Phổ điểm thi THPT môn Toán năm 2021

Kết quả phân tích phổ điểm này cho thấy:

- Có $N = 980876$ thí sinh tham gia thi bài thi Toán.
- Điểm trung bình là $m = 6,614$ điểm.
- Độ lệch tiêu chuẩn là $\sigma = 1,678$ điểm.
- Điểm trung vị là $\text{med} = 7,0$ điểm.
- Điểm số có nhiều thí sinh đạt nhất là $\text{mod} = 7,8$ điểm.
- Số thí sinh đạt điểm dưới 5 là 170802, chiếm tỉ lệ $p = 17,41\%$.

⁹Nguồn: Dữ liệu kỳ thi Trung học phổ thông quốc gia năm 2021, Bộ Giáo Dục & Đào Tạo.

3.2 Mẫu ngẫu nhiên

3.2.1 Định nghĩa

Ta đã biết có thể coi dấu hiệu nghiên cứu χ của tổng thể như một đại lượng ngẫu nhiên X có hàm phân phối xác suất là tần suất tích lũy của χ . Giả sử từ tổng thể lấy ngẫu nhiên n phần tử, và gọi X_i là giá trị của χ đo lường được trên phần tử thứ i . Khi đó, có thể xem như đã thực hiện n phép thử độc lập đối với X và X_i là các đại lượng ngẫu nhiên độc lập cùng quy luật phân phối xác suất với X .

Định nghĩa 3.2. “Mẫu ngẫu nhiên kích thước n là tập hợp n đại lượng ngẫu nhiên độc lập $X_1, X_2, \dots, X_n$ được thành lập từ đại lượng ngẫu nhiên X và có cùng quy luật phân phối xác suất đối với X .”

Mẫu ngẫu nhiên thường được kí hiệu là $W = (X_1, X_2, \dots, X_n)$.

Khi thực hiện một phép thử đối với mẫu ngẫu nhiên W ta sẽ thu được kết quả cụ thể là X_1 nhận giá trị x_1 , X_2 nhận giá trị $x_2, \dots, X_n$ nhận giá trị x_n . Tập hợp n giá trị $x_1, x_2, \dots, x_n$ tạo thành một giá trị của mẫu ngẫu nhiên, hay còn gọi là một mẫu cụ thể, kí hiệu là

$$w = (x_1, x_2, \dots, x_n).$$

Ví dụ 3.3. Gọi X là số chấm xuất hiện khi tung một súc sắc cân đối và đồng chất, X là đại lượng ngẫu nhiên có bảng phân phối xác suất như sau

X	1	2	3	4	5	6
$\mathbb{P}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Nếu tung súc sắc 3 lần và gọi X_i là số chấm xuất hiện ở lần tung thứ i thì X_i là các đại lượng ngẫu nhiên độc lập cùng phân phối với X . Khi đó, ta có mẫu ngẫu nhiên kích thước $n = 3$: $W = (X_1, X_2, X_3)$.

Thực hiện một phép thử đối với mẫu ngẫu nhiên này, tức là tung cụ thể $n = 3$ lần. Giả sử lần thứ nhất được 4 chấm, lần thứ hai được 6 chấm, lần thứ ba được 3 chấm, ta thu được mẫu cụ thể là $w = (4, 6, 3)$.

3.2.2 Các phương pháp mô tả mẫu ngẫu nhiên

Giả sử từ tổng thể với đại lượng ngẫu nhiên gốc X rút ra một mẫu ngẫu nhiên kích thước n với giá trị cụ thể là $w = (x_1, x_2, \dots, x_n)$. Giả sử giá trị x_1 xuất hiện n_1 lần, giá trị x_2 xuất hiện n_2 lần, $\dots$, giá trị x_k xuất hiện n_k lần (với $n_1 + n_2 + \dots + n_k = n$), và các x_i sắp xếp theo giá trị tăng dần. Khi đó, giá trị mẫu cụ thể w có thể được mô tả như sau

a. Bảng phân phối tần số thực nghiệm (đơn)

x_i	x_1	x_2	$\dots$	x_i	$\dots$	x_k
n_i	n_1	n_2	$\dots$	n_i	$\dots$	n_k

Hiển nhiên

$$\begin{cases} 0 \leq n_i \leq n, & \forall i = \overline{1, k} \\ \sum_{i=1}^k n_i = n. \end{cases}$$

b. Bảng phân phối tần suất thực nghiệm (đơn)

Kí hiệu f_i , ($i = \overline{1, k}$), là tần suất xuất hiện giá trị x_i trong mẫu

$$f_i = \frac{n_i}{n} \quad i = \overline{1, k}.$$

Khi đó, mẫu cụ thể còn được mô tả bởi bảng phân phối tần suất như sau

x_i	x_1	x_2	$\dots$	x_i	$\dots$	x_k
f_i	f_1	f_2	$\dots$	f_i	$\dots$	f_k

Dễ thấy

$$\begin{cases} 0 \leq f_i \leq 1, & \forall i = \overline{1, k} \\ \sum_{i=1}^k f_i = 1. \end{cases}$$

c. Bảng phân phối tần số thực nghiệm (ghép lớp)

Trong trường hợp dấu hiệu nghiên cứu có phân phối liên tục thì mẫu ngẫu nhiên còn được mô tả bởi bảng phân phối tần số ghép lớp như sau

x_i	$x_0 - x_1$	$x_1 - x_2$	$\dots$	$x_{i-1} - x_i$	$\dots$	$x_{k-1} - x_k$
n_i	n_1	n_2	$\dots$	n_i	$\dots$	n_k

Bảng trên có nghĩa là: có n_1 phần tử của mẫu w nhận giá trị trong khoảng $(x_0; x_1)$, n_2 phần tử nhận giá trị trong khoảng $(x_1; x_2)$, $\dots$, có n_k phần tử nhận giá trị trong khoảng $(x_{k-1}; x_k)$.
Hiển nhiên

$$\begin{cases} 0 \leq n_i \leq n, & \forall i = \overline{1, k}, \\ \sum_{i=1}^k n_i = n. \end{cases}$$

Tương tự, ta cũng có bảng phân phối tần suất ghép lớp như sau

x_i	$x_0 - x_1$	$x_1 - x_2$	$\dots$	$x_{i-1} - x_i$	$\dots$	$x_{k-1} - x_k$
f_i	f_1	f_2	$\dots$	f_i	$\dots$	f_k

Chú ý. Bảng phân phối ghép lớp thường được dùng trong trường hợp đại lượng ngẫu nhiên gốc X phân phối liên tục. Do đó có thể coi trong mẫu cụ thể được mô tả bởi bảng phân phối ghép lớp, trường hợp dấu hiệu nghiên cứu nhận giá trị đúng bằng x_i không xảy ra.

d. Hàm phân phối thực nghiệm

Kí hiệu $w(x)$ là tần số tích lũy của x , tức là tổng số phần tử có giá trị nhỏ hơn x trong mẫu cụ thể w

$$w(x) = \sum_{x_j < x} n_j,$$

và $F^*(x)$ là tần suất tích lũy của x

$$F^*(x) = \frac{w(x)}{n} = \sum_{x_j < x} f_j.$$

Hàm $F^*(x)$ là một hàm của x và được gọi là hàm phân phối thực nghiệm của mẫu.

3.2.3 Đồ thị của phân phối thực nghiệm

a. Đa giác tần số

Xét mẫu cụ thể được mô tả bởi bảng phân phối tần số thực nghiệm ở phần (3.2.2 a).

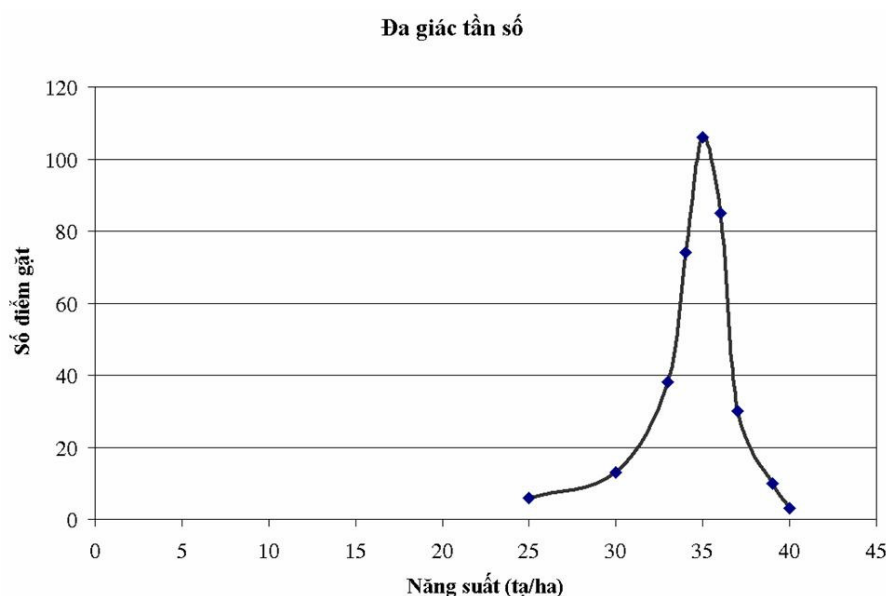
Đa giác tần số là một đường gấp khúc mà các đoạn thẳng của nó nối các điểm $(x_1, n_1); (x_2, n_2), \dots, (x_k, n_k)$ trên mặt phẳng.

Ví dụ 3.4. Gặt ngẫu nhiên 365 điểm trồng lúa của một huyện ta được các số liệu được sắp xếp thành bảng sau

Năng suất (tạ/ha)	25	30	33	34	35	36	37	39	40
Số điểm gặt	6	13	38	74	106	85	30	10	3

Như vậy, giá trị của mẫu đã được mô tả dưới dạng bảng phân phối tần số thực nghiệm.

Đa giác tần số của phân phối thực nghiệm này như sau



b. Biểu đồ tần số

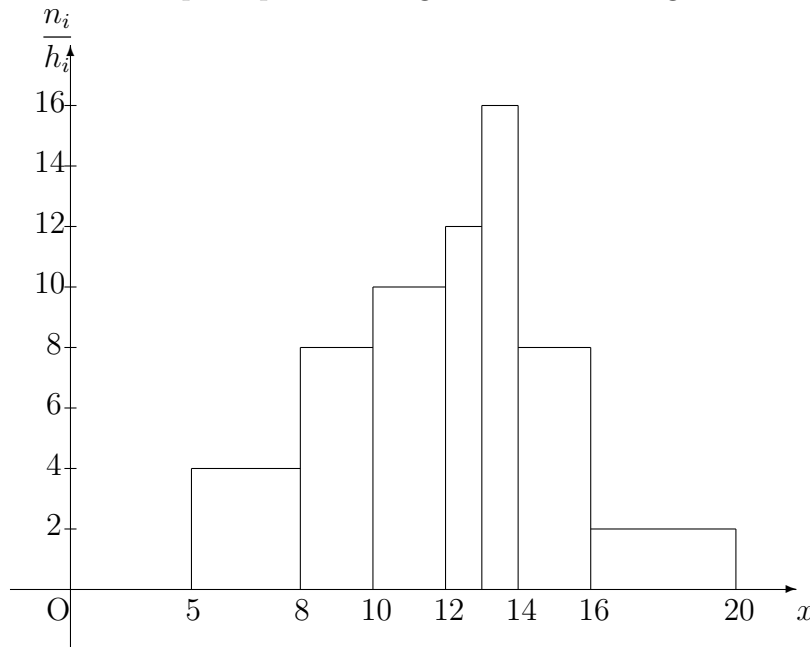
Xét mẫu ghép lớp cụ thể được mô tả bởi bảng phân phối tần số thực nghiệm ở phần (3.2.2 c).

Để vẽ biểu đồ tần số ta chia đoạn $(x_0; x_k)$ chứa tất cả các giá trị quan sát được của mẫu thành các đoạn nhỏ (x_{i-1}, x_i) , $i = \overline{1, k}$. Biểu đồ tần số là một hình bậc thang được tạo nên bởi k hình chữ nhật có đáy bằng $h_i = x_i - x_{i-1}$ và chiều cao bằng $\frac{n_i}{h_i}$. Lúc đó, diện tích hình chữ nhật thứ i đúng bằng $h_i \cdot \frac{n_i}{h_i} = n_i$ là tần số ứng với đoạn thứ i , $i = \overline{1, k}$. Vì vậy, tổng diện tích của tất cả các hình chữ nhật sẽ bằng kích thước mẫu n .

Ví dụ 3.5. Xét phân phối thực nghiệm cho ở bảng sau

x_i	5 – 8	8 – 10	10 – 12	12 – 13	13 – 14	14 – 16	16 – 20
n_i	12	16	20	12	16	16	8
$\frac{n_i}{h_i}$	4	8	10	12	16	8	2

Khi đó, biểu đồ tần số của phân phối thực nghiệm trên có dạng



3.3 Thống kê

3.3.1 Định nghĩa

Định nghĩa 3.3. “Giả sử khi nghiên cứu đại lượng ngẫu nhiên gốc X trong tổng thể, ta có mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$. Thống kê là một hàm của các giá trị $X_1, X_2, \dots, X_n$, kí hiệu là Θ . Như vậy

$$\Theta = f(X_1, X_2, \dots, X_n)."$$

Về thực chất thống kê là một hàm của các đại lượng ngẫu nhiên. Do đó, bản thân nó cũng sẽ là một đại lượng ngẫu nhiên tuân theo một quy luật phân phối xác suất nhất định và có các đặc trưng số tương ứng như kỳ vọng $\mathbb{E}(\Theta)$, phương sai $\text{var}(\Theta)$, ...

Khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w = (x_1, x_2, \dots, x_n)$ thì thống kê Θ cũng nhận một giá trị cụ thể là $\theta = f(x_1, x_2, \dots, x_n)$.

3.3.2 Trung bình mẫu

Khái niệm

Giả sử từ đại lượng ngẫu nhiên gốc X trong tổng thể lập mẫu ngẫu nhiên kích thước n là $W = (X_1, X_2, \dots, X_n)$. Trung bình mẫu là một thống kê, kí hiệu $\bar{X}$, là trung bình số học của các giá trị của mẫu xác định như sau

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Tính chất

Nếu đại lượng ngẫu nhiên gốc X có kỳ vọng toán $\mathbb{E}(X) = \mu$ và phương sai $\text{var}(X) = \sigma^2$ thì

$$\mathbb{E}(\bar{X}) = \mu; \quad \text{var}(\bar{X}) = \frac{\sigma^2}{n}.$$

3.3.3 Phương sai mẫu và phương sai hiệu chỉnh mẫu

Khái niệm

Giả sử từ đại lượng ngẫu nhiên gốc X trong tổng thể rút ra mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$. Phương sai mẫu là một thống kê phản ánh trung bình của bình phương mức độ sai lệch giữa các giá trị của mẫu với trung bình mẫu, kí hiệu S^2 , cho bởi công thức

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Ta cũng có khái niệm phương sai hiệu chỉnh mẫu S'^2 như sau

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Tính chất

- Rõ ràng,

$$S'^2 = \frac{n}{n-1} S^2.$$

- Nếu đại lượng ngẫu nhiên gốc X có phương sai $\text{var}(X) = \sigma^2$ thì

$$\begin{aligned} \mathbb{E}(S^2) &= \frac{n-1}{n} \sigma^2; \\ \mathbb{E}(S'^2) &= \sigma^2. \end{aligned}$$

3.3.4 Độ lệch tiêu chuẩn mẫu

Nếu lấy căn bậc hai của phương sai mẫu S^2 và phương sai hiệu chỉnh mẫu S'^2 ta thu được các thống kê tương ứng gọi là độ lệch tiêu chuẩn mẫu S , độ lệch tiêu chuẩn hiệu chỉnh mẫu S' .

Cụ thể,

$$S = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2},$$

$$S' = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

3.3.5 Tần suất mẫu

Khái niệm

Giả sử từ tổng thể kích thước N , trong đó M phần tử mang dấu hiệu nghiên cứu, lấy ra một mẫu ngẫu nhiên kích thước n và trong đó thấy có Y phần tử mang dấu hiệu nghiên cứu. Lúc đó, tần suất mẫu là một thống kê, kí hiệu là f và là tỉ số giữa số phần tử mang dấu hiệu nghiên cứu trong mẫu và kích thước mẫu

$$f = \frac{Y}{n}.$$

Ta chú ý rằng, tần suất mẫu là một trường hợp riêng của trung bình mẫu $\bar{X}$ khi xem dấu hiệu nghiên cứu trong tổng thể là đại lượng ngẫu nhiên X tuân theo quy luật không-một. Thật vậy, nếu gọi X_i là số phần tử mang dấu hiệu nghiên cứu khi lấy phần tử thứ i của mẫu thì ta có mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ với X_i tuân theo quy luật không-một và

$$Y = X_1 + X_2 + \dots + X_n.$$

Khi đó

$$f = \frac{X_1 + X_2 + \dots + X_n}{n} = \bar{X}.$$

Tính chất

Nếu đại lượng ngẫu nhiên gốc X tuân theo quy luật $\mathcal{A}(p)$ thì kỳ vọng toán $\mathbb{E}(X) = p$ và phương sai $\text{var}(X) = p(1-p)$, do đó

$$\mathbb{E}(f) = p; \quad \text{var}(f) = \frac{p(1-p)}{n}.$$

3.3.6 Tính giá trị cụ thể cho các thống kê

a. Xét mẫu cụ thể có bảng phân phối tần số:

x_i	x_1	x_2	$\dots$	x_i	$\dots$	x_k	$(\sum_{i=1}^k n_i = n)$
n_i	n_1	n_2	$\dots$	n_i	$\dots$	n_k	

Giá trị cụ thể của các thống kê như sau

i. Trung bình mẫu

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

ii. Phương sai mẫu, phương sai mẫu hiệu chỉnh

$$s^2 = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2;$$

$$s'^2 = \frac{1}{n-1} \sum_{i=1}^k n_i (x_i - \bar{x})^2.$$

Trong thực hành, người ta sử dụng công thức sau để tính giá trị cụ thể của phương sai mẫu và phương sai mẫu hiệu chỉnh

$$s^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2;$$

$$s'^2 = \frac{n}{n-1} s^2.$$

iii. Độ lệch tiêu chuẩn mẫu, độ lệch tiêu chuẩn hiệu chỉnh mẫu

$$s = \sqrt{s^2} = \sqrt{\frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2};$$

$$s' = \sqrt{s'^2} = \sqrt{\frac{1}{n-1} \sum_{i=1}^k n_i (x_i - \bar{x})^2}.$$

Chú ý.

- Nếu mẫu ngẫu nhiên có giá trị cụ thể cho dưới dạng liệt kê $w = (x_1, x_2, \dots, x_n)$ thì cũng có thể coi mẫu ngẫu nhiên có bảng phân phối tần số như ở mục 3.2.2 ý a) với $k = n$ và các tần số $n_1 = n_2 = \dots = n_k = 1$.
- Nếu mẫu ngẫu nhiên được cho bởi bảng phân phối tần số ghép lớp thì người ta lấy giá trị giữa của mỗi lớp $\left(\frac{x_{i-1} + x_i}{2}\right)$ đại diện cho lớp đó để tính toán các thống kê đặc trưng của mẫu.

b. Tần suất mẫu

Giả sử trong mẫu cụ thể kích thước n có k phần tử mang dấu hiệu nghiên cứu. Khi đó, giá trị cụ thể của tần suất mẫu là

$$f = \frac{k}{n}.$$

Ví dụ 3.6. Sử dụng số liệu của ví dụ 3.4: Gặt ngẫu nhiên 365 điểm trồng lúa của một huyện ta được số liệu sau

Năng suất X (tạ/ha)	25	30	33	34	35	36	37	39	40
Số điểm gặt	6	13	38	74	106	85	30	10	3

Lập bảng tính toán sau

x_i	n_i	$n_i x_i$	$n_i x_i^2$
25	6	150	3750
30	13	390	11700
33	38	1254	41382
34	74	2516	85544
35	106	3710	129850
36	85	3060	110160
37	30	1110	41070
39	10	390	15210
40	3	120	4800
	$\sum n_i = 365$	$\sum n_i x_i = 12700$	$\sum n_i x_i^2 = 443466$

Từ đó, ta có

Trung bình mẫu:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i = \frac{12700}{365} \approx 34,795 \text{ (tạ/ha)}.$$

Phương sai mẫu:

$$s^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2 = \frac{443466}{365} - 34,795^2 \approx 4,283.$$

Phương sai hiệu chỉnh mẫu:

$$s'^2 = \frac{n}{n-1} s^2 = \frac{365}{364} \cdot 4,283 \approx 4,295.$$

Độ lệch tiêu chuẩn hiệu chỉnh mẫu:

$$s' = \sqrt{s'^2} = \sqrt{4,295} \approx 2,072 \text{ (tạ/ha)}.$$

Giả sử những điểm trồng lúa có năng suất trên 36 tạ/ha được xếp vào loại A. Khi đó, tần suất mẫu của các điểm trồng lúa loại A là

$$f = \frac{k}{n} = \frac{43}{365} \approx 0,118.$$

Ví dụ 3.7. Kết quả khảo sát về doanh số bán một mặt hàng tại 10 đại lí thu được các số liệu sau (đơn vị: triệu đồng)

17,5; 19,2; 26,2; 24,5; 30,9; 31,5; 25,3; 11,9; 28,3; 25,0.

