Mục lục

Chương 1

Biến cố ngẫu nhiên và xác suất của nó

1.1. Phép thử và phân loại biến cố

1.1.1. Định nghĩa

⋆ **Định nghĩa 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, còn hiện tượng có thể xảy ra trong kết quả của phép thử đó được gọi là biến cố.

•**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 một phát súng vào bia. Việc bắn súng là phép thử, còn việc trúng vào một miền nào đó của bia là biến cố.

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. Biến cố chắc chắn được ký hiệu là U.
- Biến cố không thể có: là biến cố nhất định không xảy ra khi thực hiện một phép thử. Biến cố không thể có được ký hiệu là V.
- 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 được ký hiệu là A, B, C, ... hoặc $A_1, A_2, ...A_n, B_1, B_2, ..., B_n$.

Tất cả các biến cố ta gặp trong thực tế đều thuộc về một trong ba loại biến cố kể trên. Tuy nhiên các biến cố ngẫu nhiên là các biến cố thường gặp hơn cả.

•**Ví dụ 1.3** Tung một con xúc xắc, xét các biến cố sau đây:

U = "Xuất hiện mặt có số chấm nhỏ hơn 7". U là biến cố chắc chắn.

V = "Xuất hiện mặt có 8 chấm". V 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", $(i = 1, 2, ...6)$. A_i là các biến cố ngẫu nhiên.

1.2. Định nghĩa xác suất

1.2.1. Xác suất của biến cố

⋆ **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. Ký hiệu xác suất của biến cố A là $P(A)$.

Ta chú ý rằng, việc biến cố ngẫu nhiên xảy ra hay không xảy ra trong kết quả của phép thử là điều không thể đoán trước được, xác suất của một biến cố chỉ phản ánh khả năng khách quan xuất hiện biến cố, do những điều kiện của phép thử quy định chứ không tuỳ thuộc vào ý muốn chủ quan của con người.

1.2.2. Định nghĩa cổ điển về xác suất

a) Ví dụ mở đầu.

Giả sử thực hiện phép thử là tung một con súc sắc cân đối và đồng chất. Xét biến cố A = "Xuất hiện mặt có số chấm chẵn". Ta sẽ xác định xác suất của biến cố A.

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 chỉ xảy ra một và chỉ một kết cục trong số đó; hơn nữa chúng có khả năng xảy ra như nhau. Các kết cục thoả mãn hai điều kiện trên được gọi các kết cục duy nhất đồng khả năng.

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. Những kết cục làm cho biến cố xẩy ra được gọi là các kết cục thuận lợi cho biến cố.

Như vậy ta thấy khả năng xảy ra của biến cố A là 3 phần 6, tức là 1 phần 2. Đó là cách xác định xác suất của biến cố theo quan điểm cổ điển.

b) Định nghĩa

⋆ **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 ký 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, ta có công thức tính xác suất của biến cố A như sau:

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

c) Các tính chất của xác suất

- Xác suất của biến cố chắc chắn bằng một: P(U) = 1.
- Xác suất của biến cố không thể có bằng không: P(V) = 0.
- Xác suất của biến cố ngẫu nhiên là một số nằm trong khoảng giữa không và một:

$$0 < P(A) < 1$$

Như vậy, xác suất của một biến cố bất kỳ luôn thoả mãn điều kiện:

$$0 \leq P(A) \leq 1$$

d) Tính xác suất bằng định nghĩa cổ điển

•**Ví dụ 1.4** Một người khi gọi điện cho bạn quên mất 3 chữ số cuối và chỉ nhớ rằng chúng khác nhau. Tìm xác suất để quay ngẫu nhiên một lần được đúng số cần gọi.

Lời giải.

Gọi A là biến cố "Quay ngẫu nhiên một lần được ngay số cần gọi".

Số kết cục đồng khả năng là tất cả các cách lập nên một bộ 3 số khác nhau từ 10 số tự nhiên đầu tiên. Như vậy:

$$n = A_{10}^{3} = 10.9.8 = 720.$$

Số kết cục thuận lợi cho biến cố A chỉ có một kết cục:

$$m = 1$$

Vì vậy theo định nghĩa cổ điển, xác suất của biến cố A là:

$$P(A) = \frac{m}{n} = \frac{1}{720}.$$

•**Ví dụ 1.5** Trong bình có a quả cầu trắng và b quả cầu đen. Lấy ngẫu nhiên một quả cầu. Tìm xác suất để lấy được một quả cầu trắng.

Lời giải.

Gọi A là biến cố lấy được cầu trắng.

Khi lấy ngẫu nhiên một quả cầu, ta có thể lấy được bất kỳ quả cầu nào trong số $a + b$ quả cầu trong bình, vì vậy số kết cục duy nhất đồng khả năng là:

$$n = a + b$$

Biến cố A sẽ xảy ra khi ta lấy được một trong số a quả cầu trắng, như vậy số kết cục thuận lợi là:

$$m = a$$

Từ đó theo định nghĩa cổ điển về xác suất, ta có:

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

•**Ví dụ 1.6** 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:

a) Cả 3 sản phẩm lấy ra đều là chính phẩm.

b) 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: $P(A) = \frac{m}{n} = \frac{20}{120} = \frac{1}{6}$.

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:

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

Vậy xác suất của biến cố B là:

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

•**Ví dụ 1.7** Tung một con súc sắc hai lần. Tính xác suất để trong đó có đúng 1 lần xuất hiện mặt 6 chấm.

Lời giải.

Gọi A là biến cố: "Trong 2 lần tung súc sắc có đúng một lần xuất hiện mặt 6 chấm". Số kết cục đồng khả năng là số cách thiết lập được cặp 2 số lần lượt là số chấm xuất hiện của mỗi lần tung súc sắc. Nó bằng số cách chọn một bộ 2 số (có thể giống nhau và có quan tâm đến thứ tự) từ 6 số $\{1, 2, ..., 6\}$, vậy:

$$n = \widetilde{A}_6^2 = 6^2 = 36.$$

Số kết cục thuận lợi cho biến cố A xảy ra gồm 10 kết cục: $16, 26, 36, 46, 56, 61, 62, 63, 64, 65$, do đó: $m = 10$.

Vậy xác suất của biến cô A là: $P(A) = \frac{m}{n} = \frac{10}{36} = \frac{5}{18}$.

1.2.3. 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 cô A được tính bởi công thức:

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

Tuỳ theo Ω là đoạn thẳng, miền phẳng, mảnh mặt cong hay khối không gian mà độ đo được hiểu là độ dài, diện tích hay thể tích.

•**Ví dụ 1.8** Đường dây điện thoại ngầm nối hai trạm A, B bỗng nhiên bị đứt. Biết dây điện thoại đồng chất dài 800m chôn trong lòng đất có khả năng dứt tại mọi điểm như nhau, hãy tính xác suất dây đứt cách A không quá 100m.

Lời giải. Vì dây có thể đứt tại một điểm bất kỳ trên đoạn thẳng AB với khả năng như nhau nên có thể biểu thị tập hợp mọi kết cục đồng khả năng của phép thử bởi đoạn AB. Các kết cục thích hợp cho biến cố "dây đứt cách A không quá 100m được biểu thị bởi đoạn AC. Do đó xác suất cần tìm bằng:

$$P(A) = \frac{\text{độ dài } AC}{\text{độ dài } AB} = \frac{100}{800} = \frac{1}{8}.$$

•**Ví dụ 1.9** Hai người bạn A và B hẹn gặp nhau tại một địa điểm xác định trong vòng từ 0 giờ đến 1 giờ. Mỗi người đến chỗ hẹn vào một thời điểm bất kỳ trong khoảng thời gian trên và chờ 20 phút, nếu không gặp người kia thì bỏ đi. Tính xác suất để họ gặp nhau.

Lời giải.

Gọi x (phút) là lúc đến của A, y (phút) là lúc đến của B.

Mọi kết cục đồng khả năng là một cặp số (x,y):

$$0 \le x \le 60; 0 \le y \le 60.$$

Vậy miền kết cục đồng khả năng Ω là hình vuông OACB trên hình vẽ.

Các kết cục thuận lợi cho hai người gặp nhau là những cặp (x,y) sao cho: $|x-y| \leq 20$. Trên hình vẽ, tập hợp này ứng với miền con của hình vuông OACB gồm giữa các đường thẳng $y = x + 20$ và $y = x - 20$.

Vậy xác suất để hai người gặp nhau là:

$$P = \frac{\text{diện tích } (G)}{\text{diện tích } (\Omega)} = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}.$$

1.2.4. Định nghĩa thống kê về xác suất

a) Tần suất xuất hiện của biến cố

⋆ **Định nghĩa 1.5** Tần suất xuất hiện biến cố trong n phép thử là tỷ số giữa số phép thử trong đó biến cố xuất hiện và tổng số phép thử được thực hiện.

Như vậy, nếu ký 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:

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

•**Ví dụ 1.10** Khi kiểm tra ngẫu nhiên 80 sản phẩm do một máy sản xuất, người ta phát hiện ra 3 phế phẩm. Gọi A là biến cố "Xuất hiện phế phẩm". Vậy tần suất phát hiện phế phẩm bằng :

$$f(A) = \frac{3}{80}.$$

•**Ví dụ 1.11** Bắn 50 phát đạn vào bia thấy có 47 phát trúng. Gọi A là biến cố "Bắn trúng bia". Tần suất của việc bắn trúng bia bằng:

$$f(A) = \frac{47}{50}.$$

•**Ví dụ 1.12** Để nghiên cứu khả năng xuất hiện mặt sấp 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 0,5.

Tính ổn định của tần suất là cơ sở để đưa ra định nghĩa thống kê về xác suất.

b) Định nghĩa thống kê về xác suất

⋆ **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.

$$f \xrightarrow[n\to\infty]{\text{Hội tụ theo xác suất}} p \text{ (ký hiệu là } f \xrightarrow{p} p).$$

Như vậy về mặt thực tế, với số phép thử n đủ lớn ta có thể lấy $P(A) \approx f(A)$. 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 tuỳ ý ta luôn có:

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

1.3. Quan hệ giữa các biến cố

1.3.1. Tổng các biến cố

⋆ **Định nghĩa 1.7** Biến cố C được gọi là tổng của hai biến cố A và B, ký hiệu C = A + B, nếu C sẽ xảy ra khi có ít nhất một trong hai biến cố ấy xảy ra.

•**Ví dụ 1.13** Hai người cùng bắn vào một bia. Gọi A là biến cố "Người thứ nhất bắn trúng", B là biến cố "Người thứ hai bắn trúng", C là biến cố "Có người bắn trúng". Vậy C= A + B.

•**Ví dụ 1.14** Tung một con súc sắc. Gọi A_i là biến cố "Xuất hiện mặt i chấm" $(i = \overline{1,6})$, B là biến cố "Xuất hiện ít nhất 5 chấm". Biến cố B sẽ xảy ra khi A_5 xảy ra hoặc A_6 xảy ra. Vậy $B = A_5 + A_6$.

⋆ **Định nghĩa 1.8** Biến cố A được gọi là tổng của n biến cố $A_1, A_2, ..., A_n$, ký hiệu $\sum\limits_{i=1}^{n} A_i$, nếu A sẽ xảy ra khi có ít nhất một trong n biến cố ấy xảy ra.

1.3.2. Tích các biến cố

⋆ **Định nghĩa 1.9** Biến cố C được gọi là tích của hai biến cố A và B, ký hiệu $C = A.B$, nếu C sẽ xảy ra khi và chỉ khi cả hai biến cố A và B cùng xảy ra.

•**Ví dụ 1.15** Hai người cùng bắn vào một bia. Gọi A là biến cố "Người thứ nhất bắn trúng", B là biến cố "Người thứ hai bắn trúng", C là biến cố "Cả hai người cùng bắn trúng". Vậy $C = A.B$.

•**Ví dụ 1.16** Một mạch điện gồm hai bóng đèn mắc song song.

Gọi A là biến cố "Bóng thứ nhất bị cháy khi điện quá tải", B là biến cố "Bóng thứ hai bị cháy khi điện quá tải", C là biến cố "Mạch điện bị ngắt khi điện quá tải". Rõ ràng biến cố C chỉ xảy ra khi cả hai biến cố A và B cùng đồng thời xảy ra. Vậy $C = A.B$.

⋆ **Định nghĩa 1.10** Biến cố A được gọi là tích của n biến cố $A_1, A_2, ..., A_n$, ký hiệu $\prod\limits_{i=1}^{n} A_i$, nếu A sẽ xảy ra khi và chỉ khi cả n biến cố nói trên xảy ra.

1.3.3. Biến cố xung khắc

⋆ **Định nghĩa 1.11** Hai biến cố A và B được gọi là xung khắc với nhau nếu chúng không thể đồng thời xảy ra trong một phép thử. Trường hợp ngược lại, nếu hai biến cố đó có thể cùng xảy ra trong một phép thử thì được gọi là không xung khắc.

•**Ví dụ 1.17** Hai người cùng bắn vào một bia. Gọi A là biến cố ″Người thứ nhất bắn trúng", B là biến cố ″Người thứ hai bắn trúng". Hai biến cố A và B là không xung khắc

•**Ví dụ 1.18** Một bình có 3 loại cầu là cầu trắng, cầu xanh và cầu đỏ. Lấy ngẫu nhiên từ binh đó một quả cầu. Gọi A là biến cố ″Lấy được cầu trắng", B là biến cố ″Lấy được cầu xanh". Các biến cố A và B là xung khắc với nhau.

⋆ **Định nghĩa 1.12** Nhóm n biến cố $\{A_1, A_2, ..., A_n\}$ được gọi là xung khắc từng đôi nếu bất kỳ hai biến cố nào trong nhóm này cũng xung khắc với nhau.

Chú ý rằng việc nhận xét tính xung khắc của các biến cố chủ yếu dựa vào trực giác.

•**Ví dụ 1.19** Tung một con súc sắc. Gọi A_i là biến cố ″Xuất hiện mặt i chấm″ $(i = \overline{1,6})$. Các biến cố $\{A_1, A_2, ..., A_n\}$ là xung khắc từng đôi.

1.3.4. Nhóm đầy đủ các biến cố

⋆ **Định nghĩa 1.13** Các biến cố $\{A_1, A_2, ..., A_n\}$ được gọi là nhóm biến cố 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 $\{A_1, A_2, ..., A_n\}$ 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.

•**Ví dụ 1.20** Tung một con súc sắc.

Gọi A_i là biến cố ″Xuất hiện mặt i chấm″ $(i = \overline{1,6})$. Các biến cố $\{A_1, A_2, ..., A_n\}$ tạo nên nhóm đầy đủ các biến cố.

1.3.5. Biến cố đối lập

⋆ **Định nghĩa 1.14** Hai biến cố A và B được gọi là đối lập với nhau nếu chúng tạo nên một nhóm đầy đủ các biến cố. Ký hiệu biến cố đối lập của biến cố A là $\overline{A}$.

•**Ví dụ 1.21** Bắn một phát đạn vào bia.

Gọi A là biến cố ″Bắn trúng bia", B là biến cố ″Bắn trượt bia". Các biến cố A và B là hai biến cố đối lập nhau: $A = \overline{B}; B = \overline{A}$.

•**Ví dụ 1.22** Một hòm có a chính phẩm và b phế phẩm, lấy ngẫu nhiên n sản phẩm $(n \leq a + b)$. Gọi A là biến cố ″Lấy được ít nhất một chính phẩm". Khi đó biến cố đối lập của A là biến cố $\overline{A}$: "Trong n sản phẩm lấy ra không có chính phẩm nào (toàn phế phẩm)".

•**Ví dụ 1.23** Một xí nghiệp có 3 ôtô cùng hoạt động.

Gọi biến cố A_i là "ôtô thứ i bị hỏng", (i = 1,2,3). Hãy biểu diễn các biến cố sau theo biến cố A_1, A_2, A_3.

A = "Chỉ ôtô thứ nhất hỏng"

B = "Có đúng 1 ôtô hỏng"

C= "Có đúng 2 ôtô hỏng"

D = "Có ôtô hỏng"

Lời giải.

$A = A_1.\overline{A_2}.\overline{A_3}$

$B = A_1.\overline{A_2}.\overline{A_3} + \overline{A_1}.A_2.\overline{A3} + \overline{A_1}.\overline{A_2}.A_3$

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

$D = A_1 + A_2 + A_3.$

1.4. Định lý cộng và nhân xác suất

1.4.1. Định lý cộng xác suất (trường hợp các biến cố xung khắc)

◊ **Định lý 1.1** *Xác suất của tổng hai biến cố xung khắc bằng tổng xác suất của các biến cố đó. Như vậy, nếu A và B là hai biến cố xung khắc với nhau thì:*

$$P(A + B) = P(A) + P(B)$$

■.Ta ký hiệu:

n - số kết cục đồng khả năng có thể xảy ra khi phép thử được thực hiện.

m_1 - số kết cục thuận lợi cho biến cố A.

m_2 - số kết cục thuận lợi cho biến cố B.

Do A và B xung khắc nhau do đó không thể có kết cục thuận lợi cho cả A và B đồng thời xảy ra. Vậy số kết cục thuận lợi cho A hoặc B xảy ra bằng $m_1 + m_2$.

Vì vậy:

$$P(A + B) = \frac{m_1 + m_2}{n} = \frac{m_1}{n} + \frac{m_2}{n} = P(A) + P(B).$$

□

▽ ***Hệ quả 1.1*** *Xác suất của tổng các biến cố xung khắc từng đôi bằng tổng xác suất của các biến cố đó.*

Tức là nếu $A_1, A_2, ..., A_n$ là các biến cố xung khắc từng đôi thì:

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

▽ ***Hệ quả* 1.2** *Nếu các biến cố $A_1, A_2, ..., A_n$ tạo nên một nhóm đầy đủ các biến cố thì tổng các xác suất của chúng bằng 1.*

$$\sum_{i=1}^{n} P(A_i) = 1.$$

▽ ***Hệ quả* 1.3** *Tổng xác suất của các biến cố đối lập nhau bằng 1.*

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

•Ví dụ 1.24 Xác suất để một xạ thủ bắn bia trúng điểm 10 là 0,1; trúng điểm 9 là 0,2; trúng điểm 8 là 0,25 và ít hơn 8 điểm là 0,45. Xạ thủ ấy bắn một phát đạn. Tìm xác suất để xạ thủ được ít nhất 9 điểm.

Lời giải. Gọi A là biến cố ″Xạ thủ bắn trúng điểm 10″, B là biến cố ″Xạ thủ bắn trúng điểm 9″, C là biến cố ″Xạ thủ được ít nhất 9 điểm″. Vậy:

$$C = A + B$$

Vì A và B xung khắc nhau do đó theo định lý cộng xác suất ta có:

$$P(C) = P(A + B) = P(A) + P(B) = 0,1 + 0,2 = 0,3.$$

•Ví dụ 1.25 Xác suất để sản phẩm sản xuất ra là chính phẩm bằng 0,9. Tìm xác suất sản phẩm sản xuất ra là phế phẩm.

Lời giải. Gọi A là biến cố ″Sản phẩm sản xuất ra là phế phẩm″, khi đó $\overline{A}$ là biến cố ″Sản phẩm sản xuất ra là chính phẩm″, A và $\overline{A}$ đối lập nhau, do đó:

$$P(A) = 1 - P(\overline{A}) = 1 - 0,9 = 0,1.$$

•Ví dụ 1.26 Trong hòm có 12 chi tiết, trong đó có 5 chi tiết hỏng. Lấy ngẫu nhiên 4 chi tiết. Tính xác suất trong số 4 chi tiết lấy ra:

a) có không quá một chi tiết hỏng;

b) có chi tiết hỏng.

Lời giải.

Gọi A_i là biến cố ″Trong 4 chi tiết lấy ra có đúng i chi tiết hỏng″, $i = \overline{0,4}$.

a) Gọi A là biến cố ″Trong 4 chi tiết lấy ra có không quá một chi tiết hỏng″, vậy:

$$A = A_0 + A_1$$

Vì A_0 và A_1 xung khắc với nhau do đó:

$$P(A) = P(A_o) + P(A_1)$$

Dùng định nghĩa cổ điển về xác suất ta tính được:

$$P(A_o) = \frac{C_7^4}{C_{12}^4} = \frac{7}{99}, \qquad P(A_1) = \frac{C_5^1.C_7^3}{C_{12}^4} = \frac{35}{99}$$

Vậy:

$$P(A) = \frac{7}{99} + \frac{35}{99} = \frac{14}{33}.$$

b) Gọi B là biến cố "Trong 4 chi tiết lấy ra có chi tiết hỏng", khi đó biến cố đối lập của B là "Trong 4 chi tiết lấy ra không có chi tiết hỏng" = A_o. Vậy:

$$P(B) = 1 - P(\overline{B}) = 1 - P(A_o) = 1 - \frac{7}{99} = \frac{92}{99}.$$

1.4.2. Định lý nhân xác suất

a) Biến cố độc lập

⋆ **Định nghĩa 1.15** Hai biến cố A và B gọi là độc lập với nhau nếu việc xảy ra hay không xảy ra biến cố này không làm thay đổi xác suất xảy ra của biến cố kia và ngược lại.

Trong trường hợp 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 gọi là phụ thuộc nhau.

•**Ví dụ 1.27** Trong bình có 3 cầu trắng và 2 cầu đen. Lấy ngẫu nhiên 2 quả cầu theo theo phương thức có hoàn lại. Gọi A là biến cố "Lần thứ nhất lấy được cầu trắng", B là biến cố "Lần hai lấy được cầu trắng". Khi đó:

$$P(A) = \frac{3}{5}; \qquad P(B) = \frac{3}{5}$$

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

•**Ví dụ 1.28** Trong bình có 3 cầu trắng và 2 cầu đen. Lấy ngẫu nhiên 2 quả cầu theo theo phương thức không hoàn lại. Gọi Ả là biến cố "Lần thứ nhất lấy được cầu trắng", B là biến cố "Lần hai lấy được cầu trắng". Khi đó xác suất lấy được cầu trắng lần thứ nhất là:

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

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

+ Nếu lần thứ nhất lấy được cầu trắng (A xảy ra) thì: $P(B) = \frac{2}{4}$;

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

Vậy A và B phụ thuộc nhau.

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.

⋆ **Định nghĩa 1.16** Các biến cố $A_1, A_2, ..., A_n$ gọi là độc lập từng đôi với nhau nếu mỗi cặp hai trong n biến cố đó độc lập với nhau. Chẳng hạn ba biến cố A, B, C độc lập từng đôi với nhau nếu A độc lập với B, A độc lập với C, B độc lập với C.

•**Ví dụ 1.29** Tung một đồng xu ba lần, gọi A_i là biến cố "Được mặt sấp ở lần tung thứ i, $i = 1, 2, 3$. Rõ ràng là mỗi cặp hai trong ba biến cố đó độc lập với nhau.

Vậy A_1, A_2, A_3 độc lập từng đôi với nhau.

⋆ **Định nghĩa 1.17** Các biến cố $A_1, A_2, ..., A_n$ gọi là độc lập toàn phần với nhau nếu mỗi biến cố độc lập với một tổ hợp bất kỳ của các biến cố còn lại.

Chẳng hạn ba biến cố A, B, C độc lập toàn phần với nhau nếu A độc lập với B, A độc lập với C, A độc lập với tích BC, B độc lập với C, B độc lập với tích AC, C độc lập với tích AB.

Chú ý rằng nếu các biến cố độc lập toàn phần thì suy ra chúng độc lập từng đôi, nhưng điều ngược lại không đúng, nghĩa là nếu các biến cố độc lập từng đôi thì chúng chưa chắc độc lập toàn phần. Ta cũng quy ước rằng khi nói n biến cố độc lập thì có nghĩa là chúng độc lập toàn phần.

•**Ví dụ 1.30** Giả sử trong bình có 4 quả cầu, 1 quả sơn màu đỏ, 1 quả sơn màu xanh, 1 quả màu vàng, 1 quả sơn cả màu đỏ, xanh, vàng.

Lời giải.

Gọi A ="Lấy ngẫu nhiên 1 quả cầu từ bình được quả có màu đỏ";

B ="Lấy ngẫu nhiên 1 quả cầu từ bình được quả có màu xanh";

C ="Lấy ngẫu nhiên 1 quả cầu từ bình được quả có màu vàng";

$$P(A) = P(B) = P(C) = \frac{1}{2}.$$

Nếu biến cố A xảy ra, tức là đã lấy được quả có màu đỏ, thì vì trong 2 quả màu đỏ có 1 quả có màu sơn xanh nên xác suất của biến cố B vẫn là $P(B) = \frac{1}{2}$. Còn nếu A không xảy ra, thì trong 2 quả không có màu đỏ cũng chỉ có 1 quả màu sơn xanh nên $P(B) = \frac{1}{2}$. Vậy các biến cố A và B độc lập với nhau.

Tương tự có thể chứng tỏ A và C, B và C độc lập. Vậy ba biến cố A, B, C độc lập từng đôi.

Nếu biến cố A và B đã xảy ra, tức là quả cầu lấy ra có hai màu đỏ và xanh. Khi đó nó chỉ có thể là quả cầu có cả ba màu, vì vậy xác suất để lấy được quả cầu màu vàng là $P(C) = 1$. Nếu AB không xảy ra thì quả cầu lấy ra không phải quả có ba màu, vì vậy xác suất để lấy được quả vàng là $P(C) = \frac{1}{3}$. Vậy biến cố C không độc lập với AB. Do đó các biến cố A, B, C chỉ độc lập từng đôi chứ không độc lập toàn phần.

b) Xác suất điều kiện

⋆ **Định nghĩa 1.18** 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 ký hiệu là $P(B/A)$.

•**Ví dụ 1.31** Trong bình có 3 cầu trắng và 2 cầu đen. Lấy ngẫu nhiên 2 quả cầu theo theo phương thức có hoàn lại. Gọi A là biến cố "Lần thứ nhất lấy được cầu trắng", B là biến cố "Lần hai lấy được cầu trắng". Ta đã biết ở 1.27 là A và B độc lập. Ta có:

$$P(B/A) = \frac{3}{5}; \qquad P(B/\overline{A}) = \frac{3}{5}.$$

•**Ví dụ 1.32** Trong bình có 3 cầu trắng và 2 cầu đen. Lấy ngẫu nhiên 2 quả cầu theo theo phương thức không hoàn lại. Gọi A là biến cố "Lần thứ nhất lấy được cầu trắng", B là biến cố "Lần hai lấy được cầu trắng". Ta đã biết ở 1.28 là A và B phụ thuộc. Ta có:

$$P(B/A) = \frac{2}{4}; \qquad P(B/\overline{A}) = \frac{3}{4}.$$

Dễ có một nhận xét rằng nếu A và B độc lập thì:

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

c) Định lý nhân xác suất.

♢ **Định lý 1.2** *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:*

$$P(AB) = P(A).P(B/A) = P(B).P(A/B).$$

■.Ta ký hiệu:

n - số kết cục đồng khả năng có thể xảy ra khi phép thử được thực hiện.

m_1 - số kết cục thuận lợi cho biến cố A.

m_2 - số kết cục thuận lợi cho biến cố B.

k - số kết cục thuận lợi cho cả A và B đồng thời xảy ra.

Lúc đó:

$$P(AB) = \frac{k}{n}; \qquad P(A) = \frac{m_1}{n}; \qquad P(B) = \frac{m_2}{n}.$$

Với điều kiện biến cố A đã xảy ra thì số kết cục duy nhất đồng khả năng của phép thử đối với biến cố B là m_1, trong đó có k kết cục thuận lợi cho biến cố B xảy ra, do đó:

$$P(B/A) = \frac{k}{m_1}.$$

Như vậy: $P(AB) = \frac{k}{n} = \frac{m_1}{n}.\frac{k}{m_1} = P(A).P(B/A).$

Tương tự khi tính xác suất điều kiện đối với biến cố A, từ đó suy ra điều phải chứng minh.

□

▽ **Hệ quả 1.4** *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:*

$$P(A_1.A_2...A_n) = P(A_1).P(A_2/A_1)...P(A_n/A_1...A_{n-1}).$$

▽ **Hệ quả 1.5** *Nếu A và B độc lập thì P(AB) = P(A).P(B).*

▽ **Hệ quả 1.6** *Xác suất của tích n biến cố độc lập toàn phần bằng tích các xác suất thành phần:*

$$P(A_1.A_2...A_n) = P(A_1).P(A_2)...P(A_n).$$

d) Các ví dụ

•**Ví dụ 1.33** Trong bình có 3 cầu trắng và 2 cầu đen. Lấy ngẫu nhiên 2 quả cầu theo theo phương thức không hoàn lại. Gọi A là biến cố "Lần thứ nhất lấy được cầu trắng", B là biến cố "Lần hai lấy được cầu trắng". Ta đã biết ở 1.32 là A và B phụ thuộc và

$$P(A) = \frac{3}{5}; \qquad P(B/A) = \frac{2}{4}P(B/\overline{A}) = \frac{3}{4}.$$

Khi đó xác suất để cả hai lần đều lấy được cầu trắng là:

$$P(AB) = P(A).P(B/A) = \frac{3}{5}.\frac{2}{4} = \frac{3}{10}.$$

•**Ví dụ 1.34** Một xí nghiệp có 3 ôtô hoạt động độc lập. Xác suất để trong một ngày các ôtô bị hỏng tương úng là 0,1; 0,2; 0,15.

a) Tính xác suất để cả 3 ôtô đều hỏng.

b) Tính xác suất để trong một ngày có đúng 1 ôtô hỏng.

c) Tính xác suất để trong một ngày có ít nhất một ôtô hỏng.

Lời giải. Gọi A_i là biến cố "ôtô thứ i bị hỏng trong ngày"; $(i = \overline{1,3})$.

Theo giả thiết: $P(A_1) = 0,1; P(A_2) = 0,2; P(A_3) = 0,15.$

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

a) Gọi A là biến cố "cả 3 ôtô đều hỏng", ta có:

$$A = A_1.A_2.A_3$$

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

$$P(A) = P(A_1).P(A_2).P(A_3) = 0,1.0,2.0,15 = 0,003.$$

b) Gọi B là biến cố "trong một ngày có đúng 1 ôtô hỏng", ta có:

$$B = A_1.\overline{A_2}.\overline{A_3} + \overline{A_1}.A_2.\overline{A3} + \overline{A_1}.\overline{A_2}.A_3$$

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

$$\begin{aligned}
P(B) &= A_1.\overline{A_2}.\overline{A_3} + \overline{A_1}.A_2.\overline{A3} + \overline{A_1}.\overline{A_2}.A_3 && (1.1)\\
&= P(A_1).P(\overline{A_2}).P(\overline{A_3}) + P(\overline{A_1}).P(A_2).P(\overline{A3}) + P(\overline{A_1}).P(\overline{A_2}).P(A_3) && (1.2)\\
&= 0,1.0,8.0,85 + 0,9.0,2.0,85 + 0,9.0,8.0,15 = 0,329. && (1.3)\\
& && (1.4)
\end{aligned}$$

c) Gọi C là biến cố "trong một ngày có ít nhất một ôtô 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 ôtô đều hoạt động tốt"

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

Theo công thức nhân xác suất:

$$P(\overline{A_1}.\overline{A_2}.\overline{A_3}) = P(\overline{A_1}).P(\overline{A_2}).P(\overline{A_3}) = 0,9.0,8.0,85 = 0,612.$$

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

•**Ví dụ 1.35** 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 các biến cố, vì vậy:

$$P(A) + P(B) + P(C) = 1.$$

Vì mặt sấp và ngửa đối xứng nhau nên: P(A) = P(B), từ đó ta có: $P(A) = \dfrac{1 - P(C)}{2}$.

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 ra $m = C_6^3$, vậy:

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

Xác suất cần tìm là:

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

1.4.3. Định lý cộng xác suất (trường hợp tổng quát)

◊ **Định lý 1.3** *Xác suất của tổng hai biến cố không xung khắc bằng tổng xác suất các biến cố đó trừ đi xác suất của tích các biến cố đó.*

$$P(A + B) = P(A) + P(B) - P(AB).$$

■.Giả sử:

n là số kết cục đồng khả năng có thể xảy ra trong phép thử;

m_1 là số kết cục thuận lợi cho biến cố A xảy ra;

m_2 là số kết cục thuận lợi cho biến cố B xảy ra;

k là số kết cục thuận lợi cho biến cố tích AB xảy ra;

Khi đó số kết cục thuận lợi cho ít nhất một trong hai biến cố A hoặc B xảy ra là $m_1 + m_2 - k$.

Như vậy: $P(A + B) = \dfrac{m_1 + m_2 - k}{n} = \dfrac{m_1}{n} + \dfrac{m_2}{n} - \dfrac{k}{n} = P(A) + P(B) - P(AB)$. □

▽ ***Hệ quả* 1.7** *Nếu hai biến cố A, B phụ thuộc thì xác suất của tổng hai biến cố là:*

$$P(A + B) = P(A) + P(B) - P(A).P(B/A)$$

▽ ***Hệ quả* 1.8** *Nếu hai biến cố A, B độc lập thì xác suất của tổng hai biến cố là:*

$$P(A + B) = P(A) + P(B) - P(A).P(B)$$

▽ ***Hệ quả* 1.9** *Nếu hai biến cố A, B xung khắc thì xác suất của tổng hai biến cố là:*

$$P(A + B) = P(A) + P(B)$$

▽ ***Hệ quả* 1.10** *Xác suất của tổng n biến cố không xung khắc được xác định bằng công thức:*

$$P\left(\sum_{i-1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k) - ... + (-1)^{n-1}P(A_1A_2...A_n).$$

▽ ***Hệ quả* 1.11** *Xác suất của tích n biến cố được xác định bằng công thức:*

$$P\left(\prod_{i-1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i + A_j) + \sum_{i<j<k} P(A_i + A_j + A_k) - ... + (-1)^{n-1}P(A_1 + A_2 + ... + A_n).$$

•**Ví dụ 1.36** 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:

$$B = A_1A_2 + A_3$$

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_1A_2 và A_3, sau đó áp dụng công thức nhân xác suất cho các biến cố độc lập ta được:

$$P(B) = P(A_1.A_2) + P(A_3) - P(A_1.A_2.A_3) = P(A_1).P(A_2) + P(A_3) - P(A_1).P(A_2).P(A_3)$$

$$= 0,1.0,2 + 0,05 - 0,1.0,2.0,05 = 0,069.$$

1.4.4. Định lý liên hệ cộng và nhân xác suất

♢ **Định lý 1.4** *Xác suất của tổng n biến cố không xung khắc và độc lập toàn phần với nhau bằng một trừ đi tích xác suất của các biến cố đối lập với các biến cố đó.*

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

■.Ký hiệu $A = \sum\limits_{i=1}^{n} A_i$, khi đó: $\overline{A} = \prod\limits_{i=1}^{n} \overline{A_i}$. do đó:

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

Theo giả thiết các biến cố A_i độc lập toàn phần nên các biến cố $\overline{A_i}$ cũng độc lập toàn phần với nhau, áp dụng định lý nhân xác suất:

$$P\left(\prod_{i=1}^{n} \overline{A}_i\right) = \prod_{i=1}^{n} P(\overline{A}_i).$$

Vậy

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

▽ ***Hệ quả* 1.12** *Nếu $A_1, A_2, ..., A_n$ là các biến cố không xung khắc và độc lập toàn phần có cùng xác suất $P(A_1) = P(A_2) = ...P(A_n) = p$ thì:*

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

•**Ví dụ 1.37** Phải tung một con súc sắc bao nhiêu lần để với xác suất không nhỏ hơn 0,5 có thể hy vọng rằng trong đó có ít nhất một lần đọc 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 A là biến cố "Trong n lần tung có ít nhất một lần được mặt 6 chấm". Ta có:

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

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

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

do đó:

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

Theo giả thiết xác suất của A không nhỏ hơn 0,5, ta có bất phương trình sau:

$$1 - \left(\frac{5}{6}\right)^n \geq 0,5 \Leftrightarrow \left(\frac{5}{6}\right)^n \leq 0,5$$

$$\Leftrightarrow n.ln\left(\frac{5}{6}\right) \leq ln0,5$$

$$\Leftrightarrow n \geq \frac{ln0,5}{ln(\frac{5}{6})} \approx 3,7.$$

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

1.5. Công thức Bernoulli

1.5.1. Các phép thử độc lập

⋆ **Định nghĩa 1.19** Các phép thử được gọi là độc lập với nhau nếu xác suất xảy ra một biến cố nào đó trong từng phép thử sẽ không phụ thuộc vào việc biến cố đó có xảy ra ở các phép thử khác hay không. Khi đó ta cũng gọi các phép thử độc lập như trên là một dãy phép thử độc lập.

Chẳng hạn tung nhiều lần một đồng xu sẽ tạo nên một dãy phép thử độc lập; lấy nhiều lần sản phẩm từ một lô sản phẩm theo phương thức có hoàn lại cũng tạo nên một dãy phép thử độc lập;...

1.5.2. Công thức Bernoulli

Bài toán được gọi là tuân theo lược đồ Bernoulli nếu thoả 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 $(0 < p < 1)$, xác suất không xảy ra biến cố A đều bằng $q = 1 - p$.

♢ **Định lý 1.5** *Giả sử bài toán tuân theo lược đồ Bemoulli, khi đó xác suất để trong n phép thử độc lập nói trên, biến cố A xuất hiện đúng k lần, ký hiệu là $P_n(k)$ hoặc p_k, được tính bằng công thức Bemoulli sau đây:*

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

■.Gọi A_i là biến cố "Xảy ra biến cố A ở phép thử thứ i", $(i = \overline{1, n})$;
gọi B là biến cố "Trong n phép thử biến cố A xảy ra đúng k lần", ta có:

$$B = A_1A_2...A_k\overline{A}_{k+1}...An + ... + \overline{A}_1\overline{A}_2...\overline{A}_{n-k}A_{n-k+1}...A_n.$$

Tổng số các số hạng (mỗi số hạng là một tích n biến cố) chính là số cách chọn k phép thử trong n phép thử để biến cố A xảy ra, do đó số các số hạng của tổng trên là C_n^k. Trong mỗi số hạng trên, n biến cố là độc lập, do đó xác suất mỗi biến cố tích đều bằng $p^k.q^{n-k}$. Vì các tích biến cố đó là xung khắc từng đôi với nhau nên:

$$P_n(k) = P(B) = C_n^k p^k q^{n-k}.$$

□

Nếu gọi B_k là biến cố "Trong n phép thử biến cố A xảy ra đúng k lần" thì dễ thấy $n+1$ biến cố $B_o, B_1..., B_n$ tạo thành một nhóm đầy đủ, vì vậy:

$$p_o + p_1 + p_2 + ... + p_n = 1$$

•Ví dụ 1.38 Trong phân xưởng có 7 máy hoạt động độc lập, xác suất để trong ca mỗi máy bị hỏng đều bằng 0,1. Tìm xác suất để trong ca đó:

a) có đúng 2 máy hỏng;

b) có không ít hơn 2 máy hỏng

Lời giải. Nếu coi sự hoạt động của mỗi máy là một phép thử, ta có 7 phép thử độc lập. Trong mỗi phép thử chỉ có hai trường hợp: hoặc máy hỏng, hoặc máy chạy tốt. Xác suất hỏng của mỗi máy đều bằng 0,1. Như vậy bài toán thoả mãn lược đồ Bernoulli.
a) Xác suất để trong ca có đúng 2 máy hỏng được tính bằng công thức Bernoulli như sau:

$$P_7(2) = C_7^2.(0,1)^2.(0,9)^5 \approx 0,124.$$

b) Xác suất để trong ca có không ít hơn 2 máy hỏng (tức là có nhiều hơn hoặc bằng 2 máy hỏng) là:

$P_7(2) + P_7(3) + P_7(4) + ... + P_7(7) = 1 - P_7(0) - P_7(1)$
$= 1 - C_7^0.(0,9)^7 - C_7^1(0,l).(0,9)^8 \approx 1 - 0,8503 = 0,1497.$

1.5.3. Số lần xuất hiện chắc nhất

Giả sử bài toán tuân theo lược đồ Bernoulli, khi đó số n_o ứng với xác suất $P_n(n_o)$ lớn nhất trong số các $P_n(k), k = \overline{1,n}$ là số lần xuất hiện chắc nhất của biến cố A. Ta có công thức sau đây để xác định n_o:

$$np - q \leq n_o \leq np + p$$

■.Vì n_o là giá trị ứng với xác suất lớn nhất so với các xác suất $P_n(k)$ khác nên ta có các bất đẳng thức sau:

$$\begin{cases} P_n(n_o) & \geq P_n(n_o - 1) \\ P_n(n_o) & \geq P_n(n_o + 1) \end{cases}$$

Thay các biểu thức của xác suất vào các bất đẳng thức trên theo công thức Bernoulli: $\begin{cases} C_n^{n_o} p^{n_o} q^{n-n_o} & \geq C \\ C_n^{n_o} p^{n_o} q^{n-n_o} & \geq C \end{cases}$

$$\Leftrightarrow \begin{cases} \dfrac{n!}{n_o!(n-n_o)!} p^{n_o} q^{n-n_o} & \geq \dfrac{n!}{(n_o-1)!(n-n_o+1)!} p^{n_o-1} q^{n-n_o+1} \\ \dfrac{n!}{n_o!(n-n_o)!} p^{n_o} q^{n-n_o} & \geq \dfrac{n!}{(n_o+1)!(n-n_o-1)!} p^{n_o+1} q^{n-n_o-1} \end{cases}$$

$$\Leftrightarrow \begin{cases} \frac{p}{n_o} & \geq \frac{q}{n-n_o+1} \\ \frac{q}{n-n_o} & \geq \frac{p}{n_o+1} \end{cases} \Leftrightarrow \begin{cases} p(n-n_o+1) & \geq qn_o \\ q(n_o+1) & \geq p(n-n_o) \end{cases} \Leftrightarrow \begin{cases} n_o & \leq np+p \\ n_o & \geq np-q \end{cases} \qquad \square$$

Ta chú ý rằng, vì n_o là một số nguyên, do đó có thể xảy ra hai trường hợp:

- Nếu $np+p$ là một số nguyên thì $np-q$ cũng là một số nguyên, và $P(np+p) = P(np-q)$, khi đó số lần xuất hiện chắc nhất nhận hai giá trị $n_o = np+p$ và $n_o = np-q$.

- nếu $np+p$ không là số nguyên thì số lần xuất hiện chắc nhất n_o là giá trị nguyên duy nhất nằm trong khoảng hai số $np-q$ và $np+p$.

•**Ví dụ 1.39** Xác suất để một hành khách chậm tàu là 0,02. Tìm số hành khách chậm tàu có khả năng xảy ra nhiều nhất trong một đoàn tàu 835 khách.

Lời giải. Bài toán thoả mãn lược đồ Bemoulli với $n = 835$ phép thử; đối với mỗi phép thử (việc đến tàu của mỗi hành khách) chỉ xảy ra hai trường hợp: chậm tàu (biến cố A) hoặc không chậm tàu; $p = P(A) = 0,02$. Số hành khách chậm tàu có khả năng xảy ra nhiều nhất là số lần xuất hiện chắc nhất n_o, ta có:

$$np - q \leq n_o \leq np + p$$

$$\Leftrightarrow 835.0,02 - 0,98 \leq n_o \leq 835.0,02 + 0,02$$

$$\Leftrightarrow 15,72 \leq n_o \leq 16,72$$

Vậy $n_o = 16$, tức là số hành khách chậm tàu có khả năng xảy ra nhiều nhất là 16 người.

1.5.4. Mở rộng công thức Bernoulli

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 xung khắc $A_1, A_2, ..., A_m$ hợp thành nhóm đầy đủ.

- Trong mỗi phép thử, xác suất xuất hiện sự kiện A_i bằng p_i (không phụ thuộc chỉ số của phép thử), $p_1 + p_2 + ... + p_m = 1$.

Ký hiệu $P_n(k_1, k_2, ..., k_m)$ là xác suất để trong n phép thử đó:
biến cố A_1 xảy ra k_1 lần;
biến cố A_2 xảy ra k_2 lần;
...
biến cố A_m xảy ra k_m lần;
Ta có:

$$P_n(k_1, k_2, ..., k_m) = C_n^{k_1}.C_{n-k_1}^{k_2}...C_{n-k_1-k_2-...k_{m-1}}^{k_m}.p_1^{k_1}.p_2^{k_2}...p_m^{k_m} = \frac{n!}{k_1!k_2!...k_m!}.p_1^{k_1}.p_2^{k_2}...p_m^{k_m}. \qquad (1.5)$$

Thật vậy, gọi B là sự kiện "Trong n phép thử, biến cố A_i xảy ra đúng k_i lần". Biến cố B là tổng của các biến cố dạng:

$$\underbrace{A_1.A_1...A_1}_{k_1 \text{thừa số}}.\underbrace{A_2.A_2...A_2}_{k_2 \text{thừa số}}...\underbrace{A_m.A_m...A_m}_{k_m \text{thừa số}}$$

Vì các biến cố A_i có thể được sắp xếp theo các thứ tự khác, sao cho đảm bảo có k_i biến cố A_i nên số các biến cố dạng trên là:

$$C_n^{k_1}.C_{n-k_1}^{k_2}...C_{n-k_1-k_2-...k_{m-1}}^{k_m} = \frac{n!}{k_1!(n-k_1)!}.\frac{(n-k_1)!}{k_2!(n-k_1-k_2)!}...\frac{(n-k_1-k_2-...k_{m-1})!}{k_m!(n-k_1-k_2-...k_m)!}.$$

Vì mỗi biến cố dạng trên có cùng xác suất là $p_1^{k_1}.p_2^{k_2}...p_m^{k_m}$ và các biến cố dạng này là xung khắc nên ta có công thức 1.5. □

•Ví dụ 1.40 Tại một ga xép có 10 khách lên tàu gồm 3 toa. Xác suất một khách bất kỳ lên toa I, II và III theo thứ tự là 0,2 ; 0,5 và 0,3. Tính xác suất:

a) Có đúng 3 khách lên toa I.

b) Có 3 khách lên toa I, 5 khách lên toa II và 2 khách lên toa III;

c) 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 thực hiện 1 phép thử, như vậy ta có dãy 10 phép thử độc lập.

a) Bài toán tuân theo lược đồ Bemoulli với n = 10; p = 0,2. Xác suất cần tìm là:

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

b) Bài toán tuân theo lược đồ Bemoulli mở rộng với $n = 10; p_1 = 0,2; p_2 = 0,5; p_3 = 0,3$. Xác suất cần tìm là:

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

c) Bài toán tuân theo lược đồ Bernoulli mở rộng tương tự phần b). 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:

$$P_{10}(0,6,4) + P_{10}(0,5,5) + P_{10}(1,5,4) = C_{10}^6.(0,5)^6.(0,3)^4 + C_{10}^5.(0,5)^5.(0,3)^5 + C_{10}^1.C_9^5.0,2.(0,5)^5.(0,3)^4$$

$$\approx 0,1095.$$

1.6. Công thức đầy đủ và công thức Bayes

1.6.1. Công thức xác suất đầy đủ

♢ **Định lý 1.6** *Giả sử n biến cố $H_1, H_2, ..., H_n$ lập thành nhóm đầy đủ các biến cố; biến cố A có thể xảy ra đồng thời với một trong các biến cố $H_1, H_2, ..., H_n$. Khi đó xác suất của biến cố A được tính bằng công thức sau đây:*

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

■.Vì các biến cố $H_1, H_2, ..., H_n$ tạo nên một nhóm đầy đủ các biến cố nên:

$$A = (H_1 + H_2 + ..H_n).A = H_1A + H_2A + ... + H_nA.$$

Vì các biến cố $H_1, H_2, ..., H_n$ xung khắc từng đôi, do đó các tích $H_1A, H_2A, ..., H_nA$ cũng xung khắc từng đôi, do đó:

$$P(A) = P(H_1A) + P(H_2A) + ... + P(H_nA).$$

Áp dụng định lý nhân xác suất với các tích H_iA ta có:

$$P(A) = P(H_1)P(A/H_1) + ... + P(H_n)P(A/H_n) = \sum_{i=1}^{n} P(H_i).P(A/H_i)$$

□

•**Ví dụ 1.41** 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. Tìm xác suất để lấy ngẫu nhiên một sản phẩm từ hộp thứ hai được chính phẩm.

Lời giải. Gọi A là biến cố "Lấy được chính phẩm từ hộp thứ hai". Biến cố A có thể xảy ra đồng thời với một trong các biến cố sau đây tạo nên một nhóm đầy đủ 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.

Xác suất để lấy được một chính phẩm từ hộp một là: $P(H_1) = \dfrac{9}{10}$

Xác suất để lấy được một phế phẩm từ hộp một là: $P(H_2) = \dfrac{1}{10}$

Nếu lấy một chính phẩm từ hộp một sang hộp hai thì xác suất lấy được một chính phẩm từ hộp hai là: $P(A/H_1) = \dfrac{19}{21}$.

Nếu lấy một phế phẩm từ hộp một sang hộp hai thì xác suất lấy được một chính phẩm từ hộp hai là: $P(A/H_2) = \dfrac{18}{21}$.

Do đó theo công thức đầy đủ ta có:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) = \frac{9}{10}.\frac{19}{21} + \frac{1}{10}\frac{18}{21} = \frac{9}{10}.$$

•Ví dụ 1.42 Dây chuyền lắp ráp nhận được các chi tiết do hai máy sản xuất. Trung bình máy thứ nhất cung cấp 60% chi tiết, máy thứ hai cung cấp 40% chi tiết. Khoảng 90% chi tiết do máy thứ nhất sản xuất là đạt tiêu chuẩn, còn 85% chi tiết do máy thứ hai sản xuất là đạt tiêu chuẩn. Tính xác suất để lấy ngẫu nhiên từ dây chuyền được một sản phẩm đạt tiêu chuẩn.

Lời giải. Gọi A là biến cố "Chi tiết lấy từ dây chuyền đạt tiêu chuẩn". Biến cố A có thể xảy ra đồng thời với một trong hai biến cố sau đây tạo nên một nhóm đầy đủ các biến cố:

H_1- Chi tiết do máy thứ nhất sản xuất;

H_2 - Chi tiết do máy thứ hai sản xuất.

Theo điều kiện bài toán:

$$\begin{array}{ll} P(H_1) = 0,6; & P(A/H_1) = 0,9 \\ P(H_2) = 0,4; & P(A/H_2) = 0,85 \end{array}$$

Do đó xác suất cần tìm là:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2) = 0,6.0,9 + 0,4.0,85 = 0,88.$$

1.6.2. Công thức Bayes

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

$$P(H_k/A) = \frac{P(H_k).P(A/H_k)}{\sum\limits_{i=1}^{n} P(H_i).P(A/H_i)}; \qquad (k = \overline{1,n})$$

■.Theo định lý nhân xác suất ta có: (với $k = \overline{1,n}$)

$$P(AH_k) = P(A).P(H_k/A) = P(H_k).P(A/H_k)$$

Từ đó:

$$P(H_k/A) = \frac{P(H_k).P(A/H_k)}{P(A)}$$

Thay P(A) bằng công thức xác suất đầy đủ ta có:

$$P(H_k/A) = \frac{P(H_k).P(A/H_k)}{\sum\limits_{i=1}^{n} P(H_i).P(A/H_i)}.$$

□

•Ví dụ 1.43 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 lấy ngẫu nhiên một sản phẩm từ hộp hai thấy đó là chính phẩm, tính xác suất để sản phẩm lấy từ hộp một bỏ sang hộp hai cũng là một chính phẩm.

Lời giải. Ký hiệu các biến cố tương tự ví dụ 1.41:

A - Lấy được chính phẩm từ hộp thứ hai;

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.

Ta có $H_1, H2$ lập thành nhóm đầy đủ các biến cố và:

$$P(H_1) = \frac{9}{10}; \qquad P(A/H_1) = \frac{19}{21}$$
$$P(H_2) = \frac{1}{10}; \qquad P(A/H_2) = \frac{18}{21}$$

Ta cần tính $P(H_1/A)$, theo công thức Bayes:

$$P(H_1/A) = \frac{P(H_1).P(A/H_1)}{P(H_1).P(A/H_1) + P(H_2).P(A/H_2)}$$

$$= \frac{\frac{9}{10}.\frac{19}{21}}{\frac{9}{10}.\frac{19}{21} + \frac{1}{10}.\frac{18}{21}} = \frac{57/70}{9/10} = \frac{19}{21}.$$

•Ví dụ 1.44 Dây chuyền lắp ráp nhận được các chi tiết do hai máy sản xuất. Trung bình máy thứ nhất cung cấp 60% chi tiết, máy thứ hai cung cấp 40% chi tiết. Khoảng 90% chi tiết do máy thứ nhất sản xuất là đạt tiêu chuẩn, còn 85% chi tiết do máy thứ hai sản xuất là đạt tiêu chuẩn. Lấy ngẫu nhiên từ dây chuyền một sản phẩm, thấy nó đạt tiêu chuẩn. Tìm xác suất để sản phẩm đó do máy thứ nhất sản xuất. Hỏi sản phẩm đó có khả năng do máy nào sản xuất ra lớn hơn?

Lời giải. Ký hiệu các biến cố tương tự ví dụ 1.42:

A - Chi tiết lấy từ dây chuyền đạt tiêu chuẩn;

H_1 - Chi tiết do máy thứ nhất sản xuất;

H_2 - Chi tiết do máy thứ hai sản xuất.

