Chương 1

GIẢI TÍCH TỔ HỢP

1.1. Quy tắc nhân

Các tính chất sau của phép đếm sẽ là nền tảng của tất cả công việc của chúng ta.

Tính chất 1 (Quy tắc nhân)
Giả sử có 2 công việc được thực hiện. Nếu công việc 1 có thể thực hiện một trong m cách khác nhau và ứng với mỗi cách thực hiện công việc 1, công việc 2 có n cách thực hiện khác nhau thì có $m.n$ cách khác nhau khi thực hiện hai hai công việc.

Proof: Tính chất cơ bản có thể được chứng minh bằng cách liệt kê tất cả các cách thực hiện có thể của hai công việc như sau:

$$\begin{array}{cccc} (1,1), & (1,2), & \ldots, & (1,n) \\ (2,1), & (2,2), & \ldots, & (2,n) \\ \vdots & & & \\ (m,1), & (m,2), & \ldots, & (m,n) \end{array}$$

trong đó, chúng ta nói cách thực hiện là (i, j) nếu công việc 1 thực hiện theo cách thứ i trong m cách có thể và công việc 2 thực hiện cách thứ j trong n cách. Vì thế tập tất cả các cách có thể thực hiện bằng mn.

Ví dụ 1.1.1 Một cộng đồng nhỏ có 10 phụ nữ, mỗi người có 3 người con. Chọn một người phụ nữ và một đứa con của họ. Hỏi có bao nhiêu cách chọn ?

Giải

Ta xem việc chọn người phụ nữ như là công việc 1 và việc chọn con của họ là công việc 2. Khi đó từ tính chất cơ bản ta có $10.3 = 30$ cách chọn khác nhau.

Khi chúng ta có nhiều hơn hai công việc được thực hành, tính chất cơ bản có thể được tổng quát hoá như sau:

Tính chất 2 (Quy tắc nhân tổng quát)
Giả sử có k công việc được thực hiện. Nếu công việc 1 có thể thực hiện trong n_1 cách khác nhau và ứng với mỗi cách thực hiện công việc 1, công việc 2 có n_2 cách thực hiện khác nhau; ứng với mỗi cách thực hiện hai công việc đầu, có n_3 cách khác nhau thực hiện công việc 3, v…v .. thì có $n_1.n_2.n_3. \ldots .n_k$ cách khác nhau thực hiện k công việc đó.

Ví dụ 1.1.2 Một hội nghị học tập ở một trường đại học bao gồm 3 sinh viên năm thứ nhất, 4 sinh viên năm thứ 2, 5 sinh viên năm thứ 3 và 2 sinh viên năm cuối. Một tiểu ban gồm 4 người ở trong 4 khoá khác nhau. Hỏi có thể lập được bao nhiêu tiểu ban khác nhau?

Giải

Việc chọn một tiểu ban như là việc thực hiện 4 công việc khác nhau. Công việc i là chọn một sinh viên năm thứ i($i = 1, 2, 3, 4$). Vì thế, từ tính chất cơ bản tổng quát, chúng ta có $3.4.2.5 = 120$ tiểu ban khác nhau có thể lập.

Ví dụ 1.1.3 Số hiệu của bằng lái xe môtô gồm 7 kí tự, trong đó 3 kí tự đầu là các chữ cái và 4 kí tự sau là các chữ số. Hỏi có thể có bao nhiêu bằng lái xe môtô khác nhau ?

Giải

Áp dụng tính chất cơ bản tổng quát, chúng ta có số bằng lái khác nhau có thể có là: $26.26.26.10.10.10.10 = 175.760.000$

Nếu các chữ cái và chữ số trong số hiệu bằng khác nhau thì có bao nhiêu bằng lái khác nhau?

Ví dụ 1.1.4 Một hàm số xác định trên một tập n phần tử và chỉ nhận hai giá trị 0 và 1. Hỏi có thể lập được bao nhiêu hàm khác nhau.

Giải

Đặt các phần tử là $1, 2, 3, \ldots, n$. Vì $f(i)$ bằng 1 hoặc 0 cho mỗi $i = 1, 2, \ldots, n$ nên ta có 2^n hàm khác nhau có thể lập.

1.2. Hoán vị

Có bao nhiêu cách khác nhau khi sắp xếp có thứ tự 3 kí tự a, b, c? Bằng cách liệt kê trực tiếp chúng ta thấy có 6 cách, cụ thể là: abc, acb, bac, bca, cab và cba. Mỗi cách sắp xếp như vậy được gọi là một *hoán vị*. Vì thế có 6 hoán vị có thể của một tập 3 phần tử. Kết quả này cũng có thể suy ra từ tính chất cơ bản, vì phần tử thứ nhất trong hoán vị có thể là một trong 3 kí tự, phần tử thứ 2 trong hoán vị có thể chọn một trong 2 kí tự còn lại và phần tử thứ 3 được chọn từ một phần tử còn lại. Vì thế, có $3.2.1 = 6$ hoán vị có thể.