Ta lập bảng tính như sau

	$x_i = n_i x_i$	$x_i^2 = n_i x_i^2$
$n = 10$	17,5	306,25
	19,2	368,64
	26,2	686,44
	24,5	600,25
	30,9	954,81
	31,5	992,25
	25,3	640,09
	11,9	141,61
	28,3	800,89
	25,0	625,00
	$\sum n_i x_i =$ 240,3	$\sum n_i x_i^2 =$ 6116,23

Từ đó, có kết quả sau

$$\begin{aligned}\bar{x} &= \frac{240,3}{10} = 24,03 \text{ (triệu đồng);} \\ s^2 &= \frac{6116,23}{10} - (24,03)^2 = 34,1821; \\ s' &= \sqrt{\frac{10}{9} 34,1821} \approx 6,163 \text{ (triệu đồng).}\end{aligned}$$

Ví dụ 3.8. Đo ngẫu nhiên chiều cao X (cm) của 100 thanh niên ở một tỉnh ta thu được các số liệu sau

X	154-158	158-162	162-166	166-170	170-174	174-178	178-182
n_i	10	14	26	28	12	8	2

Ta lập bảng tính như sau

x_i	n_i	$n_i x_i$	$n_i x_i^2$
156	10	1560	243360
160	14	2240	358400
164	26	4264	699296
168	28	4704	790272
172	12	2064	355008
176	8	1408	247808
180	2	360	64800
	$\sum n_i =$ 100	$\sum n_i x_i =$ 16600	$\sum n_i x_i^2 =$ 2758944

Từ đó, có kết quả sau

$$\begin{aligned}\bar{x} &= \frac{16600}{100} = 166 \text{ (cm);} \\ s^2 &= \frac{2758944}{100} - 166^2 = 33,44; \\ s' &= \sqrt{\frac{100}{99} 33,44} \approx 5,812 \text{ (cm).}\end{aligned}$$

3.4 Mẫu ngẫu nhiên hai chiều

3.4.1 Khái niệm

Giả sử trên cùng một tổng thể phải nghiên cứu đồng thời hai dấu hiệu nghiên cứu, trong đó dấu hiệu thứ nhất có thể xem như đại lượng ngẫu nhiên X , còn dấu hiệu thứ hai có thể xem như đại lượng ngẫu nhiên Y . Hệ hai đại lượng ngẫu nhiên X, Y làm thành đại lượng ngẫu nhiên hai chiều (X, Y) .

Từ tổng thể lấy ra mẫu kích thước n , tức là thực hiện n phép thử độc lập đối với đại lượng ngẫu nhiên (X, Y) . Giả sử trên phần tử thứ i của mẫu ($i = \overline{1, n}$), đại lượng (X, Y) nhận giá trị (X_i, Y_i) , khi đó ta có n đại lượng ngẫu nhiên độc lập $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$, ở đây các đại lượng ngẫu nhiên $X_1, X_2, \dots, X_n$ độc lập và cùng phân phối xác suất với X và các đại lượng ngẫu nhiên $Y_1, Y_2, \dots, Y_n$ độc lập và cùng phân phối xác suất với Y . Bộ n đại lượng ngẫu nhiên hai chiều trên được gọi là mẫu ngẫu nhiên hai chiều, kí hiệu là

$$W = ((X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)).$$

Giả sử đại lượng ngẫu nhiên (X_i, Y_i) nhận giá trị (x_i, y_i) , $i = \overline{1, n}$, ta thu được mẫu cụ thể là

$$w = ((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)).$$

3.4.2 Phương pháp mô tả mẫu ngẫu nhiên hai chiều

Giả sử từ tổng thể rút ra một mẫu kích thước n , trong đó thành phần X nhận các giá trị $x_1, x_2, \dots, x_i, \dots, x_k$, còn thành phần Y nhận các giá trị $y_1, y_2, \dots, y_j, \dots, y_\ell$, ở đó có n_{ij} phần tử của mẫu nhận giá trị (x_i, y_j) . Khi đó, giá trị cụ thể của mẫu W được mô tả bởi bảng phân phối tần số thực nghiệm sau

X \ Y	y_1	y_2	$\dots$	y_j	$\dots$	y_ℓ	n_i
x_1	n_{11}	n_{12}	$\dots$	n_{1j}	$\dots$	$n_{1\ell}$	n_1
x_2	n_{21}	n_{22}	$\dots$	n_{2j}	$\dots$	$n_{2\ell}$	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\dots$	n_{ij}	$\dots$	$n_{i\ell}$	n_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\dots$	n_{kj}	$\dots$	$n_{k\ell}$	n_k
m_j	m_1	m_2	$\dots$	m_j	$\dots$	m_ℓ	n

trong đó n_i là tổng số phần tử của mẫu mang giá trị $X = x_i$,

m_j là tổng số phần tử của mẫu mang giá trị $Y = y_j$.

3.4.3 Một số thống kê đặc trưng của mẫu ngẫu nhiên hai chiều

Từ bảng phân phối tần số thực nghiệm của mẫu ngẫu nhiên hai chiều ta có thể rút ra

a. Bảng phân phối thực nghiệm của X

X	x_1	x_2	$\dots$	x_i	$\dots$	x_k
n_i	n_1	n_2	$\dots$	n_i	$\dots$	n_k

và tính được các thống kê đặc trưng của X là

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k n_i x_i; \quad s_X^2 = \frac{1}{n} \sum_{i=1}^k n_i (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^k n_i x_i^2 - \bar{x}^2.$$

b. Bảng phân phối thực nghiệm của Y

Y	y_1	y_2	$\dots$	y_j	$\dots$	y_ℓ
m_j	m_1	m_2	$\dots$	m_j	$\dots$	m_ℓ

và tính được các thống kê đặc trưng của Y là

$$\bar{y} = \frac{1}{n} \sum_{j=1}^{\ell} m_j y_j; \quad s_Y^2 = \frac{1}{n} \sum_{j=1}^{\ell} m_j (y_j - \bar{y})^2 = \frac{1}{n} \sum_{j=1}^{\ell} m_j y_j^2 - \bar{y}^2.$$

c. Bảng phân phối có điều kiện của Y khi $X = x_i$

$Y _{X=x_i}$	y_1	y_2	$\dots$	y_j	$\dots$	y_ℓ
n_{ij}	n_{i1}	n_{i2}	$\dots$	n_{ij}	$\dots$	$n_{i\ell}$

và tính được các thống kê tương ứng là

$$\bar{y}|_{X=x_i} = \frac{1}{n_i} \sum_{j=1}^{\ell} n_{ij} y_j;$$

$$s_{Y|X=x_i}^2 = \frac{1}{n_i} \sum_{j=1}^{\ell} n_{ij} (y_j - \bar{y}|_{X=x_i})^2 = \frac{1}{n_i} \sum_{j=1}^{\ell} n_{ij} y_j^2 - (\bar{y}|_{X=x_i})^2.$$

d. Bảng phân phối có điều kiện của X khi $Y = y_j$

$X _{Y=y_j}$	x_1	x_2	$\dots$	x_i	$\dots$	x_k
n_{ij}	n_{1j}	n_{2j}	$\dots$	n_{ij}	$\dots$	n_{kj}

và tính được các thống kê tương ứng là

$$\bar{x}|_{Y=y_j} = \frac{1}{m_j} \sum_{i=1}^k n_{ij} x_i;$$

$$s_{X|Y=y_j}^2 = \frac{1}{m_j} \sum_{i=1}^k n_{ij} (x_i - \bar{x}|_{Y=y_j})^2 = \frac{1}{m_j} \sum_{i=1}^k n_{ij} x_i^2 - (\bar{x}|_{Y=y_j})^2.$$

Ví dụ 3.9. Nghiên cứu năng suất một loại cây trồng Y (kg/ha) và kinh phí cho cải tạo đất X (triệu đồng/ha) tại một số địa phương trong vòng 10 năm ta thu được bảng số liệu sau

Y \ X	2	3	4	5	6	7
20 - 22	1	3	5			
22 - 24		2	7	8	5	
24 - 26			1	4	3	1
26 - 28				1	2	1
28 - 30					1	1

- Từ bảng trên ta có bảng phân phối tần số của X là

X	2	3	4	5	6	7
Số địa phương (n_i)	1	5	13	13	11	3

Lập bảng tính

x_i	n_i	$n_i x_i$	$n_i x_i^2$
2	1	2	4
3	5	15	45
4	13	52	208
5	13	65	325
6	11	66	396
7	3	21	147
	$n = \sum n_i = 46$	$\sum n_i x_i = 221$	$\sum n_i x_i^2 = 1125$

Từ đó, ta có kết quả sau

Lượng dầu tư trung bình $\bar{x} = \frac{221}{46} \approx 4,804$ (triệu đồng/ha)

Độ lệch tiêu chuẩn hiệu chỉnh là

$$s'_x = \sqrt{\frac{n}{n-1} \left(\frac{\sum n_i x_i^2}{n} - \bar{x}^2 \right)} = \sqrt{\frac{46}{45} \left(\frac{1125}{46} - 4,804^2 \right)} \approx 1,187 \text{ (triệu đồng/ha)}$$

- Từ bảng trên ta có bảng phân phối tần số của Y là

Y	21	23	25	27	29
Số địa phương (n_i)	9	22	9	4	2

Tương tự phần trên, ta cũng tính được

$$\bar{y} \approx 23,609 \text{ (kg/ha)}; \quad s'_y \approx 2,060 \text{ (kg/ha)}$$

- Những địa phương có lượng dầu tư không quá 6 (triệu đồng/ha) và năng suất không nhỏ hơn 24 (kg/ha) được xếp vào nhóm A. Bảng phân phối năng suất của các địa phương nhóm A là

Y_A	25	27	29
Số địa phương (n_i)	8	3	1

Tương tự phần trên, ta cũng tính được

$$\bar{y}_A \approx 25,833 \text{ (kg/ha)}; \quad s'_{y_A} \approx 1,337 \text{ (kg/ha)}.$$

- Tần suất mẫu của những địa phương nhóm A là $f = \frac{k}{n} = \frac{12}{46} \approx 0,261$.

3.5 Ước lượng tham số đại lượng ngẫu nhiên

3.5.1 Các khái niệm

Giả sử ta muốn nghiên cứu dấu hiệu χ của tổng thể, vấn đề đặt ra cần tìm các tham số đặc trưng như trung bình, phương sai hay tỉ lệ ... của dấu hiệu này.

Để giải quyết bài toán đặt ra ở trên, người ta coi dấu hiệu nghiên cứu χ như đại lượng ngẫu nhiên X nào đó với dạng quy luật phân phối được giả thiết đã biết (tuy nhiên các tham số của phân phối lại chưa được biết đầy đủ). Vì vậy, bài toán nghiên cứu tham số đặc trưng của dấu hiệu χ trở thành nghiên cứu các tham số đặc trưng của luật phân phối của X . Ví dụ, X có phân phối chuẩn $\mathcal{N}(\mu, \sigma^2)$ hay phân phối Poisson $\mathcal{P}(\lambda)$, ta phải ước lượng giá trị gần đúng của hai tham số μ, σ^2 hay λ trong các phân phối trên.

Do đó, bài toán ước lượng tham số có thể khái quát như sau: *Xét đại lượng ngẫu nhiên X với dạng quy luật phân phối xác suất đã biết song chưa biết tham số đặc trưng θ nào đó của nó, phải tìm (ước lượng) một cách gần đúng cho giá trị của θ .*

Lý thuyết mẫu được xây dựng cho bài toán ước lượng dựa vào nguyên lý quy nạp thống kê. Từ tổng thể lấy mẫu ngẫu nhiên kích thước n , từ đó lập thống kê Θ ước lượng cho θ .

3.5.2 Phân loại ước lượng

Bài toán ước lượng được chia thành hai dạng: ước lượng điểm và ước lượng khoảng.

- **Ước lượng điểm:** Việc thay thế tham số θ bởi một giá trị cụ thể của thống kê Θ là bài toán ước lượng điểm. Trong thống kê, có hai bài toán ước lượng điểm điển hình, đó là ước lượng bằng phương pháp mômen và phương pháp hàm hợp lí cực đại.
- **Ước lượng bằng khoảng tin cậy:** Xây dựng các thống kê Θ_1, Θ_2 sao cho tham số θ thuộc khoảng (Θ_1, Θ_2) với xác suất cho trước.

3.6 Bài toán ước lượng điểm

Giả sử cần ước lượng tham số θ của dấu hiệu nghiên cứu χ , ở đây θ có thể là trung bình tổng thể m , phương sai tổng thể σ^2 hay tỉ lệ tổng thể $p, \dots$. Từ tổng thể lấy mẫu ngẫu nhiên kích thước n

$$W = (X_1, X_2, \dots, X_n).$$

Xét thống kê $\Theta = g(X_1, X_2, \dots, X_n)$ phù hợp với tham số cần ước lượng θ , khi đó Θ là ước lượng điểm của θ vì ứng với mỗi giá trị cụ thể của mẫu $(x_1, x_2, \dots, x_n)$ cho một giá trị cụ thể

$\theta = g(x_1, x_2, \dots, x_n)$. Ví dụ, nếu θ là trung bình, phương sai hay tỉ lệ tổng thể thì Θ là trung bình mẫu $\bar{X}$, phương sai mẫu S^2 , tần suất mẫu f . Thống kê $g(X_1, X_2, \dots, X_n)$ để ước lượng một tham số nào đó được gọi là hàm ước lượng.

Để đánh giá chất lượng của ước lượng người ta thường xem xét độ lệch giữa Θ với θ nhưng thực tế giá trị θ chưa biết nên phải dựa vào chất lượng hàm ước lượng Θ . Rõ ràng có nhiều cách chọn hàm g , tức là có nhiều thống kê Θ nên phải có tiêu chí đánh giá hàm ước lượng sao cho nó gần với giá trị thực θ nhất. Tùy thuộc vào các tiêu chuẩn đặt ra cho ước lượng, có thể xét một số bài toán ước lượng điểm sau đây.

3.6.1 Ước lượng không chệch

Xét đại lượng ngẫu nhiên gốc X với tham số θ cần ước lượng. Gọi thống kê Θ là ước lượng điểm nào đó của θ , với mỗi mẫu, độ lệch giữa trung bình ước lượng với θ có thể xảy ra. Rõ ràng, độ lệch càng nhỏ thì ước lượng càng chính xác. Ta có định nghĩa sau

Định nghĩa 3.4. “Thống kê Θ được gọi là ước lượng không chệch của tham số θ nếu

$$\mathbb{E}(\Theta) = \theta.”$$

Ngược lại, nếu $\mathbb{E}(\Theta) \neq \theta$ thì Θ là ước lượng chệch của θ , với độ chệch $|\mathbb{E}(\Theta) - \theta|$.

Trong thực tế, việc một thống kê có kỳ vọng toán không bằng tham số cần ước lượng dẫn tới sai số có hệ thống, tức là với k mẫu lấy ra, tất cả các trung bình ước lượng này đều có giá trị lớn hơn hoặc nhỏ hơn giá trị thực của tham số. Do đó, ta phải chọn thống kê sao cho trung bình của nó càng gần tham số cần ước lượng càng tốt.

Người ta chứng minh được rằng, trung bình mẫu $\bar{X}$ là ước lượng không chệch cho trung bình tổng thể m , phương sai mẫu hiệu chỉnh S'^2 là ước lượng không chệch cho phương sai tổng thể σ^2 , tần suất mẫu là ước lượng không chệch cho tỉ lệ tổng thể p .

3.6.2 Ước lượng hiệu quả

Trong mỗi ước lượng không chệch Θ của θ , có thể tồn tại thành phần sai lệch khá lớn so với trung bình mẫu, nghĩa là phương sai $\text{var}(\Theta)$ rất lớn nên nếu lấy giá trị thống kê trên mẫu đó ước lượng cho θ sẽ dẫn tới sai số lớn, nên suy diễn thống kê không chính xác. Vì vậy, trong các ước lượng không chệch của tham số θ , cần chọn ước lượng có phương sai nhỏ nhất.

Định nghĩa 3.5. “Ước lượng không chệch Θ gọi là ước lượng hiệu quả của θ nếu phương sai của nó nhỏ nhất trong các ước lượng không chệch xét trên cùng một mẫu.”

Ước lượng hiệu quả có ý nghĩa quan trọng trong thống kê. Phương sai càng nhỏ thống kê càng hiệu quả. Để thấy điều đó xét bất đẳng thức Chebyshev

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \leq \varepsilon) \geq 1 - \frac{\text{var}(Y)}{\varepsilon^2}.$$

Do $\mathbb{E}(\Theta) = \theta$, bất đẳng thức trên trở thành

$$\mathbb{P}(|\Theta - \theta| \leq \varepsilon) \geq 1 - \frac{\text{var}(\Theta)}{\varepsilon^2}.$$

Bất đẳng thức trên cho thấy mối liên hệ giữa phương sai và xác suất để thống kê Θ gần giá trị thực θ . Qua đó, nếu phương sai càng nhỏ thì xác suất ước lượng có sai số nhỏ càng lớn, ước lượng càng đáng tin cậy. Vậy phương sai là một đại lượng phản ánh chất lượng của thống kê.

Phương sai thỏa mãn bất đẳng thức Cramer-Rao,

$$\text{var}(\Theta) \geq \frac{1}{n\mathbb{E}\left[\frac{\partial \ln f(X, \theta)}{\partial \theta}\right]^2};$$

với Θ là ước lượng không chệch bất kỳ của θ , n là kích thước mẫu (gọi tắt là cỡ mẫu), $f(x, \theta)$ là hàm mật độ xác suất của đại lượng ngẫu nhiên gốc X .

Người ta chứng minh được rằng, $\bar{X}$ là ước lượng hiệu quả của trung bình tổng thể μ , tần suất mẫu f là ước lượng hiệu quả của tỉ lệ tổng thể p .

Ví dụ 3.10. Xét biến ngẫu nhiên $X \sim \mathcal{N}(\mu, \sigma^2)$. Ta sẽ chỉ ra trung bình mẫu $\bar{X}$ là ước lượng hiệu quả của μ . Thật vậy,

$$f(x, \mu) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2};$$

lấy ln hai vế

$$\ln f(x, \mu) = -\ln \sqrt{2\pi}\sigma - \frac{(x-\mu)^2}{2\sigma^2};$$

dẫn tới

$$\frac{\partial \ln f(x, \mu)}{\partial \mu} = \frac{x-\mu}{\sigma^2}.$$

Do đó

$$\begin{aligned} n\mathbb{E}\left[\frac{\partial \ln f(X, \mu)}{\partial \mu}\right]^2 &= n\mathbb{E}\left(\frac{X-\mu}{\sigma^2}\right)^2 = \frac{n}{\sigma^4}\mathbb{E}(X-\mu)^2 \\ &= \frac{n}{\sigma^4}\text{var}(X) = \frac{n\sigma^2}{\sigma^4} = \frac{n}{\sigma^2}. \end{aligned}$$

Mặt khác, lại có $\text{var}(\bar{X}) = \frac{\sigma^2}{n}$ bằng biểu thức của vế phải bất đẳng thức Cramer-Rao nên $\bar{X}$ là ước lượng hiệu quả của μ .

3.6.3 Ước lượng vững

Theo luật yếu số lớn (hay có tên khác là luật Khinchin) trong lý thuyết xác suất, khi cỡ mẫu càng lớn trung bình mẫu càng gần trung bình tổng thể theo nghĩa xác suất. Tức là, xét mẫu $W = (X_1, X_2, \dots, X_n)$, khi đó

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\bar{X} - m| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

Yêu cầu tương tự được đặt ra với ước lượng $\Theta = f(X_1, X_2, \dots, X_n)$ đối với tham số θ của đại lượng ngẫu nhiên gốc X . Ta có định nghĩa sau

Định nghĩa 3.6. “Với mọi $\varepsilon > 0$ bất kì, nếu

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\Theta - \theta| > \varepsilon) = 0;$$

hay

$$\Theta \rightarrow \theta, \text{ theo xác suất khi } n \rightarrow \infty,$$

thì Θ gọi là ước lượng vững của θ .”

Từ định nghĩa trên, theo luật số lớn người ta cũng chứng minh được trung bình mẫu $\bar{X}$, tần suất mẫu f lần lượt là ước lượng vững của trung bình tổng thể m , tỉ lệ tổng thể p .

Ta có điều kiện đủ để thống kê không chệch là ước lượng vững như sau

Định lý 3.1. Nếu Θ là ước lượng không chệch của θ và $\lim_{n \rightarrow \infty} \text{var}(\Theta) = 0$ thì Θ là ước lượng vững.

Trong thống kê, để đánh giá chất lượng ước lượng điểm, người ta còn một số tiêu chuẩn nữa. Do khuôn khổ của cuốn sách, các tiêu chuẩn này không được trình bày ở đây, độc giả có thể tham khảo tài liệu liên quan.

3.6.4 Bài toán hợp lí cực đại

Xét đại lượng ngẫu nhiên gốc X có quy luật phân phối xác suất xác định với tham số θ nào đó. Giả sử X có hàm mật độ xác suất $f(x, \theta)$ nếu X là đại lượng ngẫu nhiên liên tục, X có phân phối xác suất $\mathbb{P}(x, \theta)$ nếu X là đại lượng ngẫu nhiên rời rạc. Ta cần xác định tham số θ trong phân phối xác suất trên. Các biểu thức và tính toán cho hai đại lượng ngẫu nhiên liên tục và rời rạc như nhau, phần sau đây trình bày cho đại lượng liên tục.