Ta có:

$$P(H_1) = 0,6; \qquad P(A/H_1) = 0,9;$$
$$P(H_2) = 0,4; \qquad P(A/H_2) = 0,85$$

Theo công thức xác suất đầy đủ:

$$P(A) = P(H_1).P(A/H_1) + P(H_2).P(A/H_2) = 0,6.0,9 + 0,4.0,85 = 0,88.$$

Xác suất để chính phẩm lấy ra do máy thứ nhất sản xuất là:

$$P(H_1/A) = \frac{P(H_1).P(A/H_1)}{P(A)} = \frac{0,6.0,9}{0,88} \approx 0,614.$$

Xác suất để chính phẩm lấy ra do máy thứ hai sản xuất là:

$$P(H_2/A) = \frac{P(H_2).P(A/H_2)}{P(A)} = \frac{0,4.0,85}{0,88} \approx 0,386.$$

Ta thấy $P(H_1/A) > P(H_2/A)$. Vậy sản phẩm đó có khả năng do máy thứ nhất sản xuất ra lớn hơn.

Bài tập chương 1

Định nghĩa cổ điển về xác suất.

Bài 1.1. Gieo một con súc sắc đối xứng và đồng chất. Tính xác suất để được:
a) Mặt sáu chấm xuất hiện.
b) Mặt có số chấm chẵn xuất hiện.
Đáp số: a) 0,167; b) 0,5.

Bài 1.2. Gieo đồng thời hai con súc sắc. Tìm xác suất để được hai mặt có:
a) Tổng số chấm bằng 7.
b) Tổng số chấm nhỏ hơn 8.
c) ít nhất một mặt 6 chấm.
d) Giả sử gieo n con súc sắc đối xứng và đồng chất, tính xác suất để được tổng số chấm là n + 1.
Đáp số: a) 0,167; b) 0,583; c) 0,306.

Bài 1.3. Gieo đồng thời hai đồng xu. Tìm xác suất để được:
a) Hai mặt cùng sấp xuất hiện.
b) Một sấp một ngửa xuất hiện.
c) Có ít nhất một mặt sấp.
Đáp số: a) 0,25; b) 0,5; c) 0,75.

Bài 1.4. Có 100 tấm bìa hình vuông như nhau được đánh số từ 1 đến 100. Lấy ngẫu nhiên một tấm bìa, tìm xác suất:
a) Được tấm bìa không có chữ số 5.
b) Được một tấm bìa có số chia hết cho 2 hoặc cho 5.
Đáp số: a) 0,81; b) 0,6.

Bài 1.5. Một người gọi điện thoại cho bạn nhưng lại quên mất 3 chữ số cuối. Tìm xác suất để

người đó quay số một lần được đúng số điện thoại của bạn nếu:
a) Người đó nhớ rằng 3 chữ số cuối khác nhau.
b) Người đó không nhớ gì thêm.
Đáp số: a) 0,001389; b) 0,001.
Bài 1.6. Mỗi vé xổ số có 5 chữ số. Tìm xác suất để một người mua 1 vé được vé:
a) Có 5 chữ số khác nhau.
b) Có 5 chữ số đều là lẻ.
Đáp số: a) 0,3024; b) 0,03125.
Bài 1.7. Một nhi đồng tập xếp chữ, em có các chữ N, Ê, H, G, H, N. Tìm xác suất để em đó trong khi sắp ngẫu nhiên được chữ NGHÊNH.
Đáp số: 0,0056.
Bài 1.8. Có 30 tấm bìa giống nhau được đánh số lần lượt từ 1 đến 30, trong đó có 15 tấm bìa có chữ ″SU", 11 tấm bìa có chữ ″ZU", 3 tấm bìa có chữ ″KI″, 1 tấm bìa có chữ ″SUZUKI". Người chơi sẽ thắng cuộc nếu ghép được (hoặc chọn được) chữ SUZUKI. Tính xác suất người chơi thắng cuộc nếu người đó được chọn:
a) 2 tấm bìa
b) 3 tấm bìa
c) 4 tấm bìa.
Đáp số: a) 0,067; b) 0,222; c) 0,368.
Bài 1.9. Trên giá sách có xếp ngẫu nhiên một tuyển tập của tác giả X gồm 12 cuốn. Tìm xác suất để các tập được xếp theo thứ tự hoặc từ trái sang phải hoặc từ phải sang trái.
Đáp số: 4,2.10-9.
Bài 1.10. Một lô hàng gồm 6 chính phẩm và 4 phế phẩm được chia ngẫu nhiên thành 2 phần bằng nhau. Tìm xác suất để mỗi phần đều có số chính phẩm như nhau.
Đáp số: 0,476.
Bài 1.11. Trong một hòm đựng 10 chi tiết đạt tiêu chuẩn và 5 chi tiết là phế phẩm, lấy ngẫu nhiên đồng thời 3 chi tiết. Tính xác suất: a) Cả 3 chi tiết lấy ra thuộc loại đạt tiêu chuẩn.
b) Trong số 3 chi tiết lấy ra có 2 chi tiết đạt tiêu chuẩn.
Đáp số: a) 0,264; b) 0,495.
Bài 1.12. Năm người A, B, C, D, E ngồi một cách ngẫu nhiên vào một chiếc ghế dài. Tìm xác suất để: a) C ngồi chính giữa.
b) A và B ngồi ở hai đầu ghế.
c) A và B ngồi cạnh nhau.
Đáp số: a) 0,2; b) 0,1; c) 0,4.
Bài 1.13. Có n người trong đó có m người trùng tên xếp một cách ngẫu nhiên thành một hàng ngang. Tìm xác suất để m người trùng tên đứng cạnh nhau.
Đáp số: $\frac{(n^{-}rn+l)!m!}{n!}$
Bài 1.14. Thang máy của một toà nhà 7 tầng xuất phát từ tầng một với 3 khách. Tìm xác suất để:

a) Tất cả cùng ra ở tầng bốn.

b) Tất cả cùng ra ở một tầng.

c) Mỗi người ra ở một tầng khác nhau.

Đáp số: a) 0,0046; b) 0,0278; c) 0,555.

Bài 1.15. Ba nữ nhân viên phục vụ A, B, c thay nhau rửa đĩa chén và giả thiết ba người này đều "khéo léo" như nhau. Trong một tháng có 4 chén bị vỡ. Tìm xác suất:

a) Chị A đánh vỡ 3 chén và chị B đánh vỡ 1 chén.

b) Một trong 3 người đánh vỡ 3 chén.

c) Một trong 3 người đánh vỡ cả 4 chén.

Đáp số: a) 0,049; b) 0,296; c) 0,037.

Bài 1.16. Có 10 khách bước ngẫu nhiên vào một cửa hàng có 3 quầy. Tìm xác suất để có

a) 3 người đến quầy số 1.

b) một quầy có 3 người và một quầy khác có 5 người.

Đáp số: a) 0,260; b) 0,256.

Định nghĩa hình học về xác suất

Bài 1.17. Một người chơi phi tiêu cắm ngẫu nhiên phi tiêu vào một điểm trong bia là hình vuông ABCD tâm O có cạnh 20cm. Tính xác suất để tiêu rơi vào một điểm:

a) nằm trong hình tròn tâm oO bán kính 5cm

b) nằm trên đường chéo AC

c) không rơi vào tâm O.

Đáp số: a) 0,196; b) 0; c) 1.

Bài 1.18. Cho một đoạn thẳng và bẻ gãy ngẫu nhiên thành 3 đoạn. Tính xác suất để 3 đoạn đó tạo thành được một tam giác.

Đáp số: 0,25.

Định nghĩa thống kê về xác suất

Bài 1.19. 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.

Đáp số: 0,517.

Bài 1.20. Tần suất xuất hiện biến cố viên đạn trúng đích là 0,85. Tìm số viên đạn bắn trúng đích của xạ thủ đó nếu người này bắn 200 viên đạn.

Đáp số: 170.

Mối quan hệ giữa các biến cố

Bài 1.21. Ba người cùng bắn vào một mục tiêu. Gọi A_k là biến cố người thứ k bắn trúng mục tiêu $(k = \overline{1,3})$. Hãy viết bằng ký hiệu các biến cố biểu thị rằng:

a) Chỉ có người thứ nhất bắn trúng mục tiêu.

b) Chỉ có một người bắn trúng mục tiêu.

c) Chỉ có 2 người bắn trúng mục tiêu.

d) Có người bắn trúng mục tiêu.

Bài 1.22. Kiểm tra theo thứ tự một lô hàng có 10 sản phẩm. Các sản phẩm đều thuộc một trong

hai loại: Tốt hoặc Xấu. Ký hiệu $\overline{A}_k(k = \overline{1,10})$ là biến cố chỉ sản phẩm kiểm tra thứ k thuộc loại Xấu. Viết bằng ký hiệu các biến cố sau:
a) Cả 10 sản phẩm đều Xấu.
b) Có ít nhất một sản phẩm Xấu.
c) Cả 6 sản phẩm kiểm tra đầu là Tốt, các sản phẩm còn lại là Xấu.
d) Các sản phẩm kiểm tra theo thứ tự chẵn là tốt, còn các sản phẩm kiểm tra theo thứ tự lẻ là xấu.

Định lý cộng và nhân xác suất

Bài 1.23. Để được nhập kho, sản phẩm của 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, tính xác suất phế phẩm được nhập kho. **Đáp số:** 0,0002.

Bài 1.24. 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.
Đáp số: 0,432.

Bài 1.25. Một hộp chứa 3 bi trắng, 7 bi đỏ và 15 bi xanh. Một hộp khác chứa 10 bi trắng, 6 bi đỏ và 9 bi xanh. Lấy ngẫu nhiên từ mỗi hộp một bi. Tìm xác suất để hai bi lấy ra là cùng màu.
Đáp số: 0,3312.

Bài 1.26. Hai người cùng bắn vào một mục tiêu. Khả năng bắn trúng của từng người là 0,8 và 0,9. Tìm xác suất: a) Chỉ có một người bắn trúng mục tiêu.
b) Có người bắn trúng mục tiêu.
c) Cả hai người bắn trượt.
Đáp số: a) 0,26; b) 0,98; c) 0,02.

Bài 1.27. Nhà máy có 3 ôtô hoạt động độc lập, xác suất để mỗi ôtô bị hỏng trong ngày làm việc lần lượt là 0,1; 0,2; 0,3. a) Tính xác suất trong ngày có chỉ ôtô thứ nhất hỏng.
b) Tính xác suất trong ngày có đúng 1 ôtô hỏng.
c) Tính xác suất trong ngày có ôtô hỏng.
d) Biết trong ngày làm việc có hai ôtô hỏng, tính xác suất 2 ôtô hỏng đó là ôtô thứ nhất và thứ hai.
Đáp số: a) 0,056; b) 0,398; c) 0,496; d) 0,152.

Bài 1.28. Trong hộp có n quả bóng bàn mới. Người ta lấy ra k quả để chơi $(k \leq \dfrac{n}{2})$ sau đó lại bỏ vào hộp. Tìm xác suất để lần sau lấy ra k quả để chơi thì lấy được toàn bóng mới.
Đáp số: $\dfrac{C_{n-k}^k}{C_n^k}$

Bài 1.29. Tín hiệu thông tin được phát 3 lần với xác suất thu được của mỗi lần là 0,4.
a) Tìm xác suất nguồn thu nhận được thông tin đó.
b) Nếu muốn xác suất thu được thông tin lên đến 0,9 thì phải phát bao nhiêu lần?
Đáp số: a) 0,784; b) 5 lần.

Bài 1.30. Hai người cùng bắn vào một mục tiêu, khả năng chỉ có một người bắn trúng là 0,38. Tìm xác suất bắn trúng của người thứ nhất, biết rằng khả năng bắn trúng của người thứ hai là

0,8Ể

Đáp số: 0,7.

Bài 1.31. rong 10 sản phẩm có 2 phế phẩm, lấy ngẫu nhiên 2 sản phẩm. Tính xác suất để cả 2 sản phẩm đều là phế phẩm trong trường hợp:

a) Lấy hoàn lại.

b) Lấy không hoàn lại.

Đáp số: a) 0,04; b) 0,022.

Bài 1.32. Xác suất để bắn một viên đạn trúng đích là 0,8. Hỏi phải bắn bao nhiêu viên đạn để với xác suất nhỏ hơn 0,4 có thể hy vọng rằng không có viên nào trượt?

Đáp số: 5 lần.

Bài 1.33. Phải gieo một con súc sắc trong bao nhiêu lần để với xác suất lớn hơn 0,5 có thể hy vọng có ít nhất một lần được mặt 6 chấm?

Đáp số: 4 lần.

Bài 1.34. Một nồi hơi được lắp 2 van bảo hiểm với xác suất hỏng của các van tương ứng là 0,1 và 0,2. Nồi hơi sẽ hoạt động an toàn khi có van không hỏngệ Tìm xác suất để nồi hơi hoạt động:

a) An toàn;

b) Mất an toàn.

Đáp số: a) 0,98; b) 0,02.

Bài 1.35. Bắn liên tiếp vào một mục tiêu cho đến khi viên đạn đầu tiên trúng mục tiêu thì dừng. Tìm xác suất sao cho phải bắn đến viên thứ 6, biết rằng xác suất trúng mục tiêu của mỗi viên đạn là 0,2 và các lần bắn độc lập với nhau.

Đáp số: 0,066.

Bài 1.36. Một thủ kho có chùm chìa khoá gồm 9 chiếc trong đó chỉ có 1 chiếc mở được cửa kho. Anh ta thử ngẫu nhiên từng chìa khoá một, chiếc nào được thử thì không thử lại. Tính xác suất anh ta mở được cửa ở lần thứ 4.

Đáp số: 0,111.

Bài 1.37. Gieo một con súc sắc đối xứng và đồng chất. Gọi A là biến cố được mặt có số chẵn chấm, B là biến cố được mặt có số chấm là bội số của 3.

a) Xét xem hai biến cố này có xung khắc không?

b) Xét xem hai biến cố này có độc lập không?

Đáp số: a) không xung khắc; b) độc lập.

Bài 1.38. Gieo hai con súc sắc đối xứng và đồng chất. Gọi A là biến cố xuất hiện khi tổng số chấm thu được là lẻ; B là biến cố được ít nhất một mặt một chấm. Tính $P(AB), P(A+B), P(\overline{AB})$.

Đáp số: 0,167; 0,639; 0,833.

Bài 1.39. Có hai bóng đèn điện với xác suất hỏng tương ứng là 0,1 và 0,2 và việc chúng hỏng là độc lập với nhau. Tính xác suất để mạch không có điện do bóng hỏng nếu chúng mắc:

a) Nối tiếp.

b) Song song.

Đáp số: a) 0,28; b) 0,02.

Bài 1.40. Có hai lô hàng: Lô 1 có 90 chính phẩm và 10 phế phẩm;
Lô 2 có 80 chính phẩm và 20 phế phẩm;
Lấy ngẫu nhiên từ mỗi lô hàng một sản phẩm, tính xác suất để:
a) Lấy được một chính phẩm.
b) Lấy được ít nhất một chính phẩm.
Đáp số: a) 0,26; b) 0,98.
Bài 1.41. Một lô hàng có 9 sản phẩm. Mỗi lần kiểm tra chất lượng lấy ngẫu nhiên 3 sản phẩm, sau khi kiểm tra xong trả lại vào lô hàng. Tính xác suất để sau 3 lần kiểm tra lô hàng, tất cả các sản phẩm đều được kiểm tra.
Đáp số: 0,0028.
Bài 1.42. Sau khi ra khỏi phòng khách, mỗi người trong N người cùng số giày xỏ ngẫu nhiên vào một đôi giày trong bóng tối. Mỗi người chỉ có thể phân biệt chiếc giày phải với chiếc giày trái, còn không thể phân biệt giày của mình với giày của người khác. Tìm xác suất để:
a) Mỗi người khách đi đúng đôi giày của mình.
b) Mỗi người khách đi đúng 2 chiếc giày của cùng một đôi (không nhất thiết là đôi giày của người đó).
Đáp số: a) $\frac{1}{(N!)^2}$, b) $\frac{1}{N!}$
Bài 1.43. Hai cầu thủ bóng rổ, mỗi người ném bóng 2 lần, xác suất ném trúng đích của mỗi cầu thủ theo thứ tự là 0,6 và 0,7. Tính xác suất:
a) Số lần ném trúng rổ của cầu thủ thứ nhất nhiều hơn số lần ném trúng rổ của cầu thủ thứ hai.
b) Số lần ném trúng rổ của hai người như nhau.
c) Biết số lần ném trúng rổ của hai người như nhau, tính xác suất để mỗi người ném trúng đúng 1 quả.
Đáp số: a) 0,2268; b) 0,3924; c) 0,513.
Bài 1.44. Hai người chơi cờ thoả thuận với nhau là ai thắng trước 3 ván thì sẽ thắng cuộc. Trận đấu bị gián đoạn khi người thứ nhất còn thiếu 1 ván thắng, người thứ hai còn thiếu 2 ván thắng. Vậy phải phân chia tiền đặt như thế nào là hợp lý nếu xác suất thắng mỗi ván của mỗi người đều bằng 0,5.
Đáp số: 0,75 : 0,25.
Công thức đầy đủ và công thức Bayes.
Bài 1.45. Hai máy cùng sản xuất một loại sản phẩm. Tỷ lệ phế phẩm của máy I là 3%, của máy II là 2%. Từ một kho gồm $\frac{2}{3}$ sản phẩm của máy I và $\frac{1}{3}$ sản phẩm của máy II ta lấy ra một số sản phẩm. Tính xác suất để sản phẩm đó là tốt.
Đáp số: 0,973.
Bài 1.46. Có hai xạ thủ loại I và tám xạ thủ loại II, xác suất bắn trúng đích của các loại xạ thủ theo thứ tự là 0,9 và 0,8.
a) Lấy ngẫu nhiên ra một xạ thủ và xạ thủ đó bắn một viên đạn. Tìm xác suất viên đạn đó trúng đích.
b) Nếu lấy ra hai xạ thủ và mỗi người bắn một viên thì khả năng cả hai viên đều trúng đích là

bao nhiêu?

Đáp số: a) 0,82; b) 0,6722.

Bài 1.47. Có hai lô sản phẩm. Lô 1 gồm toàn chính phẩm, lô 2 có tỷ lẽ phế phẩm và chính phẩm là $\frac{1}{4}$. Chọn ngẫu nhiên một lô, trong lô này lấy ngẫu nhiên 1 sản phẩm, thấy nó là chính phẩm, rồi hoàn lại sản phẩm này vào lô. Hỏi rằng nếu lấy ngẫu nhiên (cũng từ lô đã chọn) một sản phẩm thì xác suất để sản phẩm này là phế phẩm bằng bao nhiêu?

Đáp số: 0,089.

Bài 1.48. Bắn ba phát đạn vào một máy bay với xác suất trúng tương ứng là 0,4; 0,5 và 0,7. Nếu trúng một phát thì xác suất rớt máy bay là 0,2; nếu trúng 2 phát thì xác suất rớt máy bay là 0,6; còn nếu trúng cả 3 phát thì chắc chắn máy bay rơi. a) Tính xác suất máy bay rơi.

b) Giả sử máy bay đã bị bắn rơi, tính xác suất để nó trúng một phát.

Đáp số: a) 0,458; b) 0,1572.

Bài 1.49. Người ta dùng X quang để chẩn đoán bệnh A, tuy nhiên vẫn có sai sót. Nếu sau khi X quang kết luận là có bệnh thì chỉ đúng 99%, còn kết luận là không bị bệnh thì chỉ đúng được 80%. Biết tỷ lệ mắc bệnh A trên thực tế là 20,79%.

a) Tìm xác suất để một người khám bị chẩn đoán có bệnh.

b) Tìm tỷ lệ bệnh nhân không có bệnh bị chẩn đoán nhầm.

Đáp số: a) 0,01; b) 0,000126.

Bài 1.50. Có 2 lô hàng, lô 1 gồm 15 chính phẩm và 4 phế phẩm, lô 2 gồm 18 chính phẩm và 5 phế phẩmẽ Lấy ra 2 sản phẩm ở lô 1 và 3 sản phẩm ở lô 2. Sau đó chọn ngẫu nhiên 1 sản phẩm từ 5 sản phẩm ấy.

a) Tính xác suất để sản phẩm lấy ra sau cùng là chính phẩm.

b) Biết rằng sản phẩm lấy ra sau là chính phẩm. Hỏi sản phẩm ấy có khả năng thuộc lô nào nhiều hơn?

Đáp số: a) 0,785; b) lô 2.

Bài 1.51. Có hai lô hàng, lô 1 có 7 chính phẩm và 3 phế phẩm; lô 2 có 8 chính phẩm và 2 phế phẩm. Từ lô thứ nhất lấy ra hai sản phẩm, từ lô thứ hai lấy ra 3 sản phẩm; rồi trong số sản phẩm được lấy ra lại lấy tiếp ngẫu nhiên 2 sản phẩm. Tính xác suất để trong hai sản phẩm đó có ít nhất một chính phẩm.

Đáp số: 0,951.

Bài 1.52. Có hai lô sản phẩm, lô 1 có 16 chính phẩm và 4 phế phẩm; lô 2 có 15 chính phẩm và 5 phế phẩm.

a) Từ lô thứ nhất bỏ sang lô thứ hai một sản phẩm, sau đó từ lô thứ hai lấy ra một sản phẩm. Tìm xác suất để lấy được chính phẩm.

b) Từ lô thứ nhất bỏ sang lô thứ hai một sản phẩm, sau đó từ lô thứ hai bỏ sang lô thứ nhất một sản phẩm, sau đó từ lô thứ nhất lấy ra một sản phẩm. Tìm xác suất để lấy được chính phẩm.

Đáp số: a) 0ế,752; b) 0,798.

Bài 1.53. Xí nghiệp A sản xuất một loại sản phẩm với xác suất hỏng của mỗi sản phẩm bằng 0,05. Ở phân xưởng sản phẩm có thể được một trong ba nhân viên kiểm tra chất lượng với xác

suất như nhau. Xác suất phát hiện sản phẩm hỏng của ba người đó lần lượt là 0,9; 0,85; 0,8. Nếu sản phẩm không bị loại ở phân xưởng thì được chuyển đến KCS của nhà máy và ở đó, sản phẩm hỏng sẽ được phát hiện với xác suất 0,95. Tìm xác suất để sản phẩm bị loại.
Đáp số: 0,049625.
Bài 1.54. Tỷ lệ người dân nghiện thuốc lá ở một vùng là 30%. Biết rằng tỷ lệ người bị viêm họng trong số người nghiện thuốc lá là 60%, còn tỷ lệ người bị viêm họng trong số người không hút thuốc lá là 40%.
a) Lấy ngẫu nhiên một người, biết rằng người đó viêm họng, tính xác suất người đó nghiện thuốc.
b) Nếu người đó không bị viêm họng, tính xác suất người đó nghiện thuốc.
Đáp số: a) 0,391; b) 0,222.
Bài 1.55. Một cỗ máy có ba bộ phận 1, 2, 3. Xác suất hỏng của các bộ phận trong thời gian làm việc theo thứ tự là 0,2; 0,4 và 0,3. Cuối ngày làm việc được biết rằng có hai bộ phận bị hỏng, tính xác suất để hai bộ phận hỏng đó là bộ phận 1 và 2.
Đáp số: 0,298.
Bài 1.56. Trong một bệnh viện, tỷ lệ bệnh nhân các tỉnh như sau: tỉnh A-25%; tỉnh B - 35%; tỉnh C - 40%. Biết rằng tỷ lệ bệnh nhân là kỹ sư của các tỉnh là: tỉnh A - 2%; tỉnh B - 3%; tỉnh C - 3,5%.
a) Chọn ngẫu nhiên một bệnh nhân, tính xác suất để bệnh nhân đó là kỹ sư.
b) Chọn ngẫu nhiên 3 bệnh nhân, tính xác suất để trong đó có đúng 1 bệnh nhân là kỹ sư.
Đáp số: a) 0,0295; b) 0,0833.
Bài 1.57. Một người có ba chỗ ưa thích như nhau để câu cá. Xác suất để câu được cá ở những chỗ đó tương ứng là 0,6; 0,7 và 0,8. Biết rằng ở một chỗ người đó thả câu 3 lần và chỉ câu được một con cá. Tìm xác suất để cá được câu ở chỗ thứ nhất.
Đáp số: 0,503.
Bài 1.58. Máy bay đến oanh kích mục tiêu phải qua 3 tuyến phòng thủ, xác suất để mỗi tuyến phòng thủ tiêu diệt được máy bay đều bằng 0,8.
a) Tìm xác suất để máy bay rơi trước khi đến được mục tiêu.
b) Nếu muốn bảo vệ mục tiêu với xác suất không nhỏ hơn 99,99% thì phải tổ chức bao nhiêu tuyến phòng thủ?
c) Giả sử máy bay bị rơi, tìm xác suất để tuyến 1 bắn rơi.
Đáp số: a) 0,992; b) 6 tuyến; c) 0,806.
Công thức Becnoulli
Bài 1.59. Môt cầu thủ nổi tiếng về đá phạt đền, xác suất đá vào gôn là $\frac{4}{5}$. Có người cho rằng cứ "sút" 5 quả thì chắc chắn có 4 quả vào lưới, điều khẳng định đó có đúng không? Tìm xác suất để trong 5 lần sút có đúng 4 lần bóng vào lưới.
Đáp số: không đúng; 0,4096.
Bài 1.60. Một nữ công nhân quản lý 12 máy dột, 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:

a) có máy cần đến sự chăm sóc của nữ công nhân.

b) có không nhiều hơn 2 máy cần đến sự chăm sóc của nữ công nhân.

c) 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.

Đáp số: a) 0,992; b) 0,181; c) 4 máy; 0,238.

Bài 1.61. Xác suất để đường kính của một chi tiết bất kỳ rút từ một lô chi tiết là bé hơn, lớn hơn và nằm trong phạm vi cho phép tương ứng là 0,05; 0,10 và 0,85. Từ lô chi tiết đã cho rút ngẫu nhiên 100 chiếc. Tính xác suất để trong số đó có đúng 5 chi tiết mà đường kính bé hơn và 5 chi tiết mà đường kính lớn hơn phạm vi cho phép.

Đáp số: 0,00606.

Bài tập tổng hợp.

Bài 1.62. Hộp 1 có: 8 bi trắng, 10 bi đỏ. Hộp 2 có: 4 bi trắng, 16 bi đỏ.

a) Từ hộp 1 lấy ra 2 bi cho vào hộp 2 trộn đều sau đó từ hộp 2 lấy ra 2 bi. Tính xác suất để 2 bi lấy ra sau là bi đỏ.

b) Từ mỗi hộp lấy ra một bi sau đó trong 2 bi thu được lấy ngẫu nhiên ra một bi và thấy đó là bi trắng. Tính xác suất để viên bi ấy là của hộp l.

Đáp số: a) 0,598; b) 0,690.

Bài 1.63. Có 2 hộp sản phẩm. Hộp thứ nhất chứa 8 sản phẩm tốt, 2 sản phẩm xấu. Hộp thứ 2 chứa 7 sản phẩm tốt, 5 sản phẩm xấu. Người ta chuyển 1 sản phẩm từ hộp thứ nhất sang hộp thứ 2, sau đó lấy từ hộp thứ 2 ra 2 sản phẩm.

a) Tính xác suất để 2 sản phẩm lấy từ hộp 2 chỉ có 1 sản phẩm tốt.

b) Tính xác suất trong 2 sản phẩm lấy từ hộp 2 có sản phẩm tốt.

Đáp số: a) 0,518; b) 0,859.

Bài 1.64. Sinh viên phải chọn học ít nhất 1 trong 3 môn tự chọn: Toán, Lý, Hoá. Biết có 60% sinh viên học Toán, 40% học Lý, 50% học Hoá, 28% học Toán và Lý, 21% học Lý và Hoá, 20% học Toán và Hoá.

a) Tính tỷ lệ sinh viên học cả 3 môn Toán, Lý, Hoá.

b) Chọn ngẫu nhiên 1 sinh viên trong lớp học Toán, tính xác suất để sinh viên đó học Hoá.

c) Chọn ngẫu nhiên 3 sinh viên, tính xác suất để trong đó có đúng 1 sinh viên chỉ chọn học duy nhất 1 môn Toán.

Đáp số: a) 0,19; b) 0,333; c) 0,441.

Bài 1.65. Trong một kho rượu số rượu loại A và rượu loại B bằng nhau. Người ta chọn ngẫu nhiên một chai rượu trong kho và đưa cho 5 người sành rượu nếm thử, giả sử mỗi người có xác suất đoán trúng là 75%. Có 4 người kết luận chai rượu loại A và 1 người kết luận chai rượu loại B. Tính xác suất để đó là chai rượu loại A.

Đáp số: 0,964.

Bài 1.66. Bia bắn được chia làm 2 vòng, xác suất bắn trúng vòng trong là 0,7 còn trúng vòng ngoài là 0,3. Tìm xác suất sao cho bắn 3 viên đạn thì được ít ra là 29 điểm, biết rằng nếu bắn trúng vòng trong thì được 10 điểm, trúng vòng ngoài thì được 9 điểm.

ĐSễ’ 0,784.

Bài 1.67. Một người bắn 3 viên đạn. Xác suất để cả 3 viên trúng vòng 10 là 0,008, xác suất để 1 viên trúng vòng 8 là 0,15 và xác suất để 1 viên trúng vòng dưới 8 là 0,4. Tính xác suất để xạ thủ đạt ít nhất 28 điểm.

Đáp số: 0,0935.

Bài 1.68. Một nhân viên bán hàng mỗi năm đến bán ở công ty A ba lần. Xác suất lần đầu bán được hàng là 0,8. Nếu lần trước bán được hàng thì xác suất để lần sau bán được hàng là 0,9; còn nếu lần trước không bán được hàng thì xác suất để lần sau bán được hàng chỉ là 0,4. Tìm xác suất để:

a) Cả ba lần đều bán được hàng;

b) Có đúng hai lần bán được hàng.

Đáp số: a) 0,648; b) 0,176.

Bài 1.69. Một người bắn 3 viên đạn với xác suất trúng đích của mỗi viên là 0,6. Do yếu tố tâm lý nên: nếu viên thứ nhất trúng thì xác suất trúng của mỗi viên sau là 0,7; còn nếu cả hai viên đầu trúng thì xác suất trúng của viên thứ ba là 0,8; ngoài ra nếu viên thứ nhất trượt thì xác suất để cả hai viên sau trúng là 0,3. Tính xác suất:

a) Cả 3 viên cùng trúng;

b) Viên thứ hai và viên thứ ba trúng;

c) Có ít nhất 1 viên trúng.

Chương 2

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

2.1. Định nghĩa và phân loại đại lượng ngẫu nhiên

2.1.1. Định nghĩa

$\star$ **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, Z hoặc $X_1, X_2, ..., X_n; Y_1, ..., Y_m$ còn các giá trị có thể có của đại lượng ngẫu nhiên được kí hiệu là $x, x_1, x_2, ..., x_n, y, y_1, ..., y_m$.

Đại lượng X được gọi là ngẫu nhiên vì trước khi tiến hành phép thử ta chưa thể nói chắc chắn nó nhận giá trị bằng bao nhiêu mà chỉ có thể dự đoán với xác suất nhất định. Việc X nhận một giá trị nào đó $(X = x_1)$ hoặc $(X = x_2), ...(X = x_n), (a < X < b)$ là các biến cố ngẫu nhiên.

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

$\bullet$**Ví dụ 2.2** Bắn ngẫu nhiên một viên đạn vào bia. Gọi X là khoảng cách từ tâm bia đến điểm chạm vủa đạn vào bia. X cũng là một đại lượng 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 lập nên một tập 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.

Như vậy, đại lượng ngẫu nhiên sẽ là rời rạc nếu ta có thể liệt kê được tất cả các giá trị có thể có của nó. Còn đối với đại lượng ngẫu nhiên liên tục ta không thể liệt kê được tất cả các giá trị có thể có của nó.

•**Ví dụ 2.3** Một phân xưởng có 5 máy hoạt động. Gọi X là "Số máy hỏng trong một ca". X là đại lượng ngẫu nhiên rời rạc với các giá trị có thể có của nó là $0, 1, 2, 3, 4, 5$.

•**Ví dụ 2.4** 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 lập nên một tập đếm được: $0, 1, 2, 3, 4, 5, ...$

•**Ví dụ 2.5** Phép thử là bắn ngẫu nhiên một phát súng vào bia hình tròn bán kính 40 cm. 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.

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

⋆ **Định nghĩa 2.2** Bất kì một hình thức nào đó cho phép biểu diễn mối quan hệ giữa các giá trị có thể có của đại lượng ngẫu nhiên và các xác suất tương ứng đều được gọi là các quy luật phân phối xác suất của đại lượng ngẫu nhiên ấy.

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.

2.2.1. Bảng phân phối xác suất

Bảng phân phối xác suất chỉ dùng để thiết lập quy luật phân phối xác suất của đại lượng ngẫu nhiên rời rạc.

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

X	x_1	x_2	...	x_i	...	x_n
P	p_1	p_2	...	p_i	...	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:

$$\begin{cases} 0 \le p_i \le 1 \quad \text{với } \forall i = \overline{1,n} \\ \sum\limits_{i=1}^{n} p_i = 1 \end{cases}$$

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; $x_1 < x_2 < ... < x_n$.

•**Ví dụ 2.6** Tung một con súc sắc. Gọi X là "Số chấm xuất hiện". Hãy xây dựng quy luật phân phối xác suất của X.

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

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

Nhận xét thấy $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = 1.$

•**Ví dụ 2.7** Trong hộp có 10 sản phẩm trong đó có 6 chính phẩm. Lấy ngẫu nhiên 2 sản phẩm. Lập bảng phân phối xác suất của số chính phẩm được lấy ra.

Lời giải. Gọi X là số chính phẩm được lấy ra trong 2 sản phẩm. X là đại lượng ngẫu nhiên rời rạc có thể nhận các giá trị có thể có là $0, 1, 2$. Xác suất để cả hai sản phẩm lấy ra đều là phế phẩm (không có chính phẩm) là:

$$P(X = 0) = \frac{C_4^2}{C_{10}^2} = \frac{2}{15}$$

Xác suất để hai sản phẩm lấy ra có đúng 1 chính phẩm là:

$$P(X = 1) = \frac{C_4^1.C_4^1}{C_{10}^2} = \frac{8}{15}$$

Xác suất để cả hai sản phẩm lấy ra đều là chính phẩm là:

$$P(X = 2) = \frac{C_6^2}{C_{10}^2} = \frac{1}{3}$$

Vì vậy bảng phân phối xác suất của đại lượng ngẫu nhiên rời rạc X có dạng:

Nhận thấy: $\frac{2}{15} + \frac{8}{15} + \frac{1}{3} = 1.$

•**Ví dụ 2.8** Xác suất để xạ thủ bắn trúng bia là $0, 8$. Xạ thủ được phát từng viên đạn để bắn đến khi trúng bia. Xây dựng quy luật phân phối xác suất của số viên đạn được phát.

X	0	1	2
P	$\frac{2}{15}$	$\frac{8}{15}$	$\frac{1}{3}$

Lời giải. Gọi X là "Số viên đạn xạ thủ được phát". X là đại lượng ngẫu nhiên rời rạc có thể nhận các giá trị có thể có là $1, 2, ..., k, ...$.
Xác suất để được phát một viên đạn là xác suất để bắn viên đạn đầu tiên trúng bia, do đó:

$$P(X = 1) = 0,8$$

Xác suất để được phát hai viên đạn là xác suất để xảy ra đồng thời hai biến cố: phát thứ nhất bắn trượt và phát thứ hai bắn trúng. Theo định lý nhân xác suất ta có:

$$P(X = 2) = 0,2.0,8$$

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

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

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

X	1	2	3	...	k	...
P	0,8	0,2.0,8	$(0,2)^2.0,8$	...	$(0,2)^{k-1}.0,8$	...

Tổng các xác suất là tổng cấp số nhân lùi vô hạn:

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

2.2.2. Hàm phân phối xác suất

a) Định nghĩa

⋆ **Định nghĩa 2.3** Hàm phân phối xác suất của đại lượng ngẫu nhiên X, ký hiệu $F(x)$, là xác suất để đại lượng ngẫu nhiên X nhận giá trị nhỏ hơn x, với x là số thực bất kỳ.

$$F(x) = P(X < x)$$

Hàm phân phối xác suất dùng để thiết lập quy luật phân phối xác suất cho cả đại lượng ngẫu nhiên rời rạc lẫn đại lượng ngẫu nhiên liên tục.

•**Ví dụ 2.9** Đại lượng ngẫu nhiên rời rạc X có bảng phân phối xác suất như 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

$$F(x) = P(X = 1) = 0,1.$$

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

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

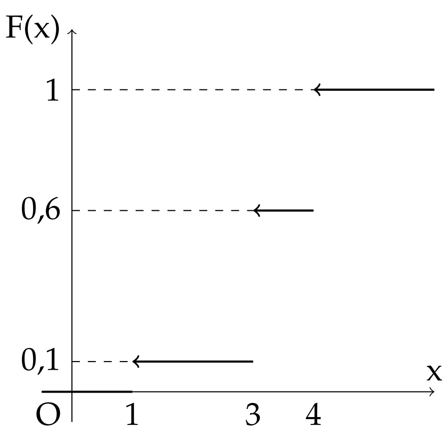
Nếu $x > 4$, biến cố $(X < x)$ sẽ xảy ra khi $(X = 1)$, hoặc khi $(X = 3)$, hoặc khi $(X = 4)$ do đó

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

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 $F(x)$ có dạng trong hình 2.1. **b) Tính chất của hàm phân phối xác suất.**



Hình 2.1:

- Tính chất 1. Hàm phân phối xác suất luôn nhận giá trị trong khoảng $[0;1]$:

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

Tính chất này suy ra trực tiếp từ định nghĩa hàm phân phối: $F(x) = P(X < x) \in [0;1]$.

- Tính chất 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).

- Tính chất 3. Hàm phân phối xác suất là hàm không giảm, tức là với $x_1 < x_2$ thì:

$$F(x_1) \leq F(x_2)$$

■.Giả sử $x_1 < x_2$, xét biến cố $(X < x_2)$. Biến cố này có thể phân tích thành hai biến cố xung khắc là $(X < x_1)$ và $(x_1 \leq X < x_2)$. Theo định lý cộng xác suất ta có:

$$P(X < x_2) = P(X < x_1) + P(x_1 \leq X < x_2)$$

Do đó:

$$F(x_2) - F(x_1) = P(X < x_2) - P(X < x_1) = P(x_1 \leq X < x_2)$$

Vì $P(x_1 \leq X < x_2)$ là một xác suất nên không âm, vì vậy ta có:

$$F(x_2) - F(x_1) \geq 0 \Leftrightarrow F(x_2) \geq F(x_1)$$

□

- Tính chất 4. Xác suất để đại lượng ngẫu nhiên liên tục X nhận một giá trị xác định bằng 0.

$$P(X = x) = 0$$

- Tính chất 5. 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:

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

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

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

- Tính chất 6. Nếu đại lượng ngẫu nhiên X chỉ nhận giá trị trong khoảng $[a;b]$ thì:

$$F(x) = 0 \text{ với } \forall x \leq a$$

$$F(x) = 1 \text{ với } \forall x > b$$

Thật vậy, với $x \leq a$, biến cố $(X < x)$ là biến cố không thể có, do đó xác suất của nó bằng 0. Còn với $x > b$, biến cố $(X < x)$ là biến cố chắc chắn, do đó xác suất của nó bằng 1.

Từ tính chất trên ta có biểu thức giới hạn sau:

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

$*$ **Chú ý:** Điều kiện để hàm số $F(x)$ là hàm phân phối xác suất của đại lượng ngẫu nhiên liên tục:

$$\begin{cases} F(x) \text{ là hàm liên tục} \\ F(x) \text{ là hàm tăng} \\ \lim\limits_{x \to -\infty} F(x) = 0; \lim\limits_{x \to +\infty} F(x) = 1. \end{cases}$$

c) 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 hàm phân phối xác suất tại mỗi điểm X cho biết có bao nhiêu phần của một đơn vị xác suất phân phối trong khoảng $(-\infty; x)$.

•**Ví dụ 2.10** Đại lượng ngẫu nhiên liên tục X có hàm phân phối xác suất như sau:

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

a) Tìm các hệ số a,b.
b) Tìm xác suất để trong kết quả của phép thử, X nhận giá trị trong khoảng $[0, \dfrac{1}{3})$.
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 $x = -1$ và $\frac{1}{3}$ vì vậy:

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

Khi đó $F(x)$ thoả 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:

$$P(0 \leq X < \frac{1}{3}) = F(\frac{1}{3}) - F(0) = 1 - \frac{3}{4} = \frac{1}{4}$$

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

$$p = P(0 < X < 1) = F(1) - F(0) = 1 - \frac{3}{4} = \frac{1}{4}$$

Vậy xác suất cần tìm bằng:

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

2.2.3. Hàm mật độ xác suất

a) Định nghĩa.

⋆ **Định nghĩa 2.4** Hàm mật độ xác suất của đại lượng ngẫu nhiên liên tục X, ký hiệu $f(x)$, là đạo hàm bậc nhất của hàm phân phối xác suất của đại lượng ngẫu nhiên đó.

$$f(x) = [F(x)]'$$

Khái niệm về hàm mật độ xác suất chỉ áp dụng đối với các đại lượng ngẫu nhiên liên tục.

b) Các tính chất của hàm mật độ xác suất.

- Tính chất 1. Hàm mật độ xác suất luôn không âm:

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

  ■.Hàm phân phối xác suất $F(x)$ là hàm không giảm, vì vậy đạo hàm bậc nhất của nó, $f(x) = [F(x)]'$, là hàm không âm. Về mặt hình học, đồ thị của hàm số $f(x)$ không nằm thấp hơn trục Ox. □

- Tính chất 2. Xác suất để đại lượng ngẫu nhiên liên tục X nhận giá trị trong khoảng $[a, b)$ bằng tích phân xác định của hàm mật độ xác suất trong khoảng đó:

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

  ■.Theo tính chất của hàm phân phối ta có:

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

Theo công thức Newton-Laibnitz:

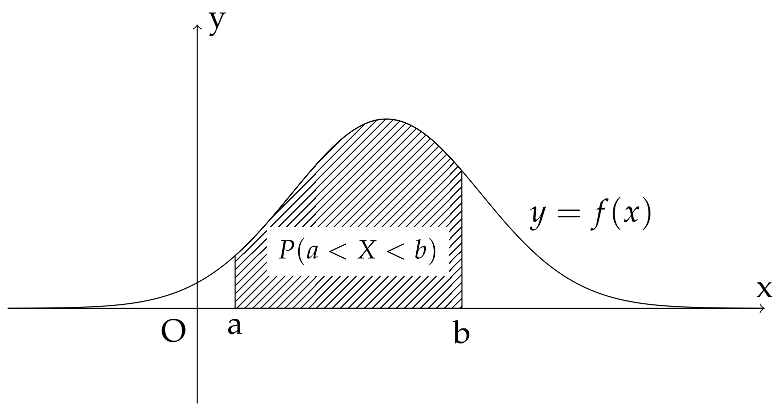
$$F(b) - F(a) = \int_a^b F'(x)dx = \int_a^b f(x)dx.$$

Từ đó suy ra điều phải chứng minh. □

Vì X là đại lượng ngẫu nhiên liên tục nên từ tính chất trên ta có:

$$P(a < X < b) = P(a \le X < b) = P(a < X \le b) = P(a \le X \le b) = \int_a^b f(x)dx$$

Về mặt hình học, kết quả trên có thể minh hoạ như sau: 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:

- Tính chất 3. Hàm phân phối xác suất của đại lượng ngẫu nhiên liên tục X bằng tích phân suy rộng của hàm mật độ xác suất trong khoảng $(-\infty; x)$.

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

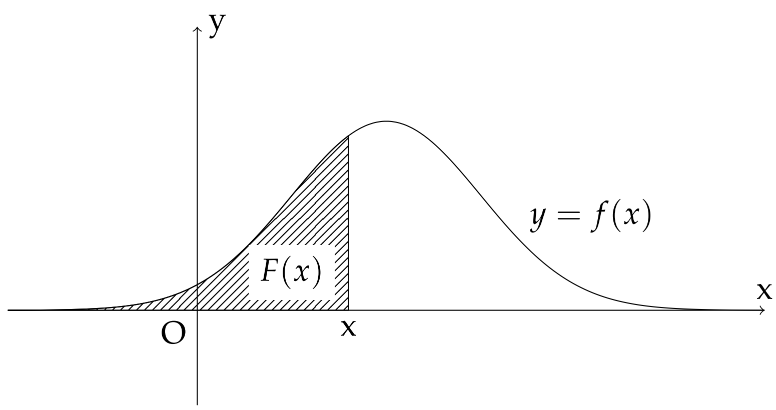
■.Từ tính chất 2, đặt $a = -\infty, b = x$, ta có:

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

□

Công thức trên cho phép tìm hàm phân phối xác suất của đại lượng ngẫu nhiên liên tục khi đã biết hàm mật độ xác suất của nó.

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 đường thẳng $x = a$.



Hình 2.3:

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

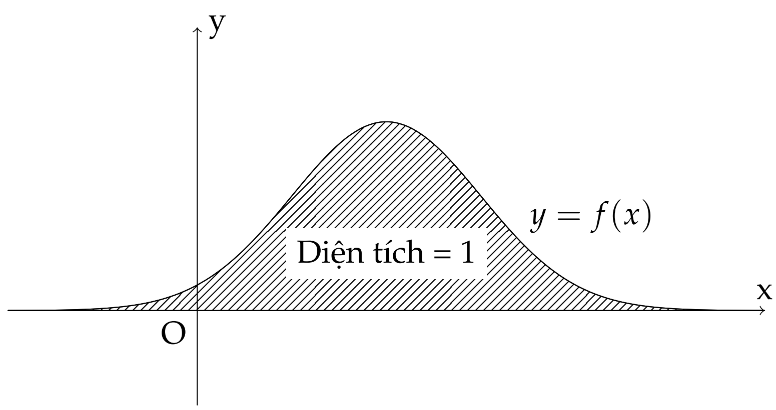
$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

■.Từ tính chất 2, đặt $a = -\infty, b = +\infty$, và chú ý rằng biến cố $(-\infty < X < +\infty)$ là biến cố chắc chắn, ta có:

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

□

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



Hình 2.4:

Ta chú ý rằng điều kiện để một hàm số $f(x)$ có thể là hàm mật độ xác suất của một đại lượng ngẫu nhiên liên tục nào đó thì nó phải thoả mãn hai điều kiện:

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

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

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

a) Tìm hệ số a,

b) Tìm xác suất để đại lượng ngẫu nhiên X nhận giá trị trong khoảng $(0, \frac{\pi}{4})$,

c) Tìm hàm phân phối $F(x)$.

Lời giải.

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

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

Ta có:

$$1 = \int\limits_{-\infty}^{+\infty} f(x)dx = \int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} a.cosxdx = a.sinx|_{-\frac{\pi}{2}}^{\frac{\pi}{2}} = 2a.$$

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

b) Theo tính chất của hàm mật độ xác suất:

$$P(0 < X < \frac{\pi}{4}) = \int\limits_{0}^{\frac{\pi}{4}} f(x)dx = \int\limits_{0}^{\frac{\pi}{4}} a.cosxdx = \frac{1}{2}a.sinx|_0^{\frac{\pi}{4}} = \frac{\sqrt{2}}{4}.$$

c) Để tìm hàm phân phối xác suất ta sử dụng tính chất:

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt$$

Với $x < -\dfrac{\pi}{2}$: $F(x) = \int\limits_{-\infty}^{x} 0dt = 0.$

Với $-\dfrac{\pi}{2} \leq x \leq \dfrac{\pi}{2}$:

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{-\frac{\pi}{2}} 0dt + \int\limits_{-\frac{\pi}{2}}^{x} \frac{1}{2}costdt = \frac{1}{2}(sinx + 1)$$

Với $x > \dfrac{\pi}{2}$:

$$F(x) = \int\limits_{-\infty}^{x} f(t)dt = \int\limits_{-\infty}^{-\frac{\pi}{2}} 0dt + \int\limits_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1}{2}costdt + \int\limits_{\frac{\pi}{2}}^{x} 0dt = 1$$

Vậy hàm phân phối xác suất của X có dạng:

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

•**Ví dụ 2.12** Hàm mật độ xác suất của đại lượng ngẫu nhiên X có dạng:

$$f(x) = \frac{1}{\pi(1 + x^2)};\ \forall x \in \mathbb{R}$$

Tìm xác suất để khi tiến hành 3 phép thử độc lập có đúng 2 lần X nhận giá trị trong khoảng $(-1, 1)$.

Lời giải. Xác suất để khi tiến hành một phép thử đại lượng ngẫu nhiên X nhận giá trị trong khoảng $(-1, 1)$ là:

$$P(-1 < X < 1) = \int\limits_{-1}^{1} f(x)dx = \int\limits_{-1}^{1} \frac{1}{\pi(1 + x^2)}dx = \frac{1}{\pi}arctanx|_{-1}^{1} = \frac{1}{2}.$$

Khi đó theo công thức Bernoulli, xác suất để trong 3 phép thử độc lập có 2 lần X nhận giá trị trong khoảng $(-1, 1)$ là:

$$p_2 = C_3^2 \left(\frac{1}{2}\right)^2 . \left(1 - \frac{1}{2}\right) = \frac{3}{8}.$$

c) 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.

Thật vậy, theo định nghĩa của hàm mật độ xác suất ta có:

$$f(x) = F'(x) = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x \leq X < x + \Delta x)}{\Delta x}.$$

Như vậy hàm mật độ xác suất tại điểm X chính là giới hạn của mật độ xác suất trung bình trên đoạn $(x; x + \Delta x)$ khi $\Delta x \to 0$.

2.3. Các tham số đặc trưng của đại lượng ngẫu nhiên

2.3.1. Kỳ vọng toán

a) Định nghĩa

⋆ **Định nghĩa 2.5** Kỳ vọng toán của đại lượng ngẫu nhiên X, ký hiệu $M(X)$ hoặc $E(X)$, là một số được xác định như sau:

- Nếu X là đại lượng ngẫu nhiên rời rạc có bảng phân phối xác suất:
  thì kỳ vọng toán $M(X)$ được xác định bằng biểu thức:

X	x_1	x_2	...	x_i	...	x_n
P	p_1	p_2	...	p_i	...	p_n

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

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

$$M(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

•**Ví dụ 2.13** Cho X là đại lượng ngẫu nhiên rời rạc có bảng phân phối xác suất: Khi đó kỳ vọng

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

toán của đại lượng ngẫu nhiên X bằng:

$$M(X) = 1.0,1 + 3.0,5 + 4.0,4 = 3,2.$$

•**Ví dụ 2.14** 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:

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{0} x.0dx + \int_{0}^{1} \frac{3}{4}x.(x^2 + 2x)dx + \int_{1}^{+\infty} x.0dx$$

$$= \frac{3}{4}\int_{0}^{1}(x^3 + 2x^2)dx = \frac{3}{4}\left(\frac{x^4}{4} + \frac{2x^3}{3}\right)\Big|_0^1 = \frac{11}{16}.$$

b) Các tính chất của kỳ vọng toán

Các chứng minh trong phần này được thực hiện với đại lượng ngầu nhiên rời rạc, nếu X là đại lượng ngầu nhiên liên tục ta cũng có các chứng minh tương tự.

- Tính chất 1. Kỳ vọng toán của hằng số bằng chính hằng số đó. Như vậy nếu C là hằng số thì: $M(C) = C$.

  ■.Thật vậy, có thể coi C là đại lượng ngẫu nhiên rời rạc đặc biệt, với một giá trị có thể có là C và xác suất tương ứng bằng 1. Khi đó theo định nghĩa kỳ vọng toán ta có:

$$M(C) = 1.C = C$$

□

- Tính chất 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.

$$M(CX) = C.M(X).$$

  ■.Thật vậy, giả sử X là đại lượng ngẫu nhiên rời rạc với bảng phân phối xác suất:

X	x_1	x_2	...	x_i	...	x_n
P	p_1	p_2	...	p_i	...	p_n

  Khi đó đại lượng ngẫu nhiên CX có bảng phân phối xác suất như sau:

CX	Cx_1	Cx_2	...	Cx_i	...	Cx_n
P	p_1	p_2	...	p_i	...	p_n

Theo định nghĩa của kỳ vọng toán ta có:

$$M(CX) = Cx_1p_1 + ... + Cx_np_n = C(x_1p_1 + ... + x_np_n) = C.M(X)$$

□

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

$$M(X+Y) = M(X) + M(Y).$$

■.Thật vậy, giả sử các đại lượng ngẫu nhiên rời rạc X và Y có các bảng phân phối xác suất như sau:

X	x_1	x_2	...	x_i	...	x_n
P	p_1	p_2	...	p_i	...	p_n

Y	y_1	y_2	...	y_i	...	y_m
P	q_1	q_2	...	q_i	...	q_m

Khi đó quy luật phân phối xác suất của tổng $X+Y$ như sau:

X+Y	x_1+y_1	x_1+y_2	...	x_i+y_j	...	x_n+y_m
P	p_{11}	p_{12}	...	p_{ij}	...	p_{nm}

trong đó p_{ij} là xác suất để tổng $X+Y$ nhận giá trị x_i+y_j.