Chúng ta định nghĩa khái niệm hoán vị một cách tổng quát như sau:

Định nghĩa 1.2.1 Cho n phần tử khác nhau. Một hoán vị của n phần tử là một cách sắp xếp có thứ tự n phần tử đã cho.

Gọi P_n là số hoán vị khác nhau có thể lập từ n phần tử đã cho. Ta có

$$P_n = n(n-1)\ldots 2.1 = n!$$

Ví dụ 1.2.5 Hỏi có bao nhiêu cách sắp xếp vị trí các cầu thủ(thủ môn, tiền vệ phải, trái,...) khác nhau trong một đội bóng gồm 9 cầu thủ?

Giải

Có $9! = 362880$ cách sắp xếp các cầu thủ.

Ví dụ 1.2.6 Một lớp học lý thuyết xác suất gồm 6 nam và 4 nữ. Một kỳ thi được tổ chức, Các sinh viên được xếp hạng theo kết quả làm bài của họ. Giải sử không có hai sinh viên nào đạt cùng một điểm.

a) Có thể có bao nhiêu cách xếp hạng khác nhau?

b) Nếu nam được xếp hạng trong nhóm nam và nữ được xếp hạng trong nhóm nữ thì có thể có bao nhiêu cách xếp hạng khác nhau?

Giải

a) Mỗi cách xếp hạng tương ứng với một cách sắp xếp có thứ tự 10 người, chúng ta có câu trả lời trong phần này là $10! = 3.628.800$.

b) Vì có 6! cách xếp hạng khác nhau trong 6 người nam và 4! cách xếp khác nhau trong 4 người nữ nên áp dụng tính chất cơ bản, chúng ta có $6!.4! = 17.280$ cách sắp xếp khác nhau có thể có.

Ví dụ 1.2.7 Cô Nga định đặt 10 cuốn sách lên một cái giá sách. Trong đó có 4 cuốn sách Toán, 3 cuốn Hoá học, 2 cuốn Lịch sử và 1 cuốn Ngoại ngữ. Cô Nga muốn sắp xếp những cuốn sách của cô các cuốn sách của minh sao cho các cuốn cùng một môn thi kề nhau. Có thể có bao nhiêu cách sắp xếp 10 cuốn sách khác nhau?

Giải

Có $4!.3!.2!.1!$ cách sắp xếp sao cho các sách Toán ở đầu hàng sau đó đến các sách Hoá rồi đến sách Sử và cuối cùng là sách Ngoại ngữ. Tương tự, với mỗi thứ tự các môn học, chúng ta có $4!.3!.2!.1!$ cách sắp xếp khác nhau. Ở đây có 4! cách sắp xếp thứ tự các môn học nên đáp án của câu hỏi là có $4!.4!.3!.2!.1! = 6912$.

Bây giờ chúng ta sẽ xác định số các hoán vị của một tập n phần tử khi mà một số phần tử trong hoán vị trùng với những phần tử khác. Để đi thẳng vào vấn đề chúng ta quan tâm, hãy xem xét ví dụ sau:

Ví dụ 1.2.8 Hỏi có bao nhiêu cách sắp xếp các kí tự khác nhau từ các ký tự $PEPPER$?

Giải

Trước hết chúng ta chú ý rằng có 6! hoán vị của các ký tự $P_1E_1P_2P_3E_2R$ khi 3 ký tự P_i và 2 ký tự E_i được xem là khác nhau. Tuy nhiên chúng ta xem xét một hoán vị bất kì trong những hoán vị này, chẳng hạn $P_1P_2E_1P_3E_2R$. Bây giờ nếu chúng ta hoán vị các ký tự P với nhau và hoán vị các kí tự E với nhau thì kết quả vẫn sẽ có dạng $PPEPER$. Đólà 3!.2! hoán vị

$$\begin{array}{ll}
P_1P_2E_1P_3E_2R & P_1P_2E_2P_3E_1R \\
P_1P_3E_1P_2E_2R & P_1P_3E_2P_3E_1R \\
P_2P_1E_1P_3E_2R & P_2P_1E_2P_3E_1R \\
P_2P_3E_1P_1E_2R & P_2P_3E_2P_1E_1R \\
P_3P_2E_1P_1E_2R & P_3P_2E_2P_1E_1R \\
P_3P_1E_1P_2E_2R & P_3P_1E_2P_2E_1R
\end{array}$$

có cùng hình thức như $PPEPER$. Vì vậy, có $6!/(3!.2!) = 60$ cách sắp xếp các kí tự khác nhau từ các ký tự $PPEPER$.