Xét hàm hợp lí theo tham số θ sau

$$L_n(x, \theta) = \prod_{i=1}^n f(x_i, \theta). \quad (3.1)$$

Do giá trị hàm hợp lí là tích các xác suất tại giá trị cụ thể của mẫu $(x_1, x_2, \dots, x_n)$ nên thống kê $\Theta = f(X_1, X_2, \dots, X_n)$ là ước lượng hợp lí cực đại nếu hàm $L_n(x, \theta)$ đạt giá trị cực đại tại $\theta = f(x_1, x_2, \dots, x_n)$. Xét hàm $\ell_n(x, \theta) := \ln L_n(x, \theta)$, khi đó điểm cực đại của $L_n(x, \theta)$ cũng là điểm cực đại của $\ell_n(x, \theta)$. Xét điều kiện cần của điểm cực đại là

$$\frac{\partial \ell_n(x, \theta)}{\partial \theta} = 0. \quad (3.2)$$

Lưu ý rằng, nghiệm của hệ này có thể không duy nhất, do đó cần xét thêm điều kiện đủ để nghiệm θ^* là điểm cực đại.¹⁰

Trong thực tế với cỡ mẫu có thể lên tới hàng trăm, thậm chí hàng nghìn và số mẫu cũng có thể lớn như thế. Vì vậy, việc giải hệ phương trình (3.2) bằng tay không hề đơn giản. Thông thường, phương pháp số được sử dụng cho việc đó, ví dụ điển hình như các thuật toán lặp được lập trình cho việc tìm ước lượng hợp lí cực đại có sẵn trong các gói lệnh của một số phần mềm thống kê.

¹⁰Người ta chứng minh được rằng nếu hệ trên có nghiệm duy nhất thì nghiệm đó luôn là điểm cực đại của hàm hợp lí (xem [12]).

Ví dụ 3.11. Giả sử chúng ta quan sát mẫu ngẫu nhiên kích thước n là $(X_1, X_2, \dots, X_n)$ ở đó các biến là độc lập cùng phân phối, $X_i \sim \mathcal{N}(\mu, \sigma^2)$. Bây giờ, ta phân tích cách tìm ước lượng hợp lý cực đại của hai tham số μ và σ^2 dựa vào n quan sát trên.

Lời giải: Ở đây vector tham số $\theta = (\mu, \sigma^2)$. Hàm mật độ của biến X_i là

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2}.$$

Hàm hợp lý cho bởi

$$L_n(x_1, x_2, \dots, x_n; \mu, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right).$$

Lấy logarit hai vế được hàm log-likelihood

$$\ell_n(\theta) = \ln L_n(x_1, x_2, \dots, x_n; \mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2.$$

Lấy đạo hàm hai vế theo μ và σ^2 và đặt chúng bằng 0

$$\begin{aligned} \frac{\partial \ell_n(\theta)}{\partial \mu} &= \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0; \\ \frac{\partial \ell_n(\theta)}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0. \end{aligned}$$

Từ hai phương trình trên nghiệm duy nhất thu được chính là ước lượng hợp lý cực đại cho hai tham số μ và σ^2

$$\begin{aligned} \mu &= \frac{1}{n} \sum_{i=1}^n x_i; \\ \sigma^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

3.7 Bài toán ước lượng khoảng

3.7.1 Bài toán tổng quát

Phương pháp ước lượng điểm ở trên chủ trương thay thế một điểm cho tham số cần tìm, chất lượng của ước lượng phụ thuộc vào mức độ sai lệch với giá trị tham số thực. Đặc biệt, khi cỡ mẫu nhỏ các ước lượng điểm thường sai lệch khá lớn. Hơn nữa, phương pháp này không đánh giá được khả năng mắc sai lầm của ước lượng là bao nhiêu. Vì vậy, người ta sử dụng phương pháp ước lượng khoảng tin cậy.

Xét tham số cần ước lượng θ của tổng thể, dựa vào mẫu thu được $W = (X_1, X_2, \dots, X_n)$, lập thống kê (hàm ước lượng) $\Theta = f(X_1, X_2, \dots, X_n, \theta)$ có phân phối xác định không phụ thuộc tham số θ . Từ thống kê đó ta tìm hai giá trị Θ_1, Θ_2 của hàm ước lượng sao cho với xác suất $1 - \alpha$ cho trước, tham số θ rơi vào khoảng (Θ_1, Θ_2) . Ta có định nghĩa sau

Định nghĩa 3.7. “Khoảng (Θ_1, Θ_2) được gọi là *khoảng tin cậy* của tham số θ nếu với xác suất $1 - \alpha$ cho trước, đẳng thức sau thỏa mãn

$$\mathbb{P}(\Theta_1 < \theta < \Theta_2) = 1 - \alpha.” \quad (3.3)$$

Khi đó, $1 - \alpha$ gọi là *độ tin cậy*, α là *mức ý nghĩa* hay khả năng mắc sai lầm, $I = \Theta_2 - \Theta_1$ là *độ dài khoảng tin cậy*.

Để giải quyết bài toán tìm hai cận của khoảng tin cậy, từ mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$, lập thống kê $\Theta = f(X_1, X_2, \dots, X_n, \theta)$ sao cho phân phối của Θ hoàn toàn xác định. Với độ tin cậy $1 - \alpha$ cho trước, ta đi tìm α_1 và α_2 là các giá trị thỏa mãn $\alpha_1 + \alpha_2 = \alpha$ ($\alpha_1 \geq 0, \alpha_2 \geq 0$). Khi đó, xét các phân vị g_{α_1} và g_{α_2} tương ứng thỏa mãn hai điều kiện sau

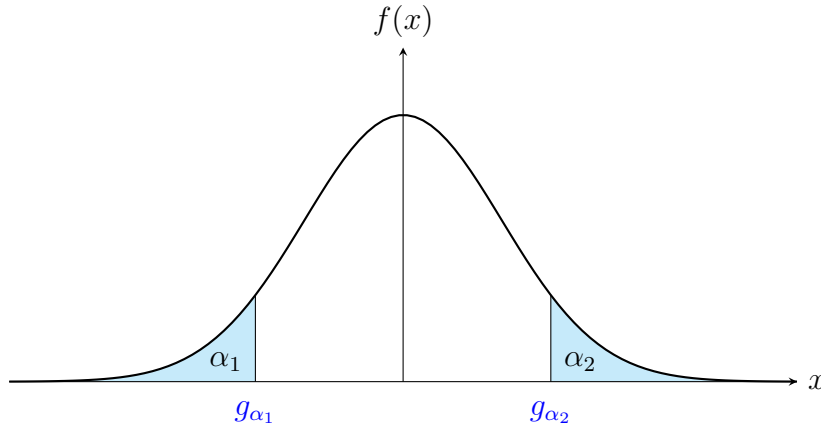
$$\mathbb{P}(\Theta < g_{\alpha_1}) = \alpha_1, \quad (3.4)$$

và

$$\mathbb{P}(\Theta > g_{\alpha_2}) = \alpha_2. \quad (3.5)$$

Từ cách xây dựng α_1, α_2 kết hợp với (3.4) và (3.5) trên

$$\mathbb{P}(g_{\alpha_1} < \Theta < g_{\alpha_2}) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha. \quad (3.6)$$



Hình 3.2 – Khoảng tin cậy $(g_{\alpha_1}; g_{\alpha_2})$ của thống kê Θ

Rõ ràng, khoảng tin cậy $(g_{\alpha_1}; g_{\alpha_2})$ chứa thống kê Θ với độ tin cậy cho trước $1 - \alpha$. Biến đổi tương đương (3.6) dẫn tới

$$\mathbb{P}(\Theta_1 < \theta < \Theta_2) = 1 - \alpha. \quad (3.7)$$

Trong thực tế, độ tin cậy thường được chọn khá cao lên tới 95%, thậm chí 98%, 99% với khả năng mắc sai lầm cho các ước lượng chính là mức ý nghĩa α , khá thấp.

3.7.2 Khoảng tin cậy cho giá trị trung bình

Giả sử biến ngẫu nhiên gốc có phân phối chuẩn $X \sim \mathcal{N}(\mu, \sigma^2)$ với tham số trung bình μ được giả thiết chưa biết, ta phải đi ước lượng μ . Để làm điều này, người ta lấy mẫu ngẫu nhiên kích thước n : $W = (X_1, X_2, \dots, X_n)$ và lập thống kê thích hợp, bài toán được chia làm ba trường hợp sau

a. Trường hợp 1: Đã biết phương sai $\sigma^2 = \sigma_0^2$

Chọn thống kê

$$G = U = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma_0}; \quad (3.8)$$

ở đó, $\bar{X}$ là trung bình mẫu. Dễ thấy, $U \sim \mathcal{N}(0; 1)$, do đó với độ tin cậy $1 - \alpha$ xác định hai giá trị α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$, từ đây tìm được hai phân vị chuẩn u_{α_1} và u_{α_2} của phân phối chuẩn hóa $\mathcal{N}(0; 1)$ sao cho $\mathbb{P}(U < u_{\alpha_1}) = \alpha_1$ và $\mathbb{P}(U > u_{1-\alpha_2}) = \alpha_2$. Từ đó

$$\mathbb{P}(u_{\alpha_1} < U < u_{1-\alpha_2}) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha.$$

Vì phân vị chuẩn $u_{\alpha_1} = -u_{1-\alpha_1}$ nên biểu thức trên trở thành

$$\mathbb{P}(-u_{1-\alpha_1} < U < u_{1-\alpha_2}) = 1 - \alpha. \quad (3.9)$$

Thay (3.8) vào (3.9) được

$$\mathbb{P}\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}}u_{1-\alpha_2} < \mu < \bar{X} + \frac{\sigma_0}{\sqrt{n}}u_{1-\alpha_1}\right) = 1 - \alpha. \quad (3.10)$$

Từ công thức trên ta thấy với độ tin cậy $1 - \alpha$, giá trị trung bình μ của phân phối gốc nằm trong khoảng

$$\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}}u_{1-\alpha_2}; \bar{X} + \frac{\sigma_0}{\sqrt{n}}u_{1-\alpha_1}\right) \quad (3.11)$$

Biểu thức (3.11) cho biết khoảng tin cậy ứng với bộ α_1, α_2 thỏa mãn $\alpha_1 + \alpha_2 = \alpha$. Do với độ tin cậy $1 - \alpha$ cho trước, có vô số cặp giá trị α_1, α_2 thỏa mãn $\alpha_1 + \alpha_2 = \alpha$, vì vậy có vô số khoảng tin cậy tương ứng. Trong thống kê, người ta xét ba loại khoảng tin cậy đặc biệt, dưới đây là chi tiết của chúng.

i. Khoảng tin cậy bên phải

Chọn $\alpha_1 = 0$ và $\alpha_2 = \alpha$, khi đó $u_{1-\alpha_1} = u_1 = +\infty$ nên khoảng tin cậy (3.11) trở thành

$$\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}}u_{1-\alpha}; +\infty\right) \quad (3.12)$$

Công thức (3.12) cho ta cách ước lượng tối thiểu của trung bình μ .

ii. Khoảng tin cậy bên trái

Chọn $\alpha_2 = 0$ và $\alpha_1 = \alpha$, khi đó $u_{1-\alpha_2} = u_1 = +\infty$ nên khoảng tin cậy (3.11) trở thành

$$\left(-\infty; \bar{X} + \frac{\sigma_0}{\sqrt{n}}u_{1-\alpha}\right) \quad (3.13)$$

Công thức (3.13) cho ta cách ước lượng tối đa của trung bình μ .

iii. Khoảng tin cậy đối xứng

Cho $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, (3.11) trở thành

$$\left(\bar{X} - \frac{\sigma_0}{\sqrt{n}}u_{1-\frac{\alpha}{2}}; \bar{X} + \frac{\sigma_0}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right), \quad (3.14)$$

ở đó $\varepsilon = \frac{\sigma_0}{\sqrt{n}}u_{1-\frac{\alpha}{2}}$ là độ chính xác của ước lượng, nó phản ánh mức độ sai lệch trung bình mẫu so với giá trị kì vọng lý thuyết với xác suất $1 - \alpha$ cho trước.

Với cùng một độ tin cậy $1 - \alpha$, khoảng tin cậy có độ dài càng nhỏ càng tốt. Người ta chứng minh được, trong các khoảng tin cậy, khoảng tin cậy đối xứng có độ dài nhỏ nhất là

$$I = \frac{2\sigma_0}{\sqrt{n}} u_{1-\frac{\alpha}{2}}. \quad (3.15)$$

Công thức (3.15) cho mối liên hệ giữa độ tin cậy $1 - \alpha$, kích thước mẫu n và độ chính xác ε (hay độ dài khoảng tin cậy $I = 2\varepsilon$). Mỗi đại lượng được hoàn toàn xác định theo hai đại lượng còn lại. Từ đó, ta có các bài toán sau:

iv. Xác định độ tin cậy của ước lượng

Khi khoảng tin cậy đối xứng có độ dài khoảng tin cậy I_0 (hoặc độ chính xác ước lượng ε_0) thì độ tin cậy $1 - \alpha$ được suy ra từ công thức sau:

$$u_{1-\frac{\alpha}{2}} = \frac{I_0 \sqrt{n}}{2\sigma_0}; \quad \text{hay} \quad u_{1-\frac{\alpha}{2}} = \frac{\varepsilon_0 \sqrt{n}}{\sigma_0}. \quad (3.16)$$

v. Xác định kích thước mẫu tối thiểu

Khi biết trước độ tin cậy $1 - \alpha$ và độ dài khoảng tin cậy I_0 (hoặc độ chính xác ước lượng ε_0) thì kích thước mẫu tối thiểu n_{tt} là số nguyên nhỏ nhất thỏa mãn công thức

$$n_{tt} \geq \frac{4\sigma_0^2}{I_0^2} u_{1-\frac{\alpha}{2}}^2; \quad \text{hay} \quad n_{tt} \geq \frac{\sigma_0^2}{\varepsilon_0^2} u_{1-\frac{\alpha}{2}}^2. \quad (3.17)$$

Bài toán xác định kích thước mẫu tối thiểu n_{tt} theo công thức (3.17) đặt ra trước khi chọn mẫu, khi công việc yêu cầu cần điều tra kích thước mẫu để đạt được độ tin cậy cho trước với độ chính xác cho trước. Rõ ràng, từ công thức trên, độ tin cậy càng cao, độ dài khoảng tin cậy càng nhỏ thì kích thước mẫu tối thiểu cần thiết càng lớn.

Ví dụ 3.12. Một phân xưởng muốn ước lượng thời gian trung bình sản xuất ra một sản phẩm, người ta lấy ngẫu nhiên 36 sản phẩm thấy thời gian trung bình tính được là 1,2 phút/sản phẩm. Biết thời gian trên tuân theo quy luật phân phối chuẩn với phương sai là 0,3 phút.

- Tìm khoảng tin cậy đối xứng với độ tin cậy 95% cho thời gian sản xuất trung bình trên.
- Muốn độ chính xác tăng gấp hai lần, và độ tin cậy không đổi thì cần nghiên cứu mẫu có kích thước tối thiểu bao nhiêu.

Lời giải:

a) Theo đề bài $\bar{x} = 1,2$; $\sigma_0 = 0,3$; $n = 36$ và $1 - \alpha = 0,95$. Tra bảng được $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$. Khoảng tin cậy đối xứng cho trung bình có được từ (3.14) là

$$\begin{aligned} \mu &\in \left(\bar{x} - \frac{\sigma_0}{\sqrt{n}} u_{1-\frac{\alpha}{2}}; \bar{x} + \frac{\sigma_0}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \right); \\ \Leftrightarrow \mu &\in \left(1,2 - \frac{0,3}{\sqrt{36}} 1,96; 1,2 + \frac{0,3}{\sqrt{36}} 1,96 \right); \\ \Leftrightarrow \mu &\in (1,102; 1,298). \end{aligned}$$

b) Ở phần a), độ chính xác của ước lượng là $\varepsilon = \frac{0,3}{\sqrt{36}} 1,96 = 0,098$, do độ chính xác tăng gấp đôi nên $\varepsilon = 0,049$. Theo (3.17), kích thước mẫu tối thiểu thỏa mãn yêu cầu là

$$n_{tt} \geq \frac{\sigma_0^2}{\varepsilon^2} u_{1,975}^2 = \frac{(0,3)^2}{(0,049)^2} (1,96)^2 = 144.$$

Số nguyên n_{tt} nhỏ nhất thỏa mãn điều kiện trên là $n_{tt} = 144$. Vậy kích thước mẫu tối thiểu phải tìm là $n_{tt} = 144$.

b. Trường hợp 2: Chưa biết phương sai σ^2 , kích thước mẫu $n \leq 30$

Trong trường hợp này, chọn thống kê

$$G = T = \frac{(\bar{X} - \mu)\sqrt{n}}{S'}. \quad (3.18)$$

Ta biết rằng, $T \sim T(n-1)$ là thống kê Student với $n-1$ bậc tự do. Việc tính toán các khoảng tin cậy tiến hành tương tự như trên. Cụ thể, với độ tin cậy cho trước $1 - \alpha$, tìm hai phân vị Student $t_{\alpha_1}^{n-1}$ và $t_{1-\alpha_2}^{n-1}$ sao cho điều kiện sau thỏa mãn

$$\begin{aligned} \mathbb{P}(T < t_{\alpha_1}^{n-1}) &= \alpha_1 \text{ và} \\ \mathbb{P}(T > t_{1-\alpha_2}^{n-1}) &= \alpha_2. \end{aligned}$$

Do đó, $\mathbb{P}(t_{\alpha_1}^{n-1} < T < t_{1-\alpha_2}^{n-1}) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha$. Vì phân vị Student có tính chất $-t_{1-\alpha_1}^{n-1} = t_{\alpha_1}^{n-1}$ nên biểu thức trên trở thành

$$\mathbb{P}(-t_{1-\alpha_1}^{n-1} < T < t_{1-\alpha_2}^{n-1}) = 1 - (\alpha_1 + \alpha_2) = 1 - \alpha. \quad (3.19)$$

Thay (3.18) vào (3.19) được

$$\mathbb{P}\left(\bar{X} - \frac{S'}{\sqrt{n}}t_{1-\alpha_2}^{n-1} < \mu < \bar{X} + \frac{S'}{\sqrt{n}}t_{1-\alpha_1}^{n-1}\right) = 1 - \alpha. \quad (3.20)$$

Cũng như trên, từ (3.20) xây dựng được các công thức khoảng tin cậy

i. Khoảng tin cậy bên phải

Cho $\alpha_1 = 0$ và $\alpha_2 = \alpha$, khi đó $t_{1-\alpha_1}^{n-1} = t_1^{n-1} = +\infty$

$$\mu \in \left(\bar{X} - \frac{S'}{\sqrt{n}}t_{1-\alpha}^{n-1}; +\infty\right). \quad (3.21)$$

ii. Khoảng tin cậy bên trái

Cho $\alpha_1 = \alpha$ và $\alpha_2 = 0$, khi đó $t_{1-\alpha_2}^{n-1} = t_1^{n-1} = +\infty$

$$\mu \in \left(-\infty; \bar{X} + \frac{S'}{\sqrt{n}}t_{1-\alpha}^{n-1}\right). \quad (3.22)$$

iii. Khoảng tin cậy đối xứng

Cho $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$

$$\mu \in \left(\bar{X} - \frac{S'}{\sqrt{n}}t_{1-\frac{\alpha}{2}}^{n-1}; \bar{X} + \frac{S'}{\sqrt{n}}t_{1-\frac{\alpha}{2}}^{n-1}\right), \quad (3.23)$$

với độ chính xác $\varepsilon = \frac{S'}{\sqrt{n}}t_{1-\frac{\alpha}{2}}^{n-1}$.

iv. Xác định độ tin cậy của ước lượng

Khi khoảng tin cậy đối xứng có độ dài khoảng tin cậy I_0 (hoặc độ chính xác ước lượng ε_0) thì độ tin cậy $1 - \alpha$ được suy ra từ công thức sau:

$$t_{1-\frac{\alpha}{2}}^{n-1} = \frac{I_0\sqrt{n}}{2S'}; \quad \text{hay} \quad t_{1-\frac{\alpha}{2}}^{n-1} = \frac{\varepsilon_0\sqrt{n}}{S'}. \quad (3.24)$$

v. Xác định kích thước mẫu tối thiểu

Khi biết trước độ tin cậy $1 - \alpha$ và độ dài khoảng tin cậy I_0 (hoặc độ chính xác ước lượng ε_0) thì kích thước mẫu tối thiểu là số nguyên n_{tt} nhỏ nhất thỏa mãn công thức

$$n_{tt} \geq \frac{4S'^2}{I_0^2} t_{1-\frac{\alpha}{2}}^{n-1}; \quad \text{hay} \quad n_{tt} \geq \frac{S'^2}{\varepsilon_0^2} (t_{1-\frac{\alpha}{2}}^{n-1})^2. \quad (3.25)$$

Ví dụ 3.13. Một người chủ trang trại muốn ước lượng trọng lượng trung bình của các con gà trong đàn gà sau khi nuôi được 1 tuần (kể từ lúc nở trứng), ông ta lấy ngẫu nhiên 5 con thấy trọng lượng của chúng lần lượt là 25, 27, 29, 32 và 33 (đơn vị gram). Với độ tin cậy 95%, hãy ước lượng trọng lượng trung bình cao nhất có thể đạt được của các con gà trong đàn gà. Giả thiết trọng lượng này tuân theo luật phân phối chuẩn.