Theo định nghĩa kỳ vọng toán ta có:

$$M(X+Y) = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i+y_j)p_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m}x_ip_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{m}y_jp_{ij}$$

$$= \sum_{i=1}^{n}x_i\left(\sum_{j=1}^{m}p_{ij}\right) + \sum_{j=1}^{m}y_j\left(\sum_{i=1}^{n}p_{ij}\right) \tag{2.1}$$

Ta sẽ chứng minh rằng $\sum\limits_{j=1}^{m} p_{ij} = p_i$. Thật vậy, biến cố $X = x_i$ sẽ xảy ra khi tổng $X+Y$ nhận các giá trị $x_i+y_1, x_i+y_2, ..., x_i+y_m$. Do đó theo định lý cộng xác suất ta có:

$$p_i = P(X=x_i) = P(X+Y = x_i+y_1) + P(X+Y = x_i+y_2) + ... + P(X+Y = x_i+y_m)$$

$$= p_{i1} + p_{i2} + ... + p_{im} = \sum_{j=1}^{m}p_{ij}.$$

Tương tự như vậy ta cũng có $\sum\limits_{i=1}^{n} p_{ij} = q_j$.
Từ đó thay vào (2.1) ta có:

$$M(X+Y) = \sum_{i=1}^{n}x_ip_i + \sum_{j=1}^{m}y_jq_j = M(X) + M(Y).$$

□

Bằng phương pháp chứng minh quy nạp, ta cũng có kết quả sau đây:

Kỳ vọng toán của tổng n đại lượng ngẫu nhiên bằng tổng các kỳ vọng toán thành phần:

$$M(X_1 + X_2 + ... + X_n) = M(X_1) + M(X_2) + ... + M(X_n).$$

- Tính chất 4. Kỳ vọng toán của tích hai đại lượng ngẫu nhiên độc lập bằng tích các kỳ vọng toán thành phần:

$$M(X.Y) = M(X).M(Y).$$

Ở đó hai đại lượng ngẫu nhiên được gọi là độc lập với nhau nếu quy luật phân phối xác suất của đại lượng ngẫu nhiên này không phụ thuộc vào việc đại lượng ngẫu nhiên kia nhận giá trị bằng bao nhiêu. Tương tự, n đại lượng ngẫu nhiên được gọi là độc lập với nhau nếu các quy luật phân phối xác suất của một số bất kỳ các đại lượng ngẫu nhiên nào đó không phụ thuộc vào việc các đại lượng ngẫu nhiên kia nhận giá trị bằng bao nhiêu.
■.Giả sử các đại lượng ngẫu nhiên X và Y độc lập, có bảng phân phối xác suất như sau:

X	x_1	x_2	...	x_i	...	x_n
P	p_1	p_2	...	p_i	...	p_n

Y	y_1	y_2	...	y_i	...	y_m
P	q_1	q_2	...	q_i	...	q_m

Khi đó quy luật phân phối xác suất của tổng $X.Y$ như sau:

X.Y	$x_1.y_1$	$x_1.y_2$	...	$x_i.y_j$	...	$x_n.y_m$
P	p_1q_1	p_1q_2	...	p_iq_j	...	p_nq_m

Theo công thức tính kỳ vọng toán:

$$M(XY) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_iy_jp_iq_j = \sum_{i=1}^{n} x_ip_i \sum_{j=1}^{m} y_jq_j = M(X).M(Y).$$

□

Bằng phương pháp chứng minh quy nạp, ta cũng có kết quả sau đây:

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

$$M(X_1.X_2....X_n) = M(X_1).M(X_2)...M(X_n).$$

- Tính chất 5. Cho X là đại lượng ngẫu nhiên và $g(x)$ là hàm số liên tục, khi đó đại lượng ngẫu nhiên $g(X)$ có kỳ vọng toán $M[g(X)]$ được xác định như sau:
+) Nếu X là đại lượng ngẫu nhiên rời rạc có phân phối xác suất $p_i = P(X = x_i);\ i = \overline{1, n}$:

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

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

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

c) Ý nghĩa của kỳ vọng toán

Kỳ vọng toán của đại lượng ngẫu nhiên X gần bằng trung bình số học của các giá trị quan sát của đại lượng ngẫu nhiên. Nó phản ánh giá trị trung tâm của phân phối xác suất của đại lượng ngẫu nhiên.

Thật vậy, giả sử đối với đại lượng ngẫu nhiên X tiến hành n phép thử độc lập trong đó có n_1 lần X nhận giá trị x_1, n_2 lần X nhận giá trị x_2, ..., n_k lần X nhận giá trị x_k. Giá trị trung bình của X trong n phép thử này là:

$$\overline{X} = \frac{x_1n_1 + x_2n_2 + ... + x_kn_k}{n} = x_1\frac{n_1}{n} + x_2\frac{n_2}{n} + ... + x_k\frac{n_k}{n}$$

ở đó $\frac{n_1}{n}, ..., \frac{n_k}{n}$ chính là tần suất xuất hiện các giá trị $x_1, x_2, ..., x_k$ trong n phép thử trên, do đó:

$$\overline{X} = x_1f_1 + x_2f_2 + ... + x_kf_k.$$

Theo định nghĩa thống kê về xác suất, khi $n \to \infty$ các tần suất trên sẽ hội tụ về xác suất tương ứng, do đó với n đủ lớn ta có:

$$\overline{X} \approx x_1p_1 + x_2p_2 + ... + x_kp_k = M(X).$$

•**Ví dụ 2.15** Tung con súc sắc n lần. Tìm kỳ vọng toán của tổng số chấm thu được.

Lời giải. Gọi $X_i, (i = \overline{1, n})$ là số chấm thu được ở lần tung thứ i và X là tổng số chấm thu được sau n lần tung. Ta có:

$$X = X_1 + X_2 + ... + X_n.$$

Theo tính chất của kỳ vọng toán:

$$M(X) = M(X_1) + M(X_2) + ... + M(X_n).$$

Mỗi đại lượng ngẫu nhiên X_i đều có bảng phân phối xác suất như sau:

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

Do đó:

$$M(X_i) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}.$$

Vì vậy kỳ vọng toán của tổng số chấm thu được sau n lần tung là:

$$M(X) = \frac{7n}{2}.$$

2.3.2. Phương sai

a) Định nghĩa

⋆ **Định nghĩa 2.6** Phương sai của đại lượng ngẫu nhiên X, ký hiệu $D(X)$, là kỳ vọng toán của bình phương sai lệch của đại lượng ngẫu nhiên so với kỳ vọng toán của nó.

$$D(X) = M[X - M(X)]^2.$$

Như vậy nếu X là đại lượng ngẫu nhiên rời rạc thì phương sai được xác định bằng công thức:

$$D(X) = \sum_{i=1}^{n} [x_i - M(X)]^2.p_i$$

còn nếu X là đại lượng ngẫu nhiên liên tục thì phương sai được xác định bằng công thức:

$$\int_{-\infty}^{+\infty} [x - M(X)]^2 f(x)dx.$$

b) Cách tính phương sai

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

$$\begin{aligned} D(X) = M[X - M(X)]^2 &= M\left(X^2 - 2X.M(X) + [M(X)]^2\right) \\ &= M(X^2) + M(-2X.M(X)) + M([M(X)]^2) \\ &= M(X^2) - 2M(X).M(X) + M([M(X)]^2) \\ &= M(X^2) - [M(X)]^2 \end{aligned}$$

Như vậy đối với đại lượng ngẫu nhiên rời rạc, ta có thể tính phương sai bằng công thức:

$$D(X) = \left(\sum_{i=1}^{n} x_i^2 p_i\right) - [M(X)]^2 = \left(\sum_{i=1}^{n} x_i^2 p_i\right) - \left(\sum_{i=1}^{n} x_i p_i\right)^2$$

còn đối với đại lượng ngẫu nhiên liên tục có hàm mật độ xác suất $f(x)$:

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x)dx - [M(X)]^2 = \left(\int_{-\infty}^{+\infty} x^2 f(x)dx\right) - \left(\int_{-\infty}^{+\infty} xf(x)dx\right)^2.$$

•Ví dụ 2.16 Cho X là đại lượng ngẫu nhiên rời rạc có bảng phân phối xác suất:

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

Tìm phương sai $D(X)$.

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

$$M(X) = 1.0,1 + 3.0,5 + 4.0,4 = 3,2.$$

Kỳ vọng toán $M(X^2)$ bằng:

$$M(X^2) = 1^2.0,1 + 3^2.0,5 + 4^2.0,4 = 11.$$

Phương sai của đại lượng ngẫu nhiên X bằng:

$$D(X) = M(X^2) - [M(X)]^2 = 11 - 3,2^2 = 0,76.$$

•Ví dụ 2.17 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 $D(X)$.

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

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-\infty}^{0} x.0dx + \int_{0}^{1} \frac{3}{4}(x^3+2x^2)dx + \int_{1}^{+\infty} x.0dx$$

$$= \frac{3}{4}\int_{0}^{1}(x^3+2x^2)dx = \frac{3}{4}\left(\frac{x^4}{4} + \frac{2x^3}{3}\right)\Bigg|_0^1 = \frac{11}{16}.$$

Kỳ vọng toán $M(X^2)$ bằng:

$$M(X^2) = \int_{-\infty}^{+\infty} x^2f(x)dx = \int_{-\infty}^{0} x^2.0dx + \int_{0}^{1} \frac{3}{4}(x^4+2x^3)dx + \int_{1}^{+\infty} x^2.0dx$$

$$= \frac{3}{4}\int_{0}^{1}(x^4+2x^3)dx = \frac{3}{4}\left(\frac{x^5}{5} + \frac{2x^4}{4}\right)\Bigg|_0^1 = \frac{21}{40}.$$

Phương sai của đại lượng ngẫu nhiên X bằng:

$$D(X) = M(X^2) - [M(X)]^2 = \frac{21}{40} - \left(\frac{11}{16}\right)^2 = \frac{67}{1280}.$$

c) Các tính chất của phương sai

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

  $$D(X) \geq 0, \ \forall X$$

  $$D(C) = 0, \ C \text{ là hằng số.}$$

  Tính chất này được suy trực tiếp từ định nghĩa phương sai.

- Tính chất 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:

  $$D(CX) = C^2.D(X).$$

  ■.Thật vậy, theo định nghĩa phương sai:

  $$D(CX) = M(CX - M(CX))^2 = M(CX - CM(X))^2$$

  $$= M[C^2(X - M(X))^2] = C^2.M[(X - M(X))^2] = C^2.D(X).$$

  □

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

  $$D(X+Y) = D(X) + D(Y).$$

  ■.Thật vậy, theo công thức tính phương sai và tính chất kỳ vọng toán của tích hai đại lượng ngẫu nhiên độc lập:

  $$\begin{aligned} D(X+Y) &= M(X+Y)^2 - [M(X+Y)]^2 \\ &= M(X^2 + 2XY + Y^2) - [M(X) + M(Y)]^2 \\ &= M(X^2) + 2M(X)M(Y) + M(Y^2) - [(M(X)^2) + 2M(X)M(Y) + (M(Y)^2)] \\ &= [M(X^2) - (M(X)^2)] + [M(Y^2) - (M(Y)^2)] \\ &= D(X) + D(Y). \end{aligned}$$

  □

Từ tính chất trên có thể chứng minh được các hệ quả sau:
+) Hệ quả 1. Phương sai của tổng một hằng số với một đại lượng ngẫu nhiên bằng phương sai của chính đại lượng ngẫu nhiên đó:

$$D(X+C) = D(X), \text{ C là hằng số.}$$

+) Hệ quả 2. Phương sai của hiệu hai đại lượng ngẫu nhiên độc lập bằng tổng các phương sai thành phần:

$$D(X-Y) = D(X) + D(Y)$$

+) Hệ quả 3. Phương sai của tổng n đại lượng ngẫu nhiên độc lập bằng tổng các phương sai thành phần:

$$D(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} D(X_i)$$

d) Ý nghĩa của phương sai

Phương sai chính là trung bình số học của bình phương các sai lệch giữa các giá trị có thể có của đại lượng ngẫu nhiên so với giá trị trung bình của các giá trị đó. Do đó phương sai phản ánh mức độ phân tán của các giá trị của đại lượng ngẫu nhiên xung quanh giá trị trung tâm của nó là kỳ vọng toán.

•**Ví dụ 2.18** Tung con súc sắc n lần, tìm phương sai của tổng số chấm thu được.

Lời giải. Gọi $X_i, (i = \overline{1,n})$ là số chấm thu được ở lần tung thứ i và X là tổng số chấm thu được sau n lần tung. Ta có:

$$X = X_1 + X_2 + ... + X_n.$$

Vì các đại lượng ngẫu nhiên $X_i, (i = \overline{1,n})$ độc lập nên theo tính chất của phương sai:

$$D(X) = D(X_1) + D(X_2) + ... + D(X_n).$$

Mỗi đại lượng ngẫu nhiên $X_i, (i = \overline{1,n})$ đều có bảng phân phối xác suất như sau: Do đó:

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

$$M(X_i) = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2} \ (\forall i = \overline{1,n})$$
$$M(X_i^2) = \frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2) = \frac{91}{6} \ (\forall i = \overline{1,n})$$
$$D(X_i) = M(X_i^2) - [M(X_i)]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \ (\forall i = \overline{1,n})$$

Vì vậy phương sai của tổng số chấm thu được sau n lần tung là:

$$D(X) = \frac{35n}{12}.$$

2.3.3. Độ lệch tiêu chuẩn

⋆ **Định nghĩa 2.7** Độ lệch tiêu chuẩn của đại lượng ngẫu nhiên X, ký hiệu σ_X, là căn bậc hai của phương sai:

$$\sigma_X = \sqrt{D(X)}$$

Ta thấy rằng đơn vị đo của độ lệch tiêu chuẩn bằng 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 vi đo của nó.

2.3.4. Mốt

⋆ **Định nghĩa 2.8** Mốt X, ký hiệu $mod(X)$, là giá trị của đại lượng ngẫu nhiên X có khả năng xuất hiện lớn nhất trong một lân cận nào đó của nó.

Đối với đại lượng ngẫu nhiên rời rạc, $mod(X)$ là giá trị của X ứng với xác suất lớn nhất.

Đối với đại lượng ngẫu nhiên liên tục, $mod(X)$ là giá trị X làm cực đại hàm mật độ.

•**Ví dụ 2.19** Đại lượng ngẫu nhiên X có bảng phân phối xác suất:

X	-2	-1	0	3
P	0,1	0,7	0,05	0,15

Suy ra $mod(X) = -1$

•**Ví dụ 2.20** Đại lượng ngẫu nhiên X có hàm mật độ xác suất:

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

Xác định $mod(X)$.

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 $x = \sqrt{2}$.
Lập bảng biến thiên của $f(x)$

x	$-\infty$	$-\sqrt{2}$	0		$\sqrt{2}$		$+\infty$
$f'(x)$			$\vert$	$+$	0	$-$	
$f(x)$	0 $-$ $-$	$- >$	0	$\nearrow$	CĐ	$\searrow$	0

Ta được $mod(X) = \sqrt{2}$.

2.3.5. Trung vị

$\star$ **Định nghĩa 2.9** Trung vị là giá trị của đại lượng ngẫu nhiên X, kí hiệu là $med(X)$, là giá trị của X thỏa mãn:

$$P(X < medX) \leq \frac{1}{2} \text{ và } P(X \leq medX) \geq \frac{1}{2}$$

Như vậy nếu X là đại lượng ngẫu nhiên liên tục thì:

$$P(X < medX) = P(X \geq medX) = \frac{1}{2}$$

và do đó $medX$ là nghiệm của phương trình $F(x) = \frac{1}{2}$.

•**Ví dụ 2.21** Đại lượng ngẫu nhiên rời rạc X có bảng phân phối xác suất:

X	-1	0	2	3	5	8
P	0,15	0,05	0,25	0,15	0,35	0,05

Khi đó $medX = 3$.

•**Ví dụ 2.22** 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 med(X).

Lời giải. Xét phương trình

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

Vậy $medX = \dfrac{1}{\sqrt{2}}$.

2.3.6. Phân vị

⋆ **Định nghĩa 2.10** Phân vị mức α của X là giá trị x_α của X thoả mãn $P(X < x_\alpha) = \alpha$.

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 $F(x) = \alpha$.

•**Ví dụ 2.23** Đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất:

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

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:

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

- Xét phương trình $F(x) = 0,95 \Leftrightarrow 1 - e^{-x} = 0,95 \Leftrightarrow e^{-x} = 0,05 \Leftrightarrow x = -ln0,05 = 2,996$. Vậy $x_{0,95} = 2,996$.

- Xét phương trình $F(x) = 0,5 \Leftrightarrow 1 - e^{-x} = 0,5 \Leftrightarrow e^{-x} = 0,5 \Leftrightarrow x = -ln0,5 = 0,693$. Vậy $x_{0,5} = 0,693$.Lưu ý rằng đây chính là $med(X)$.

- Xét phương trình $F(x) = 0,095 \Leftrightarrow 1 - e^{-x} = 0,05 \Leftrightarrow e^{-x} = 0,95 \Leftrightarrow x = -ln0,95 = 0,051$. Vậy $x_{0,05} = 0,051$.

2.4. Một số quy luật phân phối xác suất thông dụng

2.4.1. Quy luật phân phối chuẩn $N(\mu, \sigma^2)$

Định nghĩa

⋆ **Định nghĩa 2.11** 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 là μ và σ^2, nếu hàm mật độ xác suất của nó có dạng:

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

Quy luật phân phối chuẩn với hai tham số μ và σ^2 được ký hiệu là $N(\mu, \sigma^2)$.

Khảo sát hàm mật độ $f(x)$:

- Hàm số xác định với $x \in \mathbb{R}$.

- Đạo hàm bậc nhất:

$$f'(x) = -\frac{x-\mu}{\sigma^3.\sqrt{2\pi}}.e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}$$

  Như vậy khi $x = \mu$ hàm số có cực đại bằng $\dfrac{1}{\sigma.\sqrt{2\pi}}$.

- Đạo hàm bậc hai:

$$f''(x) = -\frac{1}{\sigma^3.\sqrt{2\pi}}.e^{-\dfrac{(x-\mu)^2}{2\sigma^2}}.\left\{1 - \frac{(x-\mu)^2}{\sigma^2}\right\}$$
$$f''(x) = 0 \Leftrightarrow x = \mu \pm \sigma.$$

  Như vậy hàm số có 2 điểm uốn: $\left(\mu - \sigma, \dfrac{1}{\sigma.\sqrt{2\pi e}}\right)$ và $\left(\mu + \sigma, \dfrac{1}{\sigma.\sqrt{2\pi e}}\right)$

- Giới hạn:

$$\lim_{x \to \pm\infty} f(x) = 0.$$

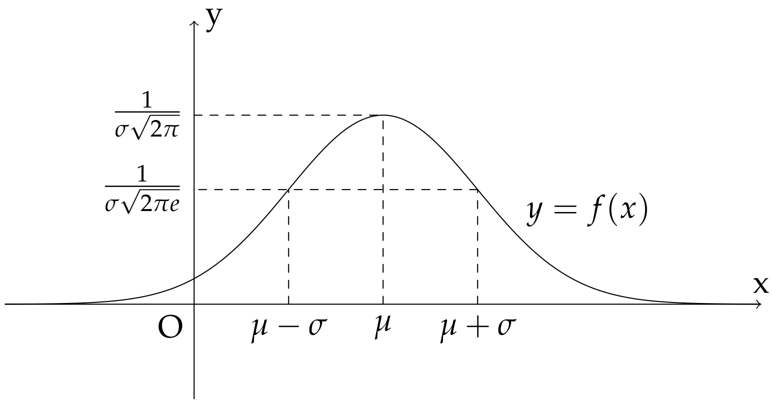
  Như vậy trục Ox là tiệm cận ngang của đồ thị hàm số.

- Đồ thị của hàm số đối xứng qua đường thẳng $x = \mu$ và có dạng như sau:

Các tham số đặc trưng của quy luật phân phối chuẩn

Cho X là đại lượng ngẫu nhiên phân phối chuẩn với hai tham số μ và σ^2.
a) Kỳ vọng toán của đại lượng ngẫu nhiên X chính là tham số μ:

$$M(X) = \mu.$$



Hình 2.5:

■.Theo định nghĩa kỳ vọng toán của đại lượng ngẫu nhiên liên tục ta có:

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx = \frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{+\infty} x.e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Thực hiện phép đổi biến số: $u = \frac{x-\mu}{\sigma}$. Khi đó cận tích phân không đổi và ta được:

$$\begin{aligned} M(X) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma u + \mu).e^{-\frac{u^2}{2}} du \\ &= \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u.e^{-\frac{u^2}{2}} du + \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du. \end{aligned}$$

Tích phân thứ nhất bằng không do hàm số dưới dấu tích phân là hàm lẻ mà cận lấy tích phân lại đối xứng qua 0. Còn tích phân thứ hai là tích phân Poisson:

$$J = \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du = \sqrt{2\pi}.$$

Do đó $M(X) = \mu$ □

b) Độ lệch tiêu chuẩn của đại lượng ngẫu nhiên X chính là tham số σ:

$$\sigma_X = \sigma \text{hay} D(X) = \sigma^2$$

■.Do $M(X) = \mu$, theo định nghĩa phương sai của đại lượng ngẫu nhiên liên tục:

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 .f(x)dx = \frac{1}{\sigma.\sqrt{2\pi}} \int_{-\infty}^{+\infty} (x-\mu)^2.e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx.$$

Thực hiện phép đổi biến số: $u = \dfrac{x-\mu}{\sigma}$. Khi đó cận tích phân không đổi và ta được:

$$
\begin{aligned}
D(X) &= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u^2.e^{-\frac{u^2}{2}} du \\
&= -\frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} u de^{-\frac{u^2}{2}} \\
&= -\frac{\sigma^2}{\sqrt{2\pi}} u.e^{-\frac{u^2}{2}} \Big|_{-\infty}^{+\infty} + \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{u^2}{2}} du \ \text{ (tích phân từng phần)} \\
&= \frac{\sigma^2}{\sqrt{2\pi}} \left(\lim_{x\to-\infty} \frac{u}{e^{\frac{u^2}{2}}} - \lim_{x\to+\infty} \frac{u}{e^{\frac{u^2}{2}}} \right) + \frac{\sigma^2}{\sqrt{2\pi}}.\sqrt{2\pi} \ \text{ (tích phân Poisson)} \\
&= 0 + \sigma^2 = \sigma^2.
\end{aligned}
$$

□

∗ **Nhận xét:** Nếu đại lượng ngẫu nhiên X phân phối chuẩn với hai tham số μ và σ^2 thì đại lượng ngẫu nhiên $U = \dfrac{X-\mu}{\sigma}$ có kỳ vọng toán $M(U) = 0$ và độ lệch chuẩn $\sigma_U = 1$.

Quy luật phân phối chuẩn hoá

a) Định nghĩa

⋆ **Định nghĩa 2.12** 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 nếu hàm mật độ xác suất của nó có dạng:

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

Hàm mật độ $f(x)$ của quy luật phân phối chuẩn hoá có đồ thị đối xứng qua trục tung Oy như sau:

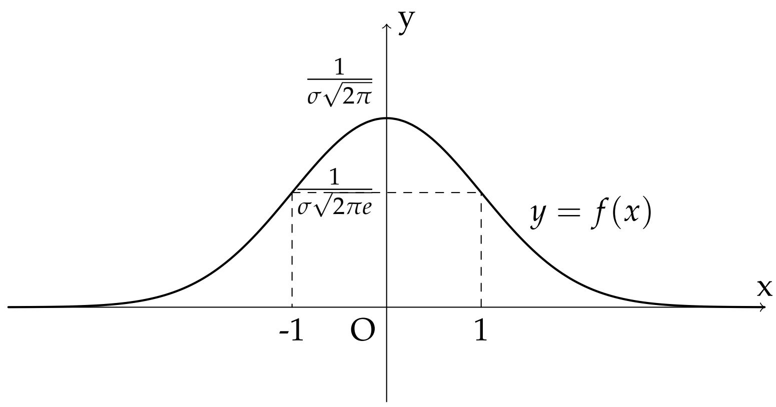
Dễ thấy:

- Hàm số $f(x)$ là hàm số chẵn: $f(-x) = f(x)$.

- $\lim\limits_{x\to+\infty} f(x) = 0$.

Các giá trị của hàm $f(x)$ được tính sẵn thành bảng (xem Phụ lục C.1).

b) Hàm phân phối xác suất của đại lượng ngẫu nhiên phân phối chuẩn hoá



Hình 2.6:

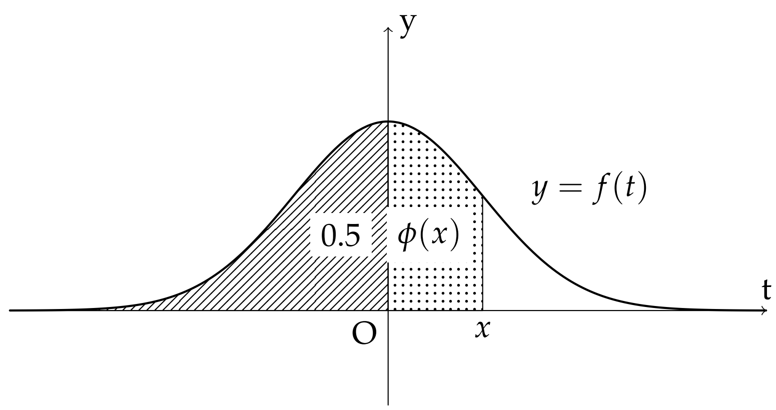
Từ hàm mật độ $f(x)$ ta có hàm phân phối xác suất $F(x)$ của đại lượng ngẫu nhiên phân phối chuẩn hoá có dạng:

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

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

Dễ thấy:

- $F(x) = 0,5 + \Phi(x)$.
- Hàm số $\Phi(x)$ là hàm số lẻ: $\Phi(-x) = -\Phi(x)$.
- $\lim\limits_{x \to +\infty} \Phi(x) = 0,5$.



Hình 2.7:

Các giá trị của hàm $\Phi(x)$ được tính sẵn thành bảng (xem Phụ lục C.2).

c) Các tham số đặc trưng của đại lượng ngẫu nhiên phân phối chuẩn hoá

Giả sử U là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn hoá. Dễ thấy hàm mật độ của nó là hàm mật độ của quy luật phân phối chuẩn với hai tham số $\mu = 0$ và $\sigma^2 = 1$. Vì vậy các tham số đặc trưng của U là:

Kỳ vọng toán: $M(U) = 0$.

Phương sai: $D(U) = 1$; độ lệch tiêu chuẩn: $\sigma_U = 1$.

Phân phối chuẩn hoá được ký hiệu là $N(0,1)$.

d) Phân vị chuẩn

⋆ **Định nghĩa 2.13** Phân vị chuẩn mức $\alpha (0 \leq \alpha \leq 1)$, ký 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:

$$P(U < u_\alpha) = \alpha.$$

Từ hàm mật độ của phân phối chuẩn hoá ta có biểu thức xác định u_α như sau:

$$P(U < u_\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u_\alpha} e^{-\frac{t^2}{2}} dt = \alpha.$$

Cho trước α, dựa vào biểu thức trên ta tính được phân vị u_α và ngược lại. Bảng tính sẵn giá trị u_α tương ứng với mức α cho trước được trình bày trong phần phụ lục (xem Phụ lục C.3).

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

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

Các công thức tính xác suất của quy luật phân phối chuẩn

Cho đại lượng ngẫu nhiên liên tục X có phân phối chuẩn với hai tham số μ và σ^2.

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

Ta có: $U = \dfrac{X-\mu}{\sigma}$ là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn hoá $N(0,1)$, do đó:

$$\begin{aligned} P(a < X < b) &= P\left(\frac{a-\mu}{\sigma} < X < \frac{b-\mu}{\sigma}\right) \\ &= F\left(\frac{b-\mu}{\sigma}\right) - F\left(\frac{a-\mu}{\sigma}\right) \quad \text{(F(x) là hàm phân phối chuẩn hoá)} \\ &= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \quad \text{(tích phân Laplace)} \end{aligned}$$

Như vậy ta thu được công thức:

$$P(a < X < b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) \tag{2.2}$$

Bảng giá trị của hàm số $\Phi(x)$ cho ở phụ lục C.2.

b) Xác suất của sự sai lệch giữa đại lượng ngẫu nhiên X và kỳ vọng toán của nó

Ta phải tính xác suất để xảy ra bất đẳng thức:$|X-\mu| < \epsilon$.

Sử dụng công thức (2.2) ta có:

$$\begin{aligned} P(|X-\mu| < \epsilon) &= P(\mu-\epsilon < X < \mu+\epsilon) \\ &= \Phi\left(\frac{\mu+\epsilon-\mu}{\sigma}\right) - \Phi\left(\frac{\mu-\epsilon-\mu}{\sigma}\right) = \Phi\left(\frac{\epsilon}{\sigma}\right) - \Phi\left(-\frac{\epsilon}{\sigma}\right) \\ &= \Phi\left(\frac{\epsilon}{\sigma}\right) + \Phi\left(\frac{\epsilon}{\sigma}\right) = 2.\Phi\left(\frac{\epsilon}{\sigma}\right) \end{aligned}$$

Như vậy ta thu được công thức:

$$P(|X-\mu| < \epsilon) = 2.\Phi\left(\frac{\epsilon}{\sigma}\right) \tag{2.3}$$

c) Quy tắc "$3-\sigma$"

Nếu trong công thức (2.3) ta đặt $\epsilon = 3\sigma$ tức là bằng 3 lần độ lệch tiêu chuẩn của X thì ta có:

$$P(|X-\mu| < 3\sigma) = 2\Phi\left(\frac{3\sigma}{\sigma}\right) = 2\Phi(3) = 2.0,49865 = 0,9973$$

Điều đó cho thấy xác suất để đại lượng ngẫu nhiên phân phối chuẩn nhận giá trị trong khoảng $(\mu-3\sigma, \mu+3\sigma)$ lên tới 99,73%, còn xác suất để nó nhận giá trị ngoài khoảng đó chỉ có 0,27%. Xuất phát từ nguyên lý xác suất nhỏ, có thể cho rằng thực tế điều đó là không thể có.

Như vậy, 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 thoả mãn điều kiện của quy tắc "ba xích ma" thì có thể xem nó là đại lượng ngẫu nhiên phân phối chuẩn.

•Ví dụ 2.24 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 $\mu = 5cm$ và độ lệch tiêu chuẩn $\sigma = 0,9cm$.
a) Tính tỷ lệ chi tiết có kích thước nằm trong khoảng từ 4 cm đến 7 cm.
b) 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.
c) 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 $\mu = 5cm$ và $\sigma = 0,9cm$.

a) Theo công thức (2.2) ta có:

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

Vậy tỷ lệ chi tiết có kích thước nằm trong khoảng từ 4cm đến 7cm là 85,329%.
b) Theo công thức (2.3) ta có xác suất để lấy ngẫu nhiên 1 chi tiết có kích thước sai lệch so với trung bình không quá 1 cm là:

$$P(|X-\mu| < 1) = 2\Phi\left(\frac{1}{0,9}\right) = 2\Phi(1,11) = 2.0,36650 = 0,733.$$

Vậy 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:

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

c) Theo quy tắc "$3-\sigma$", hầu hết chi tiết có kích thước trong khoảng:

$$(\mu - 3\sigma, \mu + 3\sigma) = (5 - 3.0,9; 5 + 3.0,9) = (2,3; 7,7)(cm).$$

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 Liapunốp 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 lý 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...

2.4.2. Quy luật không - một A(p)

Ví dụ

•**Ví dụ 2.25** Giả sử trong bình có N quả cầu trong đó có M quả cầu trắng, $N - M$ quả cầu đen. Từ bình lấy ngẫu nhiên 1 quả cầu. Gọi X là số cầu trắng lấy được.

X là đại lượng ngẫu nhiên rời rạc nhận một trong hai giá trị $0; 1$. Các xác suất tương ứng là:

$$P(X = 0) = \frac{N - M}{N}; \quad P(X = 1) = \frac{M}{N}$$

Nếu đặt $p = \frac{M}{N}$ và $q = 1 - p = \frac{N - M}{N}$ thì ta có:

$$P(X = 0) = q = p^0.q^1; \quad P(X = 1) = p = p^1.q^0$$

Do đó có thể viết quy luật phân phối xác suất của X như sau:

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

Định nghĩa

⋆ **Định nghĩa 2.14** 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 $p(0 < p < 1)$ nếu các xác suất được tính bằng công thức:

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

Quy luật không - một với tham số p được ký hiệu là $A(p)$.

Các tham số đặc trưng của quy luật không - một

Gọi X là đại lượng ngẫu nhiên tuân theo quy luật không - một với tham số p. Ta có bảng phân phối xác suất của X như sau:

X	0	1
P	q	p

trong đó $q = 1 - p$.

Theo bảng phân phối xác suất của X ta tính được các tham số đặc trưng của X như sau:

- Kỳ vọng toán:

$$M(X) = 0.q + 1.p = p$$
$$M(X^2) = 0^2.q + 1^2.p = p.$$

- Phương sai: $D(X) = M(X^2) - [M(X)]^2 = p - p^2 = pq$.

- Độ lệch chuẩn: $\sigma_X = \sqrt{D(X)} = \sqrt{pq}$.

2.4.3. Quy luật nhị thức $B(n, p)$

Ví dụ

•**Ví dụ 2.26** Giả sử trong bình có N quả cầu trong đó có M quả cầu trắng, $N - M$ quả cầu đen. Từ bình lấy ngẫu nhiên n quả cầu theo phương thức có hoàn lại. Gọi X là số cầu trắng lấy được.

Như vậy ta đã thực hiện n phép thử độc lập (mỗi phép thử là một lần lấy 1 quả cầu). Trong mỗi phép thử chỉ xảy ra hai trường hợp: hoặc lấy được cầu trắng (biến cố A) hoặc không lấy được cầu trắng (biến cô $\overline{A}$). Xác suất lấy được cầu trắng mỗi lần đều bằng $p = \dfrac{M}{N}$ và xác suất lấy được cầu đen mỗi lần đều bằng $q = \dfrac{N-M}{N} = 1 - p$. Do đó theo công thức Bernoulli xác suất để lấy được k cầu trắng là:

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

Ta cũng dễ nhận thấy là nếu gọi X_k là số cầu trắng lấy được ở lần thứ k thì X_k tuân theo quy luật không - một $A(p)$ và $X = X_1 + X_2 + ... + X_n$.

Định nghĩa

⋆ **Định nghĩa 2.15** Đại lượng ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có $X = 0, 1, 2, ..., n$ được gọi là phân phối theo quy luật nhị thức với các tham số là n và p nếu các xác suất tương ứng của nó được tính bởi công thức:

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

Quy luật nhị thức được ký hiệu là B(n,p).

∗ **Nhận xét:** Giả sử bài toán tuân theo lược đồ Bemoulli: Thực hiện dãy n phép thử độc lập; trong mỗi phép thử chỉ có hai khả năng: 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 $\overline{A}$ 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 $B(n, p)$.

Các tham số đặc trưng của quy luật nhị thức

Giả sử X là đại lượng ngẫu nhiên phân phối theo quy luật nhị thức $B(n, p)$. Bảng phân phối xác suất của X như sau:

X	0	1	2	...	k	...	n
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

a) Xác suất để đại lượng ngẫu nhiên X nhận giá trị trong khoảng $[k; k+h]$

Biến cố $(k \le X \le k+h)$ (k, h là các số tự nhiên; $k+h \le n$) có thể tách ra thành tổng của $h+1$ biến cố xung khắctừng đôi là $(X=k), (X=k+1), ..., (X=k+h)$.Do đó áp dụng định lý cộng xác suất ta có:

$$\begin{aligned} P(k \le X \le k+h) &= P(X=k) + P(X=k+1) + ... + P(X=k+h) \\ &= p_k + p_{k+1} + ... + p_{k+h} \quad \text{với} p_k = P(X=k) \end{aligned}$$

b) Kỳ vọng toán và phương sai của đại lượng ngẫu nhiên phân phối nhị thức

Từ bảng phân phối xác suất của X ta có:

$$M(X) = \sum_{k=0}^{n} k.C_n^k p^k.q^{n-k} \qquad M(X^2) = \sum_{k=0}^{n} k^2.C_n^k p^k.q^{n-k}$$

Xét tổng:

$$\sum_{k=0}^{n} C_n^k p^k.q^{n-k} = (p+q)^n \tag{2.4}$$

Đạo hàm 2 vế của (2.4) theo p ta được:

$$\begin{aligned} &\sum_{k=1}^{n} k.C_n^k p^{k-1}.q^{n-k} = n(p+q)^{n-1} \\ \Rightarrow &\sum_{k=0}^{n} k.C_n^k p^k.q^{n-k} = np(p+q)^{n-1} \end{aligned} \tag{2.5}$$

Vế trái của (2.5) là $M(X)$, còn vế phải bằng np do $p+q=l$. Vì vậy $M(X) = np$.

Đạo hàm 2 vế của (2.5) theo p ta được:

$$\begin{aligned} &\sum_{k=1}^{n} k^2.C_n^k p^{k-1}.q^{n-k} = n.(p+q)^{n-1} + n.(n-1).p.(p+q)^{n-2} \\ \Rightarrow &\sum_{k=0}^{n} k^2.C_n^k p^k.q^{n-k} = np(p+q)^{n-1} + n.(n-1).p^2.(p+q)^{n-2} \\ \Rightarrow &M(X^2) = np + n(n-1)p^2 \end{aligned}$$

Do đó:

$$D(X) = M(X^2) - [M(X)]^2 = np + n(n-1)p^2 - n^2p^2 = np(1-p) = npq$$

Ở trên ta đã tính $M(X)$ và $D(X)$ trong trường hợp $n \ge 2$. Tuy nhiên nếu $n = 1$ thì dễ thấy X tuân theo quy luật không - một $A(p)$ và các kết quả trên vẫn đúng.

Như vậy đại lượng ngẫu nhiên phân phối nhị thức $B(n,p)$ có:

- Kỳ vọng toán : $M(X) = np$.
- Phương sai : $D(X) = npq$.
- Độ lệch tiêu chuẩn : $\sigma_X = \sqrt{npq}$.
- Mod X: $np - q \leq Mod(X) \leq np + p$.

•**Ví dụ 2.27** Xác suất để một người bắn trúng bia là 0,8.
a) Giả sử người đó bắn 12 phát đạn, tính số viên đạn trúng bia có khả năng xảy ra nhiều nhất và xác suất để xảy ra điều đó.
b) Muốn số đạn trúng bia trung bình không nhỏ hơn 12 thì người đó phải bắn ít nhất bao nhiêu viên đạn?

Lời giải. Gọi X là số viên đạn mà người đó bắn trúng bia. X tuân theo quy luật nhị thức với $p = 0,8$.
a) Người đó bắn $n = 12$ viên đạn, khi đó X tuân theo quy luật nhị thức $B(n = 12; p = 0,8)$. Số viên đạn trúng bia có khả năng xảy ra nhiều nhất là $mod(X)$:

$$np - q \leq Mod(X) \leq np + p \Leftrightarrow 12.0,8 - 0,2 \leq Mod(X) \leq 12.0,8 + 0,8$$

$$\Leftrightarrow 9,4 \leq Mod(X) \leq 10,4 \Rightarrow Mod(X) = 10$$

Xác suất để người đó bắn trúng 10 viên đạn là:

$$P(X = 10) = C^{1}0_{1}2(0,8)^{1}0.(0,2)^{2} \approx 0,2835$$

b) Giả sử người đó bắn n viên đạn, khi đó X tuân theo quy luật nhị thức $B(n; p = 0,8)$. Theo giả thiết:

$$M(X) = 12 \Leftrightarrow np = 12 \Leftrightarrow n = \frac{12}{0,8} = 15$$

Vậy muốn trung bình có 12 viên đạn trúng bia thì người đó phải bắn 15 viên đạn.

Công thức xấp xỉ Laplace

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

$$\begin{cases} n & > 5 \\ \dfrac{|p-q|}{\sqrt{npq}} & < 0,3 \end{cases}$$

Lúc đó thì đại lượng ngẫu nhiên X phân phối nhị thức $B(n,p)$ có thể coi như phân phối xấp xỉ chuẩn với kỳ vọng toán $\mu = np$ và phương sai $\sigma^2 = npq$. Từ đó ta có các công thức xấp xỉ Laplace như sau:

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

$$P(a \le X \le b) \approx \Phi\left(\frac{b-np}{\sqrt{npq}}\right) - \Phi\left(\frac{a-np}{\sqrt{npq}}\right)$$

Ở đó $f(x)$ là hàm mật độ của quy luật phân phối chuẩn hoá, và $\Phi(x)$ là tích phân Laplace:

$$f(x) = \frac{1}{\sqrt{2\pi}}.e^{-\frac{x^2}{2}} \quad \text{có giá trị được cho ở phụ lục C.1.}$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \quad \text{có giá trị được cho ở phụ lục C.2.}$$

•**Ví dụ 2.28** 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.
a) Tính xác suất để có 80 sản phẩm không được kiểm tra chất lượng.
b) Tính xác suất để có từ 70 đến 100 sản phẩm không được kiểm tra chất lượng.
c) 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 $B(n = 400; p = 0,2)$. Tuy nhiên vì n khá lớn $n = 400 > 5$ và

$$\frac{|p-q|}{\sqrt{npq}} = \frac{|0,2-0,8|}{\sqrt{400.0,2.0,8}} = 0,075 < 0,3$$

nên có thể coi X tuân theo quy luật phân phối chuẩn $N(\mu = np = 80; \sigma^2 = npq = 64)$.
a) Xác suất để có 80 sản phẩm không được kiểm tra chất lượng là:

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

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

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

c) Ta phải tìm ϵ sao cho $P(|X - np| < \epsilon) = 0,9545$. Ta có:

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

Do đó;

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

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

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

2.4.4. Quy luật Poisson $P(\lambda)$

Định nghĩa

⋆ **Định nghĩa 2.16** Đại lượng ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có $X = 0, 1, ..$ được gọi là phân phối theo quy luật Poisson với tham số $\lambda (\lambda > 0)$ nếu các xác suất tương ứng của nó được tính bởi công thức:

$$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k = 0, 1, ...$$

Quy luật Poisson được ký hiệu là $P(\lambda)$.

Các tham số đặc trưng của quy luật Poisson

a) Bảng phân phối xác suất của đại lượng ngẫu nhiên X phân phối theo quy luật Poisson có dạng:

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

Chú ý rằng:

$$e^{\lambda} = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \text{ và } \lambda e^{\lambda} + e^{\lambda} = \sum_{k=1}^{+\infty} \frac{k.\lambda^{k-1}}{(k-1)!}$$

ta có các kết quả sau đây:

b) Kỳ vọng toán của đại lượng ngẫu nhiên X phân phối $P(\lambda)$:

$$\begin{aligned} M(X) &= \sum_{k=0}^{+\infty} k.P(X=k) = \sum_{k=0}^{+\infty} k.\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}.\sum_{k=1}^{+\infty}\frac{\lambda^k}{(k-1)!} \\ &= \lambda.e^{-\lambda}.\sum_{k=1}^{+\infty}\frac{\lambda^{k-1}}{(k-1)!} = \lambda.e^{-\lambda}.e^{\lambda} = \lambda \end{aligned}$$

Tương tự ta cũng có:

$$\begin{aligned} M(X^2) &= \sum_{k=0}^{+\infty} k^2.P(X=k) = \sum_{k=0}^{+\infty} k^2.\frac{\lambda^k}{k!}e^{-\lambda} = e^{-\lambda}.\sum_{k=1}^{+\infty} k.\frac{\lambda^k}{(k-1)!} \\ &= \lambda.e^{-\lambda}.\sum_{k=1}^{+\infty}\frac{k.\lambda^{k-1}}{(k-1)!} = \lambda.e^{-\lambda}.(\lambda e^{\lambda} + e^{\lambda}) = \lambda^2 + \lambda \end{aligned}$$

c) Phương sai của đại lượng ngẫu nhiên X phân phối $P(\lambda)$:

$$D(X) = M(X^2) - [M(X)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

Như vậy trong quy luật Poisson, cả kỳ vọng toán và phương sai đều bằng λ, đó là tính chất đặc biệt của quy luật Poisson.

d) Giá trị của X tương ứng với xác suất lớn nhất là:

$$\lambda - 1 \le mod(X) \le \lambda.$$

Kết quả này được chứng minh tương tự phần công thức Bemoulli:

$$\begin{cases} P(X=k) & \ge P(X=k-1) \\ P(X=k) & \ge P(X=k+1) \end{cases} \Leftrightarrow \begin{cases} \dfrac{\lambda^k}{k!} & \ge \dfrac{\lambda^{k-1}}{(k-1)!} \\ \dfrac{\lambda^k}{k!} & \ge \dfrac{\lambda^{k+1}}{(k+1)!} \end{cases} \Leftrightarrow \begin{cases} \dfrac{\lambda}{k} & \ge 1 \\ 1 & \ge \dfrac{\lambda}{k+1} \end{cases} \Leftrightarrow \begin{cases} k & \le \lambda \\ k & \ge \lambda - 1 \end{cases}$$

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 $B(n,p)$. Trong trường hợp n khá lớn và p khá nhỏ $(p < 0,1)$, phân phối nhị thức được xấp xỉ bởi phân phối Poisson.

Thật vậy, giả sử tích np luôn luôn bằng một giá trị không đổi $np = \lambda$, khi đó xác suất của quy luật nhị thức có thể viết như sau:

$$\begin{aligned} p_k = P(X=k) = C_n^k p^k q^{n-k} &= \frac{n(n-1)(n-2)...(n-k+1)}{k!} p^k q^{n-k} \\ &= \frac{n(n-1)(n-2)...(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right)\frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k} \end{aligned}$$

Vì n lớn, do đó thay cho p_k ta tìm giới hạn lim $\lim\limits_{n\to\infty} p_k$:

$$\lim_{n\to\infty} p_k = \lim_{n\to\infty} \frac{\lambda^k}{k!}\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{\frac{-n}{\lambda}.\frac{\lambda(n-k)}{-n}} = \frac{\lambda^k}{k!}.e^{-\lambda}$$

Như vậy là trong trường hợp số phép thử n rất lớn, xác suất p rất nhỏ và tích $np = \lambda$ không đổi, các xác suất p_k của công thức Bernoulli có thể thay thế bằng công thức xấp xỉ Poisson.

Mặt khác, đối với quy luật Poisson thì quá trình hội tụ của nó về quy luật chuẩn sẽ diễn ra khi λ trở nên 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 2.29** 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 $B(n = 5000, p = 0,0002)$. Tuy nhiên vì $n = 5000$ rất lớn, $p = 0,0002$ khá nhỏ và tích $np = 5000.0,0002 = 1$ không đổi nên có thể coi X phân phối theo quy luật Poisson $P(\lambda = 1)$.
Xác suất để trong một phút có không quá 2 ống sợi bị đứt là:

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

2.4.5. Quy luật siêu bội $M(N, n)$

Ví dụ

•**Ví dụ 2.30** Giả sử trong bình có N quả cầu, trong đó có M quả cầu trắng và $N - M$ quả cầu đen. Lấy ngẫu nhiên lần lượt n quả cầu theo phương thức không hoàn lại. Gọi X là số cầu trắng được lấy ra. Khi đó các phép thử không còn độc lập và ta không thể tính xác suất bằng công thức Bemoulli. Theo định nghĩa cổ điển về xác suất, xác suất để trong n quả cầu lấy ra có đúng k cầu trắng bằng:

$$P(X = k) = p_k = \frac{C_M^k . C_{N-M}^{n-k}}{C_N^n}; \quad (0 \leq k \leq M; \ 0 \leq n - k \leq N - M).$$

Định nghĩa

⋆ **Định nghĩa 2.17** Đại lượng ngẫu nhiên rời rạc X nhận một trong các giá trị có thể có $X = 0, 1, 2, ...$ được gọi là phân phối theo quy luật siêu bội với các tham số $N, M, n (n \leq N; M \leq N)$ nếu xác suất tương ứng của nó được tính bởi công thức:

$$P(X = k) = p_k = \frac{C_M^k . C_{N-M}^{n-k}}{C_N^n}; \quad max(0; n - N + M) \leq k \leq min(M; n).$$

Quy luật siêu bội được ký hiệu là $M(N, n)$.

Các tham số đặc trưng của quy luật siêu bội

Đại lượng ngẫu nhiên X phân phối theo quy luật siêu bội $M(N, n)$ có kỳ vọng toán:

$$M(X) = n.\frac{M}{N} = np \quad \text{với} p = \frac{M}{N}$$

và phương sai:

$$D(X) = n.\frac{M}{N}.\frac{N-M}{N}.\frac{N-n}{N-1} = npq.\frac{N-n}{N-1}$$

Ta chú ý rằng khi số phép thử n là rất bé so với N thì phân phối siêu bội không khác so với phân phối nhị thức vì:

$$\lim_{\frac{n}{N} \to 0} \frac{C_M^k . C_{N-M}^{n-k}}{C_N^n} = C_n^k \left(\frac{M}{N}\right)^k . \left(1 - \frac{M}{N}\right)^{n-k}.$$

2.4.6. Quy luật khi - bình phương $\chi^2(n)$

⋆ **Định nghĩa 2.18** Đại lượng ngẫu nhiên liên tục X được gọi là tuân theo quy luật phân phối khi- bình phương với n bậc tự do, ký hiệu $\chi^2(n)$, nếu hàm mật độ xác suất của nó có dạng:

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

trong đó $\Gamma(x) = \int\limits_0^\infty t^{x-1} e^{-t} dt$ là hàm Gamma; $\Gamma(x) = (x-1)\Gamma(x-1)$.
Nếu n là một số nguyên thì $\Gamma(n) = (n-1)!$.

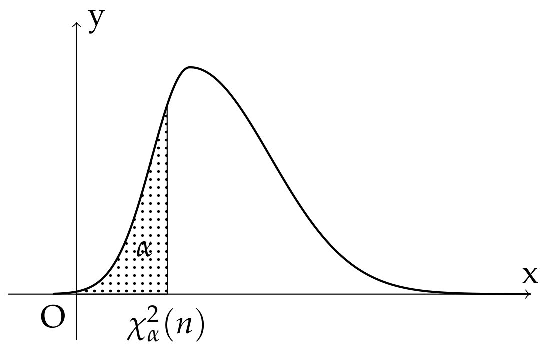
Nếu n là một số nguyên lẻ thì $\Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2} - 1\right) . \left(\frac{n}{2} - 2\right) ... \frac{3}{2} . \frac{1}{2} . \sqrt{\pi}$.

Có thể chứng minh được rằng nếu đại lượng ngẫu nhiên X phân phối theo quy luật khi - bình phương với n bậc tự do thì kỳ vọng toán: $M(X) = n$ và phương sai: $D(X) = 2n$.

Phân vị khi - bình phương mức α, ký hiệu là $\chi^2_\alpha(n)$ 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:

$$P[X < \chi^2_\alpha(n)] = \alpha$$

Giá trị của các phân vị này có thể tra ở phần phụ lục (xem Phụ lục C.5.), ý nghĩa của phân vi đươc thể hiên ở hình vẽ sau.



Hình 2.8:

Nếu có n đại lượng ngẫu nhiên độc lập $X_i, (i = \overline{1, n})$, cùng phân phối theo quy luật chuẩn hóa $N(0, 1)$ 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.

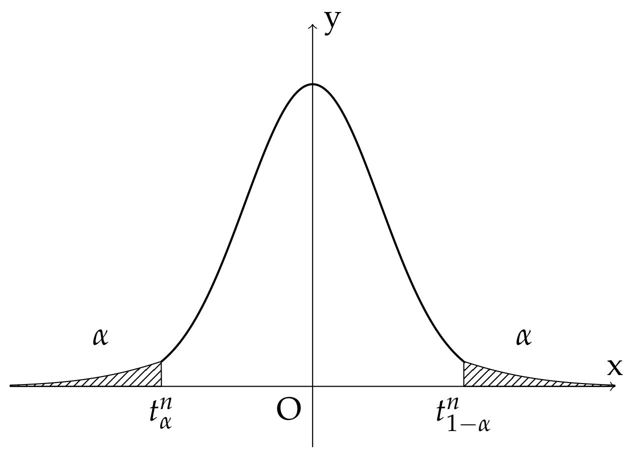
2.4.7. Quy luật Student $T(n)$

⋆ **Định nghĩa 2.19** Đại lượng ngẫu nhiên liên tục X được gọi là tuân theo quy luật phân phối Student với n bậc tự do, ký hiệu $T(n)$, nếu hàm mật độ xác suất của nó có dạng:

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

trong đó $\Gamma(x) = \int\limits_0^\infty t^{x-1}e^{-t}dt$ là hàm Gamma.

Đồ thị của hàm mật độ phân phối Student $T(n)$ có dạng như sau:



Hình 2.9:

Có thể chứng minh được rằng nếu đại lượng ngẫu nhiên X tuân theo quy luật phân phối Student với n bậc tự do thì kỳ vọng toán: $M(X) = 0$ và phương sai: $D(X) = \dfrac{n}{n-2}$.