Các hoán vị trong đó các phần tử được lặp lại như trên được gọi là hoán vị lặp. Chúng ta có định nghĩa chính xác như sau:

Định nghĩa 1.2.2 Một hoán vị chập lặp là một cách xắp xếp có thứ tự n phần tử không nhất thiết phân biệt.

Từ ví dụ (1.2.8), chúng ta chỉ ra một cách tổng quát rằng, có

$$\frac{n!}{n_1!.n_2!.\ldots n_k!}$$

hoán vị lặp khác nhau của n phần tử, trong đó n_1 phần tử như nhau, n_2 phần tử như nhau,..., n_k phần tử như nhau.

Ví dụ 1.2.9 Một vòng thi đấu cờ vua có 10 đấu thủ. Trong đó có 4 người Nga, 3 người Mỹ, 2 người Anh và 1 người Brazil. Kết quả vòng thi đấu chỉ ghi các quốc tịch của các đấu thủ theo vị trí mà họ đạt được. Hỏi có bao nhiêu kết quả có thể?

Giải
Có

$$\frac{10!}{4!.3!.2!.1!} = 12600$$

kết quả có thể.

Ví dụ 1.2.10 Có bao nhiêu tín hiệu khác nhau, trong đó mỗi tính hiệu gồm 9 cờ treo trên một hàng, được tạo ra từ một tập gồm 4 cờ trắng, 3 cờ đỏ và 2 cờ xanh nếu tất cả các cờ cùng màu là giống hệt nhau?

Giải
Có

$$\frac{9!}{4!.3!.2!} = 1260$$

tín hiệu khác nhau.

1.3. Tổ hợp

Chúng ta thường quan tâm đến việc xác định số các nhóm khác nhau gồm k phần tử được xây dựng từ một tổng thể gồm n phần tử. Ví dụ, có bao nhiêu nhóm gồm 3 chữ cái được chọn từ 5 chữ cái A, B, C, D và E? Để trả lời câu hỏi này ta lý giải như sau: Vì có năm cách chọn phần tử đầu tiên, 4 cách chọn phần tử tiếp theo và 3 cách chọn phần tử cuối cùng. Vì thế có 5.4.3 cách chọn nhóm gồm 3 phần tử khi thứ tự trong mỗi nhóm được chọn có liên quan. Tuy nhiên, vì mỗi nhóm gồm 3 phần tử, chẳng hạn nhóm gồm ba chữ cái A, B, C sẽ được đếm 6 lần(nghĩa là tất cả các hoán vị ABC, ACB, BAC, CAB và CBA sẽ được đếm khi thứ tự lựa chọn là quan trọng). Từ đó suy ra rằng số các nhóm phân biệt gồm 3 chữ cái có thể tạo ra được là

$$\frac{5.4.3}{3!} = 10$$

Mỗi nhóm con gồm 3 phần tử như trên được gọi là một tổ hợp chập 3 của 5 phần tử và số các nhóm con gồm 3 phần tử được gọi là số các tổ hợp chập 3 của 5. Ta có định nghĩa tổng quát như sau

Định nghĩa 1.3.3 Cho một tập n phần tử. Một tổ hợp chập k của n phần tử($0 \leqslant k \leqslant n$) là một tập con gồm k phần tử được lấy ra từ tập n phần tử đã cho.

Số các tổ hợp chập k của n phần tử, ký hiệu C_n^k, được xác định bởi

$$C_n^k = \frac{n(n-1)\dots(n-k+1)}{k!}$$

Cần nhấn mạnh rằng trong một tập con gồm k phần tử thì không phân biệt thứ tự của các phần tử được chọn.

Ví dụ 1.3.11 Một hội nghị gồm 3 người được thành lập từ một nhóm 20 người. Hỏi có thể thành lập được bao nhiêu hội nghị khác nhau ?

Giải
Có $C_{20}^3 = \frac{20.19.18}{3.2.1} = 1140$ hội nghị khác nhau có thể thành lập.

Ví dụ 1.3.12 Từ một nhóm gồm 5 nữ và 7 nam, hỏi có thể thành lập được bao nhiêu hội nghị khác nhau gồm 2 nữ và 3 nam? Trong trương hợp có hai người nam hận thù nhau và không chịu tham gia cùng một hội nghị thì có thể thành lập được bao nhiêu hội nghị ?