Lời giải: Gọi X (gram) là trọng lượng gà con sau 1 tuần nở, ta có $X \sim \mathcal{N}(\mu, \sigma^2)$. Từ mẫu tính được $n = 5 < 30$, $\bar{x} = 29,2$, độ lệch mẫu hiệu chỉnh $s' = 3,35$. Độ tin cậy $1 - \alpha = 0,95$ nên phân vị Student $t_{0,95}^4 = 2,132$. Do đó từ (3.23), ta có

$$\mu \in \left(-\infty; 29,2 + \frac{3,35}{\sqrt{5}} \cdot 2,132 \right) = (-\infty; 32,39).$$

Ví dụ 3.14. Cho $X(m)$ là chiều dài một loại sản phẩm được sản xuất tự động (X tuân theo quy luật phân phối chuẩn). Kiểm tra ngẫu nhiên 25 sản phẩm ta được bảng số liệu sau

Chiều dài $X(m)$	1,0	1,1	1,2	1,3	1,4
Số sản phẩm	3	6	9	5	2

- Bằng khoảng tin cậy đối xứng hãy ước lượng chiều dài trung bình của loại sản phẩm trên với độ tin cậy 95%.
- Muốn ước lượng chiều dài trung bình của loại sản phẩm trên với độ chính xác 5,63(cm) thì đạt độ tin cậy bao nhiêu?

Lời giải:

a) Đây là bài toán ước lượng bằng khoảng tin cậy đối xứng giá trị của tham số μ của phân phối $\mathcal{N}(\mu, \sigma^2)$ khi chưa biết σ^2 của X , hơn nữa kích thước mẫu $n = 25 < 30$ nên bài toán được giải như sau

Chiều dài trung bình mẫu là $\bar{x} = \frac{1}{n} \sum x_i = 1,188$ (m).

Độ lệch tiêu chuẩn mẫu hiệu chỉnh là

$$s' = \sqrt{\frac{n}{n-1} \left(\frac{\sum n_i x_i^2}{n} - \bar{x}^2 \right)} = \sqrt{\frac{25}{24} \left(\frac{35,59}{25} - 1,188^2 \right)} \approx 0,113 \text{ (m)}.$$

Từ độ tin cậy $1 - \alpha = 0,95$ suy ra $1 - \frac{\alpha}{2} = 0,975$. Tra bảng phân vị Student, ta có $t_{1-\frac{\alpha}{2}}^{n-1} = t_{0,975}^{24} = 2,064$.

Độ chính xác: $\varepsilon = \frac{s'}{\sqrt{n}} t_{1-\frac{\alpha}{2}}^{n-1} = \frac{0,113 \cdot 2,064}{\sqrt{25}} \approx 0,047$ (m).

Vậy khoảng tin cậy đối xứng của μ là

$$1,188 - 0,047 < \mu < 1,188 + 0,047 \quad \text{hay} \quad 1,141 < \mu < 1,235 \text{ (m)}.$$

b) Đây là bài toán xác định độ tin cậy của ước lượng, giả thiết đã cho độ chính xác $\varepsilon = 5,63(\text{cm}) = 0,0563(\text{m})$, và từ bảng tính phần a) ta đã có $s' \approx 0,113(\text{m})$. Từ đó, ta tính được

$$t_{1-\frac{\alpha}{2}}^{n-1} = t_{1-\frac{\alpha}{2}}^{24} = \frac{\varepsilon \sqrt{n}}{s'} = \frac{0,0563 \cdot \sqrt{25}}{0,113} \approx 2,491.$$

Từ bảng phân vị Student, ta thấy $t_{1-\frac{\alpha}{2}}^{24} \approx 2,491 \approx t_{0,99}^{24}$. Vậy

$$1 - \frac{\alpha}{2} = 0,99 \Rightarrow 1 - \alpha = 0,98.$$

Vậy độ tin cậy là $1 - \alpha = 0,98$.

c. Trường hợp 3: Chưa biết phương sai σ^2 , kích thước mẫu $n > 30$

Trong trường hợp này, thống kê (3.18) tiệm cận phân phối chuẩn hóa $\mathcal{N}(0; 1)$, mọi khoảng tin cậy với độ tin cậy $1 - \alpha$ thu được giống như ở trường hợp 2 với lưu ý phân vị Student được thay bằng phân vị chuẩn. Cụ thể như sau

i. Khoảng tin cậy bên phải

Chọn $\alpha_1 = 0$ và $\alpha_2 = \alpha$. Khi đó, $u_{\alpha_1} = u_1 = +\infty$ nên khoảng tin cậy (3.11) trở thành

$$\left(\bar{X} - \frac{S'}{\sqrt{n}} u_{1-\alpha}; +\infty \right). \quad (3.26)$$

Công thức (3.26) cho ta cách ước lượng tối thiểu của trung bình μ .

ii. Khoảng tin cậy bên trái

Chọn $\alpha_2 = 0$ và $\alpha_1 = \alpha$. Khi đó, $u_{1-\alpha_2} = u_1 = +\infty$ nên khoảng tin cậy (3.11) trở thành

$$\left(-\infty; \bar{X} + \frac{S'}{\sqrt{n}} u_{1-\alpha} \right) \quad (3.27)$$

Công thức (3.27) cho ta cách ước lượng tối đa của trung bình μ .

iii. Khoảng tin cậy đối xứng

Cho $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$, (3.11) trở thành

$$\left(\bar{X} - \frac{S'}{\sqrt{n}} u_{1-\frac{\alpha}{2}}; \bar{X} + \frac{S'}{\sqrt{n}} u_{1-\frac{\alpha}{2}} \right), \quad (3.28)$$

ở đó $\varepsilon = \frac{S'}{\sqrt{n}} u_{1-\frac{\alpha}{2}}$ là độ chính xác của ước lượng, nó phản ánh mức độ sai lệch trung bình mẫu so với giá trị kỳ vọng lý thuyết với độ tin cậy $1 - \alpha$ cho trước.

iv. Xác định độ tin cậy của ước lượng

Khi khoảng tin cậy đối xứng có độ dài khoảng tin cậy I_0 (hoặc độ chính xác ước lượng ε_0) thì độ tin cậy $1 - \alpha$ được suy ra từ công thức sau:

$$u_{1-\frac{\alpha}{2}} = \frac{I_0\sqrt{n}}{2S'}; \quad \text{hay} \quad u_{1-\frac{\alpha}{2}} = \frac{\varepsilon_0\sqrt{n}}{S'}. \quad (3.29)$$

v. Xác định kích thước mẫu tối thiểu

Khi biết trước độ tin cậy $1 - \alpha$ và độ dài khoảng tin cậy I_0 (hoặc độ chính xác ước lượng ε_0) thì kích thước mẫu n_{tt} là số nguyên thỏa mãn công thức

$$n_{tt} \geq \frac{4S'^2}{I_0^2} u_{1-\frac{\alpha}{2}}^2; \quad \text{hay} \quad n_{tt} \geq \frac{S'^2}{\varepsilon_0^2} u_{1-\frac{\alpha}{2}}^2. \quad (3.30)$$

Ví dụ 3.15. Cho $X(\text{kg})$ là trọng lượng một loại sản phẩm (X tuân theo quy luật phân phối chuẩn). Kiểm tra ngẫu nhiên trên 100 sản phẩm cùng loại người ta tính được trọng lượng trung bình $\bar{x} = 3,645(\text{kg})$ và độ lệch tiêu chuẩn điều chỉnh mẫu $s' = 0,785(\text{kg})$.

- Dựa vào số liệu trên, hãy ước lượng trọng lượng trung bình tối đa của loại sản phẩm đó với độ tin cậy 97%;
- Muốn ước lượng trọng lượng trung bình của loại sản phẩm trên với độ tin cậy 99%, độ dài khoảng tin cậy đối xứng không quá 310(g) thì cần điều tra thêm bao nhiêu sản phẩm nữa?

Lời giải: Đây là bài toán ước lượng giá trị của tham số μ của phân phối $\mathcal{N}(\mu, \sigma^2)$ khi chưa biết σ^2 của X , hơn nữa kích thước mẫu $n = 100 > 30$ nên bài toán được giải như sau

- Giả thiết đã cho $\bar{x} = 3,645(\text{kg})$ và $s' = 0,785(\text{kg})$.

Từ độ tin cậy $1 - \alpha = 0,97$ tra bảng phân vị chuẩn, ta có $u_{1-\alpha} = u_{0,97} = 1,881$.

Độ chính xác một phía là $\varepsilon = \frac{s'}{\sqrt{n}} u_{1-\alpha} = \frac{0,785 \cdot 1,881}{\sqrt{100}} \approx 0,148 (\text{kg})$.

Vậy khoảng tin cậy bên trái của μ là

$$\mu \in (-\infty; 3,645 + 0,148) \text{ hay } \mu < 3,793 (\text{kg}).$$

- Đây là bài toán xác định kích thước mẫu của ước lượng

Từ độ tin cậy $1 - \alpha = 0,99$, suy ra $1 - \frac{\alpha}{2} = 0,995$. Tra bảng phân vị chuẩn: $u_{1-\frac{\alpha}{2}} = u_{0,995} = 2,576$.

Độ dài khoảng tin cậy đối xứng: $I_0 = 2\varepsilon = 310 \text{ kg}$, suy ra độ chính xác $\varepsilon = 155(\text{g}) = 0,155(\text{kg})$.

Ta tính được: $n_{tt} \geq \left(\frac{s' \cdot u_{1-\frac{\alpha}{2}}}{\varepsilon} \right)^2 = \left(\frac{0,785 \cdot 2,576}{0,155} \right)^2 \approx 170,2$. Suy ra kích thước mẫu tối thiểu là $n_{tt} = 171$.

Vậy cần điều tra thêm ít nhất $171 - 100 = 71$ sản phẩm nữa.

3.7.3 Khoảng tin cậy cho tỉ lệ

Trên thực tế, để ước lượng trung bình μ của biến ngẫu nhiên X trong tổng thể tuân theo một quy luật phân phối nào đó khác phân phối chuẩn, người ta thường xét thống kê

$$G = \Theta = \frac{(\bar{X} - \mu)\sqrt{n}}{\sigma},$$

trong trường hợp phương sai σ^2 đã biết. Nếu phương sai chưa biết thì xét thống kê sau

$$G = \Theta = \frac{(\bar{X} - \mu)\sqrt{n}}{S'},$$

Theo định lý giới hạn trung tâm, cả hai thống kê trên đều tiệm cận phân phối chuẩn hóa $\mathcal{N}(0; 1)$ nên ta có thể dùng các thủ tục ước lượng ở trên tiến hành một cách tương tự.

Xét bài toán ước lượng khoảng cho cơ cấu tổng thể. Giả sử trong tổng thể kích thước N có M phần tử mang dấu hiệu nghiên cứu. Như vậy cơ cấu của tổng thể theo dấu hiệu nghiên cứu (hay tỉ lệ phần tử mang dấu hiệu nghiên cứu trong tổng thể) là:

$$p = \frac{M}{N}$$

Nếu lấy ngẫu nhiên ra một phần tử, và gọi X là số phần tử mang dấu hiệu nghiên cứu được lấy ra thì X là biến ngẫu nhiên phân phối theo quy luật không - một $\mathcal{A}(p)$ với kỳ vọng toán $\mathbb{E}(X) = p$ và phương sai $\text{var}(X) = \frac{p(1-p)}{n}$. Do đó bài toán ước lượng cơ cấu tổng thể chính là bài toán ước lượng kỳ vọng toán của biến ngẫu nhiên tuân theo quy luật không - một.

Xét mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$ kích thước n rút ra từ biến ngẫu nhiên gốc X , các biến X_i độc lập, cùng phân phối với X . Xét thống kê

$$f = \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Dễ thấy $\mathbb{E}(\bar{X}) = p$; $\text{var}(\bar{X}) = \frac{p(1-p)}{n}$ và f là ước lượng không chệch của cơ cấu tổng thể. Xét thống kê

$$U = \frac{f - p}{\sqrt{f(1-f)}} \sqrt{n} \sim \mathcal{N}(0, 1) \text{ khi } n \text{ đủ lớn.} \quad (3.31)$$

Với độ tin cậy $1 - \alpha$ cho trước, ta xây dựng các khoảng tin cậy cho tỉ lệ p tương tự như thống kê (3.8). Các kết quả thu được lần lượt là

i. Khoảng tin cậy bên phải

Khoảng tin cậy bên phải là

$$p \in \left(f - \sqrt{\frac{f(1-f)}{n}} u_{1-\alpha}; +\infty \right). \quad (3.32)$$

ii. Khoảng tin cậy bên trái

Khoảng tin cậy bên trái là

$$p \in \left(-\infty; f + \sqrt{\frac{f(1-f)}{n}} u_{1-\alpha} \right). \quad (3.33)$$

iii. Khoảng tin cậy đối xứng

Khoảng tin cậy đối xứng cho tỉ lệ p là

$$p \in \left(f - \sqrt{\frac{f(1-f)}{n}} u_{1-\frac{\alpha}{2}}; f + \sqrt{\frac{f(1-f)}{n}} u_{1-\frac{\alpha}{2}} \right), \quad (3.34)$$

ở đó $\varepsilon = \sqrt{\frac{f(1-f)}{n}} u_{1-\frac{\alpha}{2}}$ là độ chính xác của ước lượng. Độ dài khoảng tin cậy đối xứng là:

$$I = 2\varepsilon = 2\sqrt{\frac{f(1-f)}{n}} u_{1-\frac{\alpha}{2}}. \quad (3.35)$$

iv. Xác định độ tin cậy của ước lượng

Khi khoảng tin cậy đối xứng có độ dài khoảng tin cậy I_0 (hoặc độ chính xác ước lượng ε_0) thì độ tin cậy $1 - \alpha$ được suy ra từ công thức sau:

$$u_{1-\frac{\alpha}{2}} = \frac{I_0 \sqrt{n}}{2\sqrt{f(1-f)}}; \quad \text{hay} \quad u_{1-\frac{\alpha}{2}} = \frac{\varepsilon_0 \sqrt{n}}{\sqrt{f(1-f)}}. \quad (3.36)$$

v. Xác định cỡ mẫu tối thiểu

Khi biết trước độ tin cậy $1 - \alpha$ và độ dài khoảng tin cậy I_0 (hoặc độ chính xác ước lượng ε_0) thì kích thước mẫu tối thiểu là số nguyên n_{tt} nhỏ nhất thỏa mãn

$$n_{tt} \geq \frac{4f(1-f)}{I_0^2} u_{1-\frac{\alpha}{2}}^2; \quad \text{hay} \quad n_{tt} \geq \frac{f(1-f)}{\varepsilon_0^2} u_{1-\frac{\alpha}{2}}^2. \quad (3.37)$$

Ví dụ 3.16. Trong cuộc khảo sát về số lượng cử tri ủng hộ dự luật A ở một khu vực có số cử tri là 300000 người, người ta phỏng vấn ngẫu nhiên 400 cử tri thấy có 240 người ủng hộ.

- Hãy xác định khoảng tin cậy đối xứng số cử tri ủng hộ trong vùng với độ tin cậy 95%.
- Muốn ước lượng tỉ lệ cử tri ủng hộ dự luật với độ tin cậy 98%, độ chính xác không quá 0,05 thì cần phỏng vấn ít nhất bao nhiêu cử tri?
- Muốn ước lượng tỉ lệ cử tri ủng hộ với độ dài khoảng tin cậy không quá 6% thì độ tin cậy đạt được là bao nhiêu?

Lời giải:

a) Gọi f là tỉ lệ cử tri ủng hộ dự luật A. Ta có $f = \frac{240}{400} = 0,6$; độ tin cậy $1 - \alpha = 0,95$ nên phân vị chuẩn mức $1 - \frac{\alpha}{2} = 0,975$ bằng $u_{0,975} = 1,96$. Khoảng tin cậy cho tỉ lệ cử tri ủng hộ trong vùng là

$$\left(0,6 - \sqrt{\frac{0,6 \cdot 0,4}{400}} \cdot 1,96 ; 0,6 + \sqrt{\frac{0,6 \cdot 0,4}{400}} \cdot 1,96 \right) = (0,552 ; 0,648).$$

Suy ra số cử tri ủng hộ dự luật trong khu vực là

$$(0,552 \cdot 300000 ; 0,648 \cdot 300000) = (165600 ; 194400).$$

b) Độ tin cậy $1 - \alpha = 0,98$ nên $u_{1-\alpha/2} = u_{0,99} = 2,326$, độ chính xác ước lượng không quá $\varepsilon = 0,05$ nên để xác định số cử tri tối thiểu cần phỏng vấn, sử dụng công thức (3.35) suy ra

$$n_{tt} \geq \left(\frac{u_{1-\frac{\alpha}{2}}}{\varepsilon} \right)^2 \cdot f(1-f) = \left(\frac{2,326}{0,05} \right)^2 \cdot 0,6 \cdot 0,4 \approx 519,4.$$

Suy ra, $n_{tt} = 520$. Vậy cần phỏng vấn ít nhất 520 cử tri trong vùng.

c) Độ dài khoảng tin cậy đối xứng $I_0 = 6\%$, suy ra độ chính xác $\varepsilon = 0,03$

$$u_{1-\frac{\alpha}{2}} = \frac{\varepsilon\sqrt{n}}{\sqrt{f \cdot (1-f)}} = \frac{0,03\sqrt{400}}{\sqrt{0,6 \cdot 0,4}} \approx 1,225$$

Từ bảng phân vị chuẩn ta thấy $1,225 \approx u_{0,89}$. Vậy $1 - \frac{\alpha}{2} = 0,89$ suy ra độ tin cậy $1 - \alpha = 0,78$.

Bài tập chương 3

MẪU NGẪU NHIÊN VÀ CÁC THAM SỐ ĐẶC TRƯNG CỦA NÓ

Bài 3.1. Đo 10 lần một đại lượng ngẫu nhiên X bởi một máy không có sai số hệ thống, ta thu được các kết quả sau:

369; 378; 315; 320; 325; 420; 385; 401; 372; 383

Hãy tính $\bar{x}$; s^2 ; s'^2 .

Bài 3.2. Để kiểm tra giá cả trung bình của một loại hàng hóa trên thị trường, người ta khảo sát ngẫu nhiên tại 100 cửa hàng có cùng quy mô trên địa bàn thành phố và thu được bảng số liệu sau

Giá cả (ngàn đồng)	83	85	87	89	91	93	95	97	99	101
Số cửa hàng	6	7	12	15	30	10	8	6	4	2

- Hãy vẽ đường đa giác tần số của bảng phân phối thực nghiệm trên;
- Tìm giá cả trung bình và độ phân tán của giá cả so với giá cả trung bình của các cửa hàng nói trên.

Bài 3.3. Đo chiều cao X (cm) của 100 sinh viên của trường đại học A, ta thu được bảng số liệu

X	154-158	158-162	162-166	166-170	170-174	174-178	178-182
n_i	10	14	26	28	12	8	2

Hãy tính chiều cao trung bình và độ lệch tiêu chuẩn hiệu chỉnh mẫu.

Bài 3.4. Ở một trang trại nuôi heo, người ta áp dụng thử nghiệm một loại thuốc tăng trọng được thêm vào trong thức ăn. Sau khi nuôi 3 tháng thì thu được bảng số liệu như sau

$X(kg)$	75	77	78	79	80	81	83
n_i (số heo)	10	30	40	70	60	20	20

Còn khi không dùng thuốc tăng trọng, thì ta có lô đối chứng như sau

$X(kg)$	60	61	62	63	65	67	69	70	72
n_i (số heo)	10	20	40	50	60	30	20	10	10

Hãy tìm khối lượng trung bình và mức độ phân tán trọng lượng so với trọng lượng trung bình của hai lô heo nói trên.

Bài 3.5. Điều tra tỉ lệ thu hồi vốn nếu đầu tư vào ngành A qua một số dự án ta có bảng dữ liệu sau

Tỉ lệ (%/năm)	10-12	12-14	14-16	16-18	18-20
Số dự án	9	7	6	5	4

Hãy tính tỉ lệ thu hồi vốn trung bình khi đầu tư vào ngành A và độ lệch tiêu chuẩn hiệu chỉnh mẫu.

Bài 3.6. Cho X (%) , Y (kg/mm^2) là 2 chỉ tiêu đánh giá chất lượng của một loại sản phẩm. Kiểm tra ngẫu nhiên một lô sản phẩm ta có bảng số liệu sau

$Y \backslash X$	0-5	5-10	10-15	15-20	20-25
115-125	7				
125-135	12	8	10		
135-145		20	15	2	
145-155		19	16	9	7

- Hãy tính trung bình mẫu, phương sai mẫu của các chỉ tiêu X và Y ;
- Những sản phẩm có chỉ tiêu $X \geq 15\%$ được xếp vào loại A. Hãy tính trung bình mẫu, phương sai mẫu của chỉ tiêu Y các sản phẩm loại A.