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

$$P(X < t_\alpha^n) = \alpha$$

Giá trị của phân vị Student được cho trong phần phụ lục (xem Phụ lục C.4.), ý nghĩa của nó được thể hiện trên hình vẽ. Dễ thấy:

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

Giả sử U là đại lượng ngẫu nhiên phân phối chuẩn hoá $N(0,1)$ 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:

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

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

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

Bài tập chương 2

Bài 2.1. Một thiết bị gồm 3 bộ phận hoạt động độc lập với nhau. Xác suất trong thời gian t các bộ phận bị hỏng lần lượt là 0,4; 0,2 và 0,3.
a) Tìm quy luật phân phối xác suất của số bộ phận bị hỏng X.
b) Thiết lập hàm phân phối xác suất của X.
c) Tính xác suất trong thời gian t có không quá 2 bộ phận hỏng.
d) Tìm mod(X).
ĐS: a)

X	0	1	2	3
P	0,336	0,452	0,188	0,024

c)0,976
d)1
Bài 2.2. Một hộp có 3 quả cầu trắng và 2 quả cầu đen. Lấy ngẫu nhiên lần lượt từng quả cầu cho đến khi lấy được cầu trắng. Tìm quy luật phân phối xác suất của số cầu được lấy ra.
ĐS: **Bài 2.3.** Có hai lô sản phẩm, lồ 1 có 8 chính phẩm và 2 phế phẩm; lô 2 có 7 chính phẩm và

X	1	2	3
P	0,6	0,3	0,1

3 phế phẩm. Từ lô thứ nhất lấy ngẫu nhiên ra 2 sản phẩm bỏ sang lô thứ hai, sau đó từ lô thứ 2 lấy ra 2 sản phẩm.
a) Tìm quy luật phân phối xác suất của số chính phẩm được lấy ra.
b) Xây dựng hàm phân phối xác suất của số chính phẩm được lấy ra.
ĐS:

X	0	1	2
P	0,0639	0,4388	0,4973

Bài 2.4. Bảng phân phối xác suất của biến ngẫu nhiên X như sau:
a) Tính $M(X); D(X)$ và σ_X.
b) Tìm $mod(X)$.

X	-5	2	3	4
P	0,4	0,3	0,1	0,2

ĐS: a) 0,3; 15,21; 3,9; b) -5.

Bài 2.5. 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
P	0,3	0,5	0,2

X_2	1	4
P	0,2	0,8

a) Tính $M(X_1), M(X_2), D(X_1), D(X2)$;
b) Tính $M(X_1 + X_2), D(X_1 + X_2)$
c) Tính $M(2X_l - 3X_2), D(2X_l - 3X_2), M(X_1^2 - X_2^3)$.
ĐS: a) 3,1; 3,4; 1,09; 1,44; b) 6,5; 2,53; c) -4; 17,32; -40,7.

Bài 2.6. Đại lượng ngẫu nhiên X có kỳ vọng toán $M(X) = a$ và $D(X) = b^2$.
a) Tính $M(X - M(X))$;
b) Tìm kỳ vọng toán và phương sai của đại lượng ngẫu nhiên $U = \dfrac{X - a}{b}$.
ĐS: a) 0; b) 0; 1.

Bài 2.7. Thực hiện ba lần bắn với xác suất lần bắn thứ nhất, thứ hai và thứ ba trúng đích lần lượt là $p_1 = 0,4; p_2 = 0,3; p_3 = 0,6$. Tìm kỳ vọng toán của số lần bắn trúng đích.
ĐS: 1,3.

Bài 2.8. Đại lượng ngẫu nhiên rời rạc X nhận 3 giá trị có thể có là $x_1 = 4$với xác suất $p_1 = 0,5; X_2 = 0,6$ với xác suất $P_2 = 0,3$ và x_3 với xác suất p_3. Tìm x_3 và p_3 biết $M(X) = 8$.
ĐS: 29,1; 0,2.

Bài 2.9. Đại lượng ngẫu nhiên rời rạc X nhận các giá trị có thể có là $x_1 = -1, x_2 = 0, x_3 = 1$. Tìm các xác suất tương ứng p_1, p_2 và p_3 biết rằng $M(X) = 0,1$ và $D(X) = 0,89$.
ĐS: 0,4; 0,1; 0,5.

Bài 2.10. 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 ngừng trên đường là bao nhiêu, biết rằng mỗi khi gặp đèn đỏ người ấy phải đợi khoảng 2 phút.
ĐS: 2,2.

Bài 2.11. Đại lượng ngẫu nhiên X có bảng phân phối xác suất như sau:

X	1	4	8
P	0,3	0,1	0,6

a) Tính P(| X- M(X) | < 4)
b) Tính xác suất để trong 5 phép thử có ít nhất một lần $X < M(X)$.
ĐS: 0,7; 0,92224.

Bài 2.12. Cho X, Y là hai đại lượng ngẫu nhiên độc lập có bảng phân phối xác suất:
a) Tìm x, y, z, t biết $M(X + Y) = 1,3; M(X^2 + Y^2) = 3,3$;

X	-1	0	2
P	x	0,2	y

Y	-1	1
P	z	t

b) Lập bảng phân phối xác suất của $X^2 - 2X, XY, |X - Y|$;

c) 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.

ĐS: a) 0,3; 0,5; 0,2; 0,8; c) p = 0,175;$\sigma_{min} = 0,726$.

Bài 2.13. Cho đại lượng ngẫu nhiên rời rạc X nhận các giá trị $0, 1, 2, 3, ...$ với xác suất tương ứng:

$$P[X = n] = \frac{3}{4}q^n$$

a) Tìm q;

b) Tính xác suất trong 3 phép thử độc lập có lần $X < 3$;

c) Tính kỳ vọng toán $M(X)$ và phương sai $D(X)$.

ĐS: a) 0,25; b) $1 - 0,25^9$; c)$\frac{1}{3}; \frac{4}{9}$

Bài 2.14. 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):

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

a) Tìm hệ số k,

b) Tìm xác suất để nhu cầu về loại hàng đó không vượt quá 12 ngàn sản phẩm trong một năm.

c) Tìm nhu cầu trung bình hàng năm về loại hàng đó.

ĐS: a) $\frac{1}{450}$; b) 0,64; c) 10.

Bài 2.15. 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)

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

a) Tìm hệ số a.

b) Tìm thời gian xếp hàng trung bình.

c) Tìm xác suất để trong 3 người xếp hàng thì có không quá 2 người phải chờ quá 0,5 phút.

ĐS: a) 2; b) 0,5; c) 0,875.

Bài 2.16. Cho X là đại lượng ngẫu nhiên liên tục có hàm phân phối xác suất sau đây:

$$F(x) = a + b.arctanx$$

a) Tìm các hệ số a, b.
b) Tìm $P(0 < X < l)$.
c) Tìm giá trị x_1 thoả mãn điều kiện $P(X > x_1) = \frac{3}{4}$.
d) Tìm hàm mật độ xác suất của X.
ĐS: a)$\frac{1}{2}$; $\frac{1}{\pi}$b)0,25;c)-l;d)$\frac{1}{\pi(1+x^2)}$
Bài 2.17. Hàm phân phối xác suất của đại lượng ngẫu nhiên liên tục X có dạng:

$$F(x) = \begin{cases} 0 & \text{với } x \le 0 \\ k.sinx & \text{với } 0 < x \le \frac{\pi}{4} \\ 1 & \text{với } x > \frac{\pi}{4} \end{cases}$$

a) Tìm hệ số a.
b) Tìm hàm mật độ f(x).
c) Tìm $M(X), D(X)$.
ĐS: a) $\sqrt{2}$; c) 0,371; 0,0499.
Bài 2.18. Đại lượng ngẫu nhiên X có hàm mật độ xác suất:

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

a) Tìm hệ số a,
b) Tìm $P(0 \le X < \frac{\pi}{4})$.
c) Tìm xác suất để trong 3 phép thử có lần $X \in \left[\frac{\pi}{3}, \frac{2\pi}{3}\right]$.

ĐS: a) 0,5; b) 0,354; c) 0,1878.
Bài 2.19. Thu nhập của dân cư một vùng là đại lượng ngẫu nhiên liên tục có hàm phân phối xác suất như sau:

$$F(x) = \begin{cases} 1 - \left(\frac{x_0}{x}\right)^{\alpha} & \text{với } x \ge x_0 \\ 0 & \text{với } x < x_0 \end{cases}$$

Hãy xác định mức thu nhập sao cho lấy ngẫu nhiên một người ở vùng đó thì thu nhập của người này vượt quá mức trên với xác suất 0,5.
ĐS: $2^{\frac{1}{\alpha}} x_0$
Bài 2.20. Tỷ lệ mắc một loại bệnh trong một vùng dân cư là biến ngẫu nhiên liên tục có hàm

mật độ:

$$f(x) = \begin{cases} \dfrac{1}{20} & \text{với } x \in (5;25) \\ 0 & \text{với } x \notin (5;25) \end{cases}$$

a) Tính $P(|X - 10| > 2,5)$.

b) Tính tỷ lệ mắc bệnh trung bình và phương sai.

ĐS: a) 0,75; b) 15; 33,3.

Bài 2.21. Đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất:

$$f(x) = \frac{k}{e^x + e^{-x}} \quad (-\infty < x < \infty)$$

ĐS: $\dfrac{2}{\pi}$

Bài 2.22. Hai kiện tướng bóng bàn ngang sức thi đấu với nhau. Hỏi thắng 2 trong 4 ván dễ hơn hay thắng 3 trong 6 ván dễ hơn?

ĐS: Thắng 2 trong 4 ván dễ hơn.

Bài 2.23. 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.

a) Tính xác suất trong khoảng thời gian t có 4 máy cần đến sự chăm sóc của nữ công nhân.

b) Tính xác suất trong khoảng thời gian t có từ 3 đến 6 máy cần đến sự chăm sóc của nữ công nhân.

c) Tính số máy có khả năng cần đến sự chăm sóc của nữ công nhân nhiều nhất.

ĐS: a) 0,231; b) 0,709; c) 4 máy.

Bài 2.24. 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.

ĐS: 13 sản phẩm.

Bài 2.25. Tiến hành kiểm tra chất lượng 900 chi tiết, xác suất để một chi tiết đạt tiêu chuẩn là 0,9.

a) Tính xác suất để có 810 chi tiết đạt tiêu chuẩn.

b) Tính xác suất để có từ 800 đến 825 chi tiết đạt tiêu chuẩn.

c) 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?

ĐS: a) 0,0443; b) 0,81904; c) (792; 828).

Bài 2.26. Kích thước chi tiết do một máy sản xuất ra là đại lượng ngẫu nhiên phân phối chuẩn với $\mu = 5cm$ và $\sigma_X = 8,1mm$. a) Tìm xác suất để lấy ngẫu nhiên 1 chi tiết có kích thước từ 4cm đến 7cm.

b) Tim xác suất để lấy ngẫu nhiên 1 chi tiết có kích thước sai lệch so với kỳ vọng toán không quá 2cm.

c) Tính xác suất để khi lấy ngẫu nhiên 5 chi tiết có 2 chị tiết có kích thước lớn hơn 5cm.

ĐS: a) 0,88389; b) 0,98648; c) 0,3125.

Bài 2.27. Kích thước chi tiết là biến ngẫu nhiên phân phối chuẩn với $\mu = 50cm$. Kích thước thực tế của các chi tiết không nhỏ hơn 32cm và không lớn hơn 67cm. Tìm xác suất để lấy ngẫu nhiên 2 chi tiết có chi tiết có kích thước nhỏ hơn 40cm.

ĐS: 0,09.

Bài 2.28. Máy bay bay dọc theo cầu dài 8m, rộng 4m và ném 2 quả bom. Biết rằng khoảng cách từ điểm rơi của quả bom đến trục đối xứng theo chiều dọc và chiều ngang của cầu là các biến ngẫu nhiên độc lập, cùng phân phối chuẩn với kỳ vọng toán bằng 0 và độ lệch chuẩn tương ứng là 4m và 6m. Tìm xác suất để cầu bị trúng bom.

ĐS: 0,3439.

Bài 2.29. Khi thâm nhập một thị trường mới, doanh nghiệp chỉ dự kiến rằng doanh số hàng tháng có thể đạt được phân phối chuẩn. Xác suất để đạt được doanh số nhỏ hơn 20 triệu là 0,1; trên 50 triệu là 0,2.

a) Tính doanh số hàng tháng trung bình của doanh nghiệp đó.

b) Tính xác suất để doanh nghiệp đạt được doanh số ít nhất là 40 triệu hàng tháng.

ĐS: a) 38,113; b) 0,44828

Bài 2.30. Trọng lượng một loại sản phẩm là biến ngẫu nhiên phân phối chuẩn với trung bình 25kg và độ lệch chuẩn 0,2kg.

a) Tính tỷ lệ sản phẩm có trọng lượng lớn hơn 25,3kg.

b) Tính xác suất để khi cân thử 25 sản phẩm thì trọng lượng trung bình của 25 sản phẩm này nhỏ hơn 25,01kg.

ĐS: a) 0,06681; b) 0,09871.

Bài 2.31. Theo thống kê dân số thì một người ở độ tuổi 40 sẽ sống thêm 1 năm nữa là 0,995. Một công ty bảo hiểm nhân thọ bán bảo hiểm 1 năm cho những người ở độ tuổi đó với giá là 10 ngàn và trong trường hợp người mua bảo hiểm bị chết thì số tiền bồi thường là 1 triệu. Hỏi lợi nhuận trung bình của công ty khi bán mỗi thẻ bảo hiểm loại này là bao nhiêu?

ĐS: 5 ngàn.

Bài 2.32. 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 45000đ, bán ra 60000đ. Giả sử cửa hàng nhập 70kg thực phẩm, hãy lập bảng phân phối xác suất của lượng sản phẩm được bán ra. Tính số tiền lãi trung bình cửa hàng thu được khi đó;

c) Nếu bị ế thì cửa hàng giảm giá bán ra chỉ 30000đ thì sẽ hết hàng. Hỏi cửa hàng nên nhập 70kg hay 80kg thì tốt hơn?

ĐS: a) 68,9; b) 720000; c) nên nhập 70kg.

Bài 2.33. Một chi tiết do máy tiện sản xuất ra là một đại lượng ngẫu nhiên $X \sim N(20mm, 4mm^2)$. Chi tiết được gọi là hợp quy cách nếu X dao động từ 18mm đến 22mm. Cho máy đó sản xuất 100 chi tiết, tính xác suất để có:
a) 50 chi tiết máy hớp quy cách;
b) Không quá 80 chi tiết máy hợp quy cách.
ĐS: a) 0,105; b) 0,994
Bài 2.34. Thời gian bảo hành một sản phẩm được quy định là 3 năm. Nếu bán được 1 sản phẩm thì cửa hàng 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 tỷ lệ sản phẩm phải bảo hành không quá 10% thì phải quy định thời gian bảo hành là bao nhiêu?
ĐS: a) 24,285 ngàn; b) 2,688 năm; c) 1,896 năm.
Bài 2.35. Một người cân nhắc giữa 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 tỷ lệ bao nhiêu để mức độ rủi ro về lãi suất là nhỏ nhất?
ĐS: a) Công ty A; b) A:B = 0,297:0,703.

Chương 3

Mẫu ngẫu nhiên - Ước lượng tham số

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, ký 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, ký hiệu là χ, các dấu hiệu này có thể là định tính hoặc định lượng.

•**Ví dụ 3.1** Muốn nghiên cứu tình hình đi làm thêm của sinh viên các trường Cao đẳng, Đại học trên toàn quốc ta tiến hành tổng điều tra sinh viên toàn quốc và phân tích số liệu theo số giờ làm thêm hàng tuần. Khi đó, tổng thể nghiên cứu là tập hợp tất cả sinh viên toàn quốc, dấu hiệu nghiên cứu X là thời gian làm thêm hàng tuần của sinh viên. Kích thước tổng thể N là số sinh viên của Việt Nam.

3.1.2. Các phương pháp mô tả tổng thể

Giả sử tổng thể kích thước N, dấu hiệu nghiên cứu định lượng χ nhận các giá trị $x_1, x_2, ..., x_k$ với các tần số tương ứng $N_1, N_2, ..., N_k$ (trong tổng thể có N_i từ mà dấu hiệu nghiên cứu χ nhận giá trị bằng x_i).

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	...	x_i	...	x_k
Tần số	N_1	N_2	...	N_i	...	N_k

Hiển nhiên:

$$\begin{cases} 0 \le N_i \le N, & \forall_i = \overline{1,k} \\ \sum_{i=1}^{k} N_i = N \end{cases}$$

b) Bảng phân phối tần suất.
Ký hiệu p_i, $\left(i = \overline{1,k}\right)$ là tần suất của x_i:

$$p_i = \frac{N_i}{N} \quad i = \overline{1,k}$$

khi đó tổng thể còn được mô tả bởi bảng phân phối tần suất như sau:

Giá trị của χ	x_1	x_2	...	x_i	...	x_k
Tần số	p_1	p_2	...	p_i	...	p_k

Dễ thấy:

$$\begin{cases} 0 \le p_i \le 1 & \forall_i = \overline{1,k} \\ \sum_{i=1}^{k} p_i = 1 \end{cases}$$

c) Tần số tích luỹ - Tần suất tích luỹ.
Ký hiệu w_i $\left(i = \overline{1,k}\right)$ là tần số tích luỹ của x_i, tức là tổng số phần tử có giá trị nhỏ hơn x_i:

$$w_i = \sum_{x_j < x_i} = N_j$$

và $F(x_i)$ $\left(i = \overline{1,k}\right)$ là tần số tích luỹ của x_i:

$$F(x_i) = \frac{w_i}{N} = \sum_{x_j < x_i} p_i.$$

Việc mô tả tổng thể theo bảng phân phối tần suất và tần suất tích luỹ cho thấy dấu hiệu định lượng χ có thể mô hình hóa bằng một đại lượng ngẫu nhiên rời rạc X. Điều này cũng đúng đối với tổng thể mang dấu hiệu nghiên cứu χ phân phối liên tục. Đại lượng ngẫu nhiên X dùng để mô hình hóa dấu hiệu nghiên cứu χ được gọi đại lượng ngẫu nhiên gốc, còn 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 $x_1, x_2, ..., x_N$.

Trung bình tổng thể, ký hiệu là m, là trung bình số học của các giá trị của χ trong tổng thể:

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

Nếu tổng thể được mô tả bởi bảng phân phối tần số ở phần (3.1.2 a) 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.
Giả sử trong tổng thể kích thước N dấu hiệu định lượng χ nhận các giá trị $x_1, x_2, ..., x_N$. Phương sai tổng thể, ký 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 (3.1.2 a) thì phương sai tổng thể còn được tính bởi:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2 N_i.$$

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 công thức tính phương sai của đại lượng ngẫu nhiên, ta toàn có thể tính phương sai tổng thể bằng công thức:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} x_i^2 - \left(\frac{1}{N}\sum_{i=1}^{N} x_i\right)^2 \quad \text{hay} \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{k} x_i^2 N_i - \left(\frac{1}{N}\sum_{i=1}^{k} x_i N_i\right)^2.$$

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 tổng thể, ký hiệu p, là tỷ lệ phần tử mang dấu hiệu nghiên cứu trong tổng thể:

$$p = \frac{M}{N}.$$

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.

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 χ 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 của n đại lượng ngẫu nhiên độc lập $X_1, X_2, ..., X_n$ 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 ký hiệu là $W = (X_1, X_2, ..., X_n)$.

Khi thực hiện một phép thử đối với mẫu ngẫu nhiên W ta sẽ thu được kết quả cụ thể: X_1 nhận giá trị x_1, X_2 nhận giá trị x_2, ..., X_n nhận giá trị x_n. Tập hợp n giá trị $x_1, x_2, ..., x_n$ tạo thành một giá trị của mẫu ngẫu nhiên, hay còn gọi là một mẫu cụ thể, ký hiệu là:

$$w = (x_1, x_2, ..., x_n).$$

•**Ví dụ 3.2** Trong 1 cửa hàng băng đĩa, trên 1 kệ chứa các đĩa nhạc có các loại đĩa tương ứng với số tiền như sau:

Giá (ngàn đồng)	20	25	30	34	40
Số đĩa	35	10	25	17	13

Lấy ngẫu nhiên 1 đĩa nhạc trong giá.
Gọi X là giá tiền của đĩa nhạc này. X = $\{20; 25; 30; 34; 40\}$
Ta thấy X có quy luật phân phối xác suất như sau:

X	20	25	30	34	40
P	0,35	0,1	0,25	0,17	0,13

X đặc trưng cho dấu hiệu nghiên cứu là giá tiền của tổng thể là số đĩa trên giá.
Lấy ngẫu nhiên (có hoàn lại) 4 đĩa nhạc từ kệ.
Gọi X_i là giá tiền của đĩa nhạc thứ i lấy được, $i = 1 \div 4$
Ta thấy các X_i độc lập và có cùng quy luật phân phối xác suất giống như X.
Lập $W = (X_1, X_2, X_3, X_4)$, gọi là mẫu ngẫu nhiên.
Bây giờ ta xét xem giá cụ thể của từng đĩa lấy ra thấy như sau:

Đĩa 1: giá 20 ngàn đồng.
Đĩa 2: giá 30 ngàn đồng.
Đĩa 3: giá 20 ngàn đồng.
Đĩa 4: giá 40 ngàn đồng.
Lập $w = (x_1, x_2, x_3, x_4) = (20, 30, 20, 40)$ gọi là mẫu cụ thể.

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 mẫu ngẫu nhiên cụ thể: $w = (x_1, x_2, ..., x_n)$. Giả sử trong mẫu cụ thể đó, giá trị x_1 xuất hiện n_1 lần, giá trị x_2 xuất hiện n_2 lần,..., giá trị x_k xuất hiện n_k lần, 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	...	x_i	...	x_k
n_i	n_1	n_2	...	n_i	...	n_k

Hiển nhiên:

$$\begin{cases} 0 \leq n_i \leq n, & \forall_i = \overline{1,k} \\ \sum\limits_{i=1}^{k} n_i = n \end{cases}$$

b) Bảng phân phối suất số thực nghiệm (đơn).
Ký hiệu f_i, $\left(i = \overline{1,k}\right)$ là tần suất xuất hiện giá trị x_i trong mẫu:

$$f_i = \frac{n_i}{n} \qquad i = \overline{1,k}$$

khi đó mẫu ngẫu nhiên cụ thể còn được mô tả bởi bảng phân phối tần suất như sau:

x_i	x_1	x_2	...	x_i	...	x_k
f_i	f_1	f_2	...	f_i	...	f_k

Dễ thấy:

$$\begin{cases} 0 \leq f_i \leq 1, & \forall_i = \overline{1,k} \\ \sum\limits_{i=1}^{k} f_i = 1 \end{cases}$$

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:

x_i	$x_0 - x_1$	$x_1 - x_2$	...	$x_{i-1} - x_i$	...	$x_{k-1} - x_k$
n_i	n_1	n_2	...	n_i	...	n_k

Trong bảng trên, 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 $[x_1, x_2)$, ..., có n_k phần tử nhận giá trị trong khoảng $[x_{k-1}, x_k)$. Hiển nhiên:

$$\begin{cases} 0 \le n_i \le n, & \forall_i = \overline{1,k} \\ \sum\limits_{i=1}^{k} n_i = n \end{cases}$$

Tương tự ta cũng có bảng phân phối tần suất ghép lớp như sau:

x_i	$x_0 - x_1$	$x_1 - x_2$	...	$x_{i-1} - x_i$	...	$x_{k-1} - x_k$
f_i	f_1	f_2	...	f_i	...	f_k

Tuy nhiên nếu chú ý rằng đại lượng ngẫu nhiên gốc X phân phối liên tục thì có thể coi trong mẫu ngẫu nhiên cụ thể, trường hợp dấu hiệu nghiên cứu nhận giá trị đúng bằng x_i không xảy ra.

d) Hàm phân phối thực nghiệm.

Ký hiệu $w_i \left(i = \overline{1,k}\right)$ là tần số tích luỹ của x_i, tức là tổng số phần tử có giá trị nhỏ hơn x_i trong mẫu cụ thể w:

$$w_i = \sum_{x_j < x_i} = n_j$$

và $F^*(x_i) \left(i = \overline{1,k}\right)$ là tần số tích luỹ của x_i:

$$F^*(x_i) = \frac{w_i}{n} = \sum_{x_j < x_i} f_i.$$

Khi đó $F^*(x)$ 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 ngẫu nhiên cụ thể được mô tả bởi các bảng phân phối tần số thực nghiệm và tần suất thực nghiệm ở phần (3.2.2. a,b).

- 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), ..., trên mặt phẳng.
- Đa giác tần suất là một đường gấp khúc mà các đoạn thẳng của nó nối các điểm (x_1, f_1), (x_2, f_2), ..., (x_k, f_k) trên mặt phẳng.

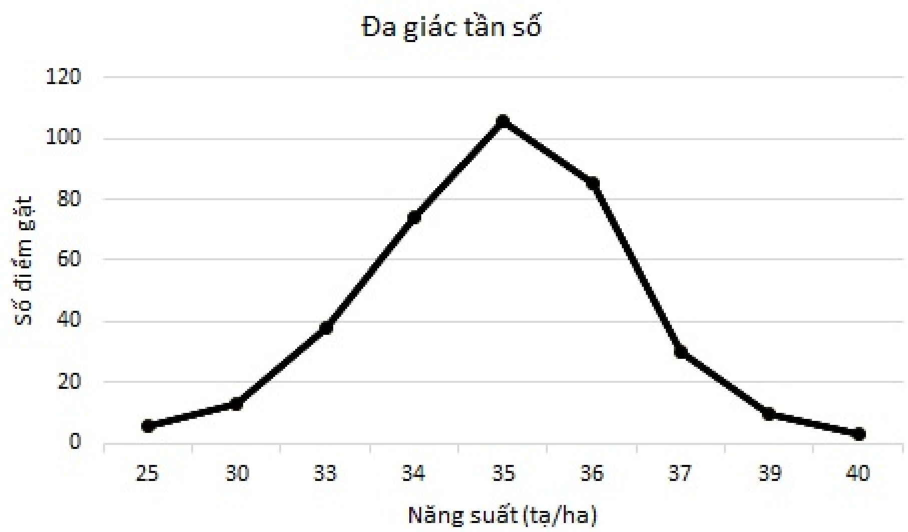
•**Ví dụ 3.3** Gặt ngẫu nhiên 365 điểm trồng lúa của một huyện ta được các số liệu được sắp xếp thành bảng sau:

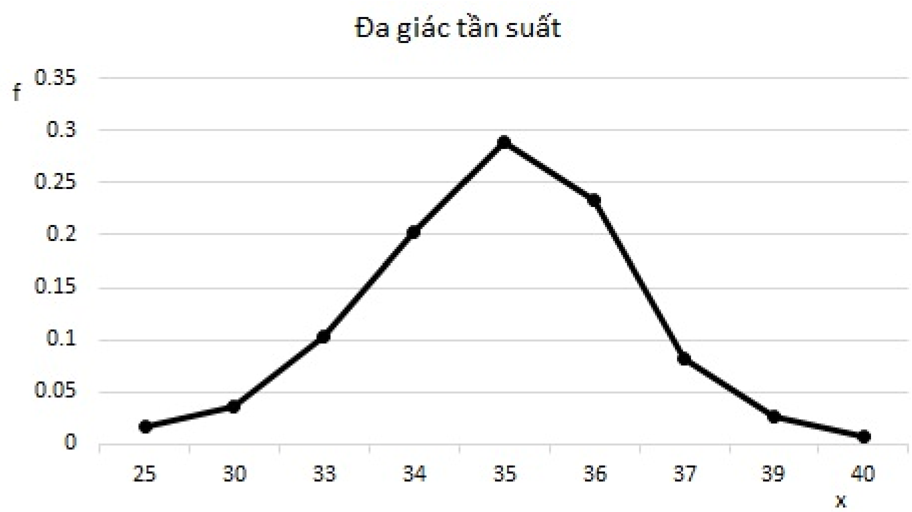
Năng suất (tạ/ha)	25	30	33	34	35	36	37	39	40
Số điểm gặt	6	13	38	74	106	85	30	10	3

Như vậy giá trị của mẫu đã được mô tả dưới dạng bảng phân phối tần số thực nghiệm. Còn bảng phân phối tần suất thực nghiệm có dạng:

x_i	25	30	33	34	35	36	37	39	40
f_i	0,016	0,036	0,104	0,203	0,290	0,233	0,082	0,027	0,008

Đa giác tần số và đa giác tần suất của phân phối thực nghiệm này như sau:





b) Biểu đồ tần số.

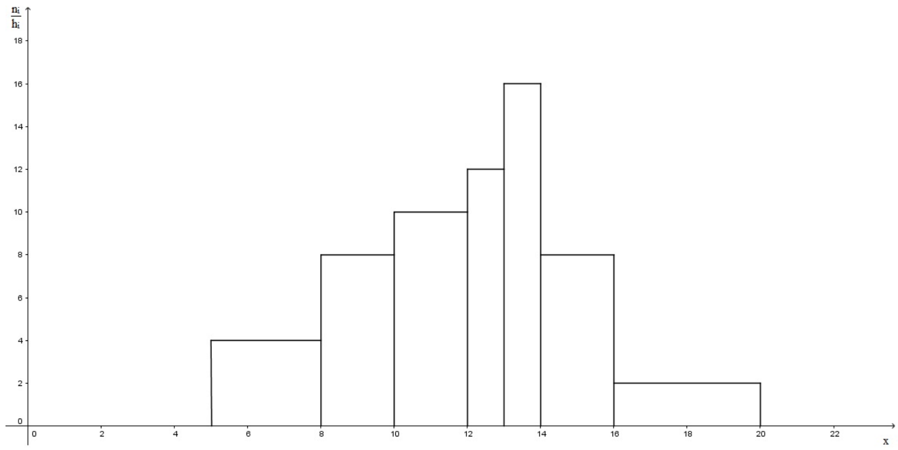
Xét mẫu ngẫu nhiên cụ thể được mô tả bởi các bảng phân phối tần số thực nghiệm và tần suất thực nghiệm ở phần (3.2.2. c).

Để vẽ biểu đồ tần số ta chia đoạn $[x_o, x_k]$ chứa tất cả các giá trị quan sát được của mẫu thành các đoạn nhỏ ứng với $[x_{i-1}, x_i]$. Biểu đồ tần suất là một bậc hình thang tạo nên bởi nhiều 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. Vì vậy tổng diện tích của tất cả các hình chữ nhật sẽ bằng kích thước mẫu n.

Tương tự biểu đồ tần suất là một hình bậc thang được tạo nên bởi các hình chữ nhật có đáy bằng $h_i = x_i - x_{i-1}$ và chiều cao bằng $\frac{f_i}{h_i}$. Lúc đó diện tích hình chữ nhật thứ i bằng f_i và tổng diện tích cả biểu đồ bằng 1.

•**Ví dụ 3.4** Vẽ biểu đồ tần số của 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



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 $W = (X_1, X_2, ..., X_n)$. Thống kê là một hàm của các giá trị $X_1, X_2, ..., X_n$, ký hiệu là G. Như vậy:

$$G = f\left(X_1, X_2, ..., X_n\right).$$

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 trung bình $M(G)$, phương sai $D(G)$...

Khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w = (x_1, x_2, ..., x_n)$ thì thống kê G cũng nhận một giá trị cụ thể là $g = f\left(x_1, x_2, ..., x_n\right)$.

3.3.2. Trung bình mẫu

a) Khái niệm. Giả sử từ đại lượng ngẫu nhiên gốc X trong tổng thể lập mẫu ngẫu nhiên kích thước $n : W = (X_1, X_2, ..., X_n)$. Trung bình mẫu là một thống kê, ký hiệu là $\overline{X}$ và là trung bình số học của các giá trị của mẫu:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

b) Tính chất
Nếu đại lượng ngẫu nhiên gốc X có kỳ vọng toán $M(X) = m$ và phương sai $D(X) = \sigma^2$ thì:

$$M(\overline{X}) = m; \qquad D(\overline{X}) = \frac{\sigma^2}{n}.$$

Thật vậy, áp dụng các tính chất của kỳ vọng toán ta có:

$$M(\overline{X}) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{1}{n}.n.m = m.$$

Áp dụng các tính chất của phương sai với lưu ý X_i là các đại lượng ngẫu nhiên độc lập ta có:

$$D(\overline{X}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) = \frac{1}{n^2}.n.\sigma^2 = \frac{\sigma^2}{n}.$$

c) Tính giá trị cụ thể.
Khi mẫu ngẫu nhiên nhận một giá trị cụ thể $w = (x_1, x_2, ..., x_n)$ thì trung bình mẫu cũng nhận giá trị cụ thể bằng:

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

Nếu mẫu ngẫu nhiên được cho bởi mảng phân phối tần số:

x_i	x_1 x_2 ... x_i ... x_k
n_i	n_1 n_2 ... n_i ... n_k

$\left(\sum_{i=1}^{k} n_i = n\right)$

thì giá trị cụ thể của trung bình mẫu được tính bởi:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} n_i x_i.$$

3.3.3. Phương sai mẫu

a) Khái niệm. Giả sử từ đại lượng ngẫu nhiên gốc X trong tổng thể lập mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$. Phương sai mẫu là một thống kê, ký hiệu S^2, và là trung bình số học của bình phương các sai lệch giữa các giá trị của mẫu và trung bình mẫu:

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

Ngoài ra người ta còn dùng hai loại phương sai sau đây:

- *Phương sai điều chỉnh mẫu*, ký hiệu là S'^2 và được xác định bằng công thức:

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

- *Phương sai* S^{*2} được xác định bằng biểu thức:

$$S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$$

trong đó m là trung bình tổng thể.

b) Tính chất.

- Phương sai mẫu có tính chất sau: Nếu đại lượng ngẫu nhiên gốc có kỳ vọng toán $M(X) = m$ và phương sai $D(X) = \sigma^2$ thì phương sai S^2 có kỳ vọng toán:

$$M\left(S^2\right) = \frac{n-1}{n}\sigma^2.$$

Thật vậy:

$$\begin{aligned}
S^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - m) - \left(\overline{X} - m\right)\right]^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - m)^2 - 2\left(X_i - m\right)\left(\overline{X} - m\right) + \left(\overline{X} - m\right)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2 - 2\left(\overline{X} - m\right)\frac{1}{n}\sum_{i=1}^{n}(X_i - m) + \left(\overline{X} - m\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2 - \left(\overline{X} - m\right)^2 \\
\Rightarrow M(S)^2 &= \frac{1}{n}\sum_{i=1}^{n}M(X_i - m)^2 - M\left(\overline{X} - m\right)^2.
\end{aligned}$$

Mặt khác:

$$M(X_i - m)^2 = D(X_i) = D(X) = \sigma^2$$

$$M\left(\overline{X} - m\right)^2 = D(\overline{X}) = \frac{\sigma^2}{n}.$$

Do đó:

$$M\left(S^2\right) = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2.$$

- Giữa phương sai mẫu và phương sai điều chỉnh mẫu có mối liên hệ:

$$S'^2 = \frac{n}{n-1}S^2.$$

Do đó từ tính chất của phương sai mẫu S^2 có thể suy ra tính chất của S'^2 như sau:

$$M\left(S'^2\right) = \sigma^2.$$

Thật vậy:

$$M\left(S'^2\right) = M\left(\frac{n}{n-1}S^2\right) = \frac{n}{n-1}M\left(S^2\right) = \frac{n}{n-1}\frac{n-1}{n}\sigma^2 = \sigma^2.$$

- Đối với phương sai S^{*2} ta cũng có tính chất tương tự sau:

$$M\left(S^{*2}\right) = \sigma^2.$$

Thật vậy:

$$M\left(S^{*2}\right) = M\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - m\right)^2\right] = \frac{1}{n}\sum_{i=1}^{n}M\left(X_i - m\right)^2 = \frac{1}{n}\sum_{i=1}^{n}D\left(X_i\right) = \sigma^2.$$

c) Tính giá trị cụ thể.

- Trên một mẫu cụ thể $w = (x_1, x_2, ..., x_n)$, phương sai mẫu nhận một giá trị cụ thể bằng:

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

Nếu mẫu ngẫu nhiên được cho bởi bảng phân phối tần số:

x_i	x_1 x_2 ... x_i ... x_k
n_i	n_1 n_2 ... n_i ... n_k

$\left(\sum_{i=1}^{k} n_i = n\right)$

thì giá trị cụ thể của phương sai mẫu được tính bởi:

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

Trên thực tế, để tiện cho việc tính toán người ta sử dụng công thức sau để tính giá trị của phương sai mẫu:

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

Thật vậy, ta có:

$$\begin{aligned} s^2 &= \frac{1}{n}\sum_{i=1}^{k}n_i\left(x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k}n_i\left(x_i^2 - 2x_i\overline{x} + \overline{x}^2\right) \\ &= \frac{1}{n}\sum_{i=1}^{k}n_i x_i^2 - 2\overline{x}\cdot\frac{1}{n}\sum_{i=1}^{k}n_i x_i + \overline{x}^2 = \frac{1}{n}\sum_{i=1}^{k}n_i x_i^2 - \overline{x}^2. \end{aligned}$$

- Giá trị cụ thể của phương sai điều chỉnh mẫu được tính bởi:

$$s'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \qquad \text{hoặc} \qquad s'^2 = \frac{1}{n-1}\sum_{i=1}^{n}n_i\left(x_i - \overline{x}\right)^2.$$

Giá trị cụ thể s'^2 có thể tính thông qua giá trị s^2 như sau:

$$s'^2 = \frac{n}{n-1}s^2 = \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2\right).$$

- Giá trị cụ thể của phương sai S^{*2} là:

$$s^{*2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 \qquad \text{hoặc} \qquad s^{*2} = \frac{1}{n}\sum_{i=1}^{n} n_i (x_i - m)^2.$$

3.3.4. Độ lệch tiêu chuẩn mẫu

Nếu lấy căn bậc hai của phương sai mẫu S^2, của phương sai điều chỉnh mẫu S'^2, của phương sai S^{*2} ta thu được các thống kê tương ứng gọi là độ lệch tiêu chuẩn mẫu S, độ lệch tiêu chuẩn điều chỉnh mẫu S', độ lệch tiêu chuẩn S^* như sau:

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

$$S'^2 = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

$$S^{*2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2}.$$

Giá trị cụ thể của chúng trên một mẫu cụ thể là những số xác định, ký hiệu là s, s', s^*.

3.3.5. Tần suất mẫu

a) 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 là một mẫu thống kê, ký hiệu là f và là tỷ 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 $\overline{X}$ 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 $W = (X_1, X_2, ..., X_n)$ với X_i tuân theo quy luật không - một và:

$$Y = X_1 + X_2 + ... + X_n.$$

Khi đó:

$$f = \frac{X_1 + X_2 + ... + X_n}{n} = \overline{X}.$$

b) Tính chất.
Nếu đại lượng ngẫu nhiên gốc X tuân theo quy luật $A(p)$ với kỳ vọng toán $M(X) = p$ và phương sai $D(X) = p(1-p)$ thì:

$$M(f) = p; \qquad D(f) = \frac{p(1-p)}{n}.$$

Thật vậy, theo kết quả phần (3.2.2):

$$M(f) = M\left(\overline{X}\right) = M(X) = p$$

$$D(f) = D\left(\overline{X}\right) = \frac{D(X)}{n} = \frac{p(1-p)}{n}.$$

c) Tính giá trị cụ thể.
Giả sử trong mẫu ngẫu nhiên kích thước n có k phần tử mang dấu hiệu nghiên cứu, khi đó giá trị cụ thể của tần suất mẫu là:

$$f = \frac{k}{n}.$$

3.3.6. Quy luật phân phối xác suất của một số thống kê đặc trưng mẫu

Quy luật phân phối xác suất của các thống kê đặc trưng mẫu phụ thuộc chặt chẽ và quy luật phân phối xác suất của đại lượng ngẫu nhiên gốc X. Trong phần này ta chỉ dừng ở việc thừa nhận các kết quả. Bạn đọc có thể tìm thấy chứng minh các kết quả này ở các tài liệu đầy đủ hơn về thống kê toán.

a) Trường hợp đại lượng ngẫu nhiên gốc X tuân theo quy luật phân phối chuẩn
Giả sử dấu hiệu nghiên cứu trong tổng thể có thể xem như một đại lượng ngẫu nhiên X tuân theo quy luật chuẩn với kỳ vọng toán $M(X) = \mu$ và phương sai $D(X) = \sigma^2$.
Để xác định quy luật phân phối xác suất của các thống kê đặc trưng mẫu có thể sử dụng định lý sau:

◊ **Định lý 3.1** *Nếu các đại lượng ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập và cùng phân phối theo quy luật chuẩn thì mọi tổ hợp tuyến tính của các đại lượng ngẫu nhiên đó cũng phân phối theo quy luật chuẩn.*

- Trung bình mẫu cũng phân phối theo quy luật chuẩn với kỳ vọng toán $M\left(\overline{X}\right) = \mu$ và phương sai $D\left(\overline{X}\right) = \dfrac{\sigma^2}{n}$.

Do đó nếu xây dựng thống kê

$$U = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\left(\overline{X} - \mu\right)\sqrt{n}}{\sigma}$$

thì thống kê U sẽ phân phối theo quy luật chuẩn hoá $N(0, 1)$.

- Từ phương sai S^{*2} xét thống kê:

$$G = \frac{nS^{*2}}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)$$

khi đó thống kê G phân phối theo quy luật "Khi bình phương" với n bậc tự do $\chi^2(n)$.

- Từ phương sai S'^2 xây dựng thống kê:

$$X^2 = \frac{(n-1)\,S'^2}{\sigma^2}$$

thì thống kê X^2 phân phối theo quy luật "Khi bình phương" với $(n-1)$ bậc tự do $\chi^2\,(n-1)$.

- Từ thống kê U và χ^2 ở trên xây dựng tiếp thống kê:

$$T = \frac{U}{\sqrt{\frac{\chi^2}{n-1}}} = \frac{\left(\overline{X} - \mu\right)\sqrt{n}}{S'}$$

thì thống kê T sẽ phân phối theo quy luật Student với $(n-1)$ bậc tự do $T\,(n-1)$.
Chú ý rằng khi n khá lớn thì quy luật Student hội tụ về quy luật chuẩn hoá, do đó với $n > 30$ thực tế có thể xem thống kê T phân phối xấp xỉ $N\,(0, 1)$.

b) Trường hợp đại lượng ngẫu nhiên gốc X tuân theo quy luật phân phối không - một
Giả sử dấu hiệu nghiên cứu trong tổng thể có thể xem như đại lượng ngẫu nhiên phân phối theo quy luật không - một với tham số $p\,(A\,(p))$.

- Trung bình mẫu chính là tần suất mẫu f với kỳ vọng toán $M(f) = p$ và phương sai $D\,(f) = \dfrac{p\,(1-p)}{n}$.
- Từ tần suất mẫu xây dựng thống kê:

$$U = \frac{f - p}{\sqrt{\frac{p\,(1-p)}{n}}} = \frac{(f-p)\sqrt{n}}{\sqrt{p\,(1-p)}}$$

thì thống kê U sẽ phân phối xấp xỉ chuẩn hoá $N\,(0, 1)$ khi n khá lớn.

- Nếu n đủ lớn thì từ thống kê U ở trên thay p bởi f và xây dựng thống kê:

$$U = \frac{(f - p)\sqrt{n}}{\sqrt{f(1-f)}}$$

thì lúc đó U cũng phân phối xấp xỉ $N(0,1)$.

3.3.7. Ví dụ

●**Ví dụ 3.5** Nghiên cứu về mức tiêu thụ điện năng (kw) của một khu dân cư, ta nghiên cứu mức tiêu thụ điện năng của 5 hộ gia đình trong một khu dân cư đó thì thu được kết quả sau:
75; 123; 97; 145; 215
a. Tính trung bình số điện tiêu thụ của các hộ gia đình. Tìm phương sai, độ lệch tiêu chuẩn hiệu chỉnh của số điện tiêu thụ.
b. Nếu quy ước những hộ có mức tiêu thụ ít hơn hoặc bằng 123 kw thì được coi là tiêu thụ điện bình thường. Tìm tỷ lệ các hộ tiêu thụ điện bình thường.

Lời giải.
(bản kèm)

●**Ví dụ 3.6** Khi kiểm tra thể lực của một nhóm sinh viên, ta có kết quả về cân nặng như sau:

kg	45	50	55	60	65
Số sv	8	14	28	18	12

a. Tìm cân nặng trung bình của nhóm sinh viên trên. Phương sai, phương sai hiệu chỉnh, độ lệch tiêu chuẩn và độ lệch tiêu chuẩn hiệu chỉnh.
b. Tìm tỷ lệ sinh viên có cân nặng lớn hơn hoặc bằng 55 kg.

Lời giải.
(bản kèm)

●**Ví dụ 3.7** Khi kiểm tra thể lực của một nhóm sinh viên, ta có kết quả về cân nặng như sau:

kg	42,5-47,5	47,5-52,5	52,5-57,5	57,5-62,5	62,5-67,5
Số sv	8	14	28	18	12

a. Tìm cân nặng trung bình của nhóm sinh viên trên. Phương sai, phương sai hiệu chỉnh, độ lệch tiêu chuẩn và độ lệch tiêu chuẩn hiệu chỉnh.
b. Tìm tỷ lệ sinh viên có cân nặng lớn hơn hoặc bằng 52,5 kg.

Lời giải.
(bản kèm)

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 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 đối với đại lượng ngẫu nhiên (X, Y). Giả sử trên phần tử thứ i của mẫu $i = \left(\overline{1, n}\right)$, đại lượng (X, Y) nhận giá trị (X_i, Y_i), khi đó ta có n đauh lượng ngẫu nhiên độc lập $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$. Nếu các đại lượng ngẫu nhiên $X_1, X_2, ..., X_n$ độc lập và cùng phân phối xác suất với X; các đại lượng ngẫu nhiên $Y_1, Y_2, ..., Y_n$ độc lập và cùng phân phối xác suất với Y thì ta có mẫu ngẫu nhiên hai chiều, ký hiệu là:

$$W = [(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)].$$

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:

$$w = [(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)].$$

3.4.2. Phương pháp mô tả 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 $x_1, x_2, ..., x_i, ..., x_k$, còn thành phần Y nhận các giá trị $y_1, y_2, ..., y_i, ..., y_h$, trong đó có n_{ij} phần tử của mẫu nhận giá trị (x_i, y_i). 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	$\cdots$	y_i	$\cdots$	y_h	n_i
x_1	n_{11}	n_{12}	$\cdots$	n_{1j}	$\cdots$	n_{1h}	n_1
x_2	n_{21}	n_{22}	$\cdots$	n_{2j}	$\cdots$	n_{2h}	n_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	n_{i1}	n_{i2}	$\cdots$	n_{ij}	$\cdots$	n_{ih}	n_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_k	n_{k1}	n_{k2}	$\cdots$	n_{kj}	$\cdots$	n_{kh}	n_k
m_j	m_1	m_2	$\cdots$	m_j	$\cdots$	m_k	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 $Y = y_i$.

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:

1. Bảng phân phối thực nghiệm của X:

X	x_1	x_2	...	x_i	...	x_k
n_i	n_1	n_2	...	n_i	...	n_k

và tính được các thống kê đặc trưng của X:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i; \qquad s_X^2 = \frac{1}{n}\sum_{i=1}^{k} n_i\left(x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2.$$

2. Bảng phân phối thực nghiệm của Y:

Y	y_1	y_2	...	y_i	...	y_h
m_j	m_1	m_2	...	m_i	...	m_h

và tính được các thống kê đặc trưng của Y:

$$\overline{y} = \frac{1}{n}\sum_{j=1}^{h} m_i y_i; \qquad s_Y^2 = \frac{1}{n}\sum_{j=1}^{h} m_j\left(y_i - \overline{y}\right)^2 = \frac{1}{n}\sum_{j=1}^{h} m_i y_i^2 - \overline{y}^2.$$

3. Bảng phân phối có điều kiện của Y khi $X = x_i$:

$Y_{X=x_i}$	y_1	y_2	...	y_i	...	y_h
n_{ij}	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ih}

và tính được các thống kê tương ứng:

$$\overline{y}_{X=x_i} = \frac{1}{n}\sum_{j=1}^{h} n_{ij} y_i; \qquad s_{Y/X=x_i}^2 = \frac{1}{n}\sum_{j=1}^{h} n_{ij}\left(y_i - \overline{y}\right)^2 = \frac{1}{n}\sum_{j=1}^{h} n_{ij} y_j^2 - \overline{y}^2.$$

4. Bảng phân phối có điều kiện của X khi $Y = y_i$:

$X_{Y=y_i}$	x_1	x_2	...	x_i	...	x_k
n_{ij}	n_{1j}	n_{2j}	...	n_{ij}	...	n_{kj}

và tính được các thống kê tương ứng:

$$\overline{x}_{Y=y_j} = \frac{1}{n}\sum_{i=1}^{k} n_{ij} x_i; \qquad s_{X/Y=y_j}^2 = \frac{1}{n}\sum_{i=1}^{k} n_{ij}\left(x_i - \overline{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{k} n_{ij} x_i^2 - \overline{x}^2.$$

●**Ví dụ 3.8** Thống kê năng suất một loại cây trồng Y(kg/ha) và lượng đầu tư cho cải tạo đất X(triệu đồng/ha) tại một số địa phương trong vòng 10 năm ta thu được bảng số liệu sau:

X \ Y	2	3	4	5	6	7
20 - 22	1	3	5			
22 - 24		1	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:

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_ix_i	$n_ix_i^2$
2	1	2	4
3	5	15	45
4	13	52	208
5	13	65	325
6	11	66	396
7	3	21	147
	$n = \sum n_i = 46$	$\sum n_ix_i = 221$	$\sum n_ix_i^2 = 1125$

Từ đó có kết quả sau:
Lượng đầu tư trung bình: $\overline{x} = \frac{221}{46} \approx 4,804$ (triệu đồng/ha)
Độ phân tán của đầu tư:

$$s'_x = \sqrt{\frac{n}{n-1}\left(\frac{\sum n_ix_i^2}{n} - \overline{x}^2\right)} = \sqrt{\frac{46}{45}\left(\frac{1125}{46} - 4,804^2\right)} \approx 1,185(\text{triệu đồng/ha}).$$

- Từ bảng trên ta có bảng phân phối tần số của Y:

Y	21	23	25	27	29
Số địa phương (n_i)	9	22	9	4	2

Tương tự phần trên ta cũng tính được:

$$\overline{y} \approx 23,609(kg/ha); \qquad s'_{y_A} \approx 1,337(kg/ha).$$

- Tần suất mẫu của những địa phương nhóm A là: $f = \dfrac{k}{n} = \dfrac{12}{46} \approx 0,261.$

3.5. Ước lượng tham số của đại lượng ngẫu nhiên

Giả sử cần phải nghiên cứu dấu hiệu định lượng χ trong tổng thể, cụ thể là cần xác định các tham số đặc trưng của tổng thể như trung bình, phương sai, cơ cấu tổng thể theo dấu hiệu nghiên cứu. Nếu xem dấu hiệu nghiên cứu như đại lượng ngẫu nhiên X và giả sử đã biết dạng phân phối xác suất của nó thì vấn đề xác định các tham số đặc trưng của tổng thể sẽ quy về bài toán xác định các tham số đặc trưng của quy luật phân phối xác suất của đại lượng ngẫu nhiên X. Chẳng hạn nếu X là đại lượng ngẫu nhiên phân phối chuẩn thì để xác định trung bình tổng thể ta sẽ ước lượng kỳ vọng toán của X, còn để xác định phương sai tổng thể ta sẽ ước lượng phương sai của X.
Như vậy bài toán ước lượng tham số có thể phát biểu như sau: *Cho đại lượng ngẫu nhiên X với quy luật phân phối xác suất đã biết song chưa biết tham số θ nào đó của nó, phải ước lượng (xác định một cách gần đúng) giá trị θ.*
Vì tổng thể thường có kích thước lớn nên việc xác định tham số trực tiếp trên tổng thể gặp nhiều khó khăn, người ta giải quyết bài toán ước lượng tham số bằng phương pháp mẫu như sau: Từ tổng thể nghiên cứu rút ra một mẫu ngẫu nhiên kích thước n và dựa vào đó xây dựng một thống kê G dùng để ước lượng θ.

3.5.1. Phương pháp ước lượng điểm

Phương pháp ước lượng điểm chủ trương dùng một giá trị để thay thế cho tham số θ chưa biết của tổng thể. Từ tổng thể lập mẫu ngẫu nhiên kích thước $n : W = (X_1, X_2, ..., X_n)$. Thông thường giá trị được chọn là một thống kê G nào đó của mẫu ngẫu nhiên. Có rất nhiều cách chọn thống kê G khác nhau tạo nên những phương pháp ước lượng điểm khác nhau. Trong giáo trình này trình bày một loại ước lượng điểm, đó là ước lượng không chênh lệch như sau:

$\star$ **Định nghĩa 3.4** Thống kê G của mẫu được gọi là ước lượng không chệch của tham số θ của đại lượng ngẫu nhiên gốc nếu $M(G) = \theta$.

Ngược lại, nếu $M(G) \neq \theta$ thì G được gọi là ước lượng chệch của θ.
Dựa vào các kết quả của phần (3.3.) có thể rút ra một số kết luận sau;

- Trung bình mẫu $\overline{X}$ là ước lượng không chệch của kỳ vọng toán m của đại lượng ngẫu nhiên gốc:
$$M\left(\overline{X}\right) = m.$$

- Tần suất mẫu f là ước lượng không chệch của tần suất tổng thể p (tham số p của đại lượng ngẫu nhiên gốc tuân theo quy luật không - một):
$$M\left(f\right) = p.$$

- Phương sai điều chỉnh mẫu S'^2 và phương sai S^{*2} là các ước lượng không chệch của phương sai σ^2 của đại lượng ngẫu nhiên gốc:

$$M\left(S'^2\right) = \sigma^2; \qquad M\left(S^{*2}\right) = \sigma^2.$$

Lưu ý rằng với k mẫu cụ thể rút ra từ tổng thể, G sẽ nhận k giá trị tương ứng là $g_1, g_2, ..., g_k$. Khi đó G là ước lượng không chệch của θ không có nghĩa mọi giá trị g_i của G đều bằng θ mà chỉ có nghĩa: trung bình các giá trị của G bằng θ, từng giá trị g_i có thể sai lệch rất lớn so với θ. Tuy nhiên chọn ước lượng không chệch G để ước lượng cho tham số θ ta sẽ loại bỏ được sai số hệ thống: tất cả các giá trị của G đều lớn hơn hoặc nhỏ hơn θ.