Giải

Vì có thể thành lập được C_5^2 nhóm gồm 2 phụ nữ và C_7^3 nhóm gồm 3 nam nên từ tính chất cơ bản ta suy ra có thể lập được $C_5^2.C_7^3 = 350$ hội nghị gồm 2 nữ và 3 nam.

Mặt khác, nếu có hai người đàn ông từ chối tham gia cùng một hội nghị thì khi đó có $C_2^0C_5^2$ cách chọn nhóm 3 người đàn ông không có hai người hận thù nhau và có $C_2^1.C_5^2$ cách chọn nhóm 3 người mỗi nhóm chứa chỉ một trong hai người đàn ông hận thù nhau. Như vậy có $C_2^0.C_5^3 + c_2^1.C_5^2 = 30$ cách chọn nhóm ba người đàn ông không có mặt cả hai người hận thù nhau trong một nhóm. Vì có C_5^2 cách chọn 2 người nữ nên trong trường hợp này có $30.C_5^2 = 300$ cách thành lập hội nghị.

Chương 2

PHÉP TÍNH XÁC SUẤT

2.1. PHÉP THỬ VÀ SỰ KIỆN

2.1.1. Phép thử và sự kiện

Định nghĩa 2.1.4 **Phép thử** là một thí nghiệm có thể lặp lại trong các điều kiện bên ngoài giống hệt nhau và kết quả là một phân tử không đoán trước được của một tập hợp các định.

Vậy dữ kiện của một phép thử gồm có: - Việc mô tả bộ máy thí nghiệm và việc chỉ dẫn các điều kiện tiến hành.

- Việc xác định tập hợp các kết quả của thí nghiệm.

Ta xét các ví du sau:

Ví dụ 2.1.13 Ta gieo một đồng tiền đồng chất xuống mặt phẳng và quan sát mặt nào xuất hiện đó là một phép thử. Phép thử có hai kết quả là đồng tiền xuất hiện mặt sấp(S) hoặc mặt ngữa(N).

Ví dụ 2.1.14 Gieo một con xúc xắc cân xứng và đồng chất trên một mặt phẳng và quan sát mặt nào xuất hiện là một phép thử. Các kết quả của phép thử là sự xuất hiện một trong 6 mặt của con xúc xắc mà ta có thể ký hiệu bằng các số trên mặt: $1, 2, 3, 4, 5, 6$.

Ví dụ 2.1.15 Trong một hộp kín có m bi đỏ, n bi xanh hoàn toàn giống nhau về kích thước, trọng lượng. Lấy ngẫu nhiên một bi và quan sát xem bi có màu gì là một phép thử. Phép thử có hai kết quả: bi lấy ra màu xanh và bi lấy ra màu đỏ.

2.1.2. Sự kiện liên kết với phép thử

Sự kiện (hay còn gọi biến cố) là một khái niệm thường gặp trong lý thuyết xác suất. Ta không có một định nghĩa chặt chẽ khái niệm này. **Sự kiện** được hiểu như là một sự việc, một hiện tượng nào đó của cuộc sống tự nhiên và xã hội.

Định nghĩa 2.1.5 **Một sự kiện lên kết với một phép thử** là sự kiện có thể xảy ra hay không xảy ra tùy thuộc vào kết quả của phép thử đó.

Sự kiện thường được ký hiệu bằng các chữ cái in hoa $A, B, C, \ldots$.

Một sự kiện xảy ra khi và chỉ khi có một kết quả cụ thể trong số những kết quả của phép thử thì được gọi là **sự kiện cơ bản** hay còn gọi là **sự kiện sơ cấp**. Tập hợp tất cả các sự kiện sơ cấp gọi là không gian sơ cấp, ký hiệu Ω.

Sự kiện tất yếu là sự kiện luôn xảy ra khi thực hiện phép thử.

Sự kiện bất khả là sự kiện không bao giờ xảy ra khi thực hiện phép thử.

Sự kiện ngâu nhiên là sự kiện có thể xảy ra hoặc không xảy ra khi thực hiện phép thử.

Ví dụ 2.1.16 Ta gieo một đồng tiền đồng chất xuống mặt phẳng và quan sát mặt nào xuất hiện. Gọi N là sự kiện xuất hiện mặt ngữa, S là sự kiện xuất hiện mặt sấp. Ta có S, N là các sự kiện sơ cấp và không gian sơ cấp là $\Omega = \{S, N\}$.

Gọi A là sự kiện không xuất hiện mặt nào cả thì A là sự kiện bất khả. Gọi B là sự kiện xuất hiện mặt nào đó của đồng tiền, B là sự kiện tất yếu.