Bài 3.7. Gọi X là "Khối lượng sản phẩm sản xuất ra" và Y là "giá thành sản phẩm", điều tra ở 30 xí nghiệp cùng quy mô và cùng sản xuất loại sản phẩm ta thu được bảng dữ liệu sau

$Y \backslash X$	50	100	150	200	250
100-120				4	4
120-140		2	6	1	1
140-160	1	4	2		
160-180	3		1		1

- Tính trung bình mẫu và độ lệch tiêu chuẩn hiệu chỉnh mẫu $\bar{x}$, s'_X , $\bar{y}$, s'_Y ;
- Những xí nghiệp có $X > 50$ và $120 \leq Y < 160$ được xếp vào loại I, hãy tính các giá trị trung bình mẫu và độ lệch tiêu chuẩn hiệu chỉnh mẫu của các xí nghiệp loại I;
- Tính tần suất mẫu của các xí nghiệp loại I.

ƯỚC LƯỢNG CÁC THAM SỐ ĐẶC TRƯNG CỦA ĐẠI LƯỢNG NGẪU NHIÊN

Ước lượng trung bình tổng thể

Bài 3.8. Một công ty điện thoại muốn ước lượng độ dài trung bình của các cuộc điện thoại đường dài cuối tuần. Kiểm tra ngẫu nhiên 50 cuộc gọi cho biết độ dài trung bình cuộc gọi là 14,5 phút, độ lệch tiêu chuẩn hiệu chỉnh là 5,6 phút. Giả thiết rằng độ dài của các cuộc điện thoại là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn. Với độ tin cậy 95%, bằng khoảng tin cậy đối xứng, hãy ước lượng độ dài trung bình của các cuộc điện thoại đường dài trong các dịp cuối tuần.

Bài 3.9. Doanh số của một cửa hàng là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn với độ lệch tiêu chuẩn hiệu chỉnh là 2 triệu đồng/tháng. Khảo sát ngẫu nhiên doanh số của 600 cửa hàng có cùng quy mô thấy doanh số trung bình là 8,5 triệu đồng. Bằng khoảng tin cậy đối xứng, hãy ước lượng doanh số trung bình của các đó với độ chính xác 96%.

Bài 3.10. Độ dày của tấm kính chắn gió ô tô là đại lượng ngẫu nhiên phân phối chuẩn với độ lệch tiêu chuẩn là 0,3 mm. Kiểm tra ngẫu nhiên 50 tấm kính chắn gió thấy độ dày trung bình của chúng là 4,05 mm.

- Bằng khoảng tin cậy đối xứng, hãy ước lượng độ dày trung bình của các tấm kính trên với độ tin cậy 95%;

- b) Muốn ước lượng độ dày trung bình của tấm kính với độ tin cậy 95% và sai số ước lượng không vượt quá 0,025 mm cần kiểm tra thêm ít nhất bao nhiêu tấm kính nữa?

Bài 3.11. Hao phí nguyên liệu cho một đơn vị sản phẩm là một biến ngẫu nhiên phân phối chuẩn. Khảo sát một số sản phẩm và thu được bảng số liệu sau

Mức hao phí nguyên liệu(gam)	Số sản phẩm tương ứng
19,5-19,7	8
19,7-19,9	8
19,9-20,1	18
20,1-20,3	2

- a) Bằng khoảng tin cậy đối xứng, hãy ước lượng mức hao phí nguyên liệu trung bình cho một đơn vị sản phẩm với độ tin cậy 95%;
- b) Với độ tin cậy 97% hãy ước lượng mức hao phí nguyên liệu tối đa cho một đơn vị sản phẩm nói trên;
- c) Muốn ước lượng mức hao phí nguyên liệu trung bình của một đơn vị sản phẩm trên với độ tin cậy 98%, độ chính xác 0,0095 gam thì cần phải khảo sát thêm ít nhất bao nhiêu sản phẩm nữa?

Bài 3.12. Chiều dài X (cm) của một loại sản phẩm là đại lượng ngẫu nhiên phân phối chuẩn với độ lệch tiêu chuẩn $\sigma_X = 3$ cm. Đo 20 sản phẩm và thấy chiều dài trung bình của 20 sản phẩm đó là $\bar{x} = 50$ cm.

- a) Bằng khoảng tin cậy đối xứng, hãy ước lượng chiều dài trung bình loại sản phẩm đó với độ tin cậy 95%;
- b) Hãy ước lượng chiều dài trung bình tối thiểu, chiều dài trung bình tối đa của loại sản phẩm trên với độ tin cậy 98%;
- c) Để ước lượng chiều dài trung bình loại sản phẩm đó với độ tin cậy 99%, độ chính xác 0,3 cm thì phải đo ít nhất bao nhiêu sản phẩm?
- d) Muốn sử dụng kết quả đo của 20 sản phẩm trên để ước lượng chiều dài trung bình của loại sản phẩm trên với độ chính xác 1,378 cm thì đạt độ tin cậy bao nhiêu?

Bài 3.13. Người ta theo dõi giá đóng cửa của một loại cổ phiếu qua một số ngày và thu được bảng số liệu sau

Giá	25-29	29-31	31-33	33-35	35-39
Số ngày	11	22	34	21	12

Giả thiết rằng giá đóng cửa của loại cổ phiếu này là biến ngẫu nhiên phân phối chuẩn. Với độ tin cậy 97%, hãy ước lượng giá đóng cửa cổ phiếu trung bình bằng khoảng tin cậy đối xứng.

Bài 3.14. Năng suất giống ngô A là đại lượng ngẫu nhiên phân phối chuẩn. Theo dõi 25 điểm thu hoạch giống ngô A ở một vùng và có bảng dữ liệu sau

Năng suất(tạ/ha)	7	9	11	13	17
Số điểm thu hoạch	2	7	12	3	1

Với độ tin cậy 95%, hãy ước lượng năng suất trung bình tối thiểu của giống ngô A ở vùng này.

Bài 3.15. Lượng điện tiêu thụ một tháng của mỗi hộ gia đình là đại lượng ngẫu nhiên phân phối chuẩn. Theo dõi lượng điện tiêu thụ của 41 hộ gia đình ở nông thôn ta có số liệu như sau.

Lượng điện tiêu thụ(kWh /tháng)	100	110	120	139	140
Số hộ gia đình	15	10	7	6	3

Với độ tin cậy 95%, bằng khoảng tin cậy đối xứng, hãy ước lượng lượng điện tiêu thụ trung bình tối đa của các hộ gia đình ở nông thôn.

Bài 3.16. Thu nhập hàng tháng của công nhân ở một khu công nghiệp là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn. Theo dõi thu nhập hàng tháng của 30 công nhân ta có kết quả sau

Thu nhập (triệu đồng)	6-8	8-10	10-12	12-14	14-16
Số công nhân	6	13	6	3	2

- Với độ tin cậy 95%, bằng khoảng tin cậy đối xứng, hãy ước lượng mức thu nhập trung bình của công nhân;
- Hãy ước lượng mức thu nhập trung bình tối thiểu, mức thu nhập trung bình tối đa của công nhân với độ tin cậy 97%;
- Muốn giữ nguyên độ tin cậy 95% và để độ chính xác của ước lượng tăng lên gấp đôi thì cần điều tra thêm ít nhất bao nhiêu công nhân nữa?

Ước lượng tỉ lệ tổng thể

Bài 3.17. Kiểm tra 200 hộp sữa của một kho hàng thấy có 8 hộp bị hỏng. Bằng khoảng tin cậy đối xứng, hãy ước lượng tỉ lệ đồ hộp bị hỏng ở kho hàng đó với độ tin cậy 95%.

Bài 3.18. Trong chiến dịch tranh cử tổng thống, trong số 1600 cử tri được phỏng vấn ngẫu nhiên có 960 người sẽ bỏ phiếu ủng hộ cho ứng cử viên A. Với độ tin cậy 99%, hãy ước lượng tỉ lệ phiếu bầu tối thiểu ủng hộ ứng cử viên A.

Bài 3.19. Để kiểm tra chất lượng của một loại sản phẩm do một nhà máy sản xuất, người ta kiểm tra ngẫu nhiên 400 sản phẩm thì thấy có 160 sản phẩm thuộc loại I. Với độ tin cậy 95%, hãy ước lượng tỉ lệ tối đa các sản phẩm thuộc loại I của nhà máy .

Bài 3.20. Tỉ lệ nảy mầm của một loại hạt giống là 90%. Nếu muốn ước lượng tỉ lệ nảy mầm của loại hạt giống trên với độ tin cậy 95% và độ chính xác 0,01 thì phải gieo ít nhất là bao nhiêu hạt?

Bài 3.21. Kiểm tra ngẫu nhiên 100 người thì thấy 20 người mắc bệnh viêm gan B. Nếu muốn sai số ước lượng không vượt quá 2% với độ tin cậy 95% thì phải kiểm tra thêm ít nhất bao nhiêu người?

Bài 3.22. Để khảo sát hiệu quả của một loại thuốc mới, người ta đem loại thuốc mới đó đi điều trị cho 50 người mắc bệnh A thì có 40 người khỏi bệnh.

- Với độ tin cậy 95%, bằng khoảng tin cậy đối xứng, hãy ước lượng tỉ lệ người khỏi bệnh khi được điều trị bằng loại thuốc đó.

- b) Để ước lượng tỉ lệ người khỏi bệnh với độ tin cậy 95%, độ chính xác là 2% thì phải theo dõi thêm ít nhất bao nhiêu người dùng thuốc nữa?

BÀI TẬP TỔNG HỢP

Bài 3.23. Để khảo sát trọng lượng X (gam) của trái quýt đường người ta cân một số trái và thu được kết quả sau (X tuân theo quy luật phân phối chuẩn).

$X(\text{gam})$	130	140	150	160	170	180	190	200
Số trái	4	9	11	16	28	15	10	7

Những trái quýt đường có trọng lượng ≥ 160 gam được xếp vào loại I.

- Bằng khoảng tin cậy đối xứng, hãy ước lượng trọng lượng trung bình của các trái quýt đường trên với độ tin cậy 95%;
- Bằng khoảng tin cậy đối xứng, hãy ước lượng tỉ lệ trái quýt đường loại I trên với độ tin cậy 97%.

Bài 3.24. Cân nặng X (kg) của một loại vật nuôi là đại lượng ngẫu nhiên phân phối chuẩn. Cân thử 100 con trong nông trại người ta thu được bảng dữ liệu sau

$X(\text{kg})$	36	42	48	54	60	66	72
Số con	15	12	25	18	10	10	10

Những con vật có cân nặng từ 60 kg trở lên được xếp vào loại I.

- Bằng khoảng tin cậy đối xứng hãy ước lượng cân nặng trung bình của vật nuôi nói trên với độ tin cậy 96%;
- Với độ tin cậy 95% hãy ước lượng cân nặng trung bình tối đa, cân nặng trung bình tối thiểu của loại vật nuôi nói trên;
- Bằng khoảng tin cậy đối xứng, hãy ước lượng tỉ lệ những con đạt loại I với độ tin cậy 95%;
- Nếu muốn ước lượng tỉ lệ những con loại I với độ tin cậy 99% và sai số ước lượng không vượt quá 10% thì cần phải kiểm tra thêm ít nhất bao nhiêu con nữa?
- Hãy ước lượng tỉ lệ tối đa vật nuôi thuộc loại I với độ tin cậy 90%.

Bài 3.25. Chỉ tiêu X của một loại sản phẩm là đại lượng ngẫu nhiên phân phối chuẩn. Quan sát 100 sản phẩm trong kho người ta thu được bảng dữ liệu sau

$X(\text{cm})$	10-14	14-18	18-22	22-26	26-30	30-34	34-38
Số sản phẩm	7	10	19	17	15	14	18

Những sản phẩm có chỉ tiêu X nhỏ hơn 18 cm được xếp vào loại II.

- Bằng khoảng tin cậy đối xứng, hãy ước lượng giá trị trung bình chỉ tiêu X của loại sản phẩm nói trên với độ tin cậy 96%.
- Để ước lượng giá trị trung bình của chỉ tiêu X với độ chính xác 1,7 cm thì đạt được độ tin cậy là bao nhiêu?
- Để ước lượng giá trị trung bình của X với độ tin cậy 99%, sai số ước lượng không vượt quá 1,5 cm thì phải điều tra thêm ít nhất bao nhiêu sản phẩm nữa?
- Với độ tin cậy 98%, bằng khoảng tin cậy đối xứng, hãy ước lượng chỉ tiêu X của những sản phẩm loại II.

- e) Để ước lượng tỉ lệ sản phẩm loại II với độ tin cậy 99%, độ chính xác ước lượng không vượt quá 10% thì cần điều tra thêm ít nhất bao nhiêu sản phẩm nữa?

Bài 3.26. Đường kính $X(\text{cm})$ và chiều cao $Y(\text{m})$ của một loại cây công nghiệp là các đại lượng ngẫu nhiên phân phối chuẩn. Quan sát 100 cây công nghiệp, người ta thu được bảng dữ liệu sau

$\begin{matrix} Y \\ \backslash X \end{matrix}$	20-22	22-24	24-26	26-28	28-30
3	2				
4	5	3			
5		5	8	4	
6			15	17	
7			10	6	7
8					12

Những cây có chiều cao ≥ 7 m được xếp vào loại I.

- Bằng khoảng tin cậy đối xứng, hãy ước lượng chiều cao trung bình của các cây nói trên với độ tin cậy 95%;
- Hãy ước lượng chiều cao trung bình tối thiểu của cây với độ tin cậy 96%;
- Muốn ước lượng đường kính trung bình của cây công nghiệp trên với độ tin cậy 95% và độ dài khoảng tin cậy là 10 mm thì cần điều tra thêm ít nhất bao nhiêu cây nữa?
- Bằng khoảng tin cậy đối xứng, hãy ước lượng tỉ lệ cây loại I với độ tin cậy 97%;
- Để ước lượng tỉ lệ cây loại I với độ chính xác là 0,0977 thì đảm bảo độ tin cậy là bao nhiêu?

Bài 3.27. Điều tra hai chỉ tiêu $X(\%)$ và $Y(\text{cm})$ của một loại sản phẩm, người ta thu được bảng số liệu sau. Giả thiết rằng X, Y tuân theo quy luật phân phối chuẩn.

$\begin{matrix} Y \\ \backslash X \end{matrix}$	0-2	2-4	4-8	8-10	10-12
100-105	5				
105-110	7	10			
110-115	3	9	16	9	
115-120		8	25	8	
120-125		15	13	17	8
125-130			15	11	9
130-135				14	6
135-140					5

Những sản phẩm có chỉ tiêu $X < 2\%$ được xếp vào loại II.

- Muốn ước lượng trung bình chỉ tiêu X với độ chính xác là 0,3% thì đạt độ tin cậy là bao nhiêu?
- Bằng khoảng tin cậy đối xứng, hãy ước lượng trung bình chỉ tiêu Y của sản phẩm loại II với độ tin cậy 95%;
- Bằng khoảng tin cậy đối xứng, hãy ước lượng tỉ lệ những sản phẩm loại II với độ tin cậy 96%.

Bài 3.28. Tỷ lệ thu nhập chi cho giáo dục $X(\%)$ và thu nhập bình quân của một hộ gia đình trong một tháng Y (triệu đồng) là các đại lượng ngẫu nhiên phân phối chuẩn. Điều tra thu nhập bình quân và tỷ lệ thu nhập chi cho giáo dục của một số hộ gia đình ta có bảng số liệu sau.

$Y \backslash X$	10	20	30	40	50
15-25	10	40	20		
25-35		40	60	20	
35-45		20	30	40	
45-55			30	30	10

Những hộ gia đình có thu nhập bình quân trên 45 triệu đồng được xếp vào hộ gia đình có thu nhập cao.

- Bằng khoảng tin cậy đối xứng, hãy ước lượng trung bình tỷ lệ thu nhập chi cho giáo dục của các hộ gia đình với độ tin cậy 95%;
- Bằng khoảng tin cậy đối xứng, hãy ước lượng tỷ lệ hộ có thu nhập cao với độ tin cậy 97%;
- Để ước lượng giá trị trung bình của tỷ lệ thu nhập chi cho giáo dục với độ chính xác 0,8% thì đảm bảo độ tin cậy bao nhiêu?

Bài 3.29. $X(\%)$ và Y (kg/mm^2) là hai chỉ tiêu chất lượng của một loại sản phẩm. Khảo sát một số sản phẩm người ta thu được kết quả sau (X, Y là các đại lượng ngẫu nhiên phân phối chuẩn).

$X \backslash Y$	0-5	5-10	10-15	15-20	20-25
120	7				
130	12	8	10		
140		20	15	2	
150		19	16	9	5
160				8	3

Những sản phẩm có chỉ tiêu $X \geq 15\%$ được xếp vào loại A.

- Bằng khoảng tin cậy đối xứng hãy ước lượng tỷ lệ sản phẩm loại A với độ tin cậy 95%;
- Bằng khoảng tin cậy đối xứng, hãy ước lượng giá trị trung bình của chỉ tiêu Y của các sản phẩm loại A với độ tin cậy 95%;
- Để ước lượng giá trị trung bình của chỉ tiêu Y với độ chính xác 1,6 kg/mm^2 thì sẽ đạt độ tin cậy là bao nhiêu?

Bài 3.30. X (cm) và Y (kg) là hai chỉ tiêu chất lượng của một loại sản phẩm. Khảo sát một số sản phẩm ta thu được bảng số liệu sau (X, Y là các đại lượng ngẫu nhiên phân phối chuẩn) Những sản phẩm có chỉ tiêu $X > 13$ (cm) $Y \geq 5$ (kg) được xếp vào loại A.

- Bằng khoảng tin cậy đối xứng, hãy ước lượng cho trung bình chỉ tiêu X với độ tin cậy 98%;
- Muốn ước lượng tỷ lệ sản phẩm loại A với độ tin cậy 95%, độ dài khoảng tin cậy là 0,3 thì cần điều tra thêm ít nhất bao nhiêu sản phẩm nữa?
- Muốn ước lượng trung bình chỉ tiêu Y với độ chính xác 539 (g) thì đảm bảo độ tin cậy bao nhiêu?

Y \ X	12	13	14	15	16
0-1	1	2	1		
1-2	1	2	3	2	
2-3	1	1	2	3	
3-4		1	4	2	
4-5		2	3	5	2
5-6		2	4	6	1
6-7			3	4	2
7-8				3	2

- Muốn ước lượng trung bình chỉ tiêu X với độ tin cậy 99%, độ chính xác 3 (mm) thì cần điều tra thêm ít nhất bao nhiêu sản phẩm nữa?
- Muốn ước lượng cho trung bình chỉ tiêu X của các sản phẩm loại A với độ chính xác 0,393 (cm) thì đảm bảo độ tin cậy bao nhiêu?
- Muốn ước lượng trung bình chỉ tiêu Y của các sản phẩm loại A với độ tin cậy 95%, độ chính xác của ước lượng là 0,25 (kg) thì cần điều tra thêm ít nhất bao nhiêu sản phẩm nữa?
- Muốn ước lượng tỉ lệ sản phẩm loại A với độ chính xác 13,89% thì đạt độ tin cậy bao nhiêu?

Bài 3.31. Nghiên cứu hai chỉ tiêu X (cm), Y (m) của một số sản phẩm cùng loại ta có bảng dữ liệu sau (Giả sử X,Y tuân theo quy luật phân phối chuẩn).

X \ Y	5	6	7	8	9
0-20	3	5	2		
20-30	1	2	3	1	1
30-40		1	2	2	1
40-50			2	4	3
50-70				3	2

Sản phẩm có chỉ tiêu $X \geq 30$ (cm) $Y \geq 7$ (m) là loại A.

- Muốn ước lượng trung bình chỉ tiêu X với độ tin cậy 96%, độ dài khoảng tin cậy là 80 (mm) thì cần nghiên cứu thêm bao nhiêu sản phẩm nữa?
- Muốn ước lượng tỉ lệ sản phẩm loại A với độ chính xác 15,26% thì đạt độ tin cậy bao nhiêu?
- Bằng khoảng tin cậy đối xứng, hãy ước lượng trung bình chỉ tiêu Y của các sản phẩm loại A với độ tin cậy 95%.

Bài 3.32. Cho X(%) là tỉ lệ tạp chất và Y (kg) là trọng lượng một loại sản phẩm. X, Y là các đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn. Điều tra trên một số sản phẩm cùng loại thu được bảng số liệu sau

Những sản phẩm có tỉ lệ tạp chất nhỏ hơn 0,4% được xếp vào loại I.

- Hãy ước lượng tỉ lệ tạp chất trung bình của các sản phẩm trên với độ tin cậy 98%;
- Bằng khoảng tin cậy đối xứng, hãy ước lượng tỉ lệ sản phẩm loại I với độ tin cậy 97%;

Y \ X	0,1-0,2	0,2-0,3	0,3-0,4	0,4-0,5	0,5-0,6
1-2	3	5	3		
2-3	2	3	6	5	
3-4		2	5	4	1
4-5			3	4	5
5-6				3	5

- c) Muốn ước lượng tỉ lệ tạp chất trung bình của các sản phẩm loại I với độ tin cậy 95%, độ chính xác 0,00015 thì cần điều tra ít nhất bao nhiêu sản phẩm nữa?
- d) Muốn ước lượng tỉ lệ sản phẩm loại I với độ dài khoảng tin cậy đối xứng 23,5% thì đạt độ tin cậy bao nhiêu?
- e) Muốn ước lượng trọng lượng trung bình của các sản phẩm trên với độ chính xác 400 (g) độ tin cậy 0,96 thì có cần điều tra thêm không?
- f) Muốn ước lượng tỉ lệ tạp chất trung bình của tất cả các sản phẩm trên với độ chính xác 0,0247% thì đạt độ tin cậy bao nhiêu?