3.5.2. Phương pháp ước lượng bằng khoảng tin cậy

Các phương pháp ước lượng điểm có một nhược điểm cơ bản là sai số của ước lượng có thể rất lớn, ngoài ra không đánh giá được khả năng mắc sai lầm khi ước lượng cũng như độ tin cậy của ước lượng bằng bao nhiêu. Do đó khi kích thước mẫu nhỏ người ta thường sử dụng phương pháp ước lượng bằng khoảng tin cậy.
Định nghĩa 5.

$\star$ **Định nghĩa 3.5** Khoảng $(G1, G2)$ của thông kê G được gọi là khoảng tin cậy của tham số θ nếu với xác suất bằng $1-\alpha$ cho trước điều kiện sau đây được thỏa mãn:

$$P(G_1 < \theta < G_2) = 1 - \alpha.$$

$I = G_2 - G_1$ được gọi là độ dài khoảng tin cậy, còn xác suất $1-\alpha$ được gọi là độ tin cậy của ước lượng (trên thực tế thường đòi hỏi độ tin cậy khá lớn. chẳng hạn $1-\alpha = 0,95$).

3.5.3. Khoảng tin cậy cho trung bình (Ước lượng kỳ vọng toán của đại lượng ngẫu nhiên phân phối theo quy luật chuẩn)

Giả sử trong tổng thể đại lượng ngẫu nhiên gốc X phân phối chuẩn $N(\mu, \sigma^2)$ nhưng chưa biết tham số μ của nó. Để ước lượng μ từ tổng thể ta lập mẫu ngẫu nhiên kích thước $n : W = (X_1, X_2, ..., X_n)$.
Ta xét các trường hợp sau đây:
1. Đã biết phương sai σ^2 của đại lượng ngẫu nhiên gốc trong tổng thể.
Xét thống kê:

$$G = U = \frac{(\overline{X} - \mu)\sqrt{n}}{\sigma}$$

trong đó $\overline{X}$ là trung bình mẫu.
Ta đã biết thống kê U phân phối chuẩn hóa $N(0,1)$. Do đó với các phân vị chuẩn u_{α_1} và $u_{1-\alpha_2}$

ta có:

$$P\left(u_{\alpha_1 < U < u_{1-\alpha_2}}\right) = (1-\alpha_2) - \alpha_1.$$

Từ đó suy ra:

$$P\left(u_{\alpha_1} < \frac{(\overline{X}-\mu)\sqrt{n}}{\sigma} < u_{1-\alpha_2}\right) = 1-(\alpha_1+\alpha_2)$$

$$\Leftrightarrow P\left(\overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\alpha_2} < \mu < \overline{X} - \frac{\sigma}{\sqrt{n}}u_{\alpha_1}\right) = 1-(\alpha_1+\alpha_2)$$

$$\Leftrightarrow P\left(\overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\alpha_2} < \mu < \overline{X} + \frac{\sigma}{\sqrt{n}}u_{1-\alpha_1}\right) = 1-(\alpha_1+\alpha_2) \quad (\text{vì } u_{\alpha_1} = -u_{1-\alpha_1})$$

Như vậy với độ tin cậy $1-\alpha$ cho trước, ta có thể chọn cặp số α_1 và α_2 thỏa mãn điều kiện $\alpha_1+\alpha_2=\alpha$ khi đó khoảng tin cậy của tham số μ là:

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\alpha_2}; \overline{X} + \frac{\sigma}{\sqrt{n}}u_{1-\alpha_2}\right).$$

a) Khoảng tin cậy đối xứng

Nếu lấy $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ thì khoảng tin cậy của μ là:

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}; \overline{X} + \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right).$$

Kí hiệu:

$$\varepsilon = \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}. \tag{3.1}$$

ε được gọi là độ chính xác của ước lượng, nó phản ánh mức độ sai lệch của trung bình mẫu so với trung bình tổng thể với xác suất $1-\alpha$ cho trước.

Khi đó khoảng tin cậy đối xứng của μ có dạng:

$$(\overline{X} - \varepsilon; \overline{X} + \varepsilon).$$

b) Khoảng tin cậy một phía

- Khoảng tin cậy bên phải

  Nếu lấy $\alpha_1 = 0$ và $\alpha_2 = \alpha$ thì $u_{1-\alpha_1} = +\infty$, do đó khoảng tin cậy của μ là:

$$\left(\overline{X} - \frac{\sigma}{\sqrt{n}}u_{1-\alpha}; +\infty\right).$$

  Biểu thức trên được sử dụng để ước lượng giá trị tối thiểu của μ.

- Khoảng tin cậy bên trái
  Nếu lấy $\alpha_1 = \alpha$ và $\alpha_2 = 0$ thì $u_{1-\alpha_2} = +\infty$, do đó khoảng tin cậy của μ là:

$$\left(-\infty; \overline{X} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha}\right).$$

Biểu thức trên được dùng để ước lượng giá trị tối đa của μ.

c) Xác định kích thước tối thiểu

Với cùng độ tin cậy $1-\alpha$ khoảng tin cậy nào có độ dài ngắn hơn sẽ tốt hơn. Trong trường hợp này độ dài khoảng tin cậy I sẽ ngắn nhất khi khoảng tin cậy là đối xứng. Lúc đó độ dài khoảng tin cậy đối xứng bằng:

$$I = 2\varepsilon = \frac{2\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}.$$

Do đó với độ tin cậy bằng $1-\alpha$ cho trước, muốn độ dài khoảng tin cậy đối xứng không vượt quá giá trị I_o cho trước thì kích thước mẫu n phải thỏa mãn điều kiện:

$$n \geq \frac{4\sigma^2}{I_o^2} u^2_{1-\frac{\alpha}{2}}.$$

Như vậy công thức xác định kích thước mẫu tối thiểu là:

$$n_{tt} = \left[\frac{4\sigma^2}{I_o^2} u^2_{1-\frac{\alpha}{2}}\right] + 1 \quad \text{hoặc} \quad n_{tt} = \left[\frac{4\sigma^2}{\varepsilon_o^2} u^2_{1-\frac{\alpha}{2}}\right] + 1.$$

Bài toán xác định kích thước tối thiểu theo công thức trên được đặt ra khi phải xác định kích thước mẫu cần điều tra để đáp ứng những yêu cầu chất lượng cho trước về độ tin cậy và độ chính xác của ước lượng.

d) Xác định độ tin cậy của ước lượng

Từ công thức 3.1 xác định độ chính xác của ước lượng, ta suy ra:

$$u_{1-\frac{\alpha}{2}} = \frac{\varepsilon\sqrt{n}}{\sigma}.$$

Từ phân vị $u_{1-\frac{\alpha}{2}}$, tra bảng phân vj chuẩn ta có thể tìm được $1-\frac{\alpha}{2}$, từ đó suy ra $1-\alpha$. Như vậy nếu biết trước độ chính xác (hoặc độ dài khoảng tin cậy đối xứng) của ước lượng thì có thể tìm được độ tin cậy của ước lượng.

Chú ý rằng các khoảng tin cậy đối xứng là khoảng tin cậy một phía trong các công thức trên chỉ là các khoảng tin cậy ngẫu nhiên. Muốn tìm khoảng tin cậy cụ thể ta phải thực hiện phép thử đối với mẫu ngẫu nhiên $W = (X_1, X_2, ..., X_n)$ để thu được mẫu cụ thể $w = (x_1, x_2, ..., x_n)$, từ đó tìm được giá trị cụ thể của trung bình mẫu $\overline{x}$. Khi đó với mẫu tin cậy $1-\alpha$, khoảng tin cậy đối xứng cụ thể của μ (qua mẫu cụ thể trên) là:

$$\left(\overline{x} - \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}; \overline{x} + \frac{\sigma}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right).$$

Tiến hành tương tự ta cũng thu được giá trị cụ thể của các khoảng tin cậy một phía.

•**Ví dụ 3.9** Trọng lượng một loại sản phẩm là đại lượng ngẫu nhiên phân phối theo quy luật chuẩn với độ lệch tiên chuẩn là $\sigma = 1$gam. Cân thử 25 sản phẩm loại này ta thu được kết quả sau:

Trọng lượng (gam)	18	19	20	21
Số sản phẩm tương ứng	3	5	15	2

a) Với độ tin cậy là 0,95 hãy tìm khoảng tin cậy đối xứng của trọng lượng trung bình của loại sản phẩm nói trên.
b) Với độ tin cậy là 98% hãy ước lượng trọng lượng trung bình tối đa của loại sản phẩm trên.
c) Muốn ướng lượng trọng lượng trung bình của loại sản phẩm trên với độ tin cậy là 98%, độ dài khoảng tin cậy đối xứng không quá 0,7gam thì cần kiểm tra ít nhất bao nhiêu sản phẩm?
d) Muốn sử dụng kết quả cân thử 25 sản phẩm trên để ước lượng trọng lượng trung bình của loại sản phẩm trên với độ chính xác 0,399 gam thì đạt độ tin cậy bao nhiêu?
Lời giải.
Gọi X là trọng lượng sản phẩm, theo giả thiết X phân phối chuẩn với $\sigma = 1$. Vậy trọng lượng trung bình của sản phẩm chính là tham số μ.
a) Đây là bài toán ước lượng tham số μ bằng khoảng tin cậy đối xứng:

$$\mu \in \left(\overline{x} - \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}; \overline{x} + \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right).$$

Từ mẫu ngẫu nhiên kích thước $n = 25$ ở trên ta tích được trung bình mẫu.

$$\overline{x} = \frac{18.3 + 19.5 + 20.15 + 21.2}{25} = \frac{491}{25} = 19,64.$$

Từ độ tin cậy $1 - \alpha = 0,95$ suy ra $1 - \frac{\alpha}{2} = 0,975$. Tra bảng phân vị chuẩn ta có $u_{1-\frac{\alpha}{2}} = u_{0,975} = 1,96$.
Độ chính xác của ước lượng:

$$\varepsilon = \frac{\sigma}{\sqrt{n}}u_{1-\frac{\alpha}{2}} = \frac{1}{\sqrt{25}}1,96 = 0,392.$$

Vậy khoảng tin cậy đối xứng của μ là:

$$\mu \in (\overline{x} - \varepsilon; \overline{x} + \varepsilon) = (19,64 - 0,392; 19,64 + 0,392) = (19,148; 20,032).$$

b) Đây là bài toán ước lượng giá trị tối đa của μ

$$\mu \in \left(-\infty; \overline{x} + \frac{\sigma}{\sqrt{n}}u_{1-\alpha}\right).$$

Từ độ tin cậy $1-\alpha = 0,98$, tra bảng phân vị chuẩn ta có $u_{0,98} = 2,054$
Giá trị tối đa của μ là:

$$\overline{x} + \frac{\sigma}{\sqrt{n}} u_{1-\alpha} = 19,64 + \frac{1.2,054}{\sqrt{25}} = 20,0508.$$

c) Đây là bài toán xác định kích thước mẫu tối thiểu:

$$n \geq \frac{4\sigma^2}{I_o^2} u^2_{1-\frac{\alpha}{2}}.$$

Từ độ tin cậy $1-\alpha = 0,98$, suy ra $1-\frac{\alpha}{2} = 0,99$. Tra bảng phân vị chuẩn: $u_{0,99} = 2,326$

$$n \geq \frac{4.1^2}{0,7^2} 2,326^2 \approx 44,2.$$

Vậy cần phải kiểm tra ít nhất 45 sản phẩm (cần kiểm tra thêm ít nhất 20 sản phẩm nữa).
d) Đây là bài toán xác định độ tin cậy:

$$u_{1-\frac{\alpha}{2}} = \frac{\varepsilon\sqrt{n}}{\sigma} = \frac{0,399\sqrt{25}}{1} = 1,995.$$

Từ bảng phân vị chuẩn suy ra $1-\frac{\alpha}{2} = 0,977$ do đó $1-\alpha = 0,954$. Vậy độ tin cậy bằng $95,4\%$.
2. Chưa biết phương sai σ^2 của đại lượng ngẫu nhiên X và kích thước mẫu $n \leq 30$
Chọn thống kê:

$$G = T = \frac{(\overline{X}-\mu)\sqrt{n}}{S'}$$

trong đó $\overline{X}$ là trung bình mẫu, và S' là độ lệch tiêu chuẩn điều chỉnh mẫu.
Ta đã biết T phân phối theo quy luật Student với $n-1$ bậc tự do. Vì vậy với các phân vị Student $t^{n-1}_{\alpha_1}$ và $t^{n-1}_{1-\alpha_2}$ ta có:

$$P\left(t^{n-1}_{\alpha_1} < T < t^{n-1}_{1-\alpha_2}\right) = (1-\alpha_2) - \alpha_1.$$

Suy ra:

$$P\left(\overline{X} - \frac{S'}{\sqrt{n}} t^{n-1}_{1-\alpha_2} < \mu < \overline{X} + \frac{S'}{\sqrt{n}} t^{n-1}_{1-\alpha_1}\right) = 1 - (\alpha_1 + \alpha_2).$$

Chọn $\alpha_1 + \alpha_2 = \alpha$, khoảng tin cậy của μ với độ tin cậy $1-\alpha$ là:

$$\left(\overline{X} - \frac{S'}{\sqrt{n}} t^{n-1}_{1-\alpha_2}; \overline{X} + \frac{S_{\prime}}{\sqrt{n}} t^{n-1}_{1-\alpha_1}\right).$$

a) Khoảng tin cậy đối xứng

Chọn $\alpha_1 = \alpha 2 = \frac{\alpha}{2}$ ta có khoảng tin cậy đối xứng:

$$\left(\overline{X} - \frac{S'}{\sqrt{n}} t^{n-1}_{1-\frac{\alpha}{2}}; \overline{X} + \frac{S'}{\sqrt{n}} t^{n-1}_{1-\frac{\alpha}{2}}\right)$$

với độ chính xác:

$$\varepsilon = \frac{S'}{\sqrt{n}} t^{n-1}_{1-\frac{\alpha}{2}}$$

và độ dài khoảng tin cậy đối xứng:

$$I = 2\varepsilon = 2\frac{S'}{\sqrt{n}} t^{n-1}_{1-\frac{\alpha}{2}}.$$

b) Khoảng tin cậy một phía

- Khoảng tin cậy bên phải
  Chọn $\alpha_1 = 0, \alpha_2 = \alpha$ ta có khoảng tin cậy bên phải dùng để ước lượng giá trị tối thiểu của μ như sau:
  $$\left(\overline{X} - \frac{S'}{\sqrt{n}} t^{n-1}_{1-\alpha}; +\infty\right).$$

- Khoảng tin cậy bên trái
  Chọn $\alpha_1 = \alpha, \alpha_2 = 0$ ta có khoảng tin cậy bên trái dùng để ước lượng giá trị tối thiểu của μ như sau:
  $$\left((-\infty; \overline{X} - \frac{S'}{\sqrt{n}} t^{n-1}_{1-\alpha}\right).$$

c) Xác định kích thước mẫu tối thiểu

Với độ tin cậy bằng $1-\alpha$ cho trước, độ dài khoảng tin cậy đối xứng không vượt quá giá trị I_o cho trước, bài toán xác định kích thước mẫu tối thiểu được giải quyết bằng phương pháp mẫu kép sau đây:

Trước hết điều tra một mẫu sơ bộ kích thước $n \geq 2 : W_1 = (X_1, X_2, ..., X_n)$, từ đó tìm được phương sai điều chỉnh mẫu:

$$S'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2; \qquad \text{với} \quad \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Giả sử để thỏa mãn độ chính xác và độ tin cậy cho trước, mẫu cần lập phải có kích thước N, khi đó giả sử lập mẫu thứ hai: $W_2 = (X_{n+1}, X_{n+2}, ..., X_N)$.

Vì thống kê:

$$T = \frac{\left(\frac{1}{N}\sum_{i=1}^{N} X_i - \mu\right)\sqrt{n}}{S'}$$

cũng phân phối theo quy luật Student $n-1$ bậc tự do nên phân vị Student $t^{n-1}_{1-\frac{\alpha}{2}}$ ta cũng có:

$$P(\sum_{i=1}^{N} X_i - \frac{S'}{\sqrt{n}} t^{n-1}_{1-\frac{\alpha}{2}} < \mu < \sum_{i=1}^{N} X_i + \frac{S'}{\sqrt{n}} t^{n-1}_{1-\frac{\alpha}{2}}) = 1-\alpha.$$

Khi đó ước lượng có độ chính xác là:

$$\varepsilon_o = \frac{S'}{\sqrt{n}} t^{n-1}_{1-\frac{\alpha}{2}}.$$

Từ đó suy ra kích thước mẫu cần tìm:

$$N \geq \left(\frac{S'}{\varepsilon_o} t^{n-1}_{1-\frac{\alpha}{2}} \right)^2 \qquad \text{hoặc} \qquad N \geq \left(\frac{2S'}{I_o} t^{n-1}_{1-\frac{\alpha}{2}} \right)^2.$$

d) Xác định độ tin cậy của ước lượng

Từ mẫu ngẫu nhiên kích thước n, muốn ước lượng μ khoảng tin cậy đối xứng, với độ chính xác ε_o, thì độ tin cậy $1-\alpha$ của ước lượng được suy ra từ biểu thức:

$$t^{n-1}_{1-\frac{\alpha}{2}} = \frac{\varepsilon_o \sqrt{n}}{S'}.$$

•**Ví dụ 3.10** Phân xưởng sản xuất tự động của nhà máy gang thép Thái Nguyên sản xuất 1 loại thép thành phẩm với chiều dài đặt trước bởi đơn đặt hàng là X (m) (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 (n_i)	3	6	9	5	2

a) Bằng khoảng tin cậy đối xứng hãy ước lượng chiều dài trung bình của loại thép trên với độ tin cậy 95%.

b) Muốn ước lượng chiều dài trung bình của loại thép đó 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 $N(\mu, \sigma^2)$ khi chưa biết σ^2 của X, hơn nữa kích thước mẫu $n = 25 < 30$ nên bài toán được giải như sau:
   Từ số liệu đã cho ta lập bảng:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
1,0	3	3,0	3,00
1,1	6	6,6	7,26
1,2	9	10,8	12,96
1,3	5	6,5	8,45
1,4	2	2,8	3,92
	$n = \sum n_i = 25$	$\sum n_i x_i = 29,7$	$\sum n_i x_i^2 = 55,59$

Chiều dài trung bình mẫu: $\overline{x} = \frac{29,7}{25} = 1,188$ (m)

Độ lệch tiêu chuẩn mẫu điều chỉnh:

$$s'_x = \sqrt{\frac{n}{n-1}\left(\frac{\sum n_i x_i^2}{n} - \overline{x}^2\right)} = \sqrt{\frac{25}{24}\left(\frac{35,59}{25} - 1,188^2\right)} \approx 0,113 \quad (m).$$

Từ độ tin cậy $1-\alpha = 0,95$ suy ra $1 - \frac{\alpha}{2} = 0,975$. Tra bảng phân vị Student $t^{n-1}_{1-\frac{\alpha}{2}} = t^{24}_{0,975} = 2,064$.

Độ chính xác: $\epsilon = \frac{s'}{\sqrt{n}} t^{n-1}_{1-\frac{\alpha}{2}} = \frac{0,113.2,064}{\sqrt{25}} \approx 0,047$ (m).

Vậy khoảng tin cậy đối xứng của μ là:

$$1,188 - 0,047 < \mu < 1,188 + 0,047 \quad \text{hay} \quad 1,141 < \mu < 1,235 \quad (m).$$

b) Đây là bài toán xác định độ tin cậy của ước lượng, giả thiết đã cho độ chính xác $\epsilon = 5,63(cm) = 0,0563(m)$, và từ bảng tính phần a) ta đã có $s' \approx 0,113(m)$. Từ đó tính được:

$$t^{n-1}_{1-\frac{\alpha}{2}} = t^{24}_{1-\frac{\alpha}{2}} = \frac{\epsilon\sqrt{n}}{s'} = \frac{0,0563.\sqrt{25}}{0,113} \approx 2,491.$$

Từ bảng phân vị Student thấy $t^{24}_{1-\frac{\alpha}{2}} \approx 2,491 \approx t^{24}_{0,99}$. Vậy:

$$1 - \frac{\alpha}{2} = 0,99 \Rightarrow 1 - \alpha = 0,98.$$

3. Chưa biết phương sai σ^2 của đại lượng ngẫu nhiên gốc X và kích thước $n > 30$

Ta chọn thống kê

$$G = T = \frac{(\overline{X} - \mu)\sqrt{n}}{S'}$$

trong đó $\overline{X}$ là trung bình mẫu, và S' là độ lệch tiêu chuẩn điều chỉnh mẫu.

Vì kích thước $n > 30$ nên thống kê T sẽ phân phối xấp xỉ chuẩn hóa $N(0,1)$. Do đó tương tự các phần trên ta cũng có:

$$P\left(u_{\alpha_1} < T < u_{1-\alpha_2}\right) = (1 - \alpha_2) - \alpha_1.$$

Từ đó suy ra:

$$P\left(\overline{X} - \frac{S'}{\sqrt{n}} u_{1-\alpha_2} < \mu < \overline{X} + \frac{S'}{\sqrt{n}} u_{1-\alpha_1}\right) = 1 - (\alpha_1 + \alpha_2).$$

Như vậy với độ tin cậy là $1-\alpha$ cho trước, ta có thể chọn cặp số α_1 và α_2 thỏa mãn điều kiện $\alpha_1 + \alpha_2 = \alpha$, khi đó khoảng tin cậy của tham số μ là:

$$\left(\overline{X} - \frac{S'}{\sqrt{n}} u_{1-\alpha_2}; \overline{X} + \frac{S'}{\sqrt{n}} u_{1-\alpha_1}\right).$$

a) Khoảng tin cậy đối xứng khi $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$:

$$\left(\overline{X} - \frac{S'}{\sqrt{n}} u_{1-\frac{\alpha}{2}}; \overline{X} + \frac{S'}{\sqrt{n}} u_{1-\frac{\alpha}{2}}\right)$$

với độ chính xác:

$$\varepsilon = \frac{S'}{\sqrt{n}} u_{1-\frac{\alpha}{2}}$$

và độ dài khoảng tin cậy đối xứng:

$$I = 2\varepsilon = \frac{2S'}{\sqrt{n}} u_{1-\frac{\alpha}{2}}.$$

b) Khoảng tin cậy một phía

- Khoảng tin cậy bên phải

$$\left(\overline{X} - \frac{S'}{\sqrt{n}} u_{1-\alpha}; +\infty\right).$$

- Khoảng tin cậy bên trái

$$\left(-\infty; \overline{X} + \frac{S'}{\sqrt{n}} u_{1-\alpha}\right).$$

c) Xác định kích thước mẫu tối thiểu

Muốn ước lượng μ với độ tin cậy $1-\alpha$, độ dài khoảng tin cậy đối xứng I_o thì kích thước mẫu N phải thỏa mãn:

$$N \geq \left(\frac{S'}{\varepsilon_o} u_{1-\frac{\alpha}{2}}\right)^2 \qquad \text{hay} \qquad N \geq \left(\frac{2S'}{I_o} u_{1-\frac{\alpha}{2}}\right)^2$$

ở đó S' là độ lệch tiêu chuẩn điều chỉnh của mẫu sơ bộ có kích thước n

d) Xác định độ tin cậy của ước lượng

Từ mẫu ngẫu nhiên kích thước n, muốn ướng lượng μ bằng khoảng tin cậy đối xứng, với độ chính xác ε_o, thì độ tin cậy $1-\alpha$ của ước lượng được suy ra từ biểu thức:

$$u_{1-\frac{\alpha}{2}} = \frac{\varepsilon_o \sqrt{n}}{S'}.$$

3.5.4. Khoảng tin cậy cho tỷ lệ (Ước lượng kỳ vọng toán của đại lượng ngẫu nhiên phân phối theo quy luật không - một)

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 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 đại lượng ngẫu nhiên phân phối theo quy luật không - một $A(p)$ với kỳ vọng toán $M(X) = p$ và phương sai $D(X) = \frac{p(1-p)}{n}$. Do đó bài toán ước lượng cơ cấu tổng thể chính là bài toán ước lượng kì vọng toán của đại lượng ngẫu nhiên tuân theo quy luật không - một.
Xét ngẫu nhiên kích thước n: $W = (X_1, X_2, ..., X_n)$. Nếu kích thước mẫu khá lớn ta có thể chọn thống kê:

$$G = U = \frac{(f-p)\sqrt{n}}{\sqrt{f(1-f)}}$$

trong đó f là tần suất mẫu.
Với n khá lớn thống kê U phân phối xấp xỉ chuẩn hóa $N(0,1)$. Do đó với các phân vị chuẩn u_{α_1} và $u_{1-\alpha_2}$ ta có:

$$P(u_{\alpha_1} < U < u_{1-\alpha_2}) = (1-\alpha_2) - \alpha_1.$$

Chọn $\alpha_1 + \alpha_2 = \alpha$ ta có:

$$P\left(-u_{1-\alpha_1} < \frac{(f-p)\sqrt{n}}{\sqrt{f(1-f)}} < u_{1-\alpha_2}\right) = 1-\alpha$$

$$\Leftrightarrow P\left(f - \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{1-\alpha_2} < p < f + \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{1-\alpha_1}\right) = 1-\alpha.$$

a) Khoảng tin cậy đối xứng
Chọn $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ ta có khoảng tin cậy đối xứng:

$$\left(f - \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{1-\frac{\alpha}{2}}; f + \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{1-\frac{\alpha}{2}}\right)$$

với độ chính xác:

$$\varepsilon = \frac{\sqrt{f(1-f)}}{\sqrt{n}}u_{1-\frac{\alpha}{2}}$$

và độ dài khoảng tin cậy đối xứng:

$$I = 2\varepsilon = \frac{2\sqrt{f(1-f)}}{\sqrt{n}} u_{1-\frac{\alpha}{2}}.$$

b) Khoảng tin cậy một phía

- Khoảng tin cậy bên phải
  Chọn $\alpha_1 = 0, \alpha_2 = \alpha$ ta có khoảng tin cậy bên phải dùng để ước lượng giá trị tối thiểu của p như sau:
  $$\left(f - \frac{\sqrt{f(1-f)}}{\sqrt{n}} u_{1-\alpha}; +\infty \right).$$

- Khoảng tin cậy bên trái
  Chọn $\alpha_1 = \alpha, \alpha_2 = 0$ ta có khoảng tin cậy bên trái dùng để ước lượng giá trị tối thiểu của p như sau:
  $$\left(-\infty; f + \frac{\sqrt{f(1-f)}}{\sqrt{n}} u_{1-\alpha} \right).$$

c) Xác định kích thước mẫu tối thiểu
Để đảm bảo độ tin cậy $1-\alpha$ cho trước và độ dài khoảng tin cậy đối xứng không vượt quá I_o cho trước thì kích thước mẫu N phải đảm bảo:

$$N \geq \frac{4f(1-f)}{I_o^2} u^2_{1-\frac{\alpha}{2}}$$

trong đó f là tần suất mẫu sơ bộ kích thước n
d) Xác định độ tin cậy của ước lượng
Từ mẫu ngẫu nhiên kích thước n, muốn ước lượng p bằng khoảng tin cậy đối xứng, với độ chính xác ε, thì độ tin cậy $1-\alpha$ của ước lượng được suy ra từ biểu thức:

$$u_{1-\frac{\alpha}{2}} = \frac{\varepsilon_o \sqrt{n}}{\sqrt{f(1-f)}}.$$

•**Ví dụ 3.11** Để điều tra số cá trong một hồ, cơ quan quản lý đánh bắt 2000 con cá, đánh dấu rồi thả xuống hồ. Lần sau bắt lại 400 con cá, được 80 con có dấu.
a) Hãy ước lượng tỷ lệ cá bị đánh dấu trong hồ bằng khoảng tin cậy đối xứng với độ tin cậy 97%;
b) Hãy ước lượng số cá tối thiểu trong hồ với độ tin cậy 95%;
c) Muốn ước lượng tỷ lệ cá bị đánh dấu trong hồ với độ tin cậy 98%, độ chính xác không quá $0,04$ thì cần đánh bắt bao nhiêu con cá nữa? d) Muốn ước lượng tỷ lệ cá bị đánh dấu trong hồ với độ dài khoảng tin cậy đối xứng không quá $5,76\%$ thì đạt độ tin cậy bao nhiêu?

Lời giải.
Mẫu ngẫu nhiên có kích thước $n = 400$ (con cá), trong đó $k = 80$ con có dấu. Do đó tần suất mẫu là $f = \frac{k}{n} = \frac{80}{400} = 0,2$.
a) Từ độ tin cậy $1-\alpha = 0,97 \Rightarrow 1-\frac{\alpha}{2} = 0,985$, tra bảng phân vi chuẩn $u_{0,985} = 2,17$;
Độ chính xác của ước lượng là: $\varepsilon = \frac{\sqrt{f(1-f)}}{\sqrt{n}} u_{1-\frac{\alpha}{2}} = \frac{\sqrt{0,2.0,8}.2,17}{\sqrt{400}} = 0,0434$.
Khoảng tin cậy đối xứng là:

$$p \in (0,2-0,0434; 0,2+0,0434) \quad \text{hay} \quad 0,1566 < p < 0,2434.$$

b)

- Ước lượng tỉ lệ cá bị đánh dấu tối đa
  Từ độ tin cậy $1-\alpha = 0,95$, tra bảng phân vị chuẩn $u_{0,95} = 1,645$.
  Độ chính xác một phía: $\varepsilon' = \frac{\sqrt{f(1-f)}}{\sqrt{n}} u_{1-\alpha} = \frac{\sqrt{0,2.0,8}.1,645}{\sqrt{400}} = 0,0329$.
  Khoảng tin cậy bên trái là: $p \in (-\infty; 0,2+0,0329)$ hay $p < 0,2329$.

- Gọi số cá trong hồ là N. Ta có: $p = \frac{2000}{N} < 0,2329$ suy ra $N > 8587,4$.
  Vậy số cá tối thiểu trong hồ là 8588 (con cá).

c) Từ độ tin cậy $1-\alpha = 0,98 \Rightarrow 1-\frac{\alpha}{2} = 0,99$ tra bảng phân vị chuẩn $u_{0,99} = 2,326$;

$$n_{tt} \geq \frac{f(1-f)(u_{0,99})^2}{\varepsilon^2} = \frac{0,2.0,8.2,326^2}{0,04^2} \approx 541,03.$$

Vậy cần đánh bắt thêm ít nhất $542 - 400 = 142$ con cá nữa.
d) Độ dài khoảng tin cậy đối xứng $I_o = 5,76\%$, suy ra độ chính xác $\varepsilon = 2,88\% = 0,0288$

$$u_{1-\frac{\alpha}{2}} = \frac{\varepsilon\sqrt{n}}{\sqrt{f(1-f)}} = \frac{0,0288\sqrt{400}}{\sqrt{0,2.0,8}} = 1,44.$$

Từ bảng phân vị chuẩn ta thấy $1,44 = u_{0,925}$. Vậy $1-\frac{\alpha}{2} = 0,925$ suy ra độ tin cậy: $1-\alpha = 0,85$.
Ví dụ 11+. Một doanh nghiệp dự định đưa một sản phẩm mới vào tiêu thụ ở một vùng dân cư có 2500000 người. Nghiên cứu thị trường đối với 3500 người thấy có 1500 người sẵn sàng mua sản phẩm đó.
a. Với độ tin cậy 95%, hãy dự đoán thị phần tiềm năng của doanh nghiệp.
b. Dự đoán số lượng khách hàng tiềm năng mà doanh nghiệp hy vọng sẽ có được ở thị trường mới là bao nhiêu?
Đây là bài toán liên quan đến dự đoán thống kê. Hoạt động dự đoán này sẽ được giải quyết theo con đường lý thuyết, nghĩa là sinh viên phải vận dụng những kiến thức về ước lượng tham số để giải quyết vấn đề. Kết quả của dự đoán sẽ được thể hiện dưới dạng khoảng. Để dự đoán thị phần tiềm năng của doanh nghiệp, sinh viên phải tìm được khoảng tin cậy cho tỷ lệ p: tỷ lệ khách hàng sẵn sàng mua sản phẩm, với điều kiện $n > 5$ và $\frac{\left|\sqrt{\frac{p}{p-1}} - \sqrt{\frac{1-p}{p}}\right|}{\sqrt{n}} < 0,3$.

Khoảng tin cậy cho p là: $\left(f - \frac{\sqrt{f(f-1)}}{\sqrt{n}}\phi_0^{-1}\left(\frac{y}{2}\right); f - \frac{\sqrt{f(f-1)}}{\sqrt{n}}\phi_0^{-1}\left(\frac{y}{2}\right)\right)$. Với tần suất mẫu $f = 0,428; \phi_0^{-1}\left(\frac{y}{2}\right) = 1,96$. Ta có: $0,412 < p < 0,444$.
Từ đó sinh viên có thể dự đoán được thị trường tiềm năng của doanh nghiệp chiếm $41,2\%$ đến $44,4\%$.
Muốn dự đoán được số lượng khách hàng tiềm năng của doanh nghiệp, sinh viên phải tiến hành phân tích, so sánh, khái quát hóa để nhận ra tỷ lệ $\frac{N}{2500000}$ thuộc khoảng tin cậy p, với N là số lượng khách hàng tiềm năng.
Khi đó, các em sẽ lập được mối quan hệ; $0,412 < \frac{N}{2500000} < 0,444$.
Suy ra: $1030000 < N < 1110000$.
Vậy số lượng khách tiềm năng mà doanh nghiệp hy vọng sẽ có được ở thị trường mới từ 1030000 đến 1110000 người.

3.5.5. Khoảng tin cậy cho phương sai (Ước lượng phương sai của đại lượng ngẫu nhiên phân phối theo quy luật chuẩn)

Giả sử trong tổng thể đại lượng ngẫu nhiên gốc X phân phối chuẩn $N(\mu, \sigma^2)$ nhưng chưa biết tham số σ^2 của nó. Để ước lượng σ^2 từ tổng thể ta lập mẫu ngẫu nhiên kích thức n: $W = (X_1, X_2, ..., X_n)$.
Ta xét các trường hợp sau đây:
1. Đã biết kỳ vọng toán μ của đại lượng ngẫu nhiên gốc X
Xét thống kê:

$$G = \frac{nS^{*2}}{\sigma^2}; \qquad \text{với} \quad S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Thống kê G ở trên phân phối theo quy luật "khi bình phương" với n bậc tự do $\chi^2(n)$. Do đó với độ tin cậy $1 - \alpha$ cho trước có thể tìm được cặp giá trị α_1 và α_2 sao cho $\alpha_1 + \alpha_2 = \alpha$. Khi đó với hai phân vị "khi bình phương" n bậc tự do tương ứng $\chi_{\alpha_1}^{2(n)}$ và $\chi_{1-\alpha_2}^{2(n)}$ ta có:

$$P\left(\chi_{\alpha_1}^{2(n)} < \frac{nS^{*2}}{\sigma^2} < \chi_{1-\alpha_2}^{2(n)}\right) = (1 - \alpha_2) - \alpha_1$$

$$\Leftrightarrow P\left(\frac{nS^{*2}}{\chi_{1-\alpha_2}^{2(n)}} < \sigma^2 < \frac{nS^{*2}}{\chi_{\alpha_1}^{2(n)}}\right) = 1 - \alpha.$$

- Chọn $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$ ta có khoảng tin cậy của σ^2 như sau:

$$\left(\frac{nS^{*2}}{\chi_{1-\frac{\alpha}{2}}^{2(n)}}; \frac{nS^{*2}}{\chi_{\frac{\alpha}{2}}^{2(n)}}\right).$$

Ta thấy rằng khoảng tin cậy này không đối xứng.

- Chọn $\alpha_1 = 0, \alpha_2 = \alpha$ ta có khoảng tin cậy bên phải dùng để ước lượng giá trị tối thiểu của σ^2 như sau:

$$\left(\frac{nS^{*2}}{\chi^{2(n)}_{1-\alpha}}; +\infty \right).$$

- Chọn $\alpha_1 = \alpha, \alpha_2 = 0$ ta có khoảng tin cậy bên trái dùng để ước lượng giá trị tối thiểu của σ^2 như sau:

$$\left(0; \frac{nS^{*2}}{\chi^{2(n)}_{\alpha}} \right).$$

●**Ví dụ 3.12** Mức hao phí nguyên liệu cho một đơn vị sản phẩm là đại lượng ngẫu nhiên phân phối chuẩn với trung bình là 20g. Để ước lượng mức độ phân tán của mức hao phí này người ta cân thử 25 sản phẩm và thu được kết quả sau:

Hao phí nguyên liệu (g)	19,5	20,0	20,5
Số sản phẩm tương ứng	5	18	2

Với độ tin cậy $1 - \alpha = 0,90$ hãy ước lượng σ^2 nếu $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$

Lời giải.

Gọi X là mức hao phí nguyên liệu cho một đơn vị sản phẩm, X tuân theo quy luật chuẩn $N(\mu = 20, \sigma^2)$. Đây là bài toán ước lượng phương sai của phân phối chuẩn khi đã biết μ.

Tra bảng phân bị Khi-bình phương:

$$\chi^{2(n)}_{\frac{\alpha}{2}} = \chi^{2(25)}_{0,05} = 14,6; \qquad \chi^{2(n)}_{1-\frac{\alpha}{2}} = \chi^{2(25)}_{0,95} = 37,7.$$

Lập bảng tính s^{*2}:

x_i	n_i	$x_i - \mu$	$n_i(x_i - \mu)$	$n_i(x_i - \mu)^2$
19,5	5	-0,5	-2,5	1,25
20,0	18	0	0	0
20,5	2	0,5	1	0,5
	$n = \sum n_i = 25$			$\sum n_i(x_i - \mu)^2 = 1,75$

$$s^{*2} = \frac{1}{n} \sum n_i(x_i - \mu)^2 = \frac{1,75}{25} = 0,07.$$

Khoảng tin cậy của σ^2 là:

$$\sigma^2 \in \left(\frac{n.s^{*2}}{\chi^{2(n)}_{1-\frac{\alpha}{2}}}; \frac{n.s^{*2}}{\chi^{2(n)}_{\frac{\alpha}{2}}} \right) = \left(\frac{25.0,07}{37,7}; \frac{25.0,07}{14,6} \right) = (0,0464; 0,1198).$$

2. Chưa biết kỳ vọng toán μ của đại lượng ngẫu nhiên gốc X

Ta chọn thông kê:

$$G = \frac{(n-1)S'^2}{\sigma^2}$$

trong đó S'^2 là phương sai điều chỉnh mẫu.

Thống kê G ở trên phân phối theo quy luật "khi bình phương" với n-1 bậc tự do $\chi^2(n-1)$.

Do đó với độ tin cậy $1-\alpha$ cho trước có thể tìm được cặp giá trị α_1 và α_2 sao cho $\alpha_1+\alpha_2=\alpha$.

Khi đó với hai phân vị "khi bình phương" $n-1$ bậc tự do tương ứng $\chi_{\alpha_1}^{2(n-1)}$ và $\chi_{1-\alpha_2}^{2(n-1)}$ ta có:

$$P\left(\chi_{\alpha_1}^{2(n-1)} < \frac{(n-1)S'^2}{\sigma^2} < \chi_{1-\alpha_2}^{2(n-1)}\right) = (1-\alpha_2)-\alpha_1$$

$$\Leftrightarrow P\left(\frac{(n-1)S'^2}{\chi_{1-\alpha_2}^{2(n-1)}} < \sigma^2 < \frac{(n-1)S'^2}{\chi_{\alpha_1}^{2(n-1)}}\right) = 1-\alpha.$$

- Chọn $\alpha_1=\alpha_2=\frac{\alpha}{2}$ ta có khoảng tin cậy của σ^2 như sau:

$$\left(\frac{(n-1)S'^2}{\chi_{1-\frac{\alpha}{2}}^{2(n-1)}}; \frac{(n-1)S'^2}{\chi_{\frac{\alpha}{2}}^{2(n-1)}}\right).$$

- Chọn $\alpha_1=0, \alpha_2=\alpha$ ta có khoảng tin cậy bên phải dùng để ước lượng giá trị tối thiểu của σ^2 như sau:

$$\left(\frac{(n-1)S'^2}{\chi_{1-\alpha}^{2(n-1)}}; +\infty\right).$$

- Chọn $\alpha_1=\alpha, \alpha_2=0$ ta có khoảng tin cậy bên trái dùng để ước lượng giá trị tối thiểu của σ^2 như sau:

$$\left(0; \frac{(n-1)S'^2}{\chi_{\alpha}^{2(n-1)}}\right).$$

•**Ví dụ 3.13** Lãi suất cổ phiếu của một công ty trong vòng 5 năm qua là:

15%; 10%; 20%; 7%; 14%;

a) Biết lãi suất cổ phiếu là biến ngẫu nhiên phân phối chuẩn, hãy ước lượng độ phân tán của lãi suất cổ phiếu của công ty đó với độ tin cậy 90%.

b) Hãy ước lượng độ phân tán tối đa của lãi suất cổ phiếu của công ty với độ tin cậy 99%.

Lời giải.

Gọi X là lãi suất cổ phiếu của công ty, X tuân theo quy luật chuẩn $N(\mu, \sigma^2)$. Đây là bài toán ước lượng σ^2 của X với μ chưa biết.

a) Từ độ tin cậy $1-\alpha = 0,9$ suy ra $\frac{\alpha}{2} = 0,05$. Tra bảng phân vị Khi-bình phương:

$$\chi^{2(n-1)}_{\frac{\alpha}{2}} = \chi^{2(4)}_{0,05} = 0,711; \qquad \chi^{2(n)}_{1-\frac{\alpha}{2}} = \chi^{2(4)}_{0,95} = 9,5.$$

Tính s'^2: Từ giả thiết dễ tính được: $\sum x_i = 66$; $\quad \sum x_i^2 = 970$. Do đó:

$$\overline{x} = \frac{66}{5} = 13,2(\%); \qquad s'^2 = \frac{n}{n-1}\left(\frac{\sum x_i^2}{n} - \overline{x}^2\right) = \frac{5}{4}\left(\frac{970}{5} - 13,2^2\right) = 24,7.$$

Khoảng tin cậy của σ^2 là:

$$\sigma^2 \in \left(\frac{(n-1)s'^2}{\chi^{2(n-1)}_{1-\frac{\alpha}{2}}}; \frac{(n-1)s'^2}{\chi^{2(n-1)}_{\frac{\alpha}{2}}}\right) = \left(\frac{4.24,7}{9,5}; \frac{4.24,7}{0,711}\right) = (10,4; 138,959).$$

b) Từ độ tin cậy $1-\alpha = 0,99$, tra bảng phân vị Khi-bình phương: $\chi^{2(n-1)}_{\alpha} = \chi^{2(4)}_{0,01} = 0,297$
Khoảng tin cậy bên trái của σ^2 là:

$$\sigma^2 \in \left(0; \frac{(n-1)s'^2}{\chi^{2(n-1)}_{\alpha}}\right) = \left(0; \frac{4.24,7}{0,297}\right) = (0; 332,660).$$

Vậy $\sigma^2 < 332,66$ hay $\sigma < 18,239\%$.

Bài tập chương 3

Mẫu ngẫu nhiên và các tham số đặc trưng của nó

Bài 3.1. Cho 8 kết quả đo đạc về một đại lượng X bởi cùng một máy không có sai lầm hệ thống:
369;378;315;420;385;401;372;383
Hãy tính $\overline{x}, s^2, s'$.
ĐS: $\overline{x} = 377,875; s^2 = 806,609; s' = 30,362$.

Bài 3.2. Theo dõi thời gian hoàn thành một sản phẩm (phút) ở hai nhóm công nhân, ta thu được số liệu sau:
Nhóm 1:

X_1 (phút)	42	44	45	58	60	64
Số người (n_i)	4	5	20	10	8	3

Nhóm 2:

X_2 (phút)	46	48	51
Số người (n_i)	2	40	8

Tính trung bình và độ lệch tiêu chuẩn mẫu điều chỉnh của hai mẫu cụ thể nói trên. Nêu nhận xét.
ĐS: $\overline{x_1} = 50,8; s'_1 = 7,735; \overline{x_2} = 48,4; s'_2 = 1,21$.

Bài 3.3. Đo chiều cao của 100 thanh niên từ 18 đến 22 tuổi ở tỉnh A, ta thu được bảng số liệu:

Chiều cao X (cm)	154-158	158-162	162-166	166-170	170-174	174-178	178-182
Số thanh niên (n_i)	10	14	26	28	12	8	2

Hãy tính chiều cao trung bình $(\overline{x})$ và độ lệch mẫu điều chỉnh (s').
ĐS: $\overline{x} = 166; s' = 5,82$.
Bài 3.4. Các kết quả đo độ bền các sợi chỉ được cho trong bảng số liệu dưới đây (đơn vị gam):

Độ bền của sợi (gam)	120-140	140-160	160-180	180-200	200-220	220-240	240-260	260-280
Số các sợi (n_i)	1	4	10	14	12	6	2	1

Hãy xác định độ bền trung bình, phương sai và độ lệch chuẩn mẫu điều chỉnh của mẫu trên.
ĐS: $\overline{x} = 195,2; s^2 = 821,96; s' = 28,8$.
Bài 3.5. Để nghiên cứu tuổi thọ (X) của một loại bóng đèn, người ta thắp thử 100 bóng và có số liệu sau:

Tuổi thọ (giờ)	Số bóng tương ứng (n_i)
1010-1030	2
1030-1050	3
1050-1070	8
1070-1090	13
1090-1110	25
1110-1130	20
1130-1150	12
1150-1170	10
1170-1190	6
1190-1210	1

a) Vẽ biểu đồ tần số và biểu đồ tần suất của bảng phân phối thực nghiệm trên.
b) Tính tuổi thọ trung bình và độ lệch tiêu chuẩn mẫu điều chỉnh của tuổi thọ loại bóng đèn nói trên.
ĐS: b) $\overline{x} = 1111,1; s' = 37$.
Bài 3.6. Sau khi cải tiến kỹ thuật, người ta thắp thử 100 bóng, kết quả là:

Tuổi thọ (giờ)	1150	1160	1170	1180	1190	1200
Số bóng (n_i)	10	15	20	30	15	10

a) Vẽ đa giác tần số, đa giác tần suất của bảng phân phối tần số trên.
b) So sánh trung bình và độ lệch tiêu chuẩn điều chỉnh của tuổi thọ loại bóng đèn nói trên trước và sau cải tiến kỹ thuật qua hai mẫu cụ thể được cho ở bài 5 và 6.
ĐS: b) $\overline{x} = 1175,5; s' = 14,3$.
Bài 3.7. Các số liệu của việc phân tích một số mẫu quặng sắt được cho ở bảng dưới đây, ở đó X (%) là hàm lượng oxyt sắt, Y (%) làm hàm lượng tạp chất.

Y \ X	3 - 9	9 - 15	15 - 21	21 - 27	27 - 33
35			5	4	2
45		1	6	3	
55	4	6	14	2	
65	5	8	3		

a) Tính $\overline{x}, s'_x, \overline{y}, s'_y$;
b) Những mẫu quặng sắt có hàm lượng tạp chất nhỏ hơn 21% được xếp và loại A, hãy tính các giá trị trung bình mẫu và độ lệch tiêu chuẩn mẫu hiệu chỉnh $\overline{x}, s'_x, \overline{y}, s'_y$ của quặng sắt loại A;
c) Tính tần suất mẫu của các quặng sắt loại A.
ĐS: a) $\overline{x} = 52,460; s'_x = 10,313; \overline{y} = 16,095; s'_y = 5,983$;
b) $\overline{x_A} = 54,808; s'_{x_A} = 9,180; \overline{y_A} = 14,192; s'_{y_A} = 4,606$;
c) $f = 0,825$.
Giải: a, ***Từ bảng trên ta có bảng tần số của** X

X x_i	35	45	55	65
Số quặng sắt n_i	11	10	26	16

Lập bảng tính:

x_i	n_i	$n_i x_i$	$n_i x_i^2$
35	11	385	13475
45	10	450	20250
55	26	1430	78650
65	16	1040	67600
	$\sum n_i = 63$	$\sum n_i x_i = 3305$	$\sum n_i x_i^2 = 179975$

Từ đó có kết quả sau:

$$\overline{x} = \frac{1}{n}\sum n_i x_i = \frac{1}{63} \cdot 3305 = 52,46.$$

$$S'_x = \sqrt{\frac{n}{n-1}\left(\frac{\sum n_i x_i^2}{n} - \overline{x}^2\right)} = \sqrt{\frac{63}{62}\left(\frac{177975}{63} - 52,46^2\right)} = 10,31.$$

***Từ bảng trên ta có bảng tần số của** Y

Y (y_i)	6	12	18	24	30
Số quặng sắt (m_j)	9	15	28	9	2

Lập bảng tính:

y_j	m_j	m_jy_j	$m_jy_j^2$
6	9	54	324
12	15	180	2160
18	28	504	9072
24	9	216	5184
30	2	60	1800
	$\sum m_j = 63$	$\sum m_jy_j = 1014$	$\sum m_jy_j^2 = 18540$

Từ đó có kết quả sau:

$$\overline{y} = \frac{1}{n}\sum m_jy_j = \frac{1}{63}\cdot 1014 = 16,095.$$

$$S'_y = \sqrt{\frac{n}{n-1}\left(\frac{\sum m_jy_j^2}{n} - \overline{y}^2\right)} = \sqrt{\frac{63}{62}\left(\frac{18540}{63} - 16,095^2\right)} = 5,983.$$

b, ***Từ bảng và các điều kiện, ta có bảng tần số của X của nhóm A:**

X(x_{i_A})	35	45	55	65
Số quặng sắt (n_{i_A})	5	7	24	16

Lập bảng tính:

x_{i_A}	n_{i_A}	$x_{i_A}n_{i_A}$	$n_{i_A}x_{i_A}^2$
35	5	175	6125
45	7	315	14175
55	24	1320	72600
65	16	1040	67600
	$\sum n_{i_A} = 52$	$\sum n_{i_A}x_{i_A} = 2850$	$\sum n_{i_A}x_{i_A}^2 = 160500$

Từ đó có kết quả sau:

$$\overline{x_A} = \frac{1}{n}\sum x_An_{i_A} = \frac{1}{52}\cdot 2850 = 54,81.$$

$$S'_{x_A} = \sqrt{\frac{n}{n-1}\left(\frac{\sum n_{i_A}x_A^2}{n} - \overline{x_A}^2\right)} = \sqrt{\frac{52}{51}\left(\frac{160500}{52} - 54,81^2\right)} = 9,17.$$

***Từ bảng và các điều kiện, ta có bảng tần số của Y của nhóm A:**

Y(y_A)	6	12	18
Số quặng sắt (m_{j_A})	9	15	28

Lập bảng tính:

y_A	m_{j_A}	$y_Am_{j_A}$	$m_{j_A}y_A^2$
6	9	54	324
12	15	180	2160
18	28	504	9072
	$\sum m_{j_A} = 52$	$\sum y_Am_{j_A} = 738$	$\sum m_{j_A}y_A^2 = 11556$

Từ đó có kết quả sau:

$$\overline{y_A} = \frac{1}{n}\sum y_A m_{j_A} = \frac{1}{52}\cdot 738 = 14,192.$$

$$S'_{y_A} = \sqrt{\frac{n}{n-1}\left(\frac{\sum m_{j_A} y_A^2}{n} - \overline{y_A}^2\right)} = \sqrt{\frac{52}{51}\left(\frac{11556}{52} - 14,192^2\right)} = 4,607.$$

c, Tần suất quặng sắt loại A là:

$$f = \frac{5+1+6+4+6+14+5+8+3}{63} = 0,825.$$

Bài 3.8. Gọi X là "khối lượng sản phẩm sản xuất ra" và Y là "giá thành sản phẩm", điều tra ở 30 xí nghiệp cùng loại sản phẩm ta có kết quả sau:

Y \ 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

a) Tính trung bình mẫu và độ lệch tiêu chuẩn mẫu điều chỉnh $\overline{x}, s'_x, \overline{y}, s'_y$

b) 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 mẫu điều chỉnh $\overline{x}, s'_x, \overline{y}, s'_y$ của các xí nghiệp loại I.

c) Tính tần suất mẫu của các xí nghiệp loại I.

ĐS: a) $\overline{x} = 155; s'_x = 66,111; \overline{y} = 136; s'_y = 21,107;$

b) $\overline{x_I} = 140,625; s'_{x_I} = 41,708; \overline{y_I} = 137; s'_{y_I} = 10;$

c) $f = 0,533.$

Ước lượng các tham số lượng đặc trưng của đại lượng ngẫu nhiên

Bài 3.9. Hao phí nguyên liệu cho một đơn vị sản phẩm là một đại lượng ngẫu nhiên tuân theo quy luật chuẩn $N\left(\mu, \sigma_X^2\right)$ với độ lệch tiêu chuẩn $\sigma_X = 0,03$. Người ta sản xuất thử 36 sản phẩm và thu được bảng số liệu sau:

Mức hao phí nguyên liệu (gam)	Số sản phẩm tương ứng (n_i)
19,5 - 19,7	8
19,7 - 19,9	8
19,9 - 20,1	18
20,1 - 20,3	2

Với độ tin cậy 95%, bằng khoảng tin cậy đối xứng hãy ước lượng mức hao phí nguyên liệu trung bình cho 1 đơn vị sản phẩm.