Ví dụ 2.1.17 Gieo một con xúc xắc cân xứng và đồng chất trên một mặt phẳng và quan sát mặt nào xuất hiện. Gọi M_i là sự kiện xuất hiện mặt i chấm ($i = \overline{1, \ldots, 6}$), M_i là các sự kiện sơ cấp. Không gian sơ cấp $\Omega = \{M_1;\ M_2;\ M_3;\ M_4;\ M_5;\ M_6\}$.

Gọi A là sự kiện xuất hiện mặt có số chấm là số chẵn. Khi đó A xảy ra khi và chỉ khi M_2 hoặc M_4 hoặc M_6 xảy ra. Ta đồng nhất sự kiện A với tập hợp $\{M_2;\ M_4;\ M_6\}$. Ta viết

$$A = \{M_2;\ M_4;\ M_6\} \subset \Omega$$

các sự kiện sơ cấp $M_2;\ M_4;\ M_6$ gọi là các sự kiện thuận lợi cho sự kiện A và A xảy ra khi và chỉ khi một trong các sự kiện sơ cấp thuộc nó xảy ra.

Tương tự, nếu gọi B là sự kiện con xúc xắc xuất hiện mặt có số chấm là số lẽ, C là sự kiện con xúc xắc xuất hiện mặt có số chấm lớn hơn 4, D là sự kiện tất yếu, E là sự kiện bất khả. Ta có:

$$B = \{M_1;\ M_3;\ M_5\} \qquad C = \{M_5;\ M_6\} \qquad D = \Omega \qquad E = \emptyset$$

Như vậy với cách ký hiệu trên ta thấy:

- Mỗi sự kiện tương ứng với một tập hợp con của không gian sơ cấp và ngược lại, một tập con của Ω xác định duy nhất một sự kiện nào đó. Như vậy, mỗi sự kiện được xem như một tập con của không gian sơ cấp.

- Nếu sự kiện $A \subset \Omega$ thì các sự kiện sơ cấp thuộc A gọi là các sự kiện thuận lợi cho sự kiện A .

2.1.3. Các phép toán và quan hệ của các sự kiện

- **Tổng:** Tổng của hai sự kiện A và B, ký hiệu $A + B$ (hoặc $A \cup B$), là một sự kiện xảy ra khi ít nhất một trong hai sự kiện A, B xảy ra.

- **Tích:** Tích của hai sự kiện A và B, ký hiệu $A.B$ (hoặc $A \cap B$), là một sự kiện xảy ra khi cả A và B đồng thời xảy ra.

- **Hiệu:** Hiệu của hai sự kiện A và B, ký hiệu $A - B$ (hay $A \setminus B$), là sự kiện xảy ra khi A xảy ra và B không xảy ra, tức là $A - B = A.\overline{B}$.

- **Đối lập:** Đối lập của A, ký hiệu $\overline{A}$, là sự kiện không xảy ra sự kiện A. Ta suy ra $\overline{\overline{A}} = A$ và $A + \overline{A} = \Omega$: sự kiện tất yếu, $A.\overline{A} = \emptyset$: sự kiện bất khả, $\overline{\Omega} = \emptyset$.

- **Xung khắc:** Hai sự kiện A và B gọi là xung khắc nếu chúng không thể xảy ra, tức $A.B = \emptyset$.

- **Kéo theo:** Sự kiện A gọi là kéo theo sự kiện B, ký hiệu $A \Rightarrow B$, nếu sự kiện A xảy ra thì sự kiện B xảy ra, tức là $A \subset B$.

- **Tương đương:** Hai sự kiện A và B gọi là tương đương, ký hiệu $A = B$, nếu sự kiện A xảy ra thì sự kiện B xảy ra và ngược lại, tức là $A \subset B$ và $B \subset A$.

Khi ta xem mỗi sự kiện như là một tập con của không gian sơ cấp Ω thì các phép toán trên các sự kiện tương ứng với các phép toán về tập hợp mà chúng ta đã quen biết và có thể minh họa chúng bằng các biểu đồ Ven.