Chương 4

Kiểm định giả thuyết thống kê

4.1	Kiểm định giả thuyết	101
4.2	Kiểm định giá trị trung bình	103
4.3	Kiểm định tỉ lệ	107
	Bài tập chương 4	109

4.1 Kiểm định giả thuyết

4.1.1 Giả thuyết thống kê

Trong nhiều lĩnh vực của đời sống chúng ta hay đưa ra các nhận định khác nhau về các đối tượng mà chúng ta quan tâm. Những nhận định đó được gọi là các *giả thuyết*. Việc xác định tính đúng sai của giả thuyết nêu ra được gọi là *kiểm định giả thuyết*.

Định nghĩa 4.1. Giả thuyết thống kê là giả thuyết về dạng phân phối xác suất của đại lượng ngẫu nhiên, về các tham số đặc trưng của đại lượng ngẫu nhiên hoặc về tính độc lập của các đại lượng ngẫu nhiên.

Giả thuyết thống kê được đưa kiểm định được gọi là *giả thuyết gốc*, ký hiệu là H_0 .

4.1.2 Cặp giả thuyết-đối thuyết

Khi đưa ra một giả thuyết thống kê H_0 , người ta thường nghiên cứu kèm theo nó một nhận định khác với giả thuyết gốc, gọi là *giả thuyết đối* (hay *đối thuyết*) và ký hiệu là H_1 . Giả thuyết H_0 và đối thuyết H_1 tạo nên một cặp giả thuyết-đối thuyết thống kê. Khi quyết định bác bỏ giả thuyết H_0 tức là ta chấp nhận đối thuyết H_1 và ngược lại.

Chẳng hạn nghiên cứu nhu cầu thị trường về một loại hàng hoá nào đó. Ta có thể đưa ra các cặp giả thuyết-đối thuyết thống kê như sau:

- H_0 : Nhu cầu trung bình về loại hàng hoá này là $\mu = 1000$ đơn vị/tháng
 - H_1 : Nhu cầu trung bình $\mu \neq 1000$ đơn vị/tháng
- hoặc

- $H_0: \mu = 1000$ đơn vị/tháng
 $H_1: \mu > 1000$ đơn vị/tháng

Bài toán kiểm định một cặp giả thuyết-đối thuyết thống kê H_0, H_1 dựa vào thông tin thực nghiệm của mẫu để đưa ra kết luận chấp nhận H_0 hay bác bỏ H_0 được gọi là kiểm định thống kê.

4.1.3 Quy tắc kiểm định giả thuyết thống kê

Nguyên tắc chung của kiểm định giả thuyết thống kê dựa trên *nguyên lý xác suất nhỏ*, tức là một biến cố có xác suất xuất hiện rất bé thì có thể coi rằng nó không xảy ra khi phép thử liên quan đến biến cố được thực hiện. Như vậy nếu giả thuyết H_0 được giả sử đúng mà dẫn đến một điều vô lý thì ta quyết định bác bỏ H_0 .

Việc kiểm định giả thuyết thống kê được tiến hành như sau:

- Từ biến ngẫu nhiên gốc X trong tổng thể, lập mẫu ngẫu nhiên kích thước n :

$$W = (X_1, X_2, \dots, X_n)$$

Để kiểm định giả thuyết H_0 liên quan đến tham số θ của tổng thể, ta chọn thống kê G ,

$$G = f(X_1, X_2, \dots, X_n)$$

(G có thể phụ thuộc tham số đã biết có mặt trong H_0) sao cho G có quy luật phân phối xác suất hoàn toàn xác định khi H_0 đúng. Thống kê G được gọi là *tiêu chuẩn kiểm định*.

- Với một xác suất khá nhỏ α cho trước (gọi là *mức ý nghĩa* của bài toán kiểm định), nếu có thể tìm được miền W_α sao cho khi H_0 đúng, xác suất để thống kê G nhận giá trị trong miền W_α bằng α :

$$\mathbb{P}(G \in W_\alpha / H_0) = \alpha$$

thì miền W_α được gọi là *miền bác bỏ* giả thuyết H_0 với mức ý nghĩa α . Vì xác suất α khá nhỏ nên có thể coi biến cố $A = (G \in W_\alpha)$ không xảy ra khi thực hiện một phép thử.

- Thực hiện một phép thử đối với mẫu ngẫu nhiên W ta thu được mẫu cụ thể:

$$w = (x_1, x_2, \dots, x_n)$$

qua đó tính được giá trị cụ thể của tiêu chuẩn kiểm định G là:

$$g = (x_1, x_2, \dots, x_n)$$

Giá trị này được gọi là giá trị quan sát của tiêu chuẩn kiểm định G .

- Quy tắc kiểm định:

Nếu $g \in W_\alpha$ thì có thể cho rằng H_0 sai và quyết định bác bỏ H_0 (chấp nhận H_1).

Nếu $g \notin W_\alpha$ thì chưa đủ cơ sở để cho rằng H_0 sai nên quyết định chấp nhận H_0 (bác bỏ H_1).

4.1.4 Các loại sai lầm

Với quy tắc kiểm định như trên có thể mắc hai loại sai lầm:

- Sai lầm loại 1: Bác bỏ giả thuyết H_0 trong khi H_0 đúng. Xác suất mắc sai lầm loại này đúng bằng mức ý nghĩa α . Thật vậy, vì nếu $G \in W_\alpha$ thì ta lập tức bác bỏ H_0 , trong khi đó nếu H_0 đúng thì vẫn có thể xảy ra $G \in W_\alpha$ với xác suất α .
- Sai lầm loại 2: Chấp nhận H_0 trong khi H_0 sai, tức là giá trị quan sát $g \notin W_\alpha$ trong khi H_1 đúng. Xác suất mắc sai lầm loại này là β :

$$P(G \notin W_\alpha / H_1) = \beta$$

Ta thấy rằng sai lầm loại 1 và sai lầm loại 2 mâu thuẫn nhau, tức là với một mẫu kích thước n xác định thì không thể cùng lúc giảm xác suất mắc hai loại sai lầm nói trên được. Khi ta giảm α đi thì sẽ đồng thời tăng β và ngược lại. Chẳng hạn giảm $\alpha = 0$ thì sẽ không bác bỏ giả thuyết nào, tức là β đạt cực đại.

Với mức ý nghĩa α cho trước và với mẫu kích thước n xác định thì trong số các miền bác bỏ W_α có thể tìm được, ta chọn ra miền bác bỏ sao cho xác suất mắc sai lầm loại 2 là nhỏ nhất:

$$\begin{cases} P(G \in W_\alpha / H_0) = \alpha \text{ cho trước} \\ P(G \in W_\alpha / H_1) = 1 - \beta \longrightarrow \max \end{cases}$$

Miền bác bỏ W_α như vậy được gọi là miền bác bỏ "tốt nhất" với mức ý nghĩa α và kích thước mẫu n xác định trước.

4.2 Kiểm định giá trị trung bình

4.2.1 Đặt vấn đề

Giả sử đại lượng ngẫu nhiên gốc X trong tổng thể phân phối theo quy luật chuẩn $N(\mu, \sigma^2)$ nhưng chưa biết kỳ vọng toán μ . Nếu có cơ sở để giả thiết rằng giá trị của μ bằng μ_o , ta lập giả thuyết thống kê $H_0 : \mu = \mu_o$. Để kiểm định giả thuyết trên, từ tổng thể lập mẫu kích thước n :

$$W = (X_1, X_2, \dots, X_n)$$

Tiêu chuẩn kiểm định G và miền bác bỏ W_α được xác định tùy theo từng trường hợp sau đây:

4.2.2 Trường hợp đã biết phương sai $\sigma^2 = \sigma_0^2$

Tiêu chuẩn kiểm định là thống kê:

$$G = U = \frac{(\bar{X} - \mu_o)\sqrt{n}}{\sigma_0}$$

Nếu giả thuyết H_0 đúng thì U là đại lượng ngẫu nhiên phân phối chuẩn hoá $N(0, 1)$. Từ đó miền bác bỏ tốt nhất W_α được xây dựng theo các đối thuyết như sau: (u_α là phân vị chuẩn mức α)

- $H_0 : \mu = \mu_o; H_1 : \mu > \mu_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ bên phải W_α như sau:

$$W_\alpha = \left\{ U = \frac{(\bar{X} - \mu_o)\sqrt{n}}{\sigma_0}; U > u_{1-\alpha} \right\} = (u_{1-\alpha}; +\infty)$$

- $H_0 : \mu = \mu_o; H_1 : \mu < \mu_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ bên trái W_α như sau:

$$W_\alpha = \left\{ U = \frac{(\bar{X} - \mu_o)\sqrt{n}}{\sigma_0}; U < -u_{1-\alpha} \right\} = (-\infty; -u_{1-\alpha})$$

- $H_0 : \mu = \mu_o; H_1 : \mu \neq \mu_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ hai phía W_α như sau:

$$W_\alpha = \left\{ U = \frac{(\bar{X} - \mu_o)\sqrt{n}}{\sigma_0}; |U| > u_{1-\frac{\alpha}{2}} \right\} = (-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$$

Chú ý: Trong bài tập cụ thể, thủ tục kiểm định giả thuyết thống kê được tiến hành như sau:

- +) Xác định cặp giả thuyết-đối thuyết thống kê H_0, H_1 .
- +) Từ mức ý nghĩa α , tra bảng phân vị chuẩn để tìm miền bác bỏ W_α .
- +) Trên mẫu cụ thể $w = (x_1, x_2, \dots, x_n)$, tính giá trị quan sát của tiêu chuẩn kiểm định:

$$u_{qs} = \frac{(\bar{x} - \mu_o)\sqrt{n}}{\sigma_0}$$

- +) So sánh u_{qs} với miền W_α để kết luận:
 - Nếu $u_{qs} \in W_\alpha$ thì bác bỏ H_0 , chấp nhận H_1 ;
 - Nếu $u_{qs} \notin W_\alpha$ thì chưa có cơ sở bác bỏ H_0 (trên thực tế ta chấp nhận H_0 , bác bỏ H_1).

Ví dụ 4.1. Bột mì được đóng bao bằng máy tự động có trọng lượng đóng bao theo quy định là 16kg và độ lệch chuẩn là 1,2kg. Lấy ngẫu nhiên 25 bao bột để kiểm tra tìm được trọng lượng đóng gói trung bình của chúng là 16,5kg. Với mức ý nghĩa 0,05 có cần dừng máy để điều chỉnh hay không? biết trọng lượng bao bột mì là biến ngẫu nhiên phân phối chuẩn.

Giải. Gọi X là trọng lượng một bao bột mì đóng bằng máy tự động với trọng lượng trung bình là μ . Ta cần kiểm định giá trị của tham số $\mu = 16$ khi đã biết $\sigma = \sigma_0 = 1,2$.

- +) Cặp giả thuyết-đối thuyết thống kê: $H_0 : \mu = \mu_o = 16; H_1 : \mu \neq 16$

Từ mẫu ngẫu nhiên ta có các kết quả: $n = 25; \bar{x} = 16,5$.

- +) Từ mức ý nghĩa $\alpha = 0,05$, suy ra $1 - \frac{\alpha}{2} = 0,975$. Tra bảng phân vị chuẩn: $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$

Miền bác bỏ hai phía là: $W_\alpha = (-\infty; -1,96) \cup (1,96; +\infty)$

- +) Giá trị quan sát: $u_{qs} = \frac{(\bar{x} - \mu_o)\sqrt{n}}{\sigma_0} = \frac{(16,5 - 16)\sqrt{25}}{1,2} = 2,083$

Ta thấy $u_{qs} \in W_\alpha$ nên có cơ sở bác bỏ H_0 , chấp nhận H_1 . Vậy có thể cần dừng máy để kiểm tra.

4.2.3 Trường hợp chưa biết phương sai σ^2 và kích thước mẫu $n < 30$

Tiêu chuẩn kiểm định là thống kê:

$$G = T = \frac{(\bar{X} - \mu_o)\sqrt{n}}{S'}$$

Nếu giả thuyết H_0 đúng thì T là đại lượng ngẫu nhiên có phân phối Student với $n - 1$ bậc tự do. Từ đó miền bác bỏ tốt nhất W_α được xây dựng theo các đối thuyết như sau: (t_α^n là phân vị Student mức α với n bậc tự do)

- $H_0 : \mu = \mu_o; H_1 : \mu > \mu_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ bên phải W_α như sau:

$$W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_o)\sqrt{n}}{S'}; T > t_{1-\alpha}^{n-1} \right\} = (t_{1-\alpha}^{n-1}; +\infty)$$

- $H_0 : \mu = \mu_o; H_1 : \mu < \mu_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ bên trái W_α như sau:

$$W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_o)\sqrt{n}}{S'}; T < -t_{1-\alpha}^{n-1} \right\} = (-\infty; -t_{1-\alpha}^{n-1})$$

- $H_0 : \mu = \mu_o; H_1 : \mu \neq \mu_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ hai phía W_α như sau:

$$W_\alpha = \left\{ T = \frac{(\bar{X} - \mu_o)\sqrt{n}}{S'}; |T| > t_{1-\frac{\alpha}{2}}^{n-1} \right\} = (-\infty; -t_{1-\frac{\alpha}{2}}^{n-1}) \cup (t_{1-\frac{\alpha}{2}}^{n-1}; +\infty)$$

Ví dụ 4.2. Thông qua một mẫu gồm 25 gia đình công nhân ở một khu công nghiệp người ta thu được kết quả là chi tiêu trung bình hàng tháng của các gia đình đó là 4,455 triệu đồng với độ lệch chuẩn mẫu là 0,3 triệu. Với mức ý nghĩa 0,05 có thể cho rằng chi tiêu trung bình hàng tháng của các gia đình công nhân ở khu công nghiệp đó là ít hơn 4,5 triệu hay không? Giả thiết mức chi tiêu của các gia đình là biến ngẫu nhiên phân phối chuẩn.

Giải. Gọi X là chi tiêu hàng tháng của một gia đình công nhân trong khu công nghiệp này. Theo giả thiết X có phân phối chuẩn với trung bình μ chưa biết. Đây là bài toán kiểm định giá trị tham số μ của X khi chưa biết σ^2 và mẫu thu được có kích thước $n = 25 < 30$.

+) Cặp giả thuyết-đối thuyết thống kê: $H_0 : \mu = \mu_o = 4,5; H_1 : \mu < 4,5$

+) Từ mức ý nghĩa $\alpha = 0,05$, tra bảng phân vị Student: $t_{1-\alpha}^{n-1} = t_{0,95}^{24} = 1,711$

Miền bác bỏ bên trái là: $W_\alpha = (-\infty; -1,711)$

+) Từ mẫu ngẫu nhiên ta có các kết quả: $n = 25 < 30; \bar{x} = 4,455; s' = 0,3$.

Giá trị quan sát: $t_{qs} = \frac{(\bar{x} - \mu_o)\sqrt{n}}{s'} = \frac{(4,455 - 4,5)\sqrt{25}}{0,3} = -0,75$

Ta thấy $t_{qs} \notin W_\alpha$ nên không có cơ sở bác bỏ H_0 . Vậy chưa thể kết luận chi tiêu trung bình của một gia đình công nhân trong khu công nghiệp này ít hơn 4,5 triệu.

4.2.4 Trường hợp chưa biết phương sai σ^2 và kích thước mẫu $n \geq 30$

Tiêu chuẩn kiểm định là thống kê:

$$G = U = \frac{(\bar{X} - \mu_o)\sqrt{n}}{S'}$$

Nếu giả thuyết H đúng thì U là đại lượng ngẫu nhiên có phân phối xấp xỉ chuẩn hoá $N(0, 1)$. Từ đó miền bác bỏ tốt nhất W_α được xây dựng theo các đối thuyết như sau: (u_α là phân vị chuẩn mức α)

- $H_0 : \mu = \mu_o; H_1 : \mu > \mu_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ bên phải W_α như sau:

$$W_\alpha = \left\{ U = \frac{(\bar{X} - \mu_o)\sqrt{n}}{S'}; U > u_{1-\alpha} \right\} = (u_{1-\alpha}; +\infty)$$

- $H_0 : \mu = \mu_o; H_1 : \mu < \mu_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ bên trái W_α như sau:

$$W_\alpha = \left\{ U = \frac{(\bar{X} - \mu_o)\sqrt{n}}{S'}; U < -u_{1-\alpha} \right\} = (-\infty; -u_{1-\alpha})$$

- $H_0 : \mu = \mu_o; H_1 : \mu \neq \mu_o$.

Lúc đó với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ hai phía W_α như sau:

$$W_\alpha = \left\{ U = \frac{(\bar{X} - \mu_o)\sqrt{n}}{S'}; |U| > u_{1-\frac{\alpha}{2}} \right\} = (-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$$

Ví dụ 4.3. Năm trước tiền lương trung bình của các cử nhân quản trị kinh doanh làm việc trong các công ty liên doanh với nước ngoài là 410 USD một tháng. Năm nay, theo điều tra lương hàng tháng của một số cử nhân kinh tế đang làm việc cho các công ty liên doanh với nước ngoài được bảng số liệu như sau:

Mức lương (USD)	Số cử nhân
370 – 390	6
390 – 410	10
410 – 430	15
430 – 450	18
450 – 470	11

Với mức ý nghĩa $\alpha = 0,02$ có thể cho rằng mức tiền lương năm trước ít hơn năm nay hay không? Giả thiết tiền lương là biến ngẫu nhiên phân phối chuẩn.

Giải. Gọi X là mức lương hàng tháng năm nay của một cử nhân kinh tế đang làm việc cho các công ty liên doanh với nước ngoài. Theo giả thiết, X có phân phối chuẩn $N(\mu, \sigma^2)$. Đây là bài toán kiểm định tham số μ khi chưa biết σ^2 và kích thước mẫu $n = 60 > 30$.

+) Cặp giả thuyết-đối thuyết thống kê là: $H_0: \mu = \mu_o = 410; H_1: \mu > 410$.

+) Từ mức ý nghĩa $\alpha = 0,02$, tra bảng phân vị chuẩn $u_{1-\alpha} = u_{0,98} = 2,054$. Miền bác bỏ một phía:

$$W_\alpha = (2,054; +\infty)$$

+) Từ mẫu ngẫu nhiên lập bảng tính:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
380	6	2280	866400
400	10	4000	1600000
420	15	6300	2646000
440	18	7920	3484800
460	11	5060	2327600
	$n = \sum n_i = 60$	$\sum n_i x_i = 25560$	$\sum n_i x_i^2 = 10924800$

$$\bar{x} = \frac{25560}{60} = 426; \quad s' = \sqrt{\frac{60}{59} \left(\frac{10924800}{60} - 426^2 \right)} \approx 24,784$$

Giá trị quan sát của tiêu chuẩn kiểm định:

$$u_{qs} = \frac{(\bar{x} - \mu_o)\sqrt{n}}{s'} = \frac{(426 - 410)\sqrt{60}}{24,784} \approx 5,000$$

Ta thấy $u_{qs} \in W_\alpha$ nên có cơ sở bác bỏ H_0 , chấp nhận H_1 . Vậy có thể cho rằng mức tiền lương năm nay đã tăng lên so với năm ngoái.

4.3 Kiểm định tỉ lệ

Giả sử trong tổng thể, đại lượng ngẫu nhiên gốc X phân phối theo quy luật không - một với kỳ vọng toán $M(X) = p$ và phương sai $D(X) = \frac{p(1-p)}{n}$. Ta đã biết p chính là cơ cấu của tổng thể theo dấu hiệu nghiên cứu, do đó bài toán kiểm định cơ cấu tổng thể cũng chính là kiểm định kỳ vọng toán của đại lượng ngẫu nhiên tuân theo quy luật không - một.

Nếu chưa biết p nhưng có cơ sở để giả thiết rằng giá trị của nó bằng p_o , ta đưa ra giả thuyết thống kê $H_0 : p = p_o$.

Từ X lập mẫu ngẫu nhiên $W = (X_1, X_2, \dots, X_n)$, và với n đủ lớn ta chọn tiêu chuẩn kiểm định là thống kê:

$$G = U = \frac{(f - p_o)\sqrt{n}}{\sqrt{p_o(1 - p_o)}}$$

Nếu H_0 đúng thì U có phân phối xấp xỉ chuẩn hoá $N(0, 1)$. Do đó tùy thuộc vào dạng của đối thuyết H_1 mà miền bác bỏ tốt nhất W_α được xây dựng, cụ thể như sau: (u_α là phân vị chuẩn mức α)

- $H_0 : p = p_o; H_1 : p > p_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ bên phải W_α như sau:

$$W_\alpha = \left\{ U = \frac{(f - p_o)\sqrt{n}}{\sqrt{p_o(1 - p_o)}}; U > u_{1-\alpha} \right\} = (u_{1-\alpha}; +\infty)$$

- $H_0 : p = p_o; H_1 : p < p_o$.