ĐS: (19,868; 19,8876).

Bài 3.10. Để định mức thời gian gia công một chi tiết máy, người ta theo dõi ngẫu nhiên quá trình gia công 25 chi tiết và thu được số liệu sau:

Thời gian gia công (phút)	Số chi tiết máy tương ứng (n_i)
15 - 17	1
17 - 19	3
19 - 21	4
21 - 23	12
23 - 25	3
25 -27	2

Bằng khoảng tin cậy đối xứng hãy ước lượng thời gian gia công trung bình một chi tiết máy với độ tin cậy $1 - \alpha = 0,95$. Giả thiết thời gian gia công chi tiết là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn.

ĐS: (20,53; 22,51). **Bài 3.11.** Cân thử 25 bao bột mỳ, người ta tính được trọng lượng trung bình một bao là $\overline{x} = 40kg$, độ lệch tiêu chuẩn mẫu điều chỉnh $s' = 5kg$. Với độ tin cậy 95%, bằng khoảng tin cậy đối xứng hãy ước lượng trọng lượng trung bình của bao bột mỳ. Giả thiết trong lượng của bao bột mỳ là đại lượng ngẫu nhiên tuân theo quy luật chuẩn.

ĐS: (37,936; 42,064).

Bài 3.12. Để ước lượng chiều dày trung bình của các tấm vật liệu do một xí nghiệp sản xuất, người ta tiến hành đo 5 tấm và thu được kết quả sau:

$2,015mm$; $2,025mm$; $2,015mm$; $2,020mm$; $2,015mm$

Dựa vào số liệu trên hãy ước lượng chiều dày trung bình các tấm vật liệu do một xí nghiệp sản xuất bằng khoảng tin cậy đối xứng với độ tin cậy $1 - \alpha = 0,95$. (Biết rằng chiều dày các tấm vật liệu là đại lượng tuân theo quy luật phân phối chuẩn).

ĐS: (2,006; 2,11).

Bài 3.13. Lấy 50 con sợi để xác định độ bền trung bình, ta có số liệu sau: (biết độ bền X - kg/cm^2 là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn).

X	0,6-0,8	0,8-1,0	1,0-1,2	1,2-1,4	1,4-1,6	1,6-1,8	1,8-2,0	2,0-2,2	2,2-2,4
n_i	1	2	7	10	11	9	6	3	1

Hãy ước lượng độ bền trung bình của loại sợi này bằng khoảng tin cậy đối xứng với hệ số tin cậy 0,98.

ĐS: (1,386; 1,614).

Bài 3.14. Để xác định giá trung bình đối với một loại hàng hoá trên thị trường, người ta điều tra ngẫu nhiên tại 100 cửa hàng thu được bảng số liệu sau đây:

Giá X (đồng)	83	85	87	89	91	93	95	97	99	101
Số cửa hàng (n_i)	6	7	12	15	30	10	8	6	4	2

Với độ tin cậy 97% hãy ước lượng giá trung bình của loại hàng đó tại thời điểm đang xét, biết rằng giá hàng hoá X là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn.

ĐS: (89,815; 91,625).

Bài 3.15. Chiều dài (X) một loại sản phẩm do một máy tự động sản xuất là đại lượng ngẫu

nhiên tân theo quy luật phân phối chuẩn với độ lệch tiêu chuẩn $\sigma_X = 3cm$.
Để ước lượng chiều dài trung bình của loại sản phẩm nói trên với độ tin cậy $1 - \alpha = 0,95$, người ta tiến hành đo 20 sản phẩm và thấy chiều dài trung bình của 20 sản phẩm đó là $\overline{x} = 50cm$.
a) Tìm khoảng tin cậy đối xứng của chiều dài trung bình loại sản phẩm đó.
b) Để ước lượng chiều dài trung bình loại sản phẩm đó với độ tin cậy $1 - \alpha = 0,99$, độ dài khoảng tin cậy đối xứng không vượt quá $0,6cm$ thì phải đeo bao nhiêu sản phẩm?
ĐS: a) (48,685; 51,315); b) 664 sp.
Bài 3.16. Kiểm tra 100 bóng điện có tuổi thọ trung bình $\overline{x} = 3000$ (giờ). Nếu ước lượng tuổi thọ trung bình của loại bóng đèn đó với độ sai lệch không quá 3,92 giờ về giá trị tuyệt đối thì độ tin cậy của ước lượng là bao nhiêu? Kết luận về tuổi thọ trung bình của loại bóng điện đó nếu tuổi thọ bóng điện là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn với $\sigma = 20$ giờ.
ĐS: a) 95%; b) (2996,08; 3003,92).
Bài 3.17. Năng suất giống ngô A ở một vùng được báo lên qua 25 điểm thu hoạch và có kết quả sau:

Năng suất (X - tạ/ha)	7	9	11	13	17
Số điểm thu hoạch (n_i)	2	7	12	3	1

Với độ tin cậy 95%, hãy tính năng suất trung bình tối thiểu của vùng này, biết rằng năng suất ngô của vùng này là đại lượng ngẫu nhiên tuân theo quy luật chuẩn.
ĐS: $\mu \geq 9,888$.
Bài 3.18. Để định mức thời gian gia công một chi tiết máy người ta theo dõi ngẫu nhiên quá trình gia công một số chi tiết và thu được bảng số liệu sau:

Thời gian X (phút)	14	16	18	20	24
Số chi tiết (n_i)	4	12	22	8	4

Với độ tin cậy $1 - \alpha = 0,99$, hãy ước lượng thời gian gia công trung bình tối đa đối với loại chi tiết đó. Giả thiết thời gian gia công chi tiết là đại lượng ngẫu nhiên tuân theo quy luật phân phối chuẩn.
ĐS: $\mu \leq 18,797$.
Bài 3.19. Hãy ước lượng tỷ lệ chính phẩm của một nhà máy bằng khoảng tin cậy đối xứng với độ tin cậy 0,95, biết rằng kiểm tra 100 sản phẩm của nhà máy thấy có 10 phế phẩm.
ĐS: (0,8412; 0,9588).
Bài 3.20. Mở thử 200 hộp của một kho đồ hộp, người ta thấy có 8 hộp bị biến chất. Với độ tin cậy 97%, bằng khoảng tin cậy đối xứng, hãy ước lượng tỷ lệ đồ hộp bị biến chất ở kho đó.
ĐS: (0,01; 0,07).
Bài 3.21. Gieo thử 400 hạt giống thì thấy có 20 hạt không nảy mầm. Tỷ lệ hạt giống không nảy mầm tối đa là bao nhiêu? Hãy kết luận với độ tin cậy 98%.
ĐS: $p \leq 0,072$. **Bài 3.22.** Trong đợt vận động bầu cử tổng thống người ta phỏng vấn ngẫu nhiên 1600 cử tri thì được biết 960 người trong đó sẽ bỏ phiếu cho ứng cử viên A. Với độ tin cậy 95%, ứng cử viên A chiếm được tối thiểu bao nhiêu % số phiếu bầu?

ĐS: $p \geq 0,5716$.

Bài 3.23. Để điều tra số cá trong một hồ, cơ quan quản lý đánh bắt 2000 con cá, đánh dấu rồi thả xuống hồ. Lần sau bắt lại 400 con, được 80 con có dấu. Bằng khoảng tin cậy đối xứng, hãy ước lượng số cá trong hồ với độ tin cậy 97%.

ĐS: (8230; 12739).

Bài 3.24. Bắt 1000 thú hiếm tại một vùng, đánh dấu rồi thả ra. Lần sau bắt lại được 300 con, có 60 con có dấu. Hãy ước lượng số tối đa thú hiếm đó với độ tin cậy 95%.

ĐS: $N \leq 6173$.

Bài 3.25. Cần phải lập một mẫu ngẫu nhiên với kích thước là bao nhiêu để tỷ lệ phế phẩm của mẫu là 0,2, độ dài khoảng tin cậy đối xứng là 0,05 và độ tin cậy của ước lượng là 0,95?

ĐS: $N \geq 983$.

Bài 3.26. Tỷ lệ nảy mầm của một loại hạt giống là 90%. Cần ước lượng tỷ lệ nảy mầm của hạt giống đó với độ tin cậy 0,95 và độ dài khoảng tin cậy không quá 2% thì phải gieo bao nhiêu hạt?

ĐS: $N \geq 3456$.

Chương 4

Số gần đúng và Sai số

4.1. Khái niệm về số gần đúng

4.1.1. Sai số tuyệt đối, sai số tương đối

Trong tính toán ta thường phải làm việc với các giá trị gần đúng của các đại lượng. Ta nói a là số gần đúng A, nếu a không sai khác A nhiều. Đại lượng $\Delta := |a - A|$ gọi là *sai số thực sự* của a. Do không biết A nên cũng không biết Δ. Tuy nhiên ta có thể tìm được số dương Δa thỏa mãn điều kiện:

$$|a - A| \leq \Delta a \tag{4.1}$$

hay $a - \Delta a \leq A \leq a + \Delta a$. Số dương Δa này gọi là sai số tuyệt đối của a. Rõ ràng nếu Δa đã là sai số tuyệt đối của a thì mọi số $\Delta' > \Delta a$ đều có thể xem là sai số tuyệt đối của a. Vì vậy trong những điều kiện cụ thể người ta chọn Δa là số dương bé nhất có thể được thỏa mãn (4.1).

Nếu a là số gần đúng của A có sai số tuyệt đối là Δa thì ta quy ước viết:

$$A = a \pm \Delta a \tag{4.2}$$

Tỷ số $\delta a = \dfrac{\Delta a}{|a|}$ được gọi là sai số tương đối của a. Ta suy ra $\Delta a = |a|\delta a$, do đó (4.2) cũng có thể viết

$$A = |a|(1 \pm \delta a) \tag{4.3}$$

⊕ **Nhận xét** Δa có cùng thứ nguyên với a, còn δa là số không có thứ nguyên và được biểu diễn bằng %.

• **Ví dụ 1** *Giả sử* $A = \pi$; $a = 3,14$. *Do* $3,14 < A < 3,15 = 3,14 + 0,01$ *nên ta có thể lấy* $\Delta a = 0,01$. *Mặt khác,* $3,14 < A < 3,142 = 3,14 + 0,002$ *do đó có thể lấy* $\Delta a = 0,002$.

• **Ví dụ 2** *Đo độ dài hai đoạn thẳng AB, CD ta được* $a = 10cm$ *và* $b = 1cm$ *với* $\Delta a = \Delta b = 0,01cm$. *Khi đó ta có* $\delta a = \dfrac{0,01}{10} = 0,1\%$ *còn* $\delta b = \dfrac{0,01}{1} = 1\%$ *hay* $\delta b = 10\delta a$.

Hiển nhiên phép đo a chính xác hơn hẳn phép đo b mặc dù $\Delta a = \Delta b$. Như vậy độ chính xác của một phép đo phản ánh qua sai số tương đối.

4.1.2. Sự làm tròn số, sai số làm tròn

Một số thập phân a đều biểu diễn được dưới dạng:

$$a = \pm(\beta_p 10^p + \beta_{p-1}10^{p-1} + \cdots + \beta_{p-s}10^{p-s})$$

trong đó β_i $(i = p, p-1, \ldots, p-s)$ là các số nguyên dương từ 0 đến 9. Chẳng hạn $a = 572,96 = 5.10^2 + 7.10^1 + 2.10^0 + 9.10^{-1} + 6.10^{-2}$, ở đây $\beta_2 = 5, \beta_1 = 7, \beta_0 = 2, \beta_{-1} = 9, \beta_{-2} = 6$.

Làm tròn a là bỏ đi một số các chữ số bên phải a để được một số $\overline{a}$ ngắn gọn hơn và gần đúng nhất với a.

⊙ **Qui tắc thu gọn:** Giả sử

$$a = \beta_p 10^p + \cdots + \beta_j 10^j + \cdots + \beta_{p-s}10^{p-s}$$

và ta giữ lại đến số hạng thứ j. Gọi phần vứt bỏ là μ, ta đặt

$$\overline{a} = \beta_p 10^p + \cdots + \beta_{j+1}10^{j+1} + \widetilde{\beta}_j 10^j,$$

trong đó:

$$\widetilde{\beta}_j := \begin{cases} \beta_j + 1 & \text{nếu} \quad 0,5.10^j < \mu < 10^j \\ \beta_j & \text{nếu} \quad 0 < \mu < 0,5.10^j \end{cases}$$

Trường hợp $\mu = 0,5.10^j$ thì $\overline{\beta}_j = \beta_j$ nếu β_j là chẵn và $\overline{\beta}_j = \beta_j + 1$ nếu β_j là lẻ vì tính toán với số chẵn thuận tiện hơn.

• **Ví dụ 3** *Thu gọn đến 2 chữ số sau dấu phẩy các số sau:*

$a = 572,96573$ $\overline{a} = 572,97;$ $b = 45,75346$ $\overline{b} = 45,75$

$c = 301,38500$ $\overline{c} = 301,38;$ $d = 432,23500$ $\overline{d} = 432,24$

Sai số làm tròn $\theta a \geq 0$ là số thỏa mãn điều kiện:

$$|\overline{a} - a| \leq \theta a$$

Vì $a = \beta_p 10^p + \cdots + \beta_j 10^j + \mu$, còn $\overline{a} = \beta_p 10^p + \cdots + \beta_{j+1}10^{j+1} + \widetilde{\beta}_j 10^j$ nên

$$|a - \overline{a}| = |(\beta_j - \widetilde{\beta}_j)10^j + \mu| < 0,5.10^j.$$

⊙ **Sai số của số đã thu gọn:** Ta có

$$\overline{a} - A = \overline{a} - a + a - A$$

do đó $|\overline{a} - A| \leq |\overline{a} - a| + |a - A| \leq \Delta a + \theta a$

Vậy có thể lấy: $\Delta\overline{a} = \Delta a + \theta a$, tức là sau khi thu gọn, sai số tuyệt đối tăng lên.

⊙ **Ảnh hưởng của sai số thu gọn:** Ta xét một ví dụ. Áp dụng công thức nhị thức Niutơn ta có công thức đúng:

$$(\sqrt{2}-1)^{10} = 3363 - 2378\sqrt{2} \tag{4.4}$$

với $\sqrt{2} = 1,41421356...$

Bây giờ ta tính hai vế của (4.4) bằng cách thay $\sqrt{2}$ bởi các số quy tròn:

$\sqrt{2}$	Vế trái	Vế phải
1,4	0,0001048576	33,8
1,41	0,00013422659	10,02
1,414	0,000147912	0,508
1,41421	0,0001866399	0,00862
1,414213563	0,00014867678	0,0001472

Sự khác biệt giữa các giá trị tính ra của hai vế chứng tỏ rằng sai số quy tròn có thể có gây ra những kết quả không mong muốn trong quá trình tính toán.

4.2. Cách viết số xấp xỉ

4.2.1. Chữ số có nghĩa

Chữ số có nghĩa là mọi chữ số khác "0" và cả "0" nếu nó kẹp giữa hai chữ số có nghĩa hoặc nó đại diện cho hàng được giữ lại.

• **Ví dụ 4** $a = 0,0030140$. *Ba chữ số "0" đầu không có nghĩa.*

4.2.2. Chữ số chắc

Giả sử a là giá trị gần đúng của A với sai số tuyệt đối Δa. Ta biết rằng mọi số thập phân a đều viết được dưới dạng: $a = \pm \sum \beta_i 10^i$, trong đó β_i là những số nguyên từ 0 đến 9. Chữ số có nghĩa β_i của a gọi là chữ số chắc, nếu $\Delta a \leq \omega \times 10^i$, trong đó ω là tham số cho trước. Tham số ω được chọn để một chữ số vốn đã chắc sau khi thu gọn vẫn là chữ số chắc. Giả sử chữ số chắc cuối cùng của a trước khi thu gọn là β_i. Để β_{i+1} và các chữ số trước nó vẫn chắc, phải có:

$$\Delta a + \theta a \leq \omega \times 10^{i+1} \Rightarrow \omega \times 10^i + 0,5 \times 10^{i+1} \leq \omega \times 10^{i+1}$$

hay $\omega \geq 5/9$. Đặc biệt β_i gọi là chữ số chắc theo nghĩa hẹp (nghĩa rộng) nếu $\omega = 0,5$ ($\omega = 1$).

• **Ví dụ 5** *Cho* $a = 65,8274$ *với* $\Delta a = 0,0043$ *thì các chữ số 6, 5, 8, 2 là chắc, còn các chữ số 7,4 là không chắc. Nếu* $\Delta a = 0,0067$ *thì các chữ số 6, 5, 8 là chắc còn các chữ số 2,7,4 là không chắc.*

Như vậy, sai số tuyệt đối Δa tách số gần đúng a thành 2 phần: phần ở bên trái gồm các chữ số chắc, phần còn lại ở bên phải gồm các chữ số không chắc. Độ chính xác của một số gần đúng được đánh giá không phải ở chỗ số ấy có nhiều chữ số mà ở chỗ số ấy có nhiều chữ số chắc.

⊙ **Chú ý:** Khi viết số gần đúng, chỉ nên giữ lại một hai chữ số không chắc để khi tính toán sai số chỉ tác động đến các chữ số không chắc mà thôi.

4.2.3. Cách viết số xấp xỉ

Cho a là số gần đúng của A với sai số tuyệt đối là Δa. Có hai cách viết số A:

a. Viết kèm theo sai số như ở công thức (4.1) hoặc (4.2).

b. Viết theo quy ước: mọi chữ số có nghĩa đều là chữ số chắc. Như vậy hai số gần đúng $a = 9$ và $b = 9,00$ khác xa nhau về độ chính xác, số a có thể sai số đến 1 đơn vị, còn số b chỉ có thể sai đến 0,01 đơn vị.

4.3. Sai số tính toán

Giả sử cần tính giá trị đại lượng $y^* = f(x_1^*, x_2^*, \dots, x_n^*)$ trong đó chỉ biết các giá trị gần đúng của đối số là $x_1, x_2, \dots, x_n$ với các sai số tương ứng là Δx_i (hay δx_i). Sai số của $y = f(x_1, x_2, \dots, x_n)$ được gọi là sai số tính toán.

Nếu f là một hàm khả vi, liên tục theo các biến x_i thì

$$y - y^* = f(x_1, x_2, \dots, x_n) - f(x_1^*, x_2^*, \dots, x_n^*) = \sum_{i=1}^{n} f'_{x_i}(\overline{x})(x_i - x_i^*)$$

trong đó $\overline{x}$ là điểm trung gian nằm giữa các điểm $(x_1, x_2, \dots, x_n)$ và $(x_1^*, x_2^*, \dots, x_n^*)$. Do f'_{x_i} liên tục và Δx_i khá bé nên ta có thể viết

$$|y - y^*| \simeq |\sum_{i=1}^{n} f'_{x_i}(x_i - x_i^*)| \leq \sum_{i=1}^{n} |f'_{x_i}|\Delta x_i =: \Delta y \tag{4.5}$$

Khi đó

$$\delta y = \frac{\Delta y}{|y|} = \sum_{i=1}^{n} \frac{|f'_{x_i}|}{|f|}\Delta x_i = \sum_{i=1}^{n} |\frac{\partial}{\partial x_i} \ln f(x_1, x_2, \dots, x_n)|\Delta x_i \tag{4.6}$$

4.3.1. Sai số các phép tính cộng trừ

Nếu $y = x_1 \pm x_2 \pm \dots \pm x_n$ thì $y'_{x_i} = \pm 1, \forall i = 1, 2, \dots, n$ nên theo (4.5) ta có:

$$\Delta y = \sum_{i=1}^{n} \Delta x_i,$$

tức là sai số tuyệt đối của tổng bằng tổng sai số tuyệt đối của các số hạng.

Giả sử $\Delta x_m = \max\limits_{1 \leq i \leq n} \Delta x_i$ và chữ số chắc cuối cùng của x_m ở hàng thứ k, nghĩa là $\Delta x_m = 10^k$. Ta có $\Delta y \geq \Delta x_m$, vì vậy khi làm phép cộng đại số, nên qui tròn các x_i đến mức giữ lại 1 hoặc 2 chữ số bên phải hàng thứ k.

⊙ **Chú ý** Trường hợp tổng đại số rất nhỏ, nghĩa là $|y| << 1$ thì sai số tương đối $\delta y = \dfrac{\Delta y}{|y|}$ trở nên rất lớn, do đó kết quả không chính xác. Vì vậy trong tính toán cần phải tránh các công thức có hiệu của hai số gần nhau.

4.3.2. Sai số các phép tính nhân chia

Giả sử

$$y = \frac{x_1 \ldots x_p}{x_{p+1} \ldots x_n}$$

Khi đó

$$\ln y = \sum_{i=1}^{p} \ln x_i - \sum_{j=p+1}^{n} \ln x_j \Longrightarrow \frac{\partial}{\partial x_i} \ln y = \pm 1$$

Theo (4.6) suy ra

$$\delta y = \sum_{i=1}^{n} \delta x_i,$$

tức là sai số tương đối của phép toán nhân chia bằng tổng sai số tương đối các số hạng.

Gọi $\delta x_m = \max\limits_{1 \le i \le n} \delta x_i$ và chữ số chắc cuối cùng của x_m ở hàng thứ k, ta thấy $\delta y \ge \delta x_m$ nên khi làm các phép tính trung gian để tính y chỉ cần lấy $k+1$, $k+2$ chữ số là đủ.

4.3.3. Sai số của phép lũy thừa, khai căn, nghịch đảo

Cho $y = x^\alpha$, khi đó $\delta y = |\dfrac{d}{dx} \ln y| \Delta x = |\alpha| \delta x$

- Nếu $\alpha > 1$ (phép luỹ thừa) thì $\delta y > \delta x$, do đó độ chính xác giảm.
- Nếu $0 < \alpha < 1$ (phép khai căn) thì $\delta y < \delta x$ hay độ chính xác tăng.
- Nếu $\alpha = -1$ ta có phép nghịch đảo, $\delta y = \delta x$ nghĩa là độ chính xác không đổi.

• **Ví dụ 6** *Một sân nhỏ hình chữ nhật có các kích thước đo được là:* $x = 2,56m \pm 0,01m$, $y = 4,2m \pm 0,02m$. *Tính chu vi và diện tích của sân?*

Giải

Chu vi của sân là $P = 2(x+y) \simeq 2(2,56+4,2) = 13,52m$. Sai số tuyệt đối $\Delta P = 2(\Delta x + \Delta y) = 0,06$. Vì $0,01 < 0,06 < 0,1$ nên P chỉ có 3 chữ số chắc và có thể viết $P \simeq 13,5m$.
Diện tích của sân: $S = xy \simeq 2,56.4,2 = 10,752$. Sai số tương đối $\delta S = \delta x + \delta y = \dfrac{1}{256} + \dfrac{1}{210}$, do đó $\Delta S = |S|.\delta S = 0,093 < 0,1$. Như vậy S có 3 chữ số chắc và có thể viết $S \simeq 10,8m^2$.

• **Ví dụ 7** *Cho diện tích hình vuông* $S = 12,34$, $\Delta S = 0,01$. *Tính cạnh* $a =$?

Ta có $a = \sqrt{S} \simeq 3,5128$. Vì $\delta S = \Delta S/|S| \simeq 0,01/12,34 \simeq 0,0008$ nên $\Delta a \simeq 3,5128 \times 0,0004 \simeq 1,4 \times 10^{-3}$. Như vậy a có 4 chữ số chắc và $a \simeq 3,513$.

• **Ví dụ 8** *Tính sai số tuyệt đối và sai số tương đối của thể tích hình cầu:* $V = \frac{1}{6}\pi d^3$, *biết* $d = 3,7 \pm 0,05$ *và* $\pi = 3,14$.

Giải

Xem π và d là đối số của V thì theo (4.5) và (4.6) ta có: $\delta V = \delta\pi + 3\delta d$ trong đó $\delta d = 0,0016/3,14 = 0,0005$ và $\delta d = 0,05/3,7 = 0,0135$. Suy ra $\delta V = 0,0005 + 3 \times 0,0135 \simeq 0,04$. Mặt khác $V = \frac{1}{6}\pi d^3 \simeq 26,5$ nên $\Delta V = |V|\delta V = 26,5.0,04 \simeq 1,1$. Vậy $V = 26,5 \pm 1,1cm^3$

4.3.4. Bài toán ngược của lý thuyết sai số

Giả sử đại lượng y tính theo công thức $y = f(x_1, x_2, \ldots, x_n)$. Hỏi phải lấy Δx_i bằng bao nhiêu để $\Delta y \leq const$ cho trước?

⊙ **Nguyên lý ảnh hưởng đều:** Ta coi $|f'_{x_i}|\Delta x_i = c(i = \overline{1,n})$, khi đó

$$\Delta y = \sum_{i=1}^{n} |f'_{x_i}|\Delta x_i = nc \Rightarrow \Delta x_i = \frac{c}{|f'_{x_i}|} = \frac{\Delta y}{n|f'_{x_i}|}$$

• **Ví dụ 9** *Một hình trụ có bán kính đáy* $R = 2m$, *chiều cao* $h = 3m$. *Hỏi* $\Delta R, \Delta h$ *phải bằng bao nhiêu để thể tích* V *được tính chính xác tới* $0,1m^3$.

Giải

Ta có $V = \pi R^2 h$, suy ra $\frac{\partial V}{\partial \pi} = R^2 h = 12$ nên $\Delta\pi = \frac{0,1}{3 \times 12} < 0,003$ Tương tự: $\frac{\partial V}{\partial R} = 2\pi R h = 37,7 \Rightarrow \Delta R = \frac{0,1}{3 \times 37,7} < 0,001$; $\frac{\partial V}{\partial h} = \pi R^2 = 12,6 \Rightarrow \Delta h = \frac{0,1}{3 \times 12,6} < 0,003$.

4.4. Sai số phương pháp và sai số tính toán

Trong thực tế khi giải một bài toán phức tạp ta thường phải thay bài toán đó bằng bài toán đơn giản hơn để có thể tính toán bằng tay hoặc bằng máy. Phương pháp thay thế như trên được gọi là **phương pháp gần đúng**. Sai số do phương pháp gần đúng tạo ra gọi là sai số phương pháp. Mặc dù bài toán đã ở dạng đơn giản, có thể tính toán bằng tay hoặc trên máy tính nhưng trong quá trình tính toán ta thường xuyên phải làm tròn các kết quả trung gian. Sai số tạo bởi tất cả các lần quy tròn như vậy được gọi là **sai số tính toán**. Để hiểu rõ hơn bản chất của sai số phương pháp và sai số tính toán ta xét ví dụ sau:

• **Ví dụ 10** *Theo khai triển Maclaurin của hàm* e^x *ta có:*

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots$$

Với $x = 1$ công thức này có thể dùng để tính giá trị của số e. Tuy nhiên đây là tổng vô hạn, còn trong thực tế ta chỉ tính được tổng

$$S_n = 1 + 1 + \frac{1}{2!} + \cdots + \frac{1}{n!},$$

nghĩa là ta đã sử dụng phương pháp gần đúng. Chẳng hạn với $n = 8$ thì sai số phương pháp tìm được dựa trên đánh giá $|e - S_n| \le \dfrac{e^{\theta x}}{(n+1)!} \le \dfrac{3}{9!} < 10^{-5}$

Khi đó $e \simeq 1 + 1 + \dfrac{1}{2!} + \cdots + \frac{1n}{8!} \simeq 2,71828$.

Bài tập chương 4

Bài 4.1. Cho $a = 12345$ với $\delta a = 0,1\%$, $b = 34,56$ với $\delta b = 0,8\%$. Xác định sai số tuyệt đối và các chữ số chắc của a và b.

Bài 4.2. Tìm số các chữ số chắc và làm tròn chỉ giữ lại 2 số không chắc:

a) $a = 57,4365$ với $\delta a = 0,5\%$.

b) $a = 1,40805$ với $\delta a = 0,6\%$.

Bài 4.3. Biết rằng $a = 12,3057$ là một số gần đúng có hai chữ số không chắc. Hãy tính sai số tuyệt đối và sai số tương đối của a.

Bài 4.4. Cho a = 23,35781 là số gần đúng với sai số tương đối là $\delta a = 1,25\%$. Hãy làm tròn a với 2 chữ số không chắc và đánh giá sai số của kết quả thu được.

Bài 4.5. Cho các số gần đúng $a = 4,7658$ và $b = 3,456$ với $\Delta a = 5.10^{-4}$ và $\Delta b = 10^{-3}$; còn $u = a.b$. Hãy tìm sai số tương đối của a và b; tính u và ước lượng sai số Δu và δu.

Bài 4.6. Tính diện tích hình chữ nhật có chiều dài $d = 40,0cm$ và chiều rộng $r = 24,0cm$. Ước lượng sai số tuyệt đối và tương đối của S nếu các chữ số biểu diễn d và r đều chắc.

Bài 4.7. Cho hình hộp chữ nhật có các cạnh d, r, h tương ứng xấp xỉ bằng 10m, 5m và 3,5m.

a) Tính thể tích V và ước lượng sai số nếu $\Delta d = \Delta r = \Delta h = 0,005m$.

b) Cần tính các cạnh với sai số như thế nào để sai số $\Delta V \le 0,1$.

Bài 4.8. Hình trụ tròn xoay có bán kính $R = 10cm$ chiều cao $h = 20cm$;

a) Tính V nếu $\Delta R = \Delta h = 0,5cm$; $p = 3,1416$ với $\Delta p = 0,5.10^{-4}$.

b) Với p như trên, cần tính R và h như thế nào để $\Delta V \le 1$.

Bài 4.9. Cho $u = a - b$ với $a = 55,23$ và $b = 55,20$; $\Delta a = \Delta b = 0,005$.

a) Tính $u, \Delta u$ và δu.

b) Giải thích vì sao người ta thường tránh trừ 2 số gần bằng nhau.

Bài 4.10. Cho $u = a/b + c$ với $a = 125, b = 0,5, c = 5$; $\Delta a = \Delta b = 0,1$; $\Delta c = 1$.

a) tính u và δu.

b) Giải thích vì sao người ta tránh chia cho số bé ở các bước trung gian.

Bài 4.11. Tính $u = a^2 b + c$ nếu $a = 4,0; b = 5,5; c = 25,48; \Delta a = \Delta b = 0,001; \Delta c = 0,01$ và thu gọn u chỉ giữ lại một chữ số không chắc.

Bài 4.12. Hãy xác định giá trị của các hàm số dưới đây cùng với sai số tuyệt đối, sai số tương đối ứng với những giá trị của các đối số cho với mọi chữ số có nghĩa đều chắc.

a) $u = \ln(x + y^2); \quad x = 0,97; y = 1,132.$

b) $u = \dfrac{x + y^2}{z}; \quad x = 3,28; y = 0,932; z = 1,132.$

Chương 5

Phép nội suy

Trong thực tế, nhiều khi ta phải sử dụng hàm $y = f(x)$ mà không biết biểu thức giải tích, chỉ biết giá trị $y_i = f(x_i)$ tại các điểm $x_i \in [a, b]$ $(i = 0, 1, \ldots, n)$. Các giá trị đó có thể nhận được qua đo đạc, thực nghiệm hoặc tính toán từ những số liệu đã cho. Cũng có trường hợp biểu thức giải tích $f(x)$ đã cho nhưng quá cồng kềnh. Khi đó dùng phép nội suy ta có thể dễ dàng tính được f tại bất kỳ $x \in [a, b]$ mà độ chính xác không kém bao nhiêu.

Mục tiêu của phép nội suy khá nhiều, nhưng chủ yếu là tìm thuật toán đơn giản tính giá trị $f(x)$ cho những x không nằm trong bảng x_i, y_i $(i = \overline{0, n})$. Một bộ số liệu x_i, y_i $(i = \overline{0, n})$ và một chương trình ngắn gọn có thể thay một bảng rất dài các giá trị $\{x_i, f(x_i)\}$. Ngoài ra sử dụng kết quả của phép nội suy, có thể tìm đạo hàm $f'(x)$ hoặc tích phân của $f(x)$ trên đoạn $[a, b]$.

5.1. Nội suy bằng đa thức đại số

Ngoài ý nghĩa lịch sử ra, đa thức đại số thường được dùng trong phép nội suy vì lý do đơn giản sau: các phép toán cộng, trừ, nhân, đạo hàm, tích phân dễ dàng được thực hiện trên đa thức. Hơn nữa nếu $P(x)$ là đa thức, còn c là hằng số thì $P(cx)$ và $P(x + c)$ cũng là đa thức.

⊙ **Bài toán nội suy** đặt ra như sau: Cho hàm số $y = f(x)$ dưới dạng bảng giá trị

x_i	x_0	x_1	$\cdots$	x_n
y_i	y_0	y_1	$\cdots$	y_n

trong đó $a \le x_0 \le x_1 \le \cdots \le x_n \le b$ gọi là các mốc nội suy và $y_i := f(x_i)$.

Hãy tìm đa thức bậc m : $P_m(x) = \sum\limits_{i=0}^{m} a_i x^i$ sao cho $P_m(x_i) = y_i \quad (i = \overline{0, n})$

⊙ **Ý nghĩa hình học của bài toán nội suy** : hãy xây dựng đường cong đại số $y = P_m(x)$ đi qua các điểm cho trước (x_i, y_i) $(i = \overline{0, n})$ từ hệ phương trình tuyến tính sau:

$$\sum_{j=0}^{m} a_j x_i^j = y_i \quad (i = \overline{0, n}) \tag{5.1}$$

Dễ thấy nếu $m < n(m > n)$ hệ nói chung vô nghiệm (vô định). Khi $m = n$ hệ (5.1) có định thức

Vandermonde

$$\Delta = \begin{vmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n \\ 1 & \dots & \dots & \dots & \dots \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{vmatrix} = \prod_{0 \le i < j \le n} (x_i - x_j) \neq 0.$$

Suy ra hệ phương trình (5.1) là hệ Crammer, do đó nó có nghiệm duy nhất.

⊙ **Chú ý** Bài toán nội suy còn được nêu dưới dạng tổng quát hơn, không những yêu cầu $P(x)$ trùng với $f(x)$ tại các mốc nội suy mà còn yêu cầu các đạo hàm cấp 1 hoặc các đạo hàm cấp cao hơn của chúng cũng trùng nhau.

5.2. Đa thức nội suy Lagrange

Sau đây ta sẽ trình bày cách xây dựng đa thức nội suy mà không cần giải hệ (5.1).

Trước hết, ta tìm đa thức $l_i(x)$ có bậc n, sao cho

$$l_i(x_j) = \begin{cases} 1 & \text{nếu} \quad i = j \\ 0 & \text{nếu} \quad i \neq j \end{cases} \qquad (i, j = \overline{0, n})$$

Dễ thấy

$$l_i(x) = A_i(x - x_0)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n).$$

Vì

$$1 = l_i(x_i) = A_i(x_i - x_0)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n).$$

nên

$$l_i(x) = \frac{(x - x_0)\dots(x - x_{i-1})(x - x_{i+1})\dots(x - x_n)}{(x_i - x_0)\dots(x_i - x_{i-1})(x_i - x_{i+1})\dots(x_i - x_n)}.$$

và gọi nó là *đa thức Lagrange cơ bản*

Đặt

$$P(x) = \sum_{i=0}^{n} y_i l_i(x), \tag{5.2}$$

Rõ ràng $P(x)$ là một đa thức bậc nhỏ hơn hoặc bằng n và ta có

$$P(x_j) = \sum_{i=0}^{n} y_i l_i(x_j) = y_j \quad (j = \overline{0, n}).$$

Như vậy $P(x)$ là đa thức nội suy cần tìm.

- **Nội suy bậc nhất**

  Với $n = 1$ ta có bảng

x_i	x_0	x_1
y_i	y_0	y_1

Đa thức nội suy (5.2) sẽ là: $P(x) = y_0 l_0(x) + y_1 l_1(x)$ trong đó

$$l_0(x) = \frac{x - x_1}{x_0 - x_1}, \quad l_1(x) = \frac{x - x_0}{x_1 - x_0}.$$

- **Nội suy bậc hai**

Với $n = 2$ ta có bảng

x_i	x_0	x_1	x_2
y_i	y_0	y_1	y_2

Đa thức nội suy (5.2) sẽ là: $P(x) = y_0 l_0(x) + y_1 l_1(x) + y_2 l_2(x)$ trong đó

$$l_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}, \quad l_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)}$$

và $l_2(x) = \dfrac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$.

• **Ví dụ 1** *Cho hàm số dưới dạng bảng*

x	-1	0	1
$f(x)$	-7	0	7

Tìm đa thức nội suy dạng Lagrange

Giải

Đa thức cần tìm có dạng:

$$P(x) = -7 \times \frac{x(x-1)}{(-1-0)(-1-1)} + 7 \times \frac{x(x+1)}{(1-0)(1+1)}$$

$$P(x) = -7 \times \frac{x(x-1)}{2} + 7 \times \frac{x(x+1)}{2} = 7x$$

⊙ **Chú ý** Nếu các mốc nội suy cách đều, tức là

$$x_{i+1} - x_i = h \quad (i = \overline{0, n-1})$$

thì đặt $t := \dfrac{x - x_0}{h}$ hay $x = x_0 + th$ ta được

$$P_i(x) = P_i(x_0 + th) = \frac{(-1)^{n-1} C_n^i}{t - i} \frac{t(t-1)\dots(t-n)}{n!}$$

Tóm lại

$$P(x_0 + th) = \frac{t(t-1)\dots(t-n)}{n!} \sum_{i=0}^{n} \frac{(-1)^{n-1} C_n^i}{t - i} y_i \tag{5.3}$$

Trong công thức (5.3), các hệ số $(-1)^{n-1} C_n^i$ không phụ thuộc vào hàm số $f(x)$, mốc nội suy và bước h. Do đó chúng được tính sẵn, lập thành bảng để sử dụng nhiều lần.

Công thức nội suy Lagrange trình bày trên có ưu điểm đơn giản nhưng nếu thêm mốc nội suy phải tính lại toàn bộ. Nhược điểm này được khắc phục trong công thức nội suy Newton.

• **Ví dụ 2** *Tìm đa thức nội suy bậc hai của hàm $y = 3^x$ trên đoạn $[-1, 1]$ tại các mốc nội suy $x_0 = -1, x_1 = 0, x_2 = 1$.*

Giải

Ta có $y_0 = 1/3, y_1 = 1, y_2 = 3$.

$$P_2(x) = \frac{1}{3} \cdot \frac{(x-0)(x-1)}{(-1-0)(-1-1)} + 1 \cdot \frac{(x+1)(x-1)}{(0+1)(0-1)} + 3 \cdot \frac{(x+1)x}{(1+1)1}$$

$$P_2(x) = 1 + \frac{4x}{3} + \frac{2x^2}{3}$$

Cho $x = 1/2$ ta được

$$3^{\frac{1}{2}} = \sqrt{3} \approx 1 + \frac{2}{3} + \frac{1}{6} \approx 1.8.$$

5.3. Sai số của phép nội suy

5.3.1. Sai số phương pháp

Giả sử $P(x)$ là đa thức nội suy bậc n của hàm $f(x)$, tức là $P(x_i) = f(x_i)$ $(i = \overline{0,n})$. Ta cố định giá trị $x \in [a, b]$ tùy ý và tìm cách ước lượng sai số $R(x) = f(x) - P(x)$. Dĩ nhiên chỉ cần xét $x \neq x_i$ $(i = \overline{0,n})$ vì $R(x_i) = 0$ $(i = \overline{0,n})$.

Xét hàm bổ trợ

$$F(x) := R(z) - kw(z),$$

trong đó $\omega(z) = \prod_{i=0}^{n}(z - x_i)$. Hằng số k chọn từ điều kiện $F(x) = 0$, nghĩa là $k = \dfrac{f(x) - P(x)}{\omega(x)}$. Mặt khác $F(x_i) = 0$ $(i = \overline{0,n})$ do đó $F(z)$ có $n + 2$ nghiệm phân biệt $x, x_0, x_1, \ldots, x_n$. Theo định lý Rolle $F'(z)$ có (n+1) nghiệm, ..., $F^{(n+1)}(z)$ có nghiệm $\xi \in [a, b]$:

$$0 = F^{(n+1)}(\xi) = f^{(n+1)}(\xi) - k(n+1)! \text{ hay } \frac{f^{(n+1)}(\xi)}{(n+1)!} = k$$

So sánh hai cách viết của k ta có:

$$R(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\omega(x) \tag{5.4}$$

Gọi $M = \sup\limits_{a \le x \le b} |f^{(n+1)}(x)|$, từ (5.4) suy ra:

$$|f(x) - P(x) \le \frac{M}{(n+1)!}|(x - x_0) \ldots (x - x_n)| \tag{5.5}$$

• **Ví dụ 3** *Ước lượng sai số phép nội suy bằng đa thức bậc 3 tính* $\sin 6°$ *với các mốc nội suy*

$$x_0 = \frac{\pi}{36}, \quad x_1 = \frac{7\pi}{180}, \quad x_2 = \frac{\pi}{20}, \quad x_3 = \frac{11\pi}{180}.$$

Giải

Ta có $f(x) = \sin x; n = 3; a = \frac{\pi}{36} = 5°; b = \frac{11\pi}{180} = 11°; x = \frac{\pi}{30} = 6°; f^{(4)}(x) = \sin x$ $\Rightarrow M = \sup_{a \le x \le b} |f^{(4)}(x)| = \sin 11° = 0.190809.$

Vậy theo công thức (5.5) ta có:

$$\begin{aligned} |\sin 6° - P(6°)| &\le \frac{0.190809}{4!}\left|\left(\frac{\pi}{30} - \frac{\pi}{36}\right)\left(\frac{\pi}{30} - \frac{7\pi}{180}\right)\left(\frac{\pi}{30} - \frac{\pi}{20}\right)\left(\frac{\pi}{30} - \frac{11\pi}{180}\right)\right| \\ &\le 1.106 \times 10^{-9}. \end{aligned}$$

⊙ **Chú ý** Từ công thức đánh giá sai số (5.5) suy ra:

1. Phần dư $R(x)$ rất lớn ngoài đoạn $[x_0, x_n]$, do đó dùng công thức nội suy để thực hiện phép ngoại suy sẽ mắc phải sai số lớn.

2. Phép nội suy có độ chính xác cao đối với các đoạn $[x_i, x_{i+1}]$ ở trung tâm và độ chính xác thấp đối với các đoạn ngoài rìa.

5.3.2. Sai số tính toán

Giả sử thay vì biết các giá trị đúng $\widetilde{y}_i = f(x_i)$ ta chỉ biết các giá trị gần đúng y_i. Khi đó thay vì đa thức nội suy

$$\widetilde{P}(x) = \sum_{i=0}^{n} \widetilde{y}_i \frac{\omega(x)}{(x - x_i)\omega'(x_i)},$$

ta có

$$P(x) = \sum_{i=0}^{n} y_i \frac{\omega(x)}{(x - x_i)\omega'(x_i)}$$

Giả sử $|y_i - \widetilde{y}_i| \le \Delta y_i$, khi đó sai số tính toán

$$|\Delta P| = |P - \widetilde{P}| \le \sum_{i=0}^{n} \Delta y_i \left|\frac{\omega(x)}{(x - x_i)\omega'(x_i)}\right|.$$

Nếu các mốc nội suy cách đều và $\Delta y_i \le \rho$ $(i = \overline{0, n})$ thì

$$|\Delta P| \le \rho \frac{|t(t-1)\ldots(t-n)|}{n!} \sum_{i=0}^{n} \frac{C_n^i}{|t - i|}$$

5.3.3. Chọn mốc nội suy tối ưu

Nếu hàm $f(x)$ đã cho thì $M = \sup_{a \le x \le b} |f^{(n+1)}(x)|$ hoàn toàn xác định. Từ công thức (**??**) suy ra sai số tuyệt đối của phép nội suy chỉ còn phụ thuộc vào $\omega(x) = (x - x_0) \dots (x - x_n)$. Vấn đề đặt ra là phải chọn các mốc nội suy $a \le x_0 < x_1 < \cdots < x_n \le b$ như thế nào để $\max_{a \le x \le b} |\omega(x)|$ nhỏ nhất ? Ta đi đến bài toán min-max sau:

$$\max_{a \le x \le b} |\omega(x)| \to \min_{a \le x_0 < x_1 < \cdots < x_n \le b}$$

⊙ **Đa thức Chebysev**

$$T_n(x) := \cos[n \arccos x] \ (|x| \le 1)$$

Đặt $\theta = \arccos x$ và để ý rằng $\cos(n \pm 1)\theta = \cos\theta \cos n\theta \mp \sin\theta \sin n\theta$, ta được $\cos(n+1)\theta + \cos(n-1)\theta = 2\cos\theta\cos n\theta$ hay:

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

Ngoài ra

$$T_1(x) = x; T_2(x) = \cos[2\arccos x] = 2x^2 - 1; T_0(x) = 1.$$

Bằng qui nạp, dễ dàng chứng minh được $T_n(x)$ là đa thức bậc n với hệ số đầu là 2^{n-1}.

Nghiệm của $T_n(x)$ là $x_i = \cos \dfrac{2i+1}{2n}\pi$ $(i = \overline{0, n-1})$ và cực trị của nó $\max_{|x| \le 1} |T_n(x)| = 1$ đạt tại $x_i = \cos \dfrac{\pi i}{n}$ $(i = \overline{0, n})$.

△ **Định lý 1** *Trong tất cả các đa thức bậc n với hệ số đầu bằng 1, đa thức Chebysev $T_n(x)/2^{n-1}$ có độ lệch (so với 0) nhỏ nhất trên đoạn $[-1, 1]$. Nghĩa là, nếu*

$$P(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$$

thì

$$\max_{|x| \le 1} |P(x)| \ge \max_{|x| \le 1} \frac{|P(x)|}{2^{n-1}} = \frac{1}{2^{n-1}}.$$

⊙ **Chọn mốc nội suy tối ưu**

i) Trong trường hợp $a = -1; b = 1$ ta lấy mốc nội suy x_i là nghiệm của đa thức Chebysev $T_{n+1}(x)$, nghĩa là: $x_i = \cos \dfrac{2i+1}{2(n+1)}\pi$ $(i = \overline{0, n})$. Khi đó $\omega(x) = (x - x_0) \dots (x - x_n) = \dfrac{T_{n+1}(x)}{2^n}$ có độ lệch nhỏ nhất, và ước lượng tốt nhất của phép nội suy là:

$$|P(x) - f(x)| \le \frac{M}{(n+1)!} |\omega(x)| \le \frac{M}{2^n (n+1)!}.$$

ii) Trong trường hợp $a < b$ bất kỳ, ta dùng phép thế biến $t = \dfrac{2x-b-a}{b-a}$ đưa đoạn $[a,b]$ về đoạn $[-1,1]$. Các mốc nội suy tối ưu là nghiệm đa thức Chebysev:

$$x_i = \frac{1}{2}\Big\{\cos\frac{2i+1}{2(n+1)}\pi + (b+a)\Big\}\ (i = \overline{0,n}).$$

Ước lượng tốt nhất của phép nội suy trong trường hợp này là:

$$|P(x) - f(x)| \leq \frac{M}{(n+1)!}\frac{(b-a)^{n+1}}{2^(2n+1)}.$$

5.4. Sai phân và các tính chất

Giả sử $f : R \to R$ là một hàm số cho trước và $h = \text{const} \neq 0$. Ta gọi sai phân cấp 1 của $f(x)$ là đại lượng $\Delta f(x) = f(x+h) - f(x)$. Tỷ sai phân cấp 1 của $f(x)$ là $\dfrac{\Delta f(x)}{h}$. Một cách tổng quát $\Delta^n f(x) := \Delta[\Delta^{n-1} f(x)]\ (n \geq 1), \Delta f^0 := f(x)$.

• **Ví dụ 4** $\Delta^2 f = \Delta(\Delta f) = \Delta(f(x+h) - f(x)) = [f(x+h+h) - f(x+h)] - [f(x+h) - f(x)] = f(x+2h) - 2f(x+h) + f(x)$.

$\Delta^3 f = \Delta(\Delta^2 f) = [f(x+2h+h) - 2f(x+h+h) + f(x+h)] - [f(x+2h) - 2f(x+h) + f(x)] = f(x+3h) - 3f(x+2h) + 3f(x+h) - f(x)$.

◇ **Tính chất**

(1) Δ là toán tử tuyến tính, nghĩa là:
$\forall \alpha, \beta \in R;\ \forall f, g \Rightarrow \Delta(\alpha f + \beta g) = \alpha\Delta f + \beta\Delta g$.

(2) Nếu $c = \text{const}$ thì $\Delta c = 0$.

(3) $\Delta^n(x^n) = n!h^n;\ \Delta^m(x^n) = 0\ (m > n)$.

(4) Nếu $P(x)$ là đa thức bậc n thì theo công thức Taylor

$$\Delta P := P(x+h) - P(x) = \sum_{i=1} n\frac{h^i}{i!}P^{(i)}(x).$$

(5) Nếu $f \in C^n[a,b]$ thì khi h đủ nhỏ $f^{(n)}(x) \simeq \dfrac{\Delta^n f(x)}{h^n}$.

5.5. Một số quy tắc nội suy hàm số trên lưới đều

5.5.1. Bảng sai phân

Giả sử hàm số $y = f(x)$ cho dưới dạng bảng $y_i = f(x_i)$ tại các mốc x_i cách đều: $x_{i+1} - x_i = h = \text{const}\ (i \geq 0)$.

Khi đó sai phân của dãy y_i được xác định như sau:

$$\begin{aligned}\Delta y_i &= y_{i+1} - y_i\\ \Delta^2 y_i &= \Delta(\Delta y_i) = \Delta y_{i+1} - \Delta y_i, \ldots\\ \Delta^n y_i &= \Delta(\Delta^{n-1} y_i) = \Delta^{n-1} y_{i+1} - \Delta^{n-1} y_i\end{aligned}$$

Ta lập bảng:

y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
y_{i-2}					
	Δy_{i-2}				
y_{i-1}		$\Delta^2 y_{i-2}$			
	Δy_{i-1}		$\Delta^3 y_{i-2}$		
y_i		$\Delta^2 y_{i-2}$		$\Delta^4 y_{i-2}$	$\ldots$
	Δy_i		$\Delta^3 y_{i-1}$		
y_{i+1}		$\Delta^2 y_i$			
	Δy_{i+1}				
y_{i+2}					
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

5.5.2. Nội suy ở đầu bảng

Mốc nội suy được sắp xếp theo thứ tự $x_0 < x_1 < \cdots < x_n$. Tìm đa thức nội suy dưới dạng

$$P(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \cdots + a_n(x - x_0)\ldots(x - x_{n-1}).$$

Cho $x = x_0$, ta được $a_0 = y_0$; $x = x_1 \Rightarrow a_1 = \dfrac{\Delta y_0}{h}$. Nói chung đặt $x = x_i$, ta có $a_i = \dfrac{\Delta^i y_0}{i!h^i}$. Khi đó ta có

$$P(x) = y_0 + \frac{\Delta y_0}{h}(x - x_0) + \frac{\Delta^2 y_0}{2!h^2}(x - x_0)(x - x_1) + \cdots + \frac{\Delta^n y_0}{n!h^n}(x - x_0)\ldots(x - x_{n-1}).$$

Đổi biến $t = \dfrac{x - x_0}{h}, x = x_0 + th$ ta được

$$P(x_0 + th) = y_0 + \frac{t}{1!}\Delta y_0 + \frac{t(t-1)}{2!}\Delta^2 y_0 + \cdots \frac{t(t-1)\ldots(t-n+1)}{n!}\Delta^n y_0. \tag{5.6}$$

Khi đó công thức sai số (5.5) trở thành

$$|f(x_0 + th) - P(x_0 + th)| \leq \frac{M}{(n+1)!}h^{n+1}t(t-1)\ldots(t-n).$$

Người ta gọi công thức (5.6) là đa thức nội suy Newton tiến xuất phát từ x_0.

5.5.3. Nội suy ở cuối bảng

Mốc nội suy được sắp xếp theo thứ tự giảm dần $x_n > x_{n-1} > \cdots > x_0$. Đặt $t = \dfrac{x - x_n}{h} \Rightarrow x = x_n + th$. Đa thức thức nội suy Newton lùi tìm dưới dạng

$$P(x) = a_0 + a_1(x - x_n) + a_2(x - x_n)(x - x_{n-1}) + \cdots + a_n(x - x_n)\ldots(x - x_1)$$

Cho $x = x_n \Rightarrow a_0 = y_n$; $x = x_{n-1} \Rightarrow a_0 + a_1(-h) = y_{n-1} \Rightarrow a_1 = \dfrac{\Delta y_{n-1}}{h}$.
Tổng quát, đặt $x = x_i$, ta được $a_i = \dfrac{\Delta^i y_{n-i}}{i!h^i}$ $(i = \overline{0, n})$.
Như vậy công thức Newton lùi sẽ có dạng

$$P(x_n + th) = y_n + \frac{t}{1!}\Delta y_{n-1} + \frac{t(t+1)}{2!}\Delta^2 y_{n-2} + \cdots + \frac{t(t+1)\ldots(t+n-1)}{n!}\Delta^n y_0 \qquad (5.7)$$

với sai số $\dfrac{M}{(n+1)!}h^{n+1}t(t+1)\ldots(t+n)$.
Công thức (5.7) được gọi là đa thức nội suy Newton lùi xuất phát từ x_n.
⊙ **Chú ý**

a) Công thức nội suy Newton tiến (lùi) chỉ là một cách viết khác của công thức Lagrange.

b) Nếu cần tính $f(x)$ tại $x \simeq x_0$ $(x \simeq x_n)$ ta nên dùng công thức nội suy Newton tiến (lùi) thì độ chính xác cao hơn.

c) Dùng công thức nội suy Newton tiến (lùi) không phải tính lại từ đầu nếu thêm mốc nội suy mới.