Ví dụ 2.1.18 Gieo hai đồng tiền cân đối và đồng chất lên mặt phẳng. Gọi:
A = Sự kiện xuất hiện mặt sấp (S) trên đồng tiền thứ 1.
B = Sự kiện xuất hiện mặt ngữa (N) trên đồng tiền thứ 2.
C = Sự kiện xuất hiện mặt ngữa (N) trên đồng tiền thứ 1.
D = Sự kiện xuất hiện ít nhất một mặt sấp (S).
E = Sự kiện xuất hiện nhiều nhất một mặt sấp (S).
a) Xác định không gian sơ cấp và biểu diễn các sự kiện trên theo ngôn ngữ tập hợp.

b) Hãy diễn tả các sự kiện sau bằng ngôn ngữ thông thường và ngôn ngữ tập hợp: $A \cup B, A \cup C, BC, BD, CE, \overline{A}, \overline{B}, \overline{D}, \overline{E}, AB \cup C$.

c) Gọi F là sự kiện không xuất hiện mặt ngữa. F tương đương với sự kiện nào.

Giải

a) Ta ký hiệu XY nghĩa là: X là mặt xuất hiện của đồng tiền thứ nhất, Y là mặt xuất hiện của đồng tiền thứ 2. X, Y nhân hai giá trị là sấp (S) và ngữa (N). Khi đó ta có không gian sơ cấp là:

$$\Omega = \{SS, SN, NN, NS\}$$

$$A = \{SS, SN\}, B = \{SN, NN\}, C = \{NN, NS\}, D = \{SS, SN, NS\}, E = \{SN, NN, NS\}$$

b) Ta có:

$A \cup B$: là sự kiện đồng tiền thứ 1 xuất hiện mặt sấp hoặc đồng tiền thứ hai xuất hiện mặt ngữa. $A \cup B = \{SS, SN, NN\}$.

$A \cup C$: là sự kiện đồng tiền thứ nhất xuất hiện mặt sấp hoặc ngữa. Đây là sự kiện tất yếu, $A \cup C = \Omega$.

BC: là sự kiện cả hai đồng tiền xuất hiện mặt ngữa, $BC = \{NN\}$.

BD: là sự kiện đồng tiền thứ 1 xuất hiện mặt sấp và đồng tiền thứ hai xuất hiện mặt ngữa. $BD = \{SN\}$.

CE: là sự kiện đồng tiền thứ 1 xuất hiện mặt ngữa (chú ý $C \Rightarrow E, CE = C$). $CE = \{NS, NN\}$.

Các trường hợp khác làm tương tự, và dành lại như một bài tập.

c) F tương đương với D.

2.2. CÁC ĐỊNH NGHĨA VỀ XÁC SUẤT

2.2.1. Định nghĩa xác suất theo lối cổ điển

Định nghĩa 2.2.6 Xét phép thử với không gian sơ cấp bao gồm n kết quả đồng khả năng. Giả sử sự kiện A bao gồm m kết quả thuận lợi cho A xảy ra. Khi đó, xác suất của sự kiện(biến cố) A, ký hiệu $P(A)$, được định nghĩa bằng công thức

$$P(A) = \frac{m}{n} = \frac{\text{Số kết quả thuận lợi cho } A \text{ xảy ra}}{\text{Tổng số kết quả của không gian sơ cấp}}$$

Ví dụ 2.2.19 Gieo đồng thời hai đồng tiền cân xứng và đồng chất. Tính xác suất để hai đồng xuất hiện khác nhau?

Giải

Ta có không gian sơ cấp $\Omega = \{(S,N);\ (S,S);\ (N,S);\ (N,N)\}$. Trong đó,$S$, N lần lượt ký hiệu cho sự xuất hiện mặt sấp và sự xuất hiện mặt ngữa và kết quả (S,N) nghĩa là đồng tiền thứ nhất xuất hiện mặt S và đồng tiền thứ hai xuất hiện mặt N, các ký hiệu khác tương tự.

Gọi A là sự kiện hai mặt đồng tiền xảy ra khác nhau, ta có:

$$A = \{(S,N);\ (N,S)\}$$

Vậy xác xuất của sự kiện A là: $\mathbf{P}_(A) = \frac{2}{4} = 0,5$.

Ví dụ 2.2.20 Một người gọi điện thoại nhưng quên mất hai số cuối của số điện thoại cần gọi mà chỉ nhớ là hai số đó khác nhau. Tìm xác suất để người đó quay ngẫu nhiên một lần trúng số cần gọi?

Giải

Gọi A là sự kiện người đó quay ngẫu nhiên một lần trúng số cần gọi.

Ta có, mỗi kết quả là một cách gọi 2 số cuối nên không gian sơ cấp có số kết quả: $n = A_{10}^2 = 90$. Trong đó số kết quả thuận lợi cho A: $m = 1$.

Vậy xác suất của sự kiện $A : P(A) = \frac{1}{90}$.

Ví dụ 2.2.21 Một hộp có 7 chính phẩm và 3 phế phẩm. Lấy ngẫu nhiên từ hộp đó 3 sản phẩm. Tìm xác suất để cả 3 sản phẩm lấy ra là chính phẩm.