Với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ bên trái W_α như sau:

$$W_\alpha = \left\{ U = \frac{(f - p_o)\sqrt{n}}{\sqrt{p_o(1 - p_o)}}; U < -u_{1-\alpha} \right\} = (-\infty; -u_{1-\alpha})$$

- $H_0 : p = p_o; H_1 : p \neq p_o$.

Lúc đó với mức ý nghĩa α cho trước có thể tìm được miền bác bỏ hai phía W_α như sau:

$$W_\alpha = \left\{ U = \frac{(f - p_o)\sqrt{n}}{\sqrt{p_o(1 - p_o)}}; |U| > u_{1-\frac{\alpha}{2}} \right\} = (-\infty; -u_{1-\frac{\alpha}{2}}) \cup (u_{1-\frac{\alpha}{2}}; +\infty)$$

Ví dụ 4.4. Tỷ lệ khách tiêu dùng một loại sản phẩm ở địa phương A là 60%. Sau một chiến dịch quảng cáo người ta cho rằng tỷ lệ này đã tăng lên. Phỏng vấn ngẫu nhiên 400 người sau chiến dịch quảng cáo thấy có 250 người có sử dụng loại sản phẩm trên. Với mức ý nghĩa 0,05 hãy kết luận về ý kiến trên.

Giải. Gọi p là tỷ lệ khách hàng tiêu dùng loại sản phẩm đó tại địa phương A sau chiến dịch quảng cáo. Đây là bài toán kiểm định tham số p của phân phối $A(p)$.

+) Cặp giả thuyết-đối thuyết thống kê là: $H_0: p = p_o = 0,6; H_1: p > 0,6$.

+) Từ mức ý nghĩa $\alpha = 0,05$, tra bảng phân vị chuẩn $u_{1-\alpha} = u_{0,95} = 1,645$. Miền bác bỏ một phía:

$$W_\alpha = (1,645; +\infty)$$

+) Từ mẫu cụ thể ta tính được tần suất mẫu: $f = \frac{k}{n} = \frac{250}{400} = 0,625$

Giá trị quan sát của tiêu chuẩn kiểm định:

$$u_{qs} = \frac{(f - p_o)\sqrt{n}}{\sqrt{p_o(1 - p_o)}} = \frac{(0,625 - 0,6)\sqrt{400}}{\sqrt{0,6 \cdot 0,4}} \approx 1,021$$

Ta thấy $u_{qs} \notin W_\alpha$ nên chưa có cơ sở để bác bỏ H_0 , tức là chưa thể cho rằng tỷ lệ khách hàng mua sản phẩm nói trên đã thực sự tăng lên sau quảng cáo.

Bài tập chương 4

KIỂM ĐỊNH GIẢ THUYẾT THỐNG KÊ

Bài 4.1. Trọng lượng sản phẩm (X) do nhà máy sản xuất ra là đại lượng ngẫu nhiên phân phối chuẩn với độ lệch tiêu chuẩn $\sigma_X = 2$ kg và trọng lượng trung bình là 20 kg. Nghi ngờ máy hoạt động không bình thường làm thay đổi trọng lượng trung bình của sản phẩm, người ta cân thử 100 sản phẩm và thu được kết quả sau:

Trọng lượng sản phẩm	19	20	21	22	23
Số sản phẩm tương ứng	10	60	20	5	5

Với mức ý nghĩa $\alpha = 0,05$, hãy kết luận về điều nghi ngờ nói trên.

Bài 4.2. Mỳ chính được đóng gói 453 gam một gói trên máy tự động. Có thể coi trọng lượng của các gói mỳ chính tuân theo quy luật phân phối chuẩn với độ lệch tiêu chuẩn 36 gam. Kiểm tra ngẫu nhiên 81 gói thấy trọng lượng trung bình là 448 gam. Với mức ý nghĩa $\alpha = 0,05$ có thể kết luận trọng lượng các gói mỳ chính có xu hướng giảm không?

Bài 4.3. Trong điều kiện chăn nuôi bình thường, lượng sữa trung bình của một con bò là 14 kg một ngày. Nghi ngờ điều kiện chăn nuôi bò kém đi làm cho lượng sữa giảm xuống. Người ta điều tra ngẫu nhiên 25 con bò và tính được lượng sữa trung bình của một con trong một ngày là 12,5 kg và độ lệch tiêu chuẩn mẫu điều chỉnh $s' = 2,5$ kg. Với mức ý nghĩa $\alpha = 0,01$ hãy kết luận về điều nghi ngờ nói trên. (giả thiết lượng sữa bò là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn)

Bài 4.4. Mức hao phí xăng (X) cho một loại xe ô tô chạy trên đoạn đường AB là đại lượng ngẫu nhiên phân phối chuẩn. Quan sát 50 chuyến xe chạy trên đường AB ta thu được bảng số liệu sau:

Mức xăng hao phí (lít)	Số chuyến xe (n_i)
48,5 – 49,0	8
49,0 – 49,5	15
49,5 – 50,0	15
50,0 – 50,5	7
50,5 – 51,0	5

Sau khi đường được tu sửa lại, người ta thấy mức hao phí xăng trung bình là 48 lít. Như vậy với mức ý nghĩa $\alpha = 0,05$, có thể kết luận việc tu sửa đường đã làm giảm mức xăng hao phí trung bình hay không?

Bài 4.5. Tỷ lệ phế phẩm do một máy tự động sản xuất là 5%. Kiểm tra ngẫu nhiên 300 sản phẩm thấy có 24 sản phẩm là phế phẩm. Từ đó có ý kiến cho rằng tỷ lệ phế phẩm do máy đó sản xuất có chiều hướng tăng lên. Hãy kết luận ý kiến nêu trên với mức ý nghĩa $\alpha = 2\%$.

Bài 4.6. Nếu áp dụng phương pháp công nghệ thứ nhất thì tỷ lệ phế phẩm là 6%; còn nếu áp dụng phương pháp công nghệ thứ hai thì trong 100 sản phẩm có 5 phế phẩm. Vậy có thể kết luận áp dụng phương pháp công nghệ thứ nhất thì tỷ lệ phế phẩm cao hơn tỷ lệ phế phẩm của phương pháp công nghệ thứ hai không?. Hãy kết luận với mức ý nghĩa $\alpha = 0,05$.

Bài 4.7. Tỷ lệ bệnh nhân khỏi bệnh T khi điều trị bằng thuốc A là 85%. Thí nghiệm dùng loại thuốc B để chữa bệnh thì trong số 900 người mắc bệnh T có 810 người được chữa khỏi. Như vậy có thể kết luận thuốc B hiệu quả hơn thuốc A hay không? Yêu cầu kết luận với mức ý nghĩa $\alpha = 0,05$.

Bài 4.8. Chủ một sân golf cho rằng trung bình một người đến chơi golf ở sân của anh ta sẽ mất một tá bóng hoặc hơn sau khi kết thúc một vòng chơi. Một tay golf nghi ngờ lời của ông chủ sân nên đã đề nghị ngẫu nhiên 15 người chơi báo lại số bóng bị mất của mình sau khi hoàn thành một vòng chơi. Kết quả thu được như sau:

1	14	8	15	17	10	12	6
14	21	15	9	11	4	8	

Giả thiết rằng số bóng bị mất của mỗi người chơi có phân phối chuẩn với độ lệch chuẩn là 3. Với mức ý nghĩa 10% có thể cho rằng số bóng golf bị mất trung bình mỗi lần chơi nhỏ hơn 12 bóng được không?

Bài 4.9. Một doanh nghiệp có hệ thống máy tính có thể xử lí 1200 hóa đơn trong một giờ. Đơn vị này muốn nhập về một hệ thống máy tính mới. Hệ thống này khi chạy kiểm tra trong 40 giờ cho thấy hóa đơn trung bình được xử lí là 1260 trên một giờ với độ lệch tiêu chuẩn là 215. Với mức ý nghĩa 5% hãy nhận định xem hệ thống mới có tốt hơn hệ thống cũ hay không?

Bài 4.10. Một nhãn hàng quảng bá sản phẩm ăn kiêng rằng sẽ giảm được trọng lượng được ít nhất 45 pound trong vòng 6 tháng (1 pound = 0,454 kg). Một mẫu gồm 28 người theo chế độ ăn kiêng này giảm trọng lượng trung bình 35 pound với độ lệch tiêu chuẩn là 20 pound. Hãy nhận định xem phương pháp ăn kiêng trên có phóng đại quá mức với mức ý nghĩa $\alpha = 0,01$ hay không?

Bài 4.11. Ở một số quốc gia, luật pháp yêu cầu người lái xe phải bật đèn pha khi lái xe dưới trời mưa. Một sĩ quan tuần tra đường cao tốc ở một quốc gia như vậy tin rằng chưa đến một phần tư tổng số người lái xe tuân theo quy tắc này. Để kiểm tra, anh ấy lấy mẫu ngẫu nhiên 200 ô tô đang chạy dưới trời mưa và đếm số ô tô bật đèn pha. Anh ta tìm thấy con số này là 41. Liệu sĩ quan này có đủ bằng chứng ở mức ý nghĩa 10% để chứng minh cho niềm tin của mình không?

Bài 4.12. Một bác sĩ tuyên bố rằng người Úc trung bình thừa cân hơn 5 kg. Để kiểm tra tuyên bố này, một mẫu ngẫu nhiên gồm 50 người Úc đã được cân và tính toán sự khác biệt giữa cân nặng thực tế và cân nặng lý tưởng của họ. Giá trị trung bình và độ lệch chuẩn của sự khác biệt đó lần lượt là 6,5 kg và 2,2 kg. Có thể kết luận với mức ý nghĩa 0,05 rằng có đủ bằng chứng cho thấy tuyên bố của bác sĩ là đúng không?

Tài liệu tham khảo

- [1] Alexander Holmes - Barbara Illowsky - Susan Dean, *Introductory Business Statistics*, Rice University, 2018.
- [2] Allan G. Bluman, *Elementary statistics: A step by step approach, 8th Edition*, Mc Graw Hill, 2012.
- [3] Barbara Illowsky - Susan Dean, *Introductory Statistics*, Rice University, 2013.
- [4] Đào Hữu Hồ - Nguyễn Văn Hữu - Hoàng Hữu Như, *Thống kê toán học*, Nhà xuất bản Đại học quốc gia Hà Nội, 2004.
- [5] Eliyathamby A. Selvanathan, Saroja Selvanathan, Gerald Keller, *Business Statistics, 8th Edition*, Australia/New Zealand, 2021.
- [6] Gerald Keller, *Statistics for Management and Economics, 11th Edition*, Cengage Learning, 2018.
- [7] Nguyễn Cao Văn - Trần Thái Ninh, *Lý thuyết Xác suất và Thống kê toán*, Nhà xuất bản giáo dục, 2002.
- [8] Nguyễn Duy Tiến - Vũ Viết Yên, *Lý thuyết Xác suất*, Nhà xuất bản giáo dục, 2001.
- [9] Nguyễn Thị Đỗ Hạnh, *Bài giảng Xác suất thống kê*, Nhà xuất bản Đại học Hàng hải Việt Nam - Tài liệu lưu hành nội bộ, 2014.
- [10] Robert M. Leekley, *Applied Statistics for Business and Economics*, CRC Press, Chapman & Hall Book, 2010.
- [11] Ron Larson, Betsy Farber, *Elementary Statistics: Picturing the World, 5th Edition*, Pearson Education, 2011.
- [12] Tống Đình Quỳ, *Giáo trình Xác suất thống kê*, Nhà xuất bản giáo dục, 1999.
- [13] Trần Tuấn Điệp - Lý Hoàng Tú, *Lý thuyết xác suất và thống kê toán học*, Nhà xuất bản giáo dục, 1999.
- [14] Trần Văn Minh - Phí Thị Vân Anh, *Xác suất thống kê và các tính toán trên Excel*, Nhà xuất bản Giao thông vận tải, 2007.
- [15] Trần Văn Minh - Phí Thị Vân Anh, *Hướng dẫn giải bài tập Xác suất thống kê với các tính toán trên Excel*, Nhà xuất bản Giao thông vận tải, 2008.

Phụ lục

A Đáp số bài tập chương 1

Bài 1.1. a) $\frac{2}{3}$; b) $\frac{1}{2}$.

Bài 1.2. a) $\frac{1}{6}$; b) $\frac{7}{12}$; c) $\frac{2}{9}$; d) $\frac{1}{6}$; e) $\frac{11}{36}$.

Bài 1.3. a) $\frac{5}{14}$; b) $\frac{25}{28}$; c) $\frac{1}{4}$.

Bài 1.4. 0,258.

Bài 1.5. a) $\frac{8}{21}$; b) $\frac{3}{7}$; c) $\frac{22}{105}$.

Bài 1.6. a) $\frac{50}{203}$; b) $\frac{185}{203}$.

Bài 1.7. a) $2,112 \cdot 10^{-4}$; b) $\frac{22}{29}$.

Bài 1.8. a) $\frac{1}{105105}$; b) $\frac{58}{91}$.

Bài 1.9. a) $\frac{3}{8}$; b) $\frac{3}{8}$.

Bài 1.10. a) $\frac{2}{275}$; b) $\frac{1}{5}$.

Bài 1.11. $\frac{139}{1152}$.

Bài 1.12. $\frac{5}{9}$.

Bài 1.13. $1 - \frac{\sqrt{3}}{2}$.

Bài 1.14. $\frac{1}{4}$.

Bài 1.15. $0,85 \cdot 200 = 170$ viên đạn.

Bài 1.16. $f = \frac{76}{147}$.

Bài 1.17. a) $\overline{A}.B.C + A.\overline{B}.C + A.B.\overline{C}$; b) $A.B.C$; c) $\overline{A}.\overline{B}.\overline{C}$;
d) $A.\overline{B}.\overline{C}$; e) $A.\overline{B}.\overline{C} + \overline{A}.B.\overline{C} + \overline{A}.\overline{B}.C$; f) $A + B + C$.

Bài 1.18.

a) $\overline{A_1}.\overline{A_2}.\overline{B_1}.\overline{B_2} + \overline{A_1}.A_2.\overline{B_1}.B_2 + A_1.\overline{A_2}.B_1.\overline{B_2} + \overline{A_1}.A_2.B_1.\overline{B_2} + A_1.\overline{A_2}.\overline{B_1}.B_2 + A_1.A_2.B_1.B_2$;

b) $\overline{A_1}.\overline{A_2}.B_1.\overline{B_2} + \overline{A_1}.\overline{A_2}.\overline{B_1}.B_2 + A_1.\overline{A_2}.B_1.B_2 + \overline{A_1}.A_2.B_1.B_2 + \overline{A_1}.\overline{A_2}.B_1.B_2$;

c) $\overline{A_1}.A_2.B_1.B_2 + A_1.\overline{A_2}.B_1.B_2 + A_1.A_2.\overline{B_1}.B_2 + A_1.A_2.B_1.\overline{B_2}$.

Bài 1.19. 0,9.

Bài 1.20. 0,0002.

Bài 1.21. 0,432.

Bài 1.22. 0,94.

Bài 1.23. 0,976.

Bài 1.24. $\frac{1}{3}$.

Bài 1.25. a) 0,056; b) 0,188; c) 0,976; d) 0,336; e) 0,452;
f) 0,788; g) 0,664.

Bài 1.26. a) 0,784; b) phải phát ít nhất 5 lần.

Bài 1.27. 0,066.

Bài 1.28. 0,51.

Bài 1.29. a) 0,14; b) 0,86; c) 0,93; d) 0,02; e) 0,05.

Bài 1.30. $\frac{15}{16}$.

Bài 1.31. $\frac{3}{8}; \frac{1}{6}$.

Bài 1.32. $\frac{5}{1764}$.

Bài 1.33. $\frac{1}{9}$.

Bài 1.34. a) 0,992; b) 0,181; c) 4; 0,238.

Bài 1.35. Cần đề nghị có quá 2 phế phẩm thì không nhận lô hàng.

Bài 1.36. a) 10,12%; b) $\frac{85}{253}$; c) Phế phẩm đó có khả năng thuộc phân xưởng I là lớn nhất.

Bài 1.37. a) $\frac{8}{15}$; b) $\frac{3}{4}$.

Bài 1.38. a) 0,17; b) 0,625; c) 0,171; d) $n_0 = 2$ và $P_{12}(2) \approx 0,296$; e) $n = 5$.

Bài 1.39. 0,204.

B Đáp số bài tập chương 2

Bài 2.1. a)

Số hệ thống bị hỏng X	0	1	2	3
$\mathbb{P}$	0,336	0,452	0,188	0,024

$$\text{b) } F(x) = \begin{cases} 0 & \text{với } x \leq 0 \\ 0,336 & \text{với } 0 < x \leq 1 \\ 0,788 & \text{với } 1 < x \leq 2 \\ 0,976 & \text{với } 2 < x \leq 3 \\ 1 & \text{với } x > 3 \end{cases} \quad \text{c) } 0,976; \quad \text{d) } \text{mod}(X) = 1.$$

Bài 2.2.

a)

X	0	1	2
$\mathbb{P}$	0,72	0,26	0,02

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

c) $\mathbb{E}(X) = 0,3$; $\text{var}(X) = 0,25$ và $\text{mod}(X) = 0$.

Bài 2.3. a) $-0,3$; $15,21$; $-3,9$. b) -5 .

Bài 2.4. a) $3,1$; $3,4$; $1,09$; $1,44$. b) $6,5$; $2,53$; c) -4 ; $17,32$; $-2,3$.

Bài 2.5. Nếu chọn phương án đầu tư sao cho tỉ lệ thu hồi vốn kỳ vọng cao hơn thì chọn phương án A, song nếu cần chọn phương án có độ rủi ro thu hồi vốn thấp hơn thì chọn phương án B.

Bài 2.6. Lợi nhuận trung bình 2 (USD). Độ rủi ro là rất cao ($\sigma_X = 89,084$).

Bài 2.7. 2,2 phút.

Bài 2.8. a) $x = 0,3$; $y = 0,5$; $z = 0,5$; $t = 0,5$.

b)

$X^2 - 2X$	-1	0
$\mathbb{P}$	0,3	0,7

XY	-2	-1	0	1	2
$\mathbb{P}$	0,25	0,15	0,2	0,15	0,25

$ X - Y $	0	1	2	3
$\mathbb{P}$	0,15	0,45	0,15	0,25

c) 0,249; 0,866.

Bài 2.9. a) $a = 2$; b) 0,5; c) 0,875.

Bài 2.10. a) $a = -\frac{1}{2}$; $b = \frac{1}{2}$; b) $\frac{7}{72}$; c) $\frac{343}{1024}$; d) $\text{med}(X) = 0$, $x_{0,95} = 0,684$; e) $-\frac{1}{12}$; $\frac{47}{144}$; 0,571.

Bài 2.11. a) $k = \frac{3}{4}$; b) $\frac{11}{16}$; $\frac{67}{1280}$; 0,229; c) $\frac{109}{864}$; d) 0,016.

Bài 2.12. a) 0,3834; b) 0,6321.

Bài 2.13. $\frac{80}{243}$.

Bài 2.14. 1900 (nghìn đồng).

Bài 2.15. a) $a = -1,5$; $b = 3$; b) $\frac{5}{16}$.

Bài 2.16. a) 1,25; b) cổ phiếu công ty M: cổ phiếu công ty N = 69,2% : 30,8%.

Bài 2.17. a) 0,231; b) 0,709; c) 3 máy.

Bài 2.18. Cần sản xuất mỗi đợt 13 sản phẩm.

Bài 2.19. a) $P(X = k) = C_{12}^k \cdot (0,3)^k \cdot (0,7)^{12-k}$; $k = 0; 1; \dots; 12$; b) 3; c) 0,0369; d) 0,993

Bài 2.20. a) 0,0443; b) 0,81904; c) [792; 828].

Bài 2.21. Vì $\mathbb{P}(X_2 > 80) > \mathbb{P}(X_1 > 80)$ nên áp dụng phương án 2.

Bài 2.22. a) 0,02275; b) 0,97725; c) 0,83999.

Bài 2.23. 381,13 triệu.

Bài 2.24. 0,6664.

Bài 2.25. 0,99865.

Bài 2.26. a) 0,10565; b) 0,30854.

Bài 2.27. a) 0,00466; b) 7950 giờ.

Bài 2.28. a) 22,167; 5,556; b) 0,30503.

Bài 2.29. a) công ty A. b) tỉ lệ $A : B = 29,7\% : 70,3\%$.

Bài 2.30. a) 68,9 kg; b) 720000 (đồng). c) nên nhập 70 kg.

Bài 2.31. a) 0,0000387; b) 0,99413.

Bài 2.32. a) 24,285 ngàn đồng; b) 2,688 năm; c) 1,896 năm.

Bài 2.33. a) 0,0819; b) 0,4013; c) 0,46483; d) 120387.

Bài 2.34. a) $k = \frac{1}{450}$; b) 0,64; c) $E(X) = 10$; $D(X) = 50$; $\sigma_X = 7,071$; d) 0,929; e) 0,107.

Bài 2.35. a) $a = 0,75$; $b = 1,5$; b) 0,640; c) $-0,85$; 0,186; d) 0,5; e) 0,853.

Bài 2.36. a) $a = 1,5$; $b = 1$; b) 0,8125; c) 944.444đ; 3.777.778đ; d) 364.000đ; e) 0,20327.