• **Ví dụ 5** *Cho một số giá trị của hàm* $\sin x$

x	0,1	0,2	0,3	0,4
$\sin x$	0,09983	0,19867	0,29552	0,38942

Hãy tính gần đúng $\sin(0,14)$ *và* $\sin(0,46)$.

Giải

Vì các mốc cách đều nên ta xây dựng đa thức nội suy Newton của hàm số dựa vào bảng giá trị đã cho. Trước hết ta lập bảng sai phân:

x	$\sin x$	Δy	$\Delta^2 y$	$\Delta^3 y$
0,1	0,09983			
		0,09884		
0,2	0,19867		-0,00199	
		0,09685		-0,00096
0,3	0,29552		-0,00295	
		0,09390		
0,4	0,38942			

Sử dụng công thức nội suy Newton tiến (5.6) ta được:

$$P(0,1+th) = 0,09983 + \frac{0,09884}{h}t - \frac{0,00199}{h^2}t(t-1) - \frac{0,00096}{h^3}t(t-1)(t-2).$$

Với $x = 0,14$, $h = 0,1$ suy ra $t = 0,4$. Khi đó:

$$\sin(0,14) \approx P(0,1+0,1.0,4) = 0,13954$$

và sai số được đánh giá bởi

$$\begin{aligned} |\sin(0,14) - P(0,14)| &\leq \frac{1}{4!}|(0,14-0,1)(0,14-0,2)(0,14-0,3)(0,14-0,4)| \\ &\leq 4,2.10^{-6}. \end{aligned}$$

Tương tự, sử dụng đa thức nội suy Newton lùi (5.7) ta cũng tính được:

$$\sin(0,46) \approx P(0,46) = 0,44394$$

với sai số $|\sin(0,46) - P(0,46)| \leq 3,8.10^{-5}$.

• **Ví dụ 6** *Giả sử hàm $f(x)$ cho dưới dạng bảng sau:*

x	0	0.1	0.2	0.3	0.4	0.5
y	0	0.1002	0.2013	0.8045	0.4108	0.5211

Tìm $f(0.14)$ theo công thức nội suy Newton tiến.

Kết quả tính toán cho thấy $f(0.14) \simeq 0.1405$. Sai số mắc phải

$$|f(0.14) - P(0.14)| < \frac{0.0001}{4!}0.4 \times 0.6 \times 1.6 \times 2.6 < \frac{1}{2}10^{-6}.$$

(Ở đây ta sử dụng hệ thức $f^{(4)}(\xi) \simeq \dfrac{\Delta^4 y_0}{h^4}$ và

$$|f(0.14) - P(0.14)| < \frac{\Delta^4 y_0}{4!h^4}h^4 \times 0.4 \times 0.6 \times 1.6 \times 2.6)$$

5.6. Một số ví dụ áp dụng sai phân và nội suy

5.6.1. Tính giá trị đa thức

Giả sử deg(P)= n. Nếu biết $P(x)$ tại $(n+1)$ điểm phân biệt, sử dụng tính chất $\Delta^m P = 0$ $(m > n)$ ta có thể tính $P(x)$ tại các điểm khác.

• **Ví dụ 7** *Tính $y = x^3$ với $x = 4, 5, 6 \ldots$.*

Lập bảng

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
0	0			
		1		
1	1		6	
		7		6
2	8		12	
		19		6
3	27		18	
		37		6
4	64		24	
		61		
5	125			

Biết $6 = \Delta^3 y_1 = \Delta^2 y_2 - \Delta^2 y_1 = \Delta^2 y_2 - 12$ suy ra $\Delta^2 y_2 = 18 = \Delta y_3 - \Delta y_2 = \Delta y_3 - 19$, hay $\Delta y_3 = 37 = y_4 - y_3 = y_4 - 27$. Vậy $y_4 = 64, \ldots$

5.6.2. Tính tổng

Giả sử cần tính $S_n = 1^2 + 2^2 + \cdots + n^2$ với $\forall n \geq 1$. Ta có: $\Delta S_n = S_{n+1} - S_n = (n+1)^2$, $\Delta^2 S_n = \Delta S_{n+1} - \Delta S_n = (n+2)^2 - (n+1)^2 = 2n+3$, $\Delta^3 S_n = \Delta^2 S_{n+1} - \Delta^2 S_n = [2(n+1)+3] - [2n+3] = 2 = \text{const}$.

Lập bảng:

n	S_n	ΔS_n	$\Delta^2 S_n$	$\Delta^3 S_n$
1	1			
		4		
2	5		5	2
		9		
3	14			

Sử dụng công thức nội suy Newton tiến $t = \dfrac{n-1}{1} = n-1$ ta được

$$S_n = 1 + 4(n-1) + \frac{5}{2!}(n-1)(n-2) + \frac{2}{3!}(n-1)(n-2)(n-3) = \frac{1}{6}n(n-1)(2n+1).$$

Hoàn toàn tương tự ta có thể tính tổng

$$\begin{aligned} S_n &= 1^3 + 2^3 + \cdots + n^3 = 1 + 8(n-1) + 19\frac{(n-1)(n-2)}{2} + \\ &+ 18\frac{(n-1)(n-2)(n-3)}{6} + 6\frac{(n-1)(n-2)(n-3)(n-4)}{24} \\ S_n &= \left[\frac{n(n+1)}{2}\right]^2 \end{aligned}$$

5.7. Nội suy trên lưới không đều

5.7.1. Tỷ sai phân

Giả sử các mốc nội suy được sắp xếp theo thứ tự: $a \le x_0 < x_1 < \cdots < x_n \le b$. Tỷ sai phân cấp một

$$\begin{aligned} f[x_0, x_1] &= \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \\ f[x_1, x_2] &= \frac{f(x_2) - f(x_1)}{x_2 - x_1}, \\ f[x_0, x_1, x_2] &= \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}. \end{aligned}$$

Một cách tổng quát:

$$f[x_0, x_1, \dots, x_k, x_{k+1}] = \frac{f[x_1, \dots, x_{k+1}] - f[x_0, \dots, x_k]}{x_{k+1} - x_0}$$

5.7.2. Công thức nội suy Newton trong trường hợp mốc không cách đều

Cho $a \le x_0 < x_1 < \cdots < x_n \le b$ là các mốc nội suy. Gọi $L_k(x)$ là đa thức nội suy Lagrange với mốc nội suy $\{x_0, \dots, x_k\}$. Khi đó

$$P_k(x) = L_k(x) - L_{k-1}(x)$$

là đa thức bậc k. Vì $P_k(x_i) = 0$ $(i = \overline{0, k-1})$ nên

$$P_k(x) = A(x - x_0) \dots (x - x_{k-1}) = A\omega_{k-1}(x).$$

Mặt khác $P_k(x_k) = y_k - L_{k-1}(x_k) = A\omega_{k-1}(x_k)$ suy ra

$$\begin{aligned} A &= \frac{y_k - L_{k-1}(x_k)}{\omega_{k-1}(x_k)} \\ &= \frac{f(x_k)}{(x_k - x_0) \dots (x_k - x_{k-1})} \\ &\quad - \frac{\sum_{i=0}^{k-1} f(x_i) \frac{(x_k - x_0) \dots (x_k - x_{i-1})(x_k - x_{i+1}) \dots (x_k - x_{k-1})}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_{k-1})}}{(x_k - x_0) \dots (x_k - x_{k-1})} \\ &= \sum_{i=0}^{k} \frac{f(x_i)}{(x_i - x_0) \dots (x_i - x_{i-1})(x_i - x_{i+1}) \dots (x_i - x_k)} \\ &= f[x_0, \dots, x_k] \end{aligned}$$

Vậy

$$\begin{aligned} L_n(x) &= f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) + \cdots \\ &\quad + f[x_0, x_1, \dots, x_n](x - x_0)(x - x_1) \dots (x - x_n) \end{aligned}$$

Trong công thức nội suy Newton này thêm mốc nội suy, ta không phải tính lại từ đầu.

5.7.3. Bài toán nội suy ngược

Giả sử ta có bảng giá trị

$$\{x_k, f(x_k)\}_{k=0}^{n}.$$

Cần tìm $\widetilde{x}$ trong khoảng (x_0, x_n) để $f(\widetilde{x}) = \widetilde{y}$ cho trước. Nếu hàm $f(x)$ đơn điệu, tức là sgn$\Delta f(x_k)$=const ($k = \overline{0, n}$) thì ta có thể xây dựng đa thức nội suy $P(y)$ dựa vào số liệu

$$\{y_k, x_k\}_{k=0}^{n},$$

trong đó $y_k := f(x_k)$. Đặt $y = \widetilde{y}$ ta tìm được $\widetilde{x} \simeq P(\widetilde{y})$.

Nếu hàm $f(x)$ bất kỳ, ta thay $f(x) \simeq P(x)$ - đa thức nội suy của hàm số $f(x)$ và giải phương trình $P(x) = \widetilde{y}$ để tìm gần đúng $\widetilde{x}$.

• Ví dụ 8

Xét công thức Newton tiến $\widetilde{t} := (x - x_0)/h$ trong trường hợp các mốc cách đều ta có

$$\widetilde{y} = y_0 + \widetilde{t}\Delta y_0 + \frac{\widetilde{t}(\widetilde{t}-1)}{2}\Delta^2 y_0 + \frac{\widetilde{t}(\widetilde{t}-1)(\widetilde{t}-2)}{6}\Delta^3 y_0 + \cdots$$

Từ đây suy ra

$$\widetilde{t} = \frac{1}{\Delta y_0}\{\widetilde{y} - y_0 - \frac{\widetilde{t}(\widetilde{t}-1)}{2}\Delta^2 y_0 - \frac{\widetilde{t}(\widetilde{t}-1)(\widetilde{t}-2)}{6}\Delta^3 y_0 - \cdots\}$$

Bỏ qua sai phân cấp hai trở đi, ta được:

$$\widetilde{t}_1 = \frac{1}{\Delta y_0}(\widetilde{y} - y_0)$$

Sử dụng $\widetilde{t}_1$ và sai phân đến cấp hai ta tìm được xấp xỉ tiếp theo:

$$\widetilde{t}_2 = \frac{1}{\Delta y_0}\{(\widetilde{y} - y_0 - \frac{\widetilde{t}_1(\widetilde{t}_1-1)}{2}\Delta^2 y_0)\}$$

Tương tự, sử dụng $\widetilde{t}_2$ và sai phân đến cấp ba, ta có:

$$\widetilde{t}_3 = \frac{1}{\Delta y_0}\{(\widetilde{y} - y_0 - \frac{\widetilde{t}_2(\widetilde{t}_2-1)}{2}\Delta^2 y_0) - \frac{\widetilde{t}_2(\widetilde{t}_2-1)(\widetilde{t}_2-2)}{6}\Delta^3 y_0)\}$$

Quá trình tính toán được tiếp tục cho đến khi $|\widetilde{t}_k - \widetilde{t}_{k+1}| < \epsilon$, trong đó $\epsilon > 0$ là độ chính xác cần đạt. Trở lại biến x ta có $\widetilde{x}_k = x_0 + \widetilde{t}_k h$.

5.8. Phương pháp bình phương bé nhất

Đa thức nội suy xét ở các tiết trước gặp phải một số khó khăn trong trường hợp:

- Khi số mốc nội suy lớn thì bậc của đa thức nội suy cũng lớn, không thuận tiện khi tính toán.

- Các số liệu x_i, y_i thường thu được bằng đo đạc, thực nghiệm nên bao giờ cũng có sai số, do đó yêu cầu $y_i = P(x_i)$ có thể không hợp lý.

Phương pháp bình phương bé nhất do Lagrange và Gauss đề xướng trong khi biểu diễn gần đúng những hàm cho bằng bảng.

5.8.1. Nội dung phương pháp

Cho bảng số liệu

x_i	x_1	x_2	$\cdots$	x_n
y_i	y_1	y_2	$\cdots$	y_n

Dựa vào nghiên cứu lý thuyết và số liệu trong bảng người ta chọn trước được hàm phụ thuộc $y = f(x, a_1, a_2, \ldots, a_k)$, trong đó $a_1, a_2, \ldots, a_k$ là các tham số. Các tham số này được xác định sao cho y_i và $f(x_i, a_1, a_2, \ldots, a_k)$ sai khác nhau nhỏ nhất . Một tiêu chuẩn để đánh giá sự sai khác đó là sử dụng hàm

$$S(a_1, a_2, \ldots, a_k) = \sum_{i=}^{n} \Big(y_i - f(x_i, a_1, \ldots, a_k)\Big)^2 \longrightarrow \min .$$

Dễ thấy hàm S đạt giá trị nhỏ nhất tại điểm tới hạn, chính là nghiệm của hệ:

$$\begin{cases} \dfrac{\partial S}{\partial a_1} & = & 0 \\ \dfrac{\partial S}{\partial a_2} & = & 0 \\ \ldots & \ldots & \ldots \\ \dfrac{\partial S}{\partial a_k} & = & 0 \end{cases} \tag{5.8}$$

Giải hệ (5.8) ta tìm được các tham số $a_i, i = \overline{1,k}$

5.8.2. Một số trường hợp áp dụng

a) y phụ thuộc x theo dạng bậc nhất $y = f(x, a, b) = a + bx$

Khi đó tổng bình phương các sai số:

$$S = \sum_{i=}^{n} \Big(y_i - f(x_i, a, b)\Big)^2 = \sum_{i=}^{n} (y_i - a - bx_i)^2$$

Ta lập hệ phương trình

$$\begin{cases} \dfrac{\partial S}{\partial a} = 0 \\ \dfrac{\partial S}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} na + b\sum x_i = \sum y_i \\ a\sum x_i + b\sum x_i^2 = \sum x_i y_i \end{cases} \tag{5.9}$$

Từ bảng số liệu $\{x_i, y_i\}$ ta tính được các tổng: $\sum x_i, \sum y_i, \sum x_i^2, \sum x_i y_i$. Hơn nữa, các x_i khác nhau nên định thức $D = n\sum x_i^2 - \left(\sum x_i\right)^2 \neq 0$, do đó hệ (5.9) có nghiệm duy nhất a và b.

• **Ví dụ 9** *Cho biết sự phụ thuộc giữa 2 đại lượng x và y có dạng $y = a + bx$ và cho bảng số liệu*

x	$-1,1$	$2,1$	$3,2$	$4,4$	$5,2$
y	$0,78$	$7,3$	$9,2$	$11,9$	$13,3$

Hãy xác định a và b bằng phương pháp bình phương bé nhất.

Giải

Ta lập bảng số

	x_i	y_i	x_i^2	$x_i y_i$
	$-1,1$	$0,78$	$1,21$	$-0,858$
	$2,1$	$7,3$	$4,41$	$15,33$
$n = 5$	$3,2$	$9,2$	$10,24$	$29,44$
	$4,4$	$11,9$	$19,36$	$52,36$
	$5,2$	$13,3$	$27,04$	$69,16$
$\sum$	$13,8$	$42,48$	$62,26$	$165,43$

Sau đó giải hệ

$$\begin{cases} 5a + 13,8b = 42,48 \\ 13,8a + 62,26b = 165,43 \end{cases} \Leftrightarrow \begin{cases} a = 2,994 \simeq 3 \\ b = 1,994 \simeq 2 \end{cases} \Rightarrow y = 3 + 2x$$

So sánh các giá trị mới của y theo hàm tìm được với các giá trị cũ theo bảng sau cho thấy phương pháp bình phương bé nhất khá chính xác.

x	$-1,1$	$2,1$	$3,2$	$4,4$	$5,2$
y cũ	$0,78$	$7,3$	$9,2$	$11,9$	$13,3$
y mới	$0,8$	$7,2$	$9,4$	$11,8$	$13,4$

b) y phụ thuộc x theo dạng bậc hai $y = f(x, a, b, c) = a + bx + cx^2$

Khi đó tổng bình phương các sai số:

$$S = \sum_{i=}^{n} \Big(y_i - f(x_i, a, b, c)\Big)^2 = \sum_{i=}^{n} (y_i - a - bx_i - cx_i^2)^2$$

Ta lập hệ phương trình

$$\left\{\begin{array}{lcl} \dfrac{\partial S}{\partial a} & = & 0 \\ \dfrac{\partial S}{\partial b} & = & 0 \\ \dfrac{\partial S}{\partial c} & = & 0 \end{array}\right. \Leftrightarrow \left\{\begin{array}{rcl} na + b\sum x_i + c\sum x_i^2 & = & \sum y_i \\ a\sum x_i + b\sum x_i^2 + c\sum x_i^3 & = & \sum x_i y_i \\ a\sum x_i^2 + b\sum x_i^3 + c\sum x_i^4 & = & \sum x_i^2 y_i \end{array}\right. \tag{5.10}$$

Từ bảng số liệu $\{x_i, y_i\}$ ta tính được các tổng:

$$\sum x_i, \quad \sum y_i, \quad \sum x_i^2, \quad \sum x_i y_i, \quad \sum x_i^3, \quad \sum x_i^4, \quad \sum x_i^2 y_i,$$

sau đó thay vào hệ (5.10) rồi giải ta được a, b và c.

c) y phụ thuộc x theo dạng hàm mũ $y = f(x, a, b) = ae^{bx}$

Lấy logarit Naper 2 vế ta được: $\ln y = \ln ae^{bx} = bx + \ln a$

Thực hiện phép đổi biến

$$\left\{\begin{array}{lcl} Y & = & \ln y \\ A & = & \ln a \end{array}\right. \Rightarrow Y = A + bx \tag{5.11}$$

Như vậy sự phụ thuộc giữa Y vào x là dạng bậc nhất. Từ bảng số liệu $\{x_i, y_i\}$ tương ứng ta lập được bảng số liệu $\{x_i, Y_i\}$ trong đó $Y_i = \ln y_i,\ i = \overline{1, n}$

Hệ phương trình để xác định A, b là

$$\left\{\begin{array}{rcl} nA + b\sum x_i & = & \sum Y_i \\ A\sum x_i + b\sum x_i^2 & = & \sum x_i Y_i \end{array}\right. \tag{5.12}$$

• **Ví dụ 10** *Cho biết cặp giá trị của x và y theo bảng sau:*

x	0,65	0,75	0,85	0,95	1,15
y	0,96	1,06	1,17	1,29	1,58

Lập công thức thực nghiệm của y dạng $y = ae^{bx}$.

Giải

Lấy logarit Naper 2 vế ta được: $\ln y = \ln ae^{bx} = bx + \ln a$

Đặt $Y = \ln y;\ A = \ln a \Rightarrow Y = A + bx$ *Ta có bảng số liệu mới*

x	$0,65$	$0,75$	$0,85$	$0,95$	$1,15$
$Y = \ln y$	$-0,04$	$0,06$	$0,18$	$0,25$	$0,46$

Lập bảng

	x_i	Y_i	x_i^2	x_iY_i
$n=5$	$0,65$	$-0,04$		
	$0,75$	$0,06$		
	$0,85$	$0,18$		
	$0,95$	$0,25$		
	$1,15$	$0,46$		
$\sum$	$4,35$	$0,89$	$3,93$	$0,92$

Giải hệ

$$\begin{cases} nA + b\sum x_i &= \sum Y_i \\ A\sum x_i + b\sum x_i^2 &= \sum x_iY_i \end{cases} \Leftrightarrow \begin{cases} 5A + 4,35b &= 0,89 \\ 4,35A + 3,93b &= 0,92 \end{cases} \Rightarrow \begin{cases} a &\simeq 0,5 \\ b &\simeq 1 \end{cases}$$

Vậy $y = \dfrac{1}{2}e^x$. *Tính lại các giá trị của y theo công thức này tại các giá trị của x rồi so sánh với các giá trị thực nghiệm*

x	$0,65$	$0,75$	$0,85$	$0,95$	$1,15$
y cũ	$0,96$	$1,06$	$1,17$	$1,29$	$1,58$
y mới	$1,00$	$3,85$	$6,50$	$9,35$	$12,05$

d) y phụ thuộc x theo dạng hàm lũy thừa $y = f(x,a,b) = ax^b$

Lấy logarit Naper 2 vế ta được: $\ln y = \ln ax^b = b\ln x + \ln a$

Thực hiện phép đổi biến

$$\begin{cases} Y &= \ln y \\ X &= \ln x \quad \Rightarrow Y = A + bX \\ A &= \ln a \end{cases} \tag{5.13}$$

Như vậy sự phụ thuộc giữa Y vào X là dạng bậc nhất. Từ bảng số liệu $\{x_i, y_i\}$ tương ứng ta lập được bảng số liệu $\{X_i, Y_i\}$ trong đó $X_i = \ln x_i,\ Y_i = \ln y_i,\ i = \overline{1,n}$

Khi đó hệ phương trình để xác định A, b là

$$\begin{cases} nA + b\sum X_i &= \sum Y_i \\ A\sum X_i + b\sum X_i^2 &= \sum X_iY_i \end{cases} \tag{5.14}$$

Bài tập chương 5

Bài 5.1. Hàm số $y = f(x)$ xác định trên $[0;5]$ và được cho bởi bảng giá trị sau

x	0	1	3	5
$y = f(x)$	1	2	1	4

Hãy xây dựng đa thức nội suy Lagrange $P(x)$ của $f(x)$ và tính gần đúng giá trị $f(2)$ bằng cách lấy $f(2) \simeq P(2)$.

Bài 5.2. Xây dựng đa thức Lagrange cho hàm $f(x) = x^3 + x^2 - 10$ tại các mốc $x = -4; -3; -1; 0$. Từ đó hãy xác định các hằng số A, B, C, D sao cho

$$\frac{x^3 + x^2 - 10}{x(x+1)(x+3)(x+4)} = \frac{A}{x} + \frac{B}{x+1} + \frac{C}{x+3} + \frac{D}{x+4}$$

Bài 5.3. Tính tổng

$$S_n = 1^3 + 2^3 + 3^3 + \cdots + n^3,$$

biết rằng S_n là một đa thức bậc 4.

Bài 5.4. Cho bảng giá trị của hàm số $y = f(x)$

x	-1	0	3	6	7
$y = f(x)$	3	-6	39	822	1611

Xây dựng đa thức nội suy Newton của hàm $f(x)$. Tính gần đúng $f(0,25)$ nhờ đa thức vừa tìm được.

Bài 5.5. Cho hàm số $y = f(x)$ xác định bởi bảng giá trị sau

x	0	1	2	3	4	5	6	7
$y = f(x)$	1,4	1,3	1,4	1,1	1,3	1,8	1,6	2,3

Tìm biểu thức của $f(x)$ bằng phương pháp bình phương tối thiểu biết rằng

a) $f(x)$ là một đa thức bậc nhất.

b) $f(x)$ là một đa thức bậc hai.

c) $f(x) = ae^{bx}$.

d) $f(x) = \ln(ax + b)$.

Chương 6

Tính gần đúng đạo hàm và tích phân

6.1. Tính gần đúng đạo hàm

Trong thực tế, nhiều khi người ta phải tìm đạo hàm một hàm số cho dưới dạng bảng, hoặc hàm cho dưới dạng giải tích nhưng phức tạp. Trong những trường hợp đó, ta thường dùng phương pháp tính gần đúng đạo hàm. Ta thay $f(x)$ bằng đa thức nội suy của nó:$f(x) = P(x) + R(x)$, trong đó phần dư

$$R(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{i=0}^{n}(x - x_i),\ \xi = \xi(x) \in (x_0, x_n).$$

Khi đó $f'(x) = P'(x) + r(x)$ với $r(x) = R'(x)$. Dễ thấy phép tính gần đúng đạo hàm như trên không chính xác bằng phép nội suy vì từ hệ thức $f(x_i) = P(x_i)$ $(i = \overline{0,n})$ không suy ra $f'(x_i) \simeq P'(x_i)$ $(i = \overline{0,n})$.

6.1.1. Sử dụng đa thức nội suy Lagrange

Ta có $P(x) = \sum\limits_{k=0}^{n} y_k P_k(x)$, trong đó $P_k(x) = \dfrac{\prod\limits_{i\neq k}(x - x_i)}{\prod\limits_{i\neq k}(x_k - x_i)}$ Vì

$$r(x) = R'(x) = \frac{d}{dx}\frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{i=0}^{n}(x - x_i) + \frac{f^{(n+1)}(\xi)}{(n+1)!}\frac{d}{dx}\prod_{i=0}^{n}(x - x_i)$$

nên $r(x_k) = \dfrac{f^{(n+1)}(\xi)}{(n+1)!}\prod\limits_{i\neq k}(x_k - x_i)$.

Xét trường hợp $n = 1$, ta có

$$P(x) = y_0\frac{x - x_1}{x_0 - x_1} + y_1\frac{x - x_0}{x_1 - x_0},\ R(x) = \frac{f''(\xi)}{2!}(x - x_0)(x - x_1).$$

Suy ra $P'(x_0) = \dfrac{y_1 - y_0}{x_1 - x_0}$ và $r(x_0) = R'(x_0) = \dfrac{f''(\xi)}{2!}(x_1 - x_0)$.

Như vậy

$$f'(x_0) = \frac{f(x_1) - f(x_0)}{x_1 - x_0} - \frac{f''(\xi)}{2!}(x_1 - x_0).$$

Từ đây ta lại nhận được khai triển Taylor:

$$f(x_1) = f(x_0) + f'(x_0)(x_1 - x_0) + \frac{f''(\xi)}{2!}(x_1 - x_0)^2.$$

6.1.2. Trường hợp các mốc nội suy cách đều

Ta có thể sử dụng các công thức nội suy Newton tiến, lùi để tính gần đúng đạo hàm. Ví dụ để tính $f'(x), f''(x)$ cho $x = x_0$, ta dùng đa thức nội suy Newton tiến

$$P(x) = y_0 + \frac{\Delta y_0}{1!}t + \frac{\Delta^2 y_0}{2!}t(t-1) + \cdots + \frac{\Delta^n y_0}{n!}t(t-1)\ldots(t-n+1)$$

trong đó $t = (x - x_0)/h$.

Dễ thấy

$$\begin{aligned}\frac{dP}{dx} = \frac{dP}{dt}\frac{dt}{dx} = \frac{1}{h}\frac{dP}{dt} \quad &= \quad \frac{1}{h}\{\Delta y_0 + \frac{2t-1}{2!}\Delta^2 y_0 + \frac{3t^2 - 6t + 2}{3!}\Delta^3 y_0 \\ &+ \quad \frac{2t^3 - 9t^2 + 11t - 3}{12}\Delta^4 y_0 + \cdots\}\end{aligned}$$

$$\frac{d^2P}{dx^2} = \frac{d}{dt}\Big(\frac{dP}{dx}\Big)\frac{dt}{dx} = \frac{1}{h}\frac{d}{dt}\Big(\frac{dP}{dx}\Big) = \frac{1}{h^2}\{\Delta^2 y_0 + (t-1)\Delta^3 y_0 + \frac{6t^2 - 18t + 11}{12}\Delta^4 y_0 + \cdots\}$$

Nói riêng

$$\begin{aligned}f'(x_0) \quad &\simeq \quad \frac{1}{h}\left\{\Delta y_0 - \frac{\Delta^2 y_0}{2} + \frac{\Delta^3 y_0}{3} - \frac{\Delta^4 y_0}{4} + \frac{\Delta^5 y_0}{5} - \cdots\right\} \\ f''(x_0) \quad &\simeq \quad \frac{1}{h^2}\left\{\Delta^2 y_0 - \Delta^3 y_0 + \frac{11}{12}\Delta^4 y_0 + \cdots\right\}\end{aligned}$$

• **Ví dụ 1** *Tìm $f'(50)$ của hàm $f(x)$ cho dưới dạng bảng sau:*

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$
50	1.6990			
		414		
55	1.7404		-36	
		378		5
60	1.7782		-31	
		347		
65	1.8129			

Ta có

$h = 5;\ f'(x_0) \simeq \frac{1}{h}\Big(\Delta y_0 - \frac{\Delta^2 y_0}{2} + \frac{\Delta^3 y_0}{3} - \cdots\Big) \Rightarrow f'(50) \simeq \frac{1}{5}(0.0414 - 0.0018 + 0.0002) =$

0.0087.

Giá trị chính xác của hàm

$$f'(x) = (\lg x)' = \left(\frac{\ln x}{\ln 10}\right)' = \frac{1}{x}\frac{1}{\ln 10} \simeq \frac{0.43429}{50} \simeq 0.0087.$$

6.2. Tính gần đúng tích phân

Trong thực tế, nhiều khi ta phải tính tích phân xác định của hàm số mà không biết nguyên hàm của nó. Nếu dùng định nghĩa tích phân

$$I = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(x_i)\Delta x_i$$

thì tổng Darboux hội tụ rất chậm, do đó để đạt được độ chính xác không cao, ta vẫn phải thực hiện một khối lượng tính toán rất lớn. Ngoài ta, trong nhiều trường hợp, hàm $f(x)$ chỉ được cho dưới dạng bảng, vì vậy khái niệm nguyên hàm trở nên vô nghĩa.

Phương pháp đơn giản nhất để tính gần đúng tích phân xác định là thay $f(x)$ bằng đa thức nội suy $P(x)$, sau đó đặt:

$$I := \int_a^b f(x)dx \simeq \int_a^b P(x)dx.$$

6.2.1. Phương pháp hình thang

a) Công thức

Ta chia đoạn $[a, b]$ thành n phần bằng nhau với các điểm chia $x_i = a + ih$ $(i = \overline{0, n})$ trong đó $h = (b - a)/n$. Khi đó theo tính chất của tích phân ta có thể viết

$$I = \int_a^b f(x)dx = \int_{x_0}^{x_1} f(x)dx + \int_{x_1}^{x_2} f(x)dx + \cdots + \int_{x_{n-1}}^{x_n} f(x)dx$$

Xét tích phân $\int_{x_0}^{x_1} f(x)dx$.

$f(x)$

y_0 y_1

x_0 h x_1

Về mặt hình học đó chính là diện tích hình thang cong giới hạn bởi các đường $y = f(x), y = 0, x = x_0, x = x_1$ (Hình vẽ)

Với giả thiết $f(x)$ liên tục, h khá bé ta có thể xấp xỉ diện tích hình thang cong bằng diện tích hình thang vuông, tức là

$$\int_{x_0}^{x_1} f(x)dx \simeq \frac{y_0+y_1}{2}h, \qquad y_i = f(x_i)$$

Tương tự, ta có

$$\int_a^b f(x)dx = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)dx \simeq \sum_{i=0}^{n-1} h\frac{y_i+y_{i+1}}{2}$$

Như vậy

$$\int_a^b f(x)dx \simeq \frac{h}{2}[y_0 + y_n + 2(y_1 + \cdots + 2y_{n-1})] \tag{6.1}$$

b) Sai số phương pháp

i) *Sai số địa phương.* Thực chất của việc thay $\int_{x_0}^{x_1} f(x)dx \simeq \frac{y_0+y_1}{2}h$ là xấp xỉ hàm $f(x)$ trên đoạn $[x_0, x_1]$ bằng đa thức nội suy bậc nhất:

$$f(x) \simeq P_1(x) = \frac{y_0(x-x_1)}{x_0-x_1} + \frac{y_1(x-x_0)}{x_1-x_0}.$$

Sai số của phép nội suy tuyến tính là:

$$|R_1(x)| = \left|\frac{f''(\xi)}{2}(x-x_0)(x-x_1)\right| \le \frac{M}{2}(x-x_0)(x_1-x)$$

trong đó $M = \max\limits_{x\in[x_0,x_1]} |f''(x)|$.

Dễ thấy

$$r_1 := |\int_{x_0}^{x_1} f(x)dx - \frac{y_0+y_1}{2}h| \le \int_{x_0}^{x_1} |R_1(x)|dx \le \frac{M}{2}\int_{x_0}^{x_1}(x-x_0)(x_1-x)dx = \frac{Mh^3}{12}$$

ii) *Sai số toàn phần.* $r = n.\frac{Mh^3}{12} = \frac{M(b-a)}{12}h^2.$

• **Ví dụ 2** *Tính* $I = \int_1^5 \frac{dx}{x}$ *bằng cách chia* $[1,5]$ *thành 4 đoạn bằng nhau*

Giải

Ở đây $a = 1, b = 5, n = 4, h = \frac{b-a}{n}, f(x) = \frac{1}{x}$, *theo* (6.1) *ta có*

$$I \simeq \frac{5-1}{2\times 4}\left(1 + \frac{2}{2} + \frac{2}{3} + \frac{2}{4} + \frac{1}{5}\right) = \frac{101}{60}.$$

Vì $f''(x) = 2x^{-3}$ *nên* $|f''(x)| \leq 2$ $(\forall x \in [1,5])$ *và như vậy*

$$r = \frac{2 \times 4}{12} \times 1^2 = \frac{2}{3} \simeq 0.66$$

6.2.2. Công thức parabol (Simpson)

a) Công thức

Chia đoạn $[a,b]$ thành $2n$ phần bằng nhau với bước $h = (b-a)/2n$. Trên mỗi đoạn $[x_{2i-2}, x_{2i}]$ $(i = \overline{1,n})$ ta thay $f(x)$ bằng đa thức nội suy bậc hai (parabol) với các mốc nội suy $x_{2i-2}, x_{2i-1}, x_{2i}$.

$$\begin{aligned} f(x) \quad &\simeq \quad P_2(x) = y_{2i-2}\frac{(x-x_{2i-1})(x-x_{2i})}{(x_{2i-2}-x_{2i-1})(x_{2i-2}-x_{2i})} \\ &+y_{2i-1}\frac{(x-x_{2i-2})(x-x_{2i})}{(x_{2i-1}-x_{2i-2})(x_{2i-1}-x_{2i})} + y_{2i}\frac{(x-x_{2i-2})(x-x_{2i-1})}{(x_{2i}-x_{2i-2})(x_{2i}-x_{2i-1})} \end{aligned}$$

Khi đó $\int\limits_{x_{2i-2}}^{x_{2i}} f(x)dx \simeq \int\limits_{x_{2i-2}}^{x_{2i}} P_2(x)dx = \frac{h}{3}(y_{2i-2} + 4y_{2i-1} + y_{2i})$.

Vì $\int\limits_a^b f(x)dx = \sum\limits_{i=1}^n \int_{x_{2i-2}}^{x_{2i}} f(x)dx \simeq \sum\limits_{i=1}^n \frac{h}{3}(y_{2i-2} + 4y_{2i-1} + y_{2i})$ nên ta có công thức sau:

$$\int\limits_a^b f(x)dx \simeq \frac{h}{3}[y_0 + y_{2n} + 4(y_1 + y_3 + \cdots + y_{2n-1}) + 2(y_2 + y_4 + \cdots + y_{2n-2})] \tag{6.2}$$

a) Sai số phương pháp

i) *Sai số địa phương.* Xét hàm $F(t) := \Phi(t) - (\frac{t}{h})^5\Phi(h)$ $(0 \leq t \leq h)$, trong đó

$$\Phi(t) = \int\limits_{x_i-t}^{x_i+t} f(x)dx - \frac{t}{3}\big[f(x_i-t) + 4f(x_i) + f(x_i+t)\big].$$

Dễ thấy $F(0) = F(h) = 0; F'(0) = F''(0) = 0$, còn

$$F^{(3)}(t) = -\frac{t}{3}\big[f^{(3)}(x_i+t) - f^{(3)}(x_i-t)\big] - \frac{60t^2}{h^5}\Phi(h),$$

hay

$$F^{(3)}(t) = -\frac{2t^2}{3}\big[f^{(4)}(\xi) + \frac{90}{h^5}\Phi(h)\big], \tag{6.3}$$

trong đó $\xi \in (x_i - t, x_i + t)$. Áp dụng định lý Rolle ta có: do $F(0) = F(h) = 0$ nên tìm được $t_1 \in (0,h)$ để $F'(t_1) = 0$. Tiếp theo $F'(0) = F'(t_1) = 0$, ta tìm được $t_2 \in (0,t_1)$ để $F''(t_2) = 0$. Cuối cùng, do $F''(0) = F''(t_2) = 0$, tồn tại $t_3 \in (0,t_2)$ sao cho $F^{(3)}(t_3) = 0$.

Từ (6.3) ta suy ra:

$$\Phi(h) = -\frac{h^5}{90} f^{(4)}(\xi), \text{ trong đó } x_i - t_3 < \xi < x_i + t_3.$$

Như vậy:

$$\int\limits_{x_i-h}^{x_i+h} f(x)dx = \frac{h}{3}\big[f(x_i-h) + 4f(x_i) + f(x_i+h)\big] - \frac{h^5}{90} f^{(4)}(\xi). \tag{6.4}$$

Đặt $M = \max\left\{|f^{(4)}(x)| : x \in [a,b]\right\}$, ta có ước lượng sau:

$$|r_{loc}| \le \frac{Mh^5}{90}.$$

Từ đây suy ra sai số toàn phần của công thức Simpson

ii) *Sai số toàn phần.*

$$|r_{glob}| \le n\frac{Mh^5}{90} = \frac{b-a}{180} Mh^4 \tag{6.5}$$

• **Ví dụ 3** *Tính tích phân* $I = \int\limits_0^1 \frac{dx}{1+x^2}$ *theo công thức parabol với* $n = 2$

Giải

Ở đây $a = 0, b = 1, h = 0,25, f(x) = \frac{1}{1+x^2}$. *Ta lập bảng giá trị*

i	x_i	$y_i = f(x_i)$
0	0,00	1,000000
1	0,25	0,941176
2	0,50	0,800000
3	0,75	0,640000
4	1,00	0,500000

$$I \simeq \frac{h}{3}(y_0 + 4y_1 + 2y_2 + 4y_3 + y_4) = 1 + 3.76471 + 1.6 + 2.56000 + 0.5 \simeq 0.785399$$

Để ý rằng $\pi = 4I$, *ta được biểu thức gần đúng* $\pi \simeq 3.141596$.

⊙ **Sai số Runge**

Phương pháp Runge trình bày dưới đây cho phép ta nhận được các ước lượng hậu nghiệm khá hiệu quả của sai số.

Gọi R là sai số của phương pháp Simson, theo (6.5) R có dạng $R \simeq Ch^4$, trong đó $C = const > 0$. Tính tích phân I hai lần theo công thức (6.2) với bước h và $h/2$, ta được

$$I = \sum_h + Ch^4 = \sum_{h/2} + C(h/2)^4.$$

Từ đây suy ra $R = Ch^4 = \left[\sum_{h/2} - \sum_h\right] \times 16/15$

6.2.3. Công thức Newton-Cotes

Giả sử phải tính tích phân $I = \int\limits_a^b f(x)dx$. Đổi biến $t = \dfrac{x-a}{b-a}$, ta được

$$I = \int\limits_a^b f(x)dx = (b-a)\int_0^1 \Phi(t)dt,$$

trong đó $\Phi(t) = f(a+(b-a)t)$. Chia $[a,b]$ thành n phần bằng nhau với bước $h = \dfrac{b-a}{n}$, ta được $x_i = a + ih$ và $t_i = \dfrac{i}{n}$ $(i = \overline{0,n})$.

Thay $I = (b-a)\int\limits_0^1 \Phi(t)dt \simeq (b-a)\int\limits_0^1 P(t)dt$, trong đó $P(t)$ là đa thức nội suy Lagrange của hàm $\Phi(t)$ với các số liệu $\{t_i, \Phi(t_i)\} = \{t_i, f(x_i)\}$. Đặt $y_i = f(x_i)$, ta có:

$$P(x) = \sum_{i=0}^n y_i l_i(x) = y_0 \frac{(t-\frac{1}{n})\cdots(t-1)}{(0-\frac{1}{n})\cdots(0-1)} + \cdots + y_n \frac{(t-0)\cdots(t-\frac{n-1}{n})}{(1-0)\cdots(1-\frac{n-1}{n})}.$$

Đặt

$$P_n^i := \frac{\int\limits_0^1 (t-0)(t-\frac{1}{n})\cdots(t-\frac{i-1}{n})(t-\frac{i+1}{n})\cdots(t-1)dt}{(\frac{i}{n}-0)(\frac{i}{n}-\frac{1}{n})\cdots(\frac{i}{n}-\frac{i-1}{n})(\frac{i}{n}-\frac{i+1}{n})\cdots(\frac{i}{n}-1)},$$

ta nhận thấy các hệ số $P_n^i (i = \overline{0,n})$ không phụ thuộc vào hàm $f(x)$ và đoạn lấy tích phân $[a,b]$, do đó chúng có thể tính sẵn, lập thành bảng và sử dụng lâu dài.

Trong biểu thức của P_i^n nếu đổi biến $z = 1-t$ ta dễ dàng chứng minh được rằng $P_n^i = P_n^{n-i}$ $(i = \overline{0,n})$.

Ngoài ra nếu đặt $f(x) = \Phi(x)$ và do đó

$$(b-a) = (b-a)\sum_{i=0}^n P_n^i \text{ hay } \sum_{i=0}^n P_n^i = 1.$$

Ta xét các trường hợp riêng:

a) Khi $n = 1, P_1^0 = P_1^1 = \dfrac{\int\limits_0^1 (t-1)dt}{0-1} = \dfrac{1}{2}$,

$$\int_a^b f(x)dx \simeq (b-a)\frac{y_0+y_1}{2}.$$

Ta được công thức hình thang (địa phương)

b) Khi $n = 2$, ta có:

$$P_2^0 = P_2^2 = \frac{\int_0^1 (t-\frac{1}{2})(t-1)dt}{(0-\frac{1}{2})(0-1)} = \frac{1}{6}; P_2^1 = \frac{\int_0^1 (t-0)(t-1)dt}{(\frac{1}{2}-0)(\frac{1}{2}-1)} = \frac{4}{6}$$

Như vậy $\int_a^b f(x)dx \simeq \dfrac{b-a}{6}(y_0 + 4y_1 + y_2)$. Đây là công thức parabol (địa phương).

Khi n lớn, các hệ số Newton-Cotes P_n^i khá phức tạp. Vì vậy ta nên chia đoạn $[0,1]$ thành một số phần bằng nhau. Sau đó áp dụng công thức Newton-Cotes với n' nhỏ hơn trên từng đoạn con.

• **Ví dụ 4** *Tính tích phân* $\int\limits_0^1 \dfrac{dx}{1+x}$ *theo công thức Newton-Cotes với* $n = 6$.

Giải

Ta có $P_6^0 = P_6^6 = \dfrac{41}{840}$; $P_6^1 = P_6^5 = \dfrac{9}{35}$; $P_6^2 = P_6^4 = \dfrac{9}{280}$; $P_6^3 = \dfrac{34}{105}$; $h = \dfrac{1-0}{6} = \dfrac{1}{6}$. *Đặt* $\widetilde{P}_n^i = 840P_6^i$. *Từ công thức Newton-Cotes suy ra*

$$I = \sum_{i=0}^{6} y_i P_6^i = \frac{1}{840}\sum_{i=0}^{6} y_i \widetilde{P}_i.$$

Bài tập chương 6

Bài 6.1. Cho hàm số $y = f(x)$ và bảng giá trị sau

x	0	1	2	3
$y = f(x)$	-6	2	-2	6

a) Tìm đa thức nội suy Lagrange của hàm số $y = f(x)$.

b) Dùng đa thức nội suy Lagrange tính gần đúng đạo hàm $f'(1)$ và $f''(2,5)$.

Bài 6.2. Xét tích phân $I = \int\limits_0^1 \dfrac{dx}{2x+1}$

a) Tính gần đúng I bằng công thức hình thang với 10 đoạn chia và đánh giá sai số.

b) Tính gần đúng I bằng công thức hình thang với sai số không quá 10^{-4}.

c) Để tính I với 10 chữ số chắc thì số đoạn chia tối thiểu là bao nhiêu?

Bài 6.3. Cho tích phân $I = \int\limits_0^1 \dfrac{\sin x}{x}dx$

a) Phải chia đoạn $[0,1]$ thành mấy đoạn con bằng nhau để khi tính I bằng công thức hình thang có sai số không quá 3.10^{-4}.

b) Với n ở câu a thì sai số là bao nhiêu khi tính I theo công thức Simson.

c) Hãy tính I với n đã chọn ở trên bằng công thức hình thang và công thức Simson đến 6 chữ số lẻ thập phân.

Bài 6.4. Lập chương trình tính tích phân $I = \int\limits_0^1 \frac{dx}{1+x^2}$ theo công thức Simson. Số liệu vào $h_0 = 0,05; \varepsilon = 10^{-4}$. Bước lưới giảm đi một nửa ($h_1 := h_0/2$) nếu kết quả bước trước so với bước sau chưa thỏa mãn điều kiện $16 \times (I_1 - I_0)/15 < \varepsilon$. Sử dụng đẳng thức $4I = \pi$ để kiểm tra kết quả tính toán. **Bài 6.5.** Trong lý thuyết số học người ta đã chứng minh rằng số các số nguyên tố trong khoảng (a, b) xấp xỉ bằng $\int\limits_a^b \frac{dx}{\ln x}$.

a) Lập chương trình tính số các số nguyên tố trong khoảng $(100, 200)$.

b) Tính tích phân $I = \int\limits_{100}^{200} \frac{dx}{\ln x}$ theo công thức Simson với $n = 100$. So sánh kết quả với phần a).

Chương 7

Giải gần đúng phương trình vi phân

7.1. Mở đầu

Trong thực tế nhiều bài toán khoa học kỹ thuật dẫn đến việc tìm nghiệm phương trình vi phân thỏa mãn một số điều kiện nào đó. Những phương trình vi phân mô tả những hệ cơ học, lý học, hóa học, sinh học nói chung rất phức tạp nên không hy vọng tìm được lời giải đúng. Chính vì vậy trong chương này này ta sẽ nghiên cứu các phương pháp giải gần đúng. Để đơn giản ta chỉ xét *bài toán Cauchy đối với phương trình vi phân cấp một* như sau:

Hãy tìm hàm $y = y(x)$ thỏa mãn:

$$\begin{cases} y'(x) &= f(x,y), x \in [x_0, X] \\ y(x_0) &= \eta \end{cases} \tag{7.1}$$

trong đó $f(x,y)$ là một hàm số đã biết của hai đối số x, y, còn η là một số thực cho trước. Nói chung có hai nhóm phương pháp giải gần đúng:

- Phương pháp giải tích: tìm nghiệm dưới dạng một biểu thức giải tích (phương pháp xấp xỉ liên tiếp Picard, phương pháp chuỗi Taylor, . . .).

- Phương pháp số: tìm nghiệm phương trình tại các điểm rời rạc (phương pháp Euler, phương pháp Euler cải tiến, phương pháp Runge-Kutta,. . .).

7.2. Phương pháp chuỗi Taylor

Xét bài toán (7.1) Ta đi tìm nghiệm $y(x)$ dưới dạng khai triển Taylor (chuỗi Taylor) tại $x = x_0$:

$$y(x) = \sum_{i=0}^{\infty} \frac{y^{(i)}(x_0)}{i!}(x - x_0)^i \tag{7.2}$$

Bây giờ ta tính các đạo hàm $y^{(i)}(x_0)$ của y tại x_0. Theo điều kiện ban đầu:

$$\begin{aligned}
&y(x_0) = y_0\\
&y'(x) = f(x,y) \Rightarrow y'(x_0) = f(x_0,y_0)\\
&y''(x) = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}y' \Rightarrow y''(x_0) = \frac{\partial f}{\partial x}(x_0,y_0) + \frac{\partial f}{\partial y}(x_0,y_0)f(x_0,y_0)
\end{aligned}$$

Một cách tương tự, để tính $y'''(x_0)$ trước hết ta phải lấy đạo hàm $y''(x)$, sau đó thay $x = x_0 \ldots$.

Về nguyên tắc, nếu f có đạo hàm riêng cấp bất kì thì có thể tính được tất cả các đạo hàm $y^{(i)}(x_0)$ và do đó xây dựng được chuỗi (7.2).

Với x khá gần x_0, tức là $|x - x_0|$ đủ bé người ta chứng minh được rằng chuỗi (7.2) hội tụ về nghiệm của bài toán (7.1). Khi đó có thể xem tổng của n số hạng đầu của (7.2) là nghiệm gần đúng của bài toán (7.1), mức độ chính xác phụ thuộc n, nó càng cao nếu n càng lớn.

• **Ví dụ 1** *Giải bài toán Côsi:* $\begin{cases} y' = x - y\,, & 0 \le x \le 1 \\ y(0) = 1 \end{cases}$

Giải

Ta có:

$$\begin{aligned}
y(0) &= 1\\
y'(0) &= 0 - 1 = -1\\
y'' &= 1 - y' \Rightarrow y''(0) = 2\\
y''' &= -y'' \Rightarrow y'''(0) = -2
\end{aligned}$$

Dễ thấy $y^{(k)}(0) = (-1)^k 2 \ (k \ge 2)$.

Vậy $y(x) = 1 - x + 2\sum\limits_{k=2}^{\infty} \dfrac{(-1)^k x^k}{k!} = 1 - x + 2\Big(\sum\limits_{k=0}^{\infty} \dfrac{(-x^k}{k!} - 1 + x\Big) = 2e^{-x} + x - 1.$

• **Ví dụ 2** *Xét bài toán*

$$\begin{cases} y' = \dfrac{y}{x+y}\,, & 1 \le x \le 2 \\ y(1) = 2 \end{cases}$$

Giải

Ta đi tìm nghiệm có dạng chuỗi Taylor. Ta có

$$\begin{cases} y'(1) = \dfrac{2}{1+2} = \dfrac{2}{3}; \\ y''(x) = (\dfrac{y}{x+y})' = \dfrac{xy' - y}{(x+y)^2} \Rightarrow y''(1) = -\dfrac{4}{27}; \\ y'''(1) = \dfrac{4}{27}\cdots \end{cases}$$

Như vậy nghiệm

$$y(x) = 2 + \frac{2}{3}(x-1) - \frac{2}{27}(x-1)^2 + \frac{2}{81}(x-1)^3 + \cdots$$

Bây giờ có thể dùng công thức này để tính gần đúng $y(x)$ tại $x = 1,1$ chẳng hạn, vì $|1,1-1| = 0,1$ khá bé nên ta bỏ qua các số hạng ở cuối và thu được:

$$y(1,1) \approx 2 + \frac{2}{3}0,1 - \frac{2}{27}(0,1)^2 + \frac{2}{81}(0,1)^3 \approx 2,06584.$$

Như vậy nội dung của phương pháp rất đơn giản nhưng tính toán rất phức tạp, hơn nữa bán kính hội tụ của chuỗi $y(x)$ rất khó xác định, do đó trong phần tiếp theo ta sẽ tập trung vào nhóm phương pháp thứ hai.

7.3. Phương pháp Euler

Các phương pháp số trong đó có phương pháp Euler cho ta nghiệm bằng số của bài toán (7.1) tại các điểm $x_i = x_0 + ih$ với bước $h > 0$. Để ý rằng theo phương pháp chuỗi Taylor,

$$y_1 = \sum_{i=0}^{\infty} \frac{y_0^{(i)}}{i!} h^i, \quad \text{với } y_0^{(i)} = y^{(i)}(x_0).$$

Như vậy để tìm nghiệm tại điểm x_1 ta phải tính giá trị của các hàm, nói chung khác nhau tại một điểm x_0. Ở đây phương pháp Euler cho phép ta tìm giá trị y_1 chỉ phải tính một hàm $f(x,y)$ tại một số điểm khác nhau.

Trở lại bài toán (7.1)

$$\begin{cases} y'(x) = f(x,y), x \in [x_0, X] \\ y(x_0) = \eta \end{cases}$$

Ta chia đoạn $[x_0, X]$ thành n đoạn con bằng nhau bởi các điểm chia: $x_0 < x_1 < \cdots < x_n = X$, trong đó

$$x_i = x_0 + ih, \quad h = \frac{X - x_0}{n}$$

Giả sử $y(x)$ là nghiệm đúng của bài toán (7.1). Mục đích của phương pháp Euler là tìm các giá trị gần đúng y_i của $y(x_i)$.

Thật vậy, khai triển Taylor hàm $y(x)$ tại x_i ta có:

$$y(x) = y(x_i) + y'(x_i)(x - x_i) + \frac{y''(c_i)}{2!}(x - x_i)^2,$$

trong đó $c_i = x_i + \theta(x - x_i), \quad 0 < \theta < 1$. Thay $x = x_{i+1} = x_i + h$, $y'(x_i) = f(x_i, y(x_i))$ vào đẳng thức trên ta được

$$y(x_{i+1}) = y(x_i) + hf(x_i, y(x_i)) + \frac{h^2}{2}y''(c_i)$$

Khi h khá bé thì có thể bỏ qua số hạng cuối cùng bên phải, đồng thời thay tương ứng các giá trị đúng $y(x_{i+1}), y(x_i)$ bằng các giá trị gần đúng y_{i+1}, y_i ta được công thức

$$\begin{cases} y_0 = y(x_0) = \eta \\ y_{i+1} = y_i + hf(x_i, y_i), \quad \forall i \geq 0, \end{cases} \tag{7.3}$$

Công thức (7.3) cho ta cách tính y_{i+1} khi đã biết y_i ($i \geq 0$) mà không phải giải một phương trình nào. Do đó phương pháp Euler là một phương pháp hiện.

⊙ **Sai số địa phương của phương pháp Euler**

$$R_i(h) = y(x_i) - y_i = \frac{h^2}{2} y''(c_i)$$

Người ta chứng minh được rằng sai số của phương pháp Euler tại điểm x_i là

$$|R_i(h)| = |y(x_i) - y_i| \leq Mh,$$

trong đó M là hằng số không phụ thuộc h. Điều này chứng tỏ khi $h \to 0$ thì $y_i \to y(x_i)$ tại mọi điểm x_i cố định.