Giải

Mỗi kết kết quả là một cách lấy ra 3 sản phẩm khác nhau từ 10 sản phẩm nên không gian sơ cấp có số kết quả là: $n = C_{10}^3 = 120$. Số kết quả thuận lợi cho A là số cách lấy ra 3 chính phẩm từ 7 chính phẩm: $m = C_7^3 = 35$.

Vậy xác suất của sự kiện A là $P(A) = \frac{35}{120} = \frac{7}{24}$.

Từ định nghĩa cổ điển của xác suất, ta dễ dàng suy ra được các tính chất sau:

- $0 \leqslant P(A) \leqslant 1$;
- $P(\Omega) = 1;\ P(\emptyset) = 0$;
- Nếu A, B xung khắc ($AB = \emptyset$) thì $P(A+B) = P(A) + P(B)$;
- $P(\overline{A}) = 1 - P(A)$;
- Nếu $A \Rightarrow B$ thì $P(A) \leqslant P(B)$.

Định nghĩa cổ điển về xác suất chỉ áp dụng cho các phép thử có hữu hạn kết quả đồng khả năng. Trong thực tế, có những phép thử có vô số kết quả đồng khả năng. Khi đó, định nghĩa cổ điển về xác suất không áp dụng được. Để khắc phục hạn chế đó, người ta đưa ra định nghĩa hình học của xác suất như sau:

Xét một phép thử có vô hạn các kết quả đồng khả năng. Mỗi kết quả của phép thử được biểu diễn mỗi một điểm trong mặt phẳng(hoặc trong không gian). Giả sử tất cả các kết quả của phép thử được biểu diễn bỡi một miền hình học G(chẳng hạn đoạn thẳng, một miền mặt cong hoặc một khối không gian ...), Còn tập các kết quả thuận lợi cho sự kiện A bỡi miền con nào đó $S \subset G$. Khi đó

$$P(A) = \frac{\text{Độ đo } S}{\text{Độ đo } G}$$

Ở đây tùy thuộc vào S và G mà độ đo có thể là độ dài, diện tích hoặc thể tích và luôn giả sử rằng S và G đều là các tập đo được và độ đo của G khác không.

Ví dụ 2.2.22 Đường dây điện thoại ngầm nối một tổng đài với một trạm dài 1 km. Tính xác suất để dây đứt tại nơi cách tổng đài không quá $100m$.

Giải

Rõ ràng nếu dây điện thoại là đồng chất thì khả năng nó bị đứt tại một điểm bất kỳ là như nhau nên không gian sơ cấp có thể biểu diễn bằng một đoạn thẳng $MN = 1km$ nối tổng đài với trạm. Gọi A là sự kiện dây đứt tại nơi cách tổng đài không quá $100m$. Sự kiện A được biểu diễn bằng một đoạn thẳng MK có độ dài 100m. Từ đó $P(A) = \frac{100}{1000} = 0,1$.

2.2.2. Định nghĩa xác suất theo lối thống kê

Điều kiện đồng khả năng của các kết quả của một phép thử không phải lúc nào cũng được đảm bảo. Có nhiều hiện tượng xảy ra không theo các yêu cầu của định nghĩa cổ điển, chẳng hạn, tính xác suất một đứa trẻ sắp sinh là con trai, ngày mai trời mua lúc 5 giờ,.... Có một cách khác để xác định xác suất của một sự kiện như sau:

Xét một phép thử và sự kiện A liên kết với phép thử đó. Giả sử, phép thử được thực hiện n lần và có m lần xuất hiện sự kiện A. Khi đó m được gọi là *tần số* xuất hiện của sự kiện A và tỉ số $\frac{m}{n}$ được gọi là *tần suất* xuất hiện sự kiện A. Tương tự, nếu phép thử thực hiện lại lần thứ 2, thứ 3, ...thì tần suất xuất hiện sự kiện A tương ứng là $\frac{m_1}{n_1}$ và $\frac{m_2}{n_2}$.

Trên cơ sở quan sát lâu dài các thực nghiệm khác nhau, người ta nhận thấy rằng tần suất xuất hiện một sự kiện có tính ổn định, thay đổi rất ít trong các loạt phép thử khác nhau và thay đổi xung quanh một hằng số xác định. Sự khác biệt càng ít khi số phép thử càng lớn. Nói cách khác, khi số phép thử tăng lên vô hạn, tần suất xuất hiện sự kiện A dần đến một số xác định, số đó gọi là xác suất của sự kiện A

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

Trong thực tế, xác suất của sự kiện A được lấy gần đúng bằng tần suất xuất hiện của sự đó khi số lần thực hiện phép thử đủ lớn.

2.3. CÁC ĐỊNH LÝ CƠ BẢN VỀ PHÉP TÍNH XÁC SUẤT

2.3.1. Định lý công xác suất

Định lý 2.3.1 *Nếu A, B là hai sự kiện xung khắc (nghĩa là $A \cap B = \emptyset$) thì*

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

Proof: Tính chất này là một hệ quả của định nghĩa xác suất theo phương pháp tiên đề. Ở đây, đưa ra chứng minh cho trường hợp phép thử có hữu hạn kết quả đồng khả năng(định nghĩa cổ điển của xác suất).

Giả sử không gian sơ cấp có n kết quả đồng khả năng. Gọi m_A là số kết quả thuận lợi cho sự kiện A xảy ra, m_B là số kết quả thuận lợi cho sự kiện B xảy ra. Vì A và B xung khắc nên không có kết quả nào thuận lợi cho cả A và B nên số kết quả thuận lợi cho $A+B$ xảy ra là $m_A + m_B$. Vì thế ta có

$$P(A+B) = \frac{m_A + m_B}{n} \qquad P(A) = \frac{m_A}{n} \qquad P(B) = \frac{m_B}{n}$$

Từ đó ta có $P(A+B) = P(A) + P(B)$.

Hệ quả 2.3.1 *1. Nếu $A, \overline{A}$ là hai sự kiện đối lập thì $P(\overline{A}) = 1 - P(A)$.*

2. Nếu $A_1, A_2, \ldots, A_n$ là n sự kiện đôi một xung khắc thì

$$P(A_1 + A_2 + \ldots + A_n) = P(A_1) + P(A_2) + \ldots P(A_n)$$

Ví dụ 2.3.23 Một hộp có 6 bi đỏ và 4 bi xanh. Lấy ngâu nhiên ra 3 bi, tính xác suất để

a) 3 bi lấy ra cùng màu.

b) 3 bi lấy ra có ít nhất một bi đỏ.

Giải

a) Gọi A là sự kiện 3 bi lấy ra cùng màu; B là sự kiện 3 bi lấy ra màu xanh và C là sự kiện 3 bi lấy ra màu đỏ. Ta có $A = B + C$, hai sự kiện B, C xung khắc nên ta có

$$P(A) = P(B) + P(C) = \frac{C_4^3}{C_{10}^3} + \frac{C_6^3}{C_{10}^3} = \frac{1}{30} + \frac{1}{6} = \frac{1}{5}$$

b) Gọi A_i là sự kiện lấy ra được i bi đỏ (i=1,2,3), gọi D là sự lấy ra ít nhất một bi đỏ. Ta có $D = A_1 + A_2 + A_3$, trong đó A_1, A_2, A_3 đôi một xung khắc nên

$$P(D) = P(A_1) + P(A_2) + P(A_3) = \frac{C_6^1.C_4^2}{C_{10}^3} + \frac{C_6^2.C_4^1}{C_{10}^3} + \frac{C_6^3.C_4^0}{C_{10}^3} = \frac{29}{30}$$

Chú ý bài này cũng có thể được giải như sau:$P(D) = 1 - P(\overline{D}) = 1 - \frac{C_4^3}{C_{10}^3} = \frac{29}{30}$. Ở đây $\overline{D}$ là sự kiện 3 viên bi lấy ra màu xanh.

Ví dụ 2.3.24 Một lô hàng gồm 10 sản phẩm, trong đó có 2 phế phẩm. Lấy ngẫu nhiên không hoàn lại từ lô hàng ra 6 sản phẩm. Tìm xác suất để trong 6 sản phẩm lấy ra có nhiều nhất là 1 phế phẩm?

Giải

Gọi A là sự kiện lấy ra 6 sản phẩm và không có phế phẩm; B là sự kiện lấy ra 6 sản phẩm và có đúng 1 phế phẩm; C là sự kiện lấy ra 6 sản phẩm và có nhiều nhất là 1 phế phẩm. Ta có $C = A + B$, trong đó hai sự kiện A, B xung khắc. Áp dụng công thức cộng, ta có

$$P(C) = P(A) + P(B) = \frac{C_8^6}{C_{10}^6} + \frac{C_8^5 C_2^1}{C_{10}^6} = \frac{2}{15} + \frac{8}{15} = \frac{2}{3}$$

Định lý 2.3.2 (Định lý cộng mở rộng) *Nếu A, B là hai sự kiện bất kỳ liên kết với cùng một phép thử thì*

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

Một cách tổng quát: Nếu