C Đáp số bài tập chương 3

Bài 3.1. $\bar{x} = 366,8$; $s^2 = 1137,16$; $s'^2 = 1263,511$.

Bài 3.2. b) $\bar{x} = 90,72$; $s = 4,147$.

Bài 3.3. $\bar{x} = 166$; $s' = 5,812$.

Bài 3.4. Lô 1: $\bar{x} = 79,16$; $s = 1,759$. Lô 2: $\bar{x} = 64,64$; $s = 3,018$.

Bài 3.5. $\bar{x} = 14,226$; $s' = 2,813$.

Bài 3.6. a) $\bar{x} = 10,10$; $s_X^2 = 26,64$; $\bar{y} = 140,56$; $s_Y^2 = 86,886$. b) $\bar{y}_A = 148,889$; $s_{Y_A}^2 = 9,877$.

Bài 3.7. a) $\bar{x}=155$; $s'_X=66,111$; $\bar{y}=136$; $s'_Y=21,107$; b) $\bar{x}_I = 140,625$; $s'_{X_I} = 41,708$; $\bar{y}_I = 137,5$; $s'_{Y_I} = 10$; c) $f = 0,533$.

Bài 3.8. (12,948; 16,052).

Bài 3.9. (8,332; 8,668).

Bài 3.10. a) (3,967; 4,133); b) 504 tấm.

Bài 3.11. a) (19,819; 19,937); b) $(-\infty; 19,935)$; c) 1928 sản phẩm.

Bài 3.12. a) (48,685; 50,315); b) $\mu \geq 48,622$, $\mu \leq 51,378$; c) 664 sản phẩm; d) 96%.

Bài 3.13. (31,434; 32,626).

Bài 3.14. $\mu \geq 9,888$.

Bài 3.15. $\mu \leq 118,367$.

Bài 3.16. a) (8,954; 10,646); b) 8,991; 10,609; c) 90 công nhân.

Bài 3.17. $0,013 < p < 0,067$.

Bài 3.18. $p \geq 0,572$.

Bài 3.19. $p \leq 0,440$.

Bài 3.20. $n \geq 3458$.

Bài 3.21. 1537.

Bài 3.22. a) $(0, 69; 0, 91)$; b) 1487.

Bài 3.23. a) $\mu \in (163, 994; 171, 006)$; b) $p \in (0, 667; 0, 853)$.

Bài 3.24. a) $\mu \in (49, 688; 54, 232)$; b) tối đa 53,78 kg; tối thiểu 50,14 kg; c) $p \in (0, 2102; 0, 3898)$; d) 40 con; e) 35,87%.

Bài 3.25. a) $\mu \in (23, 954; 27, 006)$; b) 97,8%; c) cần điều tra thêm 63 sản phẩm; d) $\mu \in (13, 082; 15, 624)$; e) không cần điều tra thêm.

Bài 3.26. a) $\mu \in (5, 839; 6, 331)$; b) chiều cao tối thiểu: 5,866 m; c) không cần điều tra thêm; d) $p \in (0, 264; 0, 48)$; e) 95%.

Bài 3.27. a) 85,4%; b) $(104, 775; 108, 891)$; c) $p \in (0, 034; 0, 106)$.

Bài 3.28. a) $(28, 793; 30, 635)$; b) $(0, 154; 0, 246)$; c) 91,2%.

Bài 3.29. a) $(0, 1336; 0, 2694)$; b) $(150, 879; 155, 787)$; c) 92,4%.

Bài 3.30. a) $(14, 059; 14, 649)$; b) 0; c) 96,8%; d) 13; e) 98%; f) 14; g) 97,6%.

Bài 3.31. a) 39; b) 94%; c) $(7, 750; 8, 460)$.

Bài 3.32. a) $(0, 344; 0, 418)$; b) $(0, 401; 0, 683)$; c) 65; d) 93%; e) 0; f) 88,4%.

D Đáp số bài tập chương 4

Bài 4.1. Chưa có cơ sở KL như vậy.

Bài 4.2. Chưa có cơ sở KL như vậy.

Bài 4.3. Điều nghi ngờ trên là đúng.

Bài 4.4. Có thể KL như vậy.

Bài 4.5. Ý kiến trên là đúng.

Bài 4.6. Không có cơ sở KL như vậy.

Bài 4.7. Có thể KL như vậy.

Bài 4.8. Có thể KL như vậy.

Bài 4.9. Hệ thống mới tốt hơn hệ thống cũ.

Bài 4.10. Quảng cáo quá mức.

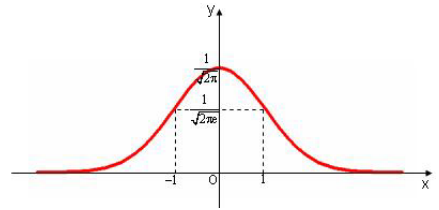
Bài 4.11. Không thể KL như vậy.

Bài 4.12. Có thể KL như vậy.

E Bảng tra thống kê

E.1 Hàm mật độ của phân phối chuẩn hóa

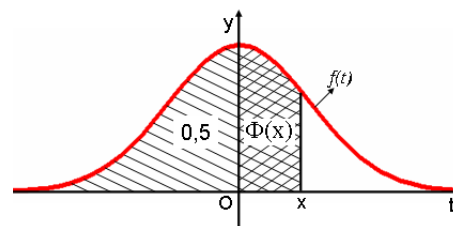
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$



x	0	1	2	3	4	5	6	7	8	9	x
0,0	0,39894	0,39892	0,39886	0,39876	0,39862	0,39844	0,39822	0,39797	0,39767	0,39733	0,0
0,1	0,39695	0,39654	0,39608	0,39559	0,39505	0,39448	0,39387	0,39322	0,39253	0,39181	0,1
0,2	0,39104	0,39024	0,38940	0,38853	0,38762	0,38667	0,38568	0,38466	0,38361	0,38251	0,2
0,3	0,38139	0,38023	0,37903	0,37780	0,37654	0,37524	0,37391	0,37255	0,37115	0,36973	0,3
0,4	0,36827	0,36678	0,36526	0,36371	0,36213	0,36053	0,35889	0,35723	0,35553	0,35381	0,4
0,5	0,35207	0,35029	0,34849	0,34667	0,34482	0,34294	0,34105	0,33912	0,33718	0,33521	0,5
0,6	0,33322	0,33121	0,32918	0,32713	0,32506	0,32297	0,32086	0,31874	0,31659	0,31443	0,6
0,7	0,31225	0,31006	0,30785	0,30563	0,30339	0,30114	0,29887	0,29659	0,29431	0,29200	0,7
0,8	0,28969	0,28737	0,28504	0,28269	0,28034	0,27798	0,27562	0,27324	0,27086	0,26848	0,8
0,9	0,26609	0,26369	0,26129	0,25888	0,25647	0,25406	0,25164	0,24923	0,24681	0,24439	0,9
1,0	0,24197	0,23955	0,23713	0,23471	0,23230	0,22988	0,22747	0,22506	0,22265	0,22025	1,0
1,1	0,21785	0,21546	0,21307	0,21069	0,20831	0,20594	0,20357	0,20121	0,19886	0,19652	1,1
1,2	0,19419	0,19186	0,18954	0,18724	0,18494	0,18265	0,18037	0,17810	0,17585	0,17360	1,2
1,3	0,17137	0,16915	0,16694	0,16474	0,16256	0,16038	0,15822	0,15608	0,15395	0,15183	1,3
1,4	0,14973	0,14764	0,14556	0,14350	0,14146	0,13943	0,13742	0,13542	0,13344	0,13147	1,4
1,5	0,12952	0,12758	0,12566	0,12376	0,12188	0,12001	0,11816	0,11632	0,11450	0,11270	1,5
1,6	0,11092	0,10915	0,10741	0,10567	0,10396	0,10226	0,10059	0,09893	0,09728	0,09566	1,6
1,7	0,09405	0,09246	0,09089	0,08933	0,08780	0,08628	0,08478	0,08329	0,08183	0,08038	1,7
1,8	0,07895	0,07754	0,07614	0,07477	0,07341	0,07206	0,07074	0,06943	0,06814	0,06687	1,8
1,9	0,06562	0,06438	0,06316	0,06195	0,06077	0,05959	0,05844	0,05730	0,05618	0,05508	1,9
2,0	0,05399	0,05292	0,05186	0,05082	0,04980	0,04879	0,04780	0,04682	0,04586	0,04491	2,0
2,1	0,04398	0,04307	0,04217	0,04128	0,04041	0,03955	0,03871	0,03788	0,03706	0,03626	2,1
2,2	0,03547	0,03470	0,03394	0,03319	0,03246	0,03174	0,03103	0,03034	0,02965	0,02898	2,2
2,3	0,02833	0,02768	0,02705	0,02643	0,02582	0,02522	0,02463	0,02406	0,02349	0,02294	2,3
2,4	0,02239	0,02186	0,02134	0,02083	0,02033	0,01984	0,01936	0,01888	0,01842	0,01797	2,4
2,5	0,01753	0,01709	0,01667	0,01625	0,01585	0,01545	0,01506	0,01468	0,01431	0,01394	2,5
2,6	0,01358	0,01323	0,01289	0,01256	0,01223	0,01191	0,01160	0,01130	0,01100	0,01071	2,6
2,7	0,01042	0,01014	0,00987	0,00961	0,00935	0,00909	0,00885	0,00861	0,00837	0,00814	2,7
2,8	0,00792	0,00770	0,00748	0,00727	0,00707	0,00687	0,00668	0,00649	0,00631	0,00613	2,8
2,9	0,00595	0,00578	0,00562	0,00545	0,00530	0,00514	0,00499	0,00485	0,00470	0,00457	2,9
3,0	0,00443	0,00430	0,00417	0,00405	0,00393	0,00381	0,00370	0,00358	0,00348	0,00337	3,0
3,1	0,00327	0,00317	0,00307	0,00298	0,00288	0,00279	0,00271	0,00262	0,00254	0,00246	3,1
3,2	0,00238	0,00231	0,00224	0,00216	0,00210	0,00203	0,00196	0,00190	0,00184	0,00178	3,2
3,3	0,00172	0,00167	0,00161	0,00156	0,00151	0,00146	0,00141	0,00136	0,00132	0,00127	3,3
3,4	0,00123	0,00119	0,00115	0,00111	0,00107	0,00104	0,00100	0,00097	0,00094	0,00090	3,4
3,5	0,00087	0,00084	0,00081	0,00079	0,00076	0,00073	0,00071	0,00068	0,00066	0,00063	3,5
3,6	0,00061	0,00059	0,00057	0,00055	0,00053	0,00051	0,00049	0,00047	0,00046	0,00044	3,6
3,7	0,00042	0,00041	0,00039	0,00038	0,00037	0,00035	0,00034	0,00033	0,00031	0,00030	3,7
3,8	0,00029	0,00028	0,00027	0,00026	0,00025	0,00024	0,00023	0,00022	0,00021	0,00021	3,8
3,9	0,00020	0,00019	0,00018	0,00018	0,00017	0,00016	0,00016	0,00015	0,00014	0,00014	3,9
x	0	1	2	3	4	5	6	7	8	9	x

E.2 Hàm tích phân Laplace

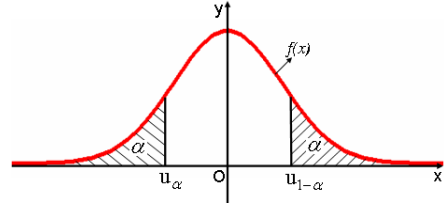
$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt.$$



x	0	1	2	3	4	5	6	7	8	9	x
0,0	0,00000	0,00399	0,00798	0,01197	0,01595	0,01994	0,02392	0,02790	0,03188	0,03586	0,0
0,1	0,03983	0,04380	0,04776	0,05172	0,05567	0,05962	0,06356	0,06749	0,07142	0,07535	0,1
0,2	0,07926	0,08317	0,08706	0,09095	0,09483	0,09871	0,10257	0,10642	0,11026	0,11409	0,2
0,3	0,11791	0,12172	0,12552	0,12930	0,13307	0,13683	0,14058	0,14431	0,14803	0,15173	0,3
0,4	0,15542	0,15910	0,16276	0,16640	0,17003	0,17364	0,17724	0,18082	0,18439	0,18793	0,4
0,5	0,19146	0,19497	0,19847	0,20194	0,20540	0,20884	0,21226	0,21566	0,21904	0,22240	0,5
0,6	0,22575	0,22907	0,23237	0,23565	0,23891	0,24215	0,24537	0,24857	0,25175	0,25490	0,6
0,7	0,25804	0,26115	0,26424	0,26730	0,27035	0,27337	0,27637	0,27935	0,28230	0,28524	0,7
0,8	0,28814	0,29103	0,29389	0,29673	0,29955	0,30234	0,30511	0,30785	0,31057	0,31327	0,8
0,9	0,31594	0,31859	0,32121	0,32381	0,32639	0,32894	0,33147	0,33398	0,33646	0,33891	0,9
1,0	0,34134	0,34375	0,34614	0,34849	0,35083	0,35314	0,35543	0,35769	0,35993	0,36214	1,0
1,1	0,36433	0,36650	0,36864	0,37076	0,37286	0,37493	0,37698	0,37900	0,38100	0,38298	1,1
1,2	0,38493	0,38686	0,38877	0,39065	0,39251	0,39435	0,39617	0,39796	0,39973	0,40147	1,2
1,3	0,40320	0,40490	0,40658	0,40824	0,40988	0,41149	0,41309	0,41466	0,41621	0,41774	1,3
1,4	0,41924	0,42073	0,42220	0,42364	0,42507	0,42647	0,42785	0,42922	0,43056	0,43189	1,4
1,5	0,43319	0,43448	0,43574	0,43699	0,43822	0,43943	0,44062	0,44179	0,44295	0,44408	1,5
1,6	0,44520	0,44630	0,44738	0,44845	0,44950	0,45053	0,45154	0,45254	0,45352	0,45449	1,6
1,7	0,45543	0,45637	0,45728	0,45818	0,45907	0,45994	0,46080	0,46164	0,46246	0,46327	1,7
1,8	0,46407	0,46485	0,46562	0,46638	0,46712	0,46784	0,46856	0,46926	0,46995	0,47062	1,8
1,9	0,47128	0,47193	0,47257	0,47320	0,47381	0,47441	0,47500	0,47558	0,47615	0,47670	1,9
2,0	0,47725	0,47778	0,47831	0,47882	0,47932	0,47982	0,48030	0,48077	0,48124	0,48169	2,0
2,1	0,48214	0,48257	0,48300	0,48341	0,48382	0,48422	0,48461	0,48500	0,48537	0,48574	2,1
2,2	0,48610	0,48645	0,48679	0,48713	0,48745	0,48778	0,48809	0,48840	0,48870	0,48899	2,2
2,3	0,48928	0,48956	0,48983	0,49010	0,49036	0,49061	0,49086	0,49111	0,49134	0,49158	2,3
2,4	0,49180	0,49202	0,49224	0,49245	0,49266	0,49286	0,49305	0,49324	0,49343	0,49361	2,4
2,5	0,49379	0,49396	0,49413	0,49430	0,49446	0,49461	0,49477	0,49492	0,49506	0,49520	2,5
2,6	0,49534	0,49547	0,49560	0,49573	0,49585	0,49598	0,49609	0,49621	0,49632	0,49643	2,6
2,7	0,49653	0,49664	0,49674	0,49683	0,49693	0,49702	0,49711	0,49720	0,49728	0,49736	2,7
2,8	0,49744	0,49752	0,49760	0,49767	0,49774	0,49781	0,49788	0,49795	0,49801	0,49807	2,8
2,9	0,49813	0,49819	0,49825	0,49831	0,49836	0,49841	0,49846	0,49851	0,49856	0,49861	2,9
3,0	0,49865	0,49869	0,49874	0,49878	0,49882	0,49886	0,49889	0,49893	0,49896	0,49900	3,0
3,1	0,49903	0,49906	0,49910	0,49913	0,49916	0,49918	0,49921	0,49924	0,49926	0,49929	3,1
3,2	0,49931	0,49934	0,49936	0,49938	0,49940	0,49942	0,49944	0,49946	0,49948	0,49950	3,2
3,3	0,49952	0,49953	0,49955	0,49957	0,49958	0,49960	0,49961	0,49962	0,49964	0,49965	3,3
3,4	0,49966	0,49968	0,49969	0,49970	0,49971	0,49972	0,49973	0,49974	0,49975	0,49976	3,4
3,5	0,49977	0,49978	0,49978	0,49979	0,49980	0,49981	0,49981	0,49982	0,49983	0,49983	3,5
3,6	0,49984	0,49985	0,49985	0,49986	0,49986	0,49987	0,49987	0,49988	0,49988	0,49989	3,6
3,7	0,49989	0,49990	0,49990	0,49990	0,49991	0,49991	0,49992	0,49992	0,49992	0,49992	3,7
3,8	0,49993	0,49993	0,49993	0,49994	0,49994	0,49994	0,49994	0,49995	0,49995	0,49995	3,8
3,9	0,49995	0,49995	0,49996	0,49996	0,49996	0,49996	0,49996	0,49996	0,49997	0,49997	3,9
x	0	1	2	3	4	5	6	7	8	9	x

E.3 Phân vị chuẩn

$$\mathbb{P}(U < u_\alpha) = \alpha.$$



α	u_α	α	u_α	α	u_α	α	u_α	α	u_α
0,50	0,000	0,840	0,994	0,880	1,175	0,920	1,405	0,960	1,751
0,51	0,025	0,841	0,999	0,881	1,180	0,921	1,412	0,961	1,762
0,52	0,050	0,842	1,003	0,882	1,185	0,922	1,419	0,962	1,774
0,53	0,075	0,843	1,007	0,883	1,190	0,923	1,426	0,963	1,787
0,54	0,100	0,844	1,011	0,884	1,195	0,924	1,433	0,964	1,799
0,55	0,126	0,845	1,015	0,885	1,200	0,925	1,440	0,965	1,812
0,56	0,151	0,846	1,019	0,886	1,206	0,926	1,447	0,966	1,825
0,57	0,176	0,847	1,024	0,887	1,211	0,927	1,454	0,967	1,838
0,58	0,202	0,848	1,028	0,888	1,216	0,928	1,461	0,968	1,852
0,59	0,228	0,849	1,032	0,889	1,221	0,929	1,468	0,969	1,866
0,60	0,253	0,850	1,036	0,890	1,227	0,930	1,476	0,970	1,881
0,61	0,279	0,851	1,041	0,891	1,232	0,931	1,483	0,971	1,896
0,62	0,305	0,852	1,045	0,892	1,237	0,932	1,491	0,972	1,911
0,63	0,332	0,853	1,049	0,893	1,243	0,933	1,499	0,973	1,927
0,64	0,358	0,854	1,054	0,894	1,248	0,934	1,506	0,974	1,943
0,65	0,385	0,855	1,058	0,895	1,254	0,935	1,514	0,975	1,960
0,66	0,412	0,856	1,063	0,896	1,259	0,936	1,522	0,976	1,977
0,67	0,440	0,857	1,067	0,897	1,265	0,937	1,530	0,977	1,995
0,68	0,468	0,858	1,071	0,898	1,270	0,938	1,538	0,978	2,014
0,69	0,496	0,859	1,076	0,899	1,276	0,939	1,546	0,979	2,034
0,70	0,524	0,860	1,080	0,900	1,282	0,940	1,555	0,980	2,054
0,71	0,553	0,861	1,085	0,901	1,287	0,941	1,563	0,981	2,075
0,72	0,583	0,862	1,089	0,902	1,293	0,942	1,572	0,982	2,097
0,73	0,613	0,863	1,094	0,903	1,299	0,943	1,580	0,983	2,120
0,74	0,643	0,864	1,098	0,904	1,305	0,944	1,589	0,984	2,144
0,75	0,674	0,865	1,103	0,905	1,311	0,945	1,598	0,985	2,170
0,76	0,706	0,866	1,108	0,906	1,317	0,946	1,607	0,986	2,197
0,77	0,739	0,867	1,112	0,907	1,323	0,947	1,616	0,987	2,226
0,78	0,772	0,868	1,117	0,908	1,329	0,948	1,626	0,988	2,257
0,79	0,806	0,869	1,122	0,909	1,335	0,949	1,635	0,989	2,290
0,800	0,842	0,870	1,126	0,910	1,341	0,950	1,645	0,990	2,326
0,805	0,860	0,871	1,131	0,911	1,347	0,951	1,655	0,991	2,366
0,810	0,878	0,872	1,136	0,912	1,353	0,952	1,665	0,992	2,409
0,815	0,896	0,873	1,141	0,913	1,359	0,953	1,675	0,993	2,457
0,820	0,915	0,874	1,146	0,914	1,366	0,954	1,685	0,994	2,512
0,825	0,935	0,875	1,150	0,915	1,372	0,955	1,695	0,995	2,576
0,830	0,954	0,876	1,155	0,916	1,379	0,956	1,706	0,996	2,652
0,835	0,974	0,877	1,160	0,917	1,385	0,957	1,717	0,997	2,748
0,837	0,982	0,878	1,165	0,918	1,392	0,958	1,728	0,998	2,878
0,839	0,990	0,879	1,170	0,919	1,398	0,959	1,739	0,999	3,090
α	u_α	α	u_α	α	u_α	α	u_α	α	u_α