• **Ví dụ 3** *Giải phương trình sau bằng phương pháp Euler:*

$$\begin{cases} y' = y - \dfrac{2x}{y}, \quad 0 \leq x \leq 1 \\ y(0) = 1, \end{cases}$$

với $h = 0,1$.

Giải

Nghiệm đúng của bài toán này là $y = \sqrt{2x+1}$.
Theo (7.3) *ta có:*

$$\begin{cases} x_i = ih = 0,1i; \\ y_0 = 1; \\ y_{i+1} = y_i + h(y_i - \dfrac{2x_i}{y_i}), \quad i \geq 0. \end{cases}$$

Kết quả tính toán ghi thành bảng sau.

i	x_i	y_i	Nghiệm đúng $y(x_i)$
0	0,0	1	1
1	0,1	1,1	1,095445
2	0,2	1,191818	1,183216
3	0,3	1,277438	1,264911
4	0,4	1,358213	1,31641
5	0,5	1,435133	1,414214
6	0,6	1,508966	1,483249
7	0,7	1,580338	1,549193
8	0,8	1,649783	1,612452
9	0,9	1,717779	1,673320
10	1	1,784771	1,732051

7.4. Phương pháp Euler cải tiến

Để tăng độ chính xác của phương pháp Euler người ta làm như sau: Theo công thức Newton-Lepnitz ta có

$$y(x_2) - y(x_1) = \int_{x_1}^{x_2} y'(x)\mathrm{d}x.$$

Tính gần đúng tích phân ở vế phải bằng công thức hình thang ta có:

$$\int_{x_1}^{x_2} y'(x)\mathrm{d}x \approx \frac{h}{2}\left[y'(x_1) + y'(x_2)\right] = \frac{h}{2}\left[f(x_1, y(x_1)) + f(x_2, y(x_2))\right],$$

trong đó $h = x_2 - x_1$. Thay x_1 bằng x_i và x_2 bằng x_{i+1} ta được

$$\begin{cases} y_{i+1} = y_i + \dfrac{h}{2}\left[f(x_i, y_i) + f(x_{i+1}, y_{i+1})\right], \quad \forall i \geq 0, \\ y_0 = y(x_0) = \eta \end{cases} \tag{7.4}$$

Người ta chứng minh được rằng sai số của công thức (7.4) tại điểm x_i là:

$$|y_i - y(x_i)| \leq Mh^2,$$

trong đó M là hằng số không phụ thuộc h. Như vậy công thức (7.4) chính xác hơn công thức (7.3). Tuy nhiên nó có nhược điểm là y_{i+1} xuất hiện ở cả hai vế, do đó khi đã biết y_i ta vẫn phải giải một phương trình đối với y_{i+1}. Để khắc phục điều này người ta đã cải tiến công thức (7.4) bằng cách kết hợp với phương pháp Euler như sau:

$$\left\{ \begin{array}{rcl} y_0 & = & y(x_0) = \eta \\ y_{i+1}^{(0)} & = & y_i + hf(x_i, y_i) \\ y_{i+1}^{(m)} & = & y_i + \dfrac{h}{2}\left[f(x_i, y_i) + f(x_{i+1}, y_{i+1}^{(m-1)})\right] \end{array} \right. \tag{7.5}$$

Quá trình lặp trong (7.5) sẽ dừng lại khi hai giá trị $y_{i+1}^{(m)}$ và $y_{i+1}^{(m-1)}$ gần nhau đến mức cần thiết, nghĩa là $\left|y_{i+1}^{(m)} - y_{i+1}^{(m-1)}\right| \leq \varepsilon$ (cho trước). Công thức (7.5) được gọi là công thức Euler cải tiến.

• **Ví dụ 4** *Cho phương trình* $\begin{cases} y' = x + y \\ y(0) = 1 \end{cases}$ *với* $x \in [0; 0,5]$. *Tính* $y(0,25)$ *bằng phương pháp Euler cải tiến với* $h = 0,25$ *và sai số không quá* 10^{-3}.

Giải

Với $h = 0,25$ *ta có* $x_1 = x_0 + h = 0,25$. *Áp dụng công thức Euler cải tiến ta được*

$$\begin{cases} y_1^{(0)} &= y_0 + hf(x_0, y_0) = 1 + 0,25(0+1) = 1,250000 \\ y_1^{(1)} &= y_0 + \dfrac{h}{2}\left[f(x_0, y_0) + f(x_0, y_1^{(0)})\right] = 1,312500 \\ y_1^{(2)} &= y_0 + \dfrac{h}{2}\left[f(x_0, y_0) + f(x_0, y_1^{(1)})\right] = 1,320313 \\ y_1^{(3)} &= y_0 + \dfrac{h}{2}\left[f(x_0, y_0) + f(x_0, y_1^{(2)})\right] = 1,321289 \end{cases}$$

Ta thấy $\left|y_1^{(3)} - y_1^{(2)}\right| = 9,76.10^{-4} < 10^{-3}$, *do đó ta lấy* $y_1 \approx y_1^{(3)} = 1,321289$

7.5. Phương pháp Runge-Kutta

Để nhận được các công thức có độ chính xác cao hơn công thức Euler, Euler cải tiến người ta dùng khai triển Taylor nghiệm $y(x)$ của bài toán (7.1) tại x_i với nhiều số hạng hơn. Thật vậy, ta có

$$y(x) = y(x_i) + y'(x_i)(x - x_i) + \frac{y''(x_i)}{2!}(x - x_i)^2 + \frac{y'''(c_i)}{3!}(x - x_i)^3,$$

trong đó $c_i = x_i + \theta(x - x_i), \quad 0 < \theta < 1$.

Thay $x = x_{i+1} = x_i + h$, ta được:

$$y(x_{i+1}) = y(x_i) + hy'(x_i) + \frac{h^2}{2}y''(x_i) + \frac{h^3}{6}y'''(c_i),$$

trong đó

$$\begin{cases} y'(x_i) = f(x_i, y(x_i)) \\ y''(x_i) = f'_x(x_i, y(x_i)) + f'_y(x_i, y(x_i))y'(x_i) \end{cases}$$

Thay các giá trị đúng $y(x_{i+1})$, $y(x_i)$, $y'(x_i)$ bằng các giá trị gần đúng y_{i+1}, y_i, y'_i ta được công thức

$$y_{i+1} = y_i + hy'_i + \frac{h^2}{2}\left[f'_x(x_i, y_i) + f'_y(x_i, y_i)y'_i\right] + O(h^3), \quad \forall i \geq 0, \tag{7.6}$$

Để tránh tính trực tiếp $f'_x(x_i, y_i)$ và $f'_y(x_i, y_i)$, Runge và Kutta đã làm như sau:

Đặt

$$y_{i+1} = y_i + r_1 k_1^{(i)} + r_2 k_2^{(i)}, \tag{7.7}$$

với

$$\begin{cases} k_1^{(i)} = hf(x_i, y_i) \\ k_2^{(i)} = hf(x_i + \alpha h, y_i + \beta k_1^{(i)}) \end{cases} \tag{7.8}$$

trong đó α, β, r_1, r_2 được chọn sao cho y_{i+1} trong (**??**) và (7.6) trùng nhau đến 3 số hạng đầu. Sử dụng công thức Taylor của hàm 2 biến, ta có:

$$\begin{aligned} f(x_i + \alpha h, y_i + \beta k_1^{(i)}) &= f(x_i, y_i) + \alpha h f'_x(x_i, y_i) + \beta k_1^{(i)} f'_y(x_i, y_i) + O(h^2) \\ &= y'_i + \alpha h f'_x(x_i, y_i) + \beta k_1^{(i)} f'_y(x_i, y_i) + O(h^2). \end{aligned}$$

Khi đó

$$\begin{cases} k_1^{(i)} = hf(x_i, y_i) = hy'_i \\ k_2^{(i)} = hf(x_i + \alpha h, y_i + \beta k_1^{(i)}) = hy'_i + \alpha h^2 f'_x(x_i, y_i) + \beta h^2 y'_i f'_y(x_i, y_i) + O(h^3). \end{cases}$$

Suy ra

$$\begin{aligned} y_{i+1} &= y_i + r_1 h y'_i + r_2 \left[h y'_i + \alpha h^2 f'_x(x_i, y_i) + \beta h^2 y'_i f'_y(x_i, y_i) \right] + O(h^3) \\ &= y_i + (r_1 + r_2) h y'_i + h^2 \left[\alpha r_2 f'_x(x_i, y_i) + \beta r_2 f'_y(x_i, y_i) y'_i \right] + O(h^3) \end{aligned} \tag{7.9}$$

So sánh các hệ số lũy thừa của h trong (7.6) và (7.9) ta nhận được

$$\begin{cases} r_1 + r_2 = 1 \\ \alpha r_2 + \beta r_2 = \dfrac{1}{2} \end{cases}$$

Hệ này gồm 3 phương trình, 4 ẩn số nên vô số nghiệm. Ở đây ta chỉ xét hai họ nghiệm đơn giản của hệ.

(1) $r_1 = 0, r_2 = 1, \alpha = \beta = \dfrac{1}{2}$. Khi đó bỏ qua số hạng $O(h^3)$ các công thức (7.7), (7.8) có dạng

$$\begin{cases} y_0 = y(x_0) = \eta \\ k_1^{(i)} = hf(x_i, y_i) \\ k_2^{(i)} = hf(x_i + h/2, y_i + k_1^{(i)}/2) \\ y_{i+1} = y_i + k_2^{(i)}, \forall i \geq 0 \end{cases} \tag{7.10}$$

(2) $r_1 = r_2 = 1/2, \alpha = \beta = 1$. Khi đó bỏ qua số hạng $O(h^3)$ các công thức (7.7), (7.8) trở thành

$$\begin{cases} y_0 = y(x_0) = \eta \\ k_1^{(i)} = hf(x_i, y_i) \\ k_2^{(i)} = hf(x_i + h, y_i + k_1^{(i)}) \\ y_{i+1} = y_i + \dfrac{1}{2}(k_1^{(i)} + k_2^{(i)}), \forall i \geq 0 \end{cases} \tag{7.11}$$

Trong cả hai trường hợp trên người ta chứng minh được rằng sai số tại điểm x_i thỏa mãn

$$|y_i - y(x_i)| \leq Mh^2,$$

trong đó M là hằng số dương không phụ thuộc vào h, tức là phương pháp của Runge-Kutta có độ chính xác cấp hai.

Hoàn toàn tương tự, nếu trong khai triển Taylor của $y(x_{i+1})$ tại x_i ta bỏ qua số hạng $O(h^4)$ thì ta nhận được công thức Runge-Kutta có độ chính xác cấp ba, nghĩa là

$$|y_i - y(x_i)| \leq Mh^3,$$

trong đó

$$\begin{cases} y_0 = y(x_0) = \eta \\ k_1^{(i)} = hf(x_i, y_i) \\ k_2^{(i)} = hf(x_i + h/2, y_i + k_1^{(i)}/2) \\ k_3^{(i)} = hf(x_i + h, y_i - k_1^{(i)} + 2k_2^{(i)}) \\ y_{i+1} = y_i + \dfrac{1}{6}(k_1^{(i)} + 4k_2^{(i)} + k_3^{(i)}), \forall i \geq 0 \end{cases} \tag{7.12}$$

Nếu bỏ qua số hạng $O(h^5)$ thì ta nhận được công thức Runge-Kutta có độ chính xác cấp bốn, nghĩa là

$$|y_i - y(x_i)| \leq Mh^4,$$

trong đó

$$\begin{cases} y_0 = y(x_0) = \eta \\ k_1^{(i)} = hf(x_i, y_i) \\ k_2^{(i)} = hf(x_i + h/2, y_i + k_1^{(i)}/2) \\ k_3^{(i)} = hf(x_i + h/2, y_i + k_2^{(i)}/2) \\ k_4^{(i)} = hf(x_i + h/2, y_i + k_3^{(i)}) \\ y_{i+1} = y_i + \dfrac{1}{6}(k_1^{(i)} + 2k_2^{(i)} + 2k_3^{(i)} + k_4^{(i)}), \forall i \geq 0 \end{cases} \tag{7.13}$$

• **Ví dụ 5** *Cho bài toán Cô si $y' = x + y, y(0) = 1$. Hãy tìm nghiệm gần đúng bằng phương pháp Runge-Kutta* (7.13) *trên* $[0; 0,5]$ *với bước* $h = 0,1$.

Giải

Ta có $x_i = 0,1i$ *với* $i = 0,1,2,3,4,5$

$$\begin{cases} y_0 = 1 \\ k_1^{(0)} = 0,1(0+1) = 0,1 \\ k_2^{(0)} = 0,1\left[(0+0,05)+(1+0,05)\right] = 0,11 \\ k_3^{(0)} = 0,1\left[(0+0,05)+(1+0,055)\right] = 0,1105 \\ k_4^{(0)} = 0,1\left[(0+0,1)+(1+0,1105)\right] = 0,12105 \end{cases}$$

Từ đó

$$y_1 = 1 + \frac{1}{6}(0,1 + 2.0,11 + 2.0,1105 + 0,12105) = 1,1103.$$

Tương tự ta có thể tính y_2, y_3, y_4, y_5. *Dễ thấy* $y(0,5) \approx y_5 = 1,7974$.
Nghiệm đúng của bài toán đã cho là $y = 2e^x - x - 1$ *suy ra* $y(0,5) = 2e^{0,5} - 0,5 - 1 = 1,79744$.
Như vậy kết quả nhận được đúng đến 4 số lẻ thập phân.

Bài tập chương 7

Bài 7.1. Giải phương trình sau bằng phương pháp Euler

$$y' = \frac{xy}{2}; x \in [0,1]; y(0) = 1; h = 0,1$$

Bài 7.2. Giải phương trình sau bằng phương pháp Euler

$$y' = x^2 + y^2; x \in [0,1]; y(0) = 1; h = 0,2$$

Bài 7.3. Giải phương trình sau bằng phương pháp Euler cải tiến và so sánh kết quả với nghiệm đúng:

$$y' = x^2 + y^2; x \in [0,1]; y(0) = 1; h = 0,2$$

Bài 7.4. Giải phương trình sau bằng phương pháp Runge-Kutta

$$y' = y - \frac{2x}{y}; x \in [0,1]; y(0) = 1; h = 0,2$$

Phụ lục A

Giải tích tổ hợp

A.1. Các quy tắc đếm

A.1.1. Quy tắc cộng

Nếu có m_1 cách chọn đối tượng x_1, m_2 cách chọn đối tượng x_2,..., m_n cách chọn đối tượng x_n; và cách chọn x_i không trùng cách chọn $x_j (i \neq j)$ thì có:

$$m_1 + m2 + ... + m_n$$

cách chọn một trong các đối tượng $x_1, x_2, ..., x_n$.

A.1.2. Quy tắc nhân

Giả sử một phép chọn được thực hiện qua n bước liên tiếp. Bước 1 có m_1 cách thực hiện, bước 2 có m_2 cách,..., bước n có m_n cách thực hiện. Khi đó số cách thực hiện phép chọn là:

$$m_1.m_2...m_n.$$

A.2. Hoán vị - Chỉnh hợp - Tổ hợp

A.2.1. Chỉnh hợp (chỉnh hợp không lặp)

⋆ **Định nghĩa A.1** Cho tập hợp E gồm n phần tử. Chỉnh hợp chập k của n phần tử đã cho là một dãy có thứ tự gồm k $(0 < k \leq n)$ phần tử khác nhau lấy từ n phần tử đã cho.

•**Ví dụ A.1** Cho tập hợp E = a,b,c. Các chỉnh hợp chập 2 của 3 phần tử của E là:

$$ab, ac, ba, bc, ca, cb$$

Vậy có 6 chỉnh hợp chập 2 của 3.

Công thức tính. Ký hiệu số chỉnh hợp chập k của n là $A_n^k, (0 < k \leq n)$.

Để có một chỉnh hợp chập k ta có thể tiến hành chọn theo các bước như sau: Chọn phần tử đầu theo n cách, sau đó chọn phần tử thứ hai theo n- 1 cách, chọn phần tử đứng thứ k theo n — k + 1 cách. Vậy số chỉnh hợp chập k của n là:

$$A_n^k = \underbrace{n(n-1)...(n-k+1)}_{\text{k thừa số}} = \frac{n!}{(n-k)!}.$$

•**Ví dụ A.2** Có bao nhiêu số tự nhiên có 3 chữ số khác nhau mà chữ số nào cũng là số lẻ?

Lời giải. Mỗi số tự nhiên thoả mãn yêu cầu là một bộ 3 chữ số khác nhau (có kể thứ tự) chọn từ 5 chữ số $\{1, 3, 5, 7, 9\}$. Vậy số các số tự nhiên thoả mãn là:

$$A_5^3 = 5.4.3 = 60(\text{ số }).$$

A.2.2. Chỉnh hợp lặp

⋆ **Định nghĩa A.2** Cho tập hợp E gồm n phần tử. Chỉnh hợp lặp chập k của n phần tử đã cho là một dãy có thứ tự gồm k (k > 0) phần tử khác nhau lấy từ n phần tử đã cho, trong đó mỗi phần tử có thể có mặt một hay nhiều lần trong dãy tạo thành.

•**Ví dụ A.3** Cho tập hợp $E = \{a, b, c\}$. Các chỉnh hợp lặp chập 2 của 3 phần tử của E là:

$$aa, ab, ac, ba, bb, bc, ca, cb, cc$$

Vậy có 9 chỉnh hợp lặp chập 2 của 3.

Công thức tính. Ký hiệu số chỉnh hợp lặp chập k của n là $\widetilde{A}_n^k$, (k > 0).

Để có một chỉnh hợp lặp chập k ta có thể tiến hành chọn theo các bước như sau: Chọn phần tử đầu theo n cách, sau đó chọn phần tử thứ hai theo n cách, ..., chọn phần tử đứng thứ k theo n cách. Vậy số chỉnh hợp lặp chập k của n là:

$$\widetilde{A}_n^k = \underbrace{n.n...n}_{\text{k thừa số}} = n^k.$$

•**Ví dụ A.4** Để truyền tin bảng tín hiệu moóc-xơ gồm hai ký hiệu chấm (.) và vạch (-), người ta mã hoá mỗi chữ cái của bảng chữ cái thành một dãy có thứ tự gồm không quá 4 ký hiệu. Biết rằng cùng một ký hiệu có thể có mặt nhiều lần trong dãy có thứ tự tạo thành. Hỏi có thể mã hoá được bao nhiêu chữ cái?

Lời giải. Mỗi nhóm có thứ tự gồm k ký hiệu $(1 \leq k \leq 4)$ tạo nên chính là một chỉnh hợp lặp chập k từ 2 phần tử đã cho (hai phần tử này là các ký hiệu (.) và (-)). Vì vậy số chữ cái mã hoá được là:

$$\widetilde{A}_2^1 + \widetilde{A}_2^2 + \widetilde{A}_2^3 + \widetilde{A}_2^4 + = 2^1 + 2^2 + 2^3 + 2^4 = 30.$$

Như vậy nếu bảng chữ cái của một thứ tiếng nào đó gồm không quá 30 chữ cái thì ta có thể mã hoá theo cách trên.

A.2.3. Hoán vị

⋆ **Định nghĩa A.3** Cho tập hợp E gồm n phần tử. Hoán vị của n phần tử là một dãy có thứ tự gồm đủ n phần tử đã cho.

•**Ví dụ A.5** Cho tập hợp $E = \{a, b, c\}$. Các hoán vị của 3 phần tử của E là:

$$abc, acb, bac, bca, cab, cba$$

Vậy có 6 hoán vị của 3 phần tử.

Công thức tính. Ký hiệu số hoán vị của n phần tử là P_n.

Dễ thấy một hoán vị của n phần tử chính là một chỉnh hợp chập n của n phần tử. Vì vậy ta tính được số hoán vị của n phần tử là:

$$P_n = A_n^n = n(n-1)...1 = n!$$

•**Ví dụ A.6** Một bàn có 5 học sinh ngồi. Khi đó có tất cả: $P_5 = 5! = 120$ cách sắp xếp chỗ ngồi.

A.2.4. Tổ hợp

⋆ **Định nghĩa A.4** Cho tập hợp E gồm n phần tử. Tổ hợp chập k từ n phần tử $(k \leq n)$ là tập con gồm k phần tử (phân biệt, không kể thứ tự) của tập n phần tử đã cho.

•**Ví dụ A.7** ho tập hợp $E = \{a, b, c\}$. Các tổ hợp chập 2 của 3 phần tử là:

$$\{a, b\}; \{b, c\}; \{a, c\}.$$

Vậy có 3 tổ hợp chập 2 của 3.

Công thức tính. Ký hiệu số tổ hợp chập k của n là $C_n^k, (0 \leq k \leq n)$.

Giả sử từ n phần tử đã cho ta tạo nên C_n^k tổ hợp. Đem mỗi tổ hợp chập k này hoán vị theo mọi cách ta sẽ được $k!$ chỉnh hợp chập k. Mặt khác, hai chỉnh hợp chập k tạo ra từ hai tổ hợp chập k phân biệt là khác nhau. Vậy số chỉnh hợp chập k của n phần tử đã cho là:

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

Vì vậy:

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

Tính chất thường gặp về tổ hợp:

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

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

•Ví dụ A.8 Có 10 đội bóng thi đấu với nhau theo thể thức đấu vòng tròn một lượt. Hỏi phải tổ chức bao nhiêu trận đấu?

Lời giải. Mỗi trận đấu ứng với một nhóm gồm 2 đội (phân biệt, không kể thứ tự) từ 10 đội. Vì vậy phải tổ chức tất cả:

$$C_{10}^2 = \frac{10.9}{2.1} = 45 \,(\text{trận đấu}).$$

Bài tập phụ lục A

Bài A.1. Một nhóm sinh viên gồm 3 sinh viên nữ và 4 sinh viên nam. Hỏi có bao nhiêu cách chọn từ nhóm sinh viên đó:
a) 3 sinh viên vào Ban chấp hành Đoàn.
b) 3 sinh viên vào Ban cán sự lớp, trong đó có 1 lớp trưởng, 1 lớp phó phụ trách học tập, 1 lớp phó phụ trách đời sống.
c) 2 sinh viên nam và 2 sinh viên nữ đi tập văn nghệ.
d) 3 sinh viên đi dự Đại hội, trong đó có ít nhất 1 sinh viên nữ.
Đáp số: a) 35; b) 210; c) 18; d) 31.
Bài A.2. Biển số xe ôtô của Hải Phòng có dạng: 16 - 2 chữ cái - 4 chữ số.
a) Hỏi trên lý thuyết có thể lập được bao nhiêu biển xe? (các chữ cái được lấy từ bảng chữ cái tiếng Anh gồm 26 chữ cái).
b) Trong đó có bao nhiêu biển xe mà tổng 4 chữ số là một số có tận cùng là 9?
Đáp số: a) $26^2.10^4$; b) 676000.
Bài A.3. Có bao nhiêu cách xếp chỗ cho 5 người A, B, C, D, E:
a) vào một chiếc ghế dài sao cho A, B ở hai đầu ghế?
b) vào một chiếc ghế dài sao cho A, B, C luôn ở cạnh nhau?
c) vào một bàn tròn sao cho A, B luôn ở cạnh nhau?
Đáp số: a) 12; b) 36; c) 12.
Bài A.4. Có 20 điểm phân biệt trong mặt phẳng trong đó có 8 điểm thẳng hàng. Nối 20 điểm đó lại với nhau. Hỏi:
a) có bao nhiêu đường thẳng phân biệt?
b) có bao nhiêu tam giác?
Đáp số: a) 163; b) 1084.
Bài A.5. Cho phương trình: x + y + z = 12
a) Tìm số nghiệm nguyên dương của phương trình sao cho x, y, z đôi một phân biệt.
b) Tìm số nghiệm nguyên dương của phương trình.
c) Tìm số nghiệm nguyên không âm của phương trình.
Đáp số: a) 42; b) 55; c) 91.
Bài A.6. Cho tập hợp $A = \{0, 1, 2, 3, 4, 5, 6\}$. Từ A có thể tạo ra bao nhiêu:
a) số tự nhiên chẵn 4 chữ số?
b) số tự nhiên chẵn 4 chữ số đôi một phân biệt.
c) số tự nhiên 7 chữ số, trong đó chữ số 2 có mặt đúng 2 lần, số 3 có mặt đúng 3 lần, các chữ số khác có mặt không quá 1 lần?
Đáp số: a) 1176; b) 420; c) 3960.
Bài A.7. Thầy giáo có 12 cuốn sách đôi một khác nhau trong đó có 5 cuốn sách Toán, 4 cuốn sách tiếng Anh và 3 cuốn sách Triết học. Thầy muốn lấy ra 6 cuốn và đem tặng cho 6 sinh viên

khác nhau mỗi em một cuốn.

a) Giả sử thầy giáo chỉ muốn tặng các cuốn sách thuộc hai thể loại Toán và tiếng Anh, hỏi có tất cả bao nhiêu cách tặng?

b) Giả sử thầy giáo muốn rằng sau khi tặng sách xong, mỗi một trong ba thể loại đều còn lại ít nhất một cuốn. Hỏi có tất cả bao nhiêu cách tặng?

Đáp số: a) 60480; b) 579600.

Bài A.8. Một bàn dài có hai dãy ghế đối diện nhau, mỗi dãy gồm 6 ghế. Người ta muốn xếp chỗ ngồi cho 6 sinh viên lớp A và 6 sinh viên lớp B vào bàn nói trên. Hỏi có bao nhiêu cách xếp trong mỗi trường hợp sau:

a) Bất cứ 2 sinh viên nào ngồi cạnh nhau hoặc đối diện nhau thì khác lớp nhau.

b) Bất cứ 2 sinh viên nào ngồi đối diện nhau thì khác lớp nhau.

Đáp số: a) 1036800; b) 33177600.

Phụ lục B

Sử dụng CNTT giải toán thống kê

Tất cả các SV ở năm thứ nhất đều đã được trang bị những kiến thức cơ bản của Tin học. Việc giải các bài toán bằng máy tính điện tử (MTĐT) cũng đã trở nên quen thuộc đối với các em.

Trong quá trình ngồi trên ghế nhà trường, SV không chỉ tiếp thu các tri thức thông qua bài giảng của GV mà họ còn có thể tự mình tìm kiếm, kiến tạo ra những tri thức mới thông qua sách báo, tài liệu, mạng internet... Việc GV dành ra thời gian để hướng dẫn cho SV giải các bài toán thống kê bằng cách sử dụng các công cụ của CNTT sẽ tăng thêm khả năng hứng thú, tính tò mò trong khoa học, kích thích hoạt động tìm tòi khám phá của SV, từ đó dẫn tới việc hình thành kiến thức và kỹ năng mới.

Sau đây chúng tôi trình bày cách thức thể hiện của một số công cụ CNTT hỗ trợ cho việc giải các bài toán thống kê, kết hợp cùng các ví dụ tương ứng.

B.1. Đối với máy tính điện tử cầm tay

B.1.1. Tính các đặc trưng của mẫu

•**Ví dụ B.1** Gặt ngẫu nhiên 365 điểm trồng lúa của một huyện ta có bảng số liệu sau:

X	25	30	33	34	35	36	37	39	40
n	6	13	38	74	106	85	30	10	3

Tính $\overline{x}, S^2, \hat{S}^2, S, \hat{S}$.

Đối với Casio FX570ES

Nhập các số liệu

- Ấn các phím: Shift Mode 4 1
- Ấn các phím: Mode 3
- Ấn các phím: 1 1 - Var

Hiện ra bảng:

x	freq

Dùng phím Replay để di chuyển qua lại giữa 2 cột x và freq

Nhập cột x:

25 =
30 =
33 =
34 =
35 =
36 =
37 =
39 =
40 =

Nhập cột freq: Dùng phím Replay di chuyển qua cột freq, dòng 25

6 =
13 =
38 =
74 =
106 =
85 =
30 =
10 =
3 =

Nhập xong ấn phím AC *Xem kết quả:*

- Ấn các phím: Shift 1 5 1 = $\rightarrow n = 365$ (cỡ mẫu).
- Ấn các phím: Shift 1 5 2 = $\rightarrow \overline{X} = 34,795$ (tạ/ha).
- Ấn các phím: Shift 1 5 4 = $\rightarrow \hat{S} = 2,072$ (độ lệch chuẩn mẫu đã hiệu chỉnh).
- Ấn các phím: Shift 1 5 3 = $\rightarrow S = 2,069$ (độ lệch chuẩn mẫu chưa hiệu chỉnh).

$*$ **Chú ý:** khi nhập số liệu ta có thể dùng các phím để:

- Xoá 1 dòng dữ liệu: di chuyển đến dòng cần xoá và ấn Del

- Chèn thêm 1 dòng dữ liệu: di chuyển đến dòng cần chèn và nhấn các phím
  Shift 1 3 1 Ins

- Xoá toàn bộ nội dung đã nhập
  Shift 1 3 2 Del - A

Đối với các loại máy $fx - 500MS, fx - 350MS, fx - 350TL, fx - 570MS$
Nhập các số liệu:

- Ấn các phím: MODE 2 để chuyển máy sang chế độ SD
  hoặc MODE MODE 2 để chuyển máy sang chế độ SD

- Ấn các phím: 25 Shift , 6 data (Phím M_+)

- Ấn các phím: 30 Shift , 13 data

- Ấn các phím: 33 Shift , 38 data

- Ấn các phím: 34 Shift , 74 data

- Ấn các phím: 35 Shift , 106 data

- Ấn các phím: 36 Shift , 85 data

- Ấn các phím: 37 Shift , 30 data

- Ấn các phím: 39 Shift , 10 data

- Ấn các phím: 40 Shift , 3 data

Xem kết quả:

- Ấn các phím: Shift S - VAR 1 = $\rightarrow \overline{X} = 34,795$

- Ấn các phím: Shift S - VAR 2 = $\rightarrow \hat{S} = 2,072$

- Ấn các phím: Shift S - VAR 3 = $\rightarrow S = 2,069$

$*$ **Chú ý:** Xoá SD bằng lệnh: Shift MODE 2 =

Đối với các loại máy $fx - 220MS, fx - 500A, fx - 95, fx - 82Super$
Nhập số liệu:

- Ấn các phím: MODE . (Dấu chấm) để đưa máy về chế độ SD

- Ấn các phím: Shift Sac 26×6 data (Phím M_+)

- Ấn các phím: 30×13 data

- Ấn các phím: 33×38 data

- Ấn các phím: 34×74 data
- Ấn các phím: 35×106 data
- Ấn các phím: 36×85 data
- Ấn các phím: 37×30 data
- Ấn các phím: 39×10 data
- Ấn các phím: 40×3 data

Xem kết quả:

- Ấn các phím: Shift $n \to 365$
- Ấn các phím: Shift $\overline{X} \to 34,795$
- Ấn các phím: Shift $\hat{S} \to 2,072$
- Ấn các phím: Shift $S \to 2,069$

$*$ **Chú ý:** Xoá SD bằng lệnh: MODE 0

B.1.2. Bài toán tìm hàm hồi quy

Vào **REG** ấn:
MODE **3** (máy fx 500MS, 95MS)
Máy khác
MODE **MODE** **2** (fx 570MS)
- Màn hình hiện:

Lin	Log	Exp →
1	2	3

► ↑ ↓ ◄

← Pwr	Inv	Quad
1	2	3

- Ấn số tương ứng ta sẽ và chức năng muốn chọn.
$\boxed{1}$ (**Lin**) (tuyến tính) $y = A + Bx$
$\boxed{2}$ (**Log**) (logait) $y = A + B \ln x$
$\boxed{3}$ (**Exp**) (mũ) $y = Ae^{bx}$ $(\ln y = \ln A + Bx)$
$\boxed{►}$ $\boxed{1}$ (**Pwr**) (luỹ thừa) $y = Ae^{b}$ $(\ln y = \ln A + B \ln x)$

▶ 2 (**Inv**) (nghịch đảo) $y = A + B.1/x$
▶ 3 (**Quad**) (bậc hai) $y = A + Bx + Cx^2$
Trước khi tính toán phải ấn
SHIFT **CLR** 1 (**Sel**) =
để xoá bộ nhớ của thống kê.
- Nhập dữ liệu theo cú pháp: <dữ liệu>'<dữ liệu y>**DT**.
- Các kết quả theo dữ liệu đã nhập được gọi theo bảng sau:

Giá trị cần gọi	Ấn
$\sum x^2$	SHIFT S.SUM 1
$\sum x$	SHIFT S.SUM 2
n	SHIFT S.SUM 3
$\sum y^2$	SHIFT S.SUM ▶ 1
$\sum y$	SHIFT S.SUM ▶ 2
$\sum xy$	SHIFT S.SUM ▶ 3
$\sum x$	SHIFT S.SUM ▶ ▶ 1
$\sum x^2 y$	SHIFT S.SUM ▶ ▶ 2
$\sum x^4$	SHIFT S.SUM ▶ ▶ 3
$\bar{x}$	SHIFT S.VAR 1
$x\sigma_n$	SHIFT S.VAR 2
$x\sigma_{n-1}$	SHIFT S.VAR 3
$\bar{y}$	SHIFT S.VAR ▶ 1
$y\sigma_n$	SHIFT S.VAR ▶ 2
$y\sigma_{n-1}$	SHIFT S.VAR ▶ 3
Hệ số A	SHIFT S.VAR ▶ ▶ 1
Hệ số B	SHIFT S.VAR ▶ ▶ 2

- Trừ $y = A + Bx + Cx^2$

Gọi	Ấn
Hệ số tương quan r	SHIFT S.VAR ▶ ▶ 3
Số dự đoán $\hat{x}$	SHIFT S.VAR ▶ ▶ ▶ 1
Số dự đoán $\hat{y}$	SHIFT S.VAR ▶ ▶ ▶ 2

- Riêng với $y = A + Bx + Cx^2$ thì theo bảng sau:

Gọi	Ấn
Hệ số C	SHIFT S.VAR ▶ ▶ 3
Số dự đoán $\hat{x}_1$	SHIFT S.VAR ▶ ▶ ▶ 1
Số dự đoán $\hat{x}_2$	SHIFT S.VAR ▶ ▶ ▶ 2
Số dự đoán $\hat{y}$	SHIFT S.VAR ▶ ▶ ▶ 3

- Các giá trị này có thể dùng như các biến trong các biểu thức tính.

* *Hồi quy tuyến tính* $y = A + Bx$

•Ví dụ B.2 Áp suất theo nhiệt độ trong bảng sau:

Nhiệt độ	10°C	15°C	20°C	25°C	30°C
Áp suất	1003hpa	1005hpa	1010hpa	1011hpa	1014hpa

Hãy dùng Hồi quy tuyến tính $y = A + Bx$ để tính A, B và hệ số tương quan r, áp suất ở 18°C. Tìm nhiệt độ khi áp suất 1000hpa, hệ số tới hạn r^2 và số hiệp biến $\left(\frac{\sum xy - n\bar{x}.\bar{y}}{n-1}\right)$.

Vào **Mode RED**

Ấn 1 (**Lin**)

SHIFT **CLR** 1 (**Sel**) = (xoá nhớ)

10 • 1003 **DT**

n = **REG**

1

(Khi ấn **DT** dữ liệu được nhập và màn hình hiện giá trị của n).

Ấn tiếp

15 ' 1005 **DT**

20 ' 1010 **DT** 25 ' 1011 **DT**

30 ['] 1014 [DT]

Hệ số $A = 997.4$

[SHIFT] [S.VAR] [▶] [▶] [1] [=]

Hệ số $B = 0.56$

[SHIFT] [S.VAR] [▶] [▶] [2] [=]

Hệ số tương quan $r = 0.982607368$

[SHIFT] [S.VAR] [▶] [▶] [3] [=]

Tìm áp suất ở 18°C = 1007.48

18 [SHIFT] [S.VAR] [▶] [▶] [▶] [2] [=]

Nhiệt độ ở 1000 hpa = 4.642857143

1000 [SHIFT] [S.VAR] [▶] [▶] [▶] [1] [=]

$r^2 = 0.965517241$

[SHIFT] [S.VAR] [▶] [▶] [3] [x^2] [=]

Số hiệp biến = 35

[(] [SHIFT] [S.SUM] [▶] [3] [-] [SHIFT] [S.SUM] [3] [x] [SHIFT]
[S.VAR] [1] [x] [SHIFT] [S.VAR] [▶] [1] [)] [÷] [(] [SHIFT]
[S.SUM] [3] [-] [1] [)] [=]

* *Hồi quy bậc hai:* $A + Bx + Cx^2$.

•**Ví dụ B.3** Cho bảng sau

x_i	29	50	74	103	118
y_i	1.6	23.5	38.0	46.4	48

Theo công thức hồi quy bậc hai hãy tìm các hệ số A, B, C sau đó tìm $\hat{y}$ với $x = 16$ và $\hat{x}$ với $y = 20$.

Ở **Mode RED**

Ấn ▶ 3 (**Quad**)

[SHIFT] [CLR] [1] (**Sel**) [=] (xoá nhớ thống kê)

29 ['] 1.6 [DT] 50 ['] 23.5 [DT]
74 ['] 38.0 [DT] 103 ['] 46.4 [DT]
118 ['] 48.0

Tính hệ số $A = -35.59856934$

[SHIFT] [S.VAR] [▶] [▶] [1] [=]

Tính hệ số $B = 1.4959856934$

[SHIFT] [S.VAR] [▶] [▶] [2] [=]

Tính hệ số $C = -6.71629667 \times 10^{-3}$

SHIFT S.VAR ▶ ▶ 3 =

Tính $\hat{y}$ khi $x_i = 16$ (=13.38291067)

16 SHIFT S.VAR ▶ ▶ ▶ 3 =

Tính $\hat{x}_1$ khi $y_i = 20$ (=47.14556728)

20 SHIFT S.VAR ▶ ▶ ▶ 1 =

Tính $\hat{x}_2$ khi $y_i = 20$ (=175.5872105)

20 SHIFT S.VAR ▶ ▶ ▶ 2 =

Chú ý về nhập dữ liệu

- Ấn **DT DT** để nhập dữ liệu hai lần.
- Dùng phím **SHIFT ;** để nhập nhiều dữ liệu giống nhau. Ví dụ nhập "20, 30" năm lần thì ấn **20 ′ 30 SHIFT ; 5 DT**.
- Kết quả được gọi không cần thứ tự như bảng trên.
- Vẫn có phần chú ý như ở **SD**.
- Xem thêm phần tính $P(t), Q(t), R(t)$ ở phần bổ sung cho máy Casio fx 570 MS.
- Các số nhớ **A, B, C, D, E, F, X, Y** không dùng được thống kê **SD** và hồi quy **RED**. Khi vào các chương trình này, các giá trị số nhớ được gán trước bị xoá hết.

B.2. Dùng phần mềm Excel

B.2.1. Tính toán trong bài toán ước lượng

•**Ví dụ B.4** Bảng số liệu sau cho độ lệch X của 200 chi tiết máy với kích thước chuẩn. Với mức ý nghĩa $\alpha = 0,05$ hãy kiểm tra giả thiết độ lệch X là biến ngẫu nhiên có phân phối chuẩn $N\left(a\sigma^2\right)$.

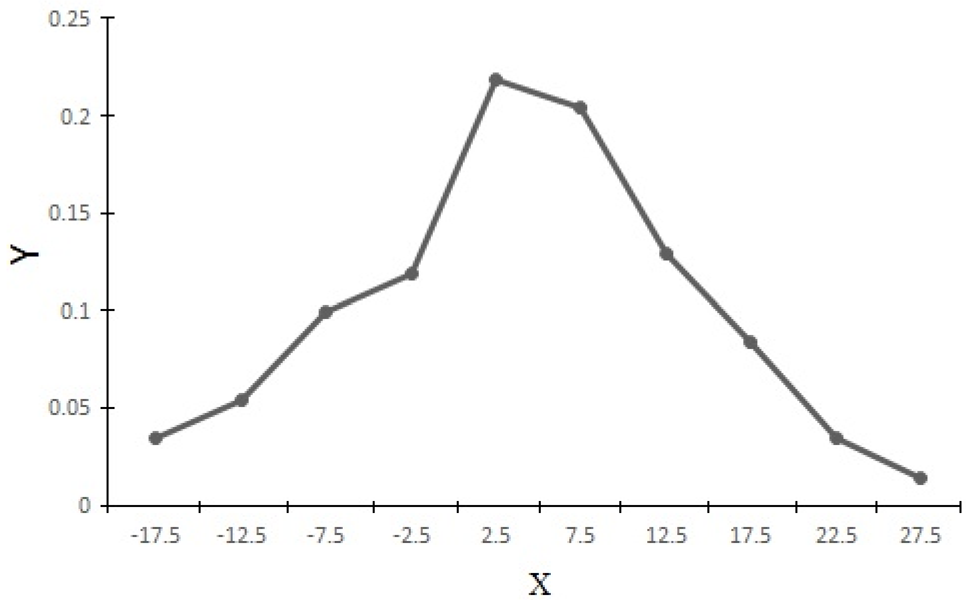
Thứ tự	Khoảng (x_{i-1}, x_i)	n_i	Thứ tự	Khoảng (x_{i-1}, x_i)	n_i
1	(-20, -15)	7	6	(5, 10)	41
2	(-15, -10)	11	7	(10, 15)	26
3	(-10, -5)	20	8	(15, 20)	17
4	(-5, 0)	24	9	(20, 25)	7
5	(0, 5)	44	10	(25, 30)	3

Lời giải. Giả thiết $H_0 : X \sim N\left(a, \sigma^2\right)$, đối thiết $H_1 : X \sim N\left(a, \sigma^2\right)$

+) Lập bảng Excel ước lượng các tham số $a = \overline{X}$ và $\sigma = S'$ và vẽ đa giác tần suất:

TT	A	B	C	D	E	F	G
1	TT	$(x_{i-1}, x_i$	x_i	n_i	f_i	$n_i \cdot x_i$	$n_i\left(x_i - \overline{X}\right)^2$
2	1	(-20, -15)	-17.5	7	0.035	-122.5	3250.818
3	2	(-15, -10)	-12.5	11	0.055	-137.5	3012.928
4	3	(-10, -5)	-7.5	20	0.1	-150	2668.05
5	4	(-5, 0)	-2.5	24	0.12	-60	1029.66
6	5	(0, 5)	2.5	44	0.22	110	105.71
7	6	(5, 10)	7.5	41	0.205	307.5	488.0025
8	7	(10, 15)	12.5	26	0.13	325	1856.465
9	8	(15, 20)	17.5	1	0.085	297.5	3075.343
10	9	(20, 25)	22.5	7	0.035	157.5	2382.818
11	10	(25, 30)	27.5	3	0.015	82.5	1649.708
12				200	1	810	19519.5
13			$\overline{X} = 810/200 =$			4.05	
14			$S' = \sqrt{19519.5/199} =$				9.903936

Chọn cột E làm trục tâm, cột C làm trục hoành ta có đa giác tần suất. Đồ thị cho ta hình dạng của hàm mật độ của phân phối chuẩn.



+) Thực hiện chia khoảng: Vì $n_{10} = 3 < 5$ nên ta ghép khoảng 10 vào khoảng 9, vậy ta có bảng:

Thứ tự	Khoảng (x_{i-1}, x_i)	n_i	Thứ tự	Khoảng (x_{i-1}, x_i)	n_i
1	< -15	7	6	(5, 10)	41
2	(-15, -10)	11	7	(10, 15)	26
3	(-10, -5)	20	8	(15, 20)	17
4	(-5, 0)	24	9	(20, 25)	10
5	(0 ,5)	44			

Như vậy ta có k = 9, r = 2.

Lập bảng Excel tính Q_n và $\chi^2_{6,0.05}$, trong bảng Excel cột B ta chỉ ghi các cận dưới của các khoảng.

TT	A	B	C	D	E
1			n_i	p_i	$(n_i - np_i)^2/np_i$
2	1	-15	7	0.02721	0.44604605
3	2	-10	11	0.050794	0.06966124
4	3	-5	20	0.102413	0.01136968
5	4	0	24	0.160879	2.07749498
6	5	5	44	0.196912	0.5413946
7	6	10	41	0.187795	0.31523907
8	7	15	26	0.139551	0.13073038
9	8	20	17	0.080798	0.04371067
10	9		10	0.053648	0.04960009
11			200	$Q_n =$	3.68524676
12				$\chi^2_{6,0.05}$	12.5915774

Để tính các xác suất p_i $\left(i = \overline{1,9}\right)$ ta thực hiện các lệnh sau:

D2 = NORMDIST(B2,4.05,9.903936,TRUE)

D3 = NORMDIST(B3,4.05,9.903936,TRUE)-NORMDIST(B3,4.05,9.903936,TRUE)

Sau đó dùng Autofill từ D4 đến D9

D10 = 1-NORMDIST(B9,4.05,9.903936,TRUE)

Ta cũng có thể tra bảng hàm Laplace tính các xác suất P_i trên các khoảng

$$(x_{i-1}, x_i)\ \left(i = \overline{1,9}\right) \text{ với: } P_i = \phi\left(\frac{x_i - 4,05}{9,9039}\right) - \phi\left(\frac{x_{i-1} - 4,05}{9,9039}\right).$$

Vậy $Q_n = 3,68524776 < \chi^2_{6,0.05} = 12,59$ chấp nhận H_0, hay X có phân phối chuẩn N(4,3;94,26).

B.2.2. Tính toán các đặc trưng của mẫu

•**Ví dụ B.5** Đo chiều cao của 100 sinh viên năm thứ nhất ta được bảng thống kê (tính bằng cm).

X	150-154	154-158	158-162	162-166	166-170	170-174	174-178
SV	10	14	24	28	12	8	4

Chọn $x_0 = 160$ và lập bảng tính Excel ta được:

TT	A	B	C	D	E
1	X	n_i	$x_i - x_0$	$n_i(x_i - x_0)$	$n_i(x_i - x_0)^2$
2	152	10	-8	-80	640
3	156	14	-4	-56	224
4	160	24	0	0	0
5	164	28	4	112	448
6	168	12	8	96	768
7	172	8	12	96	1152
8	176	4	16	64	1024
9		n = 100		B = 232	C = 4256
10	$\overline{X} = \frac{B}{n} + x_0 =$			162.32	
11	$S^2 = \frac{C}{n} - (\overline{X} - x_0)^2 =$			37.1176	
12	S =			6.09734	

•Ví dụ B.6 Hai máy A và B cùng sản xuất một chi tiết có độ dài X. Để kiểm tra độ chính xác của mỗi máy, định kỳ 6 tháng một lần người ta lấy mỗi máy một lô gồm 1000 sản phẩm và kiểm tra độ dài của chúng. Nếu độ lệch trung bình của mỗi lô vượt quá 5 mm thì máy cần duy tu lại. Bảng thống kê một lần của hai máy được cho dưới đây:

Độ dài	1222 1226	1226 1230	1230 1234	1234 1238	1238 1242	1242 1246	1246 1250	1250 1254	1254 1256
n(A)	5	20	80	190	360	300	30	10	5
m(B)	70	90	110	140	170	170	140	90	20

Hãy kiểm tra xem máy nào đã cần duy tu lại.

Lời giải. Chọn $x_0 = 1236$ và lập bảng tính theo Excel:

X	x_i	$x_i - x_0$	n_iA	m_iB	$n_i(x_i - x_0)$	$m_i(x_i - x_0)$	$n_i(x_i - \overline{A})^2$	$m_i(x_i - \overline{B})^2$
1222-1226	1224	-12	5	70	-60	-840	720	10080
1226-1230	1228	-8	20	90	-160	-720	1280	5760
1230-1234	1232	-4	80	110	-320	-440	1280	1760
1234-1238	1236	0	190	140	0	0	0	0
1238-1242	1240	4	360	170	1440	680	5760	2720
1242-1246	1244	8	300	170	2400	1360	19200	10880
1246-1250	1248	12	30	140	360	1680	4320	20160
1250-1254	1252	16	10	90	160	1440	2560	23040
1254-1258	1256	20	5	20	100	400	2000	8000
			1000	1000	B1=3920	B2=3560	C1=37120	C2=82400
$\overline{A} = \frac{B_1}{n} + x_0 =$				1239.92	$\overline{B} = \frac{B_2}{n} + x_0 =$			1239.56
$S_A^2 = \frac{C_2}{n} - \left(\overline{A} - x_0\right)^2 =$				21.7536	$S_A'^2 = \frac{n}{n-1} \cdot S_A^2 =$			21.77538
$S_B^2 = \frac{C_2}{n} - \left(\overline{B} - x_0\right)^2 =$				69.73	$S_B'^2 = \frac{n}{n-1} \cdot S_B^2 =$			69.7962
$S_A' =$				4.6664	$S_B' =$			8.3544

Như vậy máy A không cần duy tu còn máy B phải duy tu.

B.2.3. Các phân phối xác suất

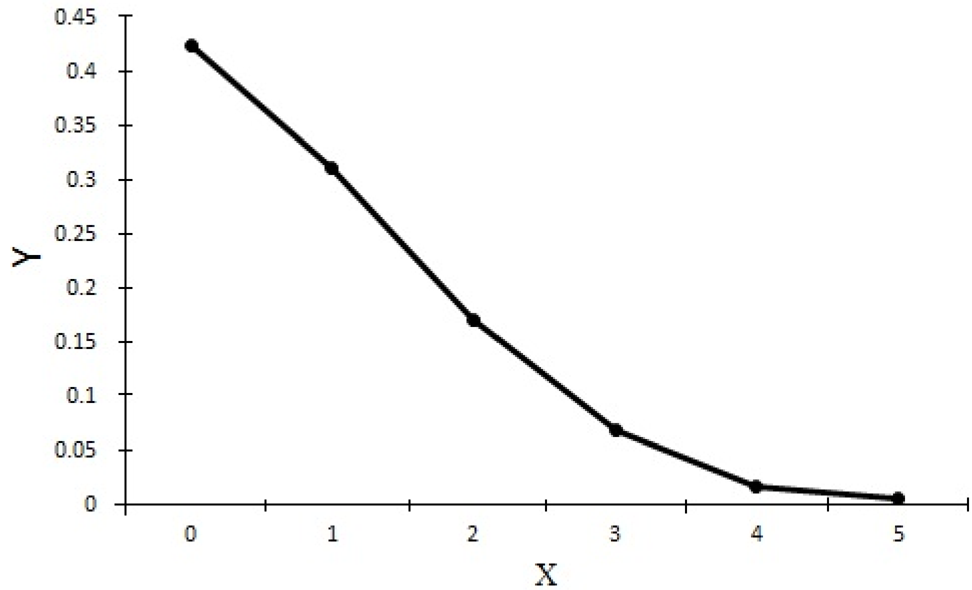
•Ví dụ B.7 Quan sát số lượng ô tô vào xúc tại một máy xác trong khoảng thời gian định sẵn tại một mỏ than người ta thu được bảng kết quả sau:

Số ô tô	0	1	2	3	4	5
Số khoảng thời gian	75	55	30	12	3	1

Hãy kiểm tra xem số liệu trên có tuân theo quy luật phân phối Poát-Xông $p_m = \frac{\lambda^m}{m!}e^{-\lambda}$ không? Trong đó P_m là xác suất trong khoảng thời gian trên có m ô tô xuất hiện tại máy xúc.

Lời giải. Lập bảng Excel ước lượng tham số $\lambda = \overline{X}$ và vẽ đa giác tần suất, gọi số ô tô là X, số khoảng thời gian là n_i:

X	0	1	2	3	4	5	
n_i	75	55	30	12	3	1	176
f_i	0.426136	0.3125	0.170455	0.068182	0.017045	0.005682	1
	$\lambda \approx \overline{X} =$			0.954545			



Đa giác tần suất có dạng hàm mật độ của phân phối Poát-Xông. Các xác suất $p_m = \frac{1}{m!}e^{-0.621212}$ lập bảng tính Q_n ta được:

TT	A	B	C	D
1	X	n_i	p_i	$(n_i - np_i)^2/np_i$
2	0	75	0.384987	0.774082
3	1	55	0.367488	1.448107
4	2	30	0.175392	0.024461
5	3	12	0.055806	0.482998
6	4	3	0.013317	0.183674
7	5	1	0.002542	0.682274
8		176	$Q_n =$	3.595594
9		$\chi^2_{4,0.05} = CHIINV(0.05,4) =$		9.487728

Để tính các p_i trong Excel ta thực hiện:
C2 = POISSON(A2,0.954545,FALSE)
Sau đó Autofill từ C3 đến C7.
Số bậc tự do k = 6, r = 1 ta có k - r - 1 = 6 - 1 - 1 = 4.

$$\chi^2_{4,0.05} = 9,487728 > Q_n = 3,595594.$$

Vậy bảng số liệu thu được tuân theo phân phối Poát-Xông.

Trong phần này chúng tôi đã giới thiệu cách sử dụng một số máy tính điện tử cá nhân thông dụng và phần mềm Excel để giải một số bài toán trong thống kê. Với sự phát triển mạnh mẽ của công nghệ thông tin, ngày càng nhiều thế hệ máy tính và phần mềm mới ra đời, đó sẽ là một thuận lợi lớn cho SV. Do đó yêu cầu sử dụng được MT ĐT các nhân và các phần mềm hỗ trợ giải toán là một điều cần thiết. Các công cụ này sẽ giúp chúng ta giải quyết nhanh chóng các bài toán nói chung và thống kê nói riêng, góp phần nâng cao tính chính xác và kịp thời; đồng thời tiết kiệm thời gian tính toán cho một bài toán, rèn luyện tư duy thuật giải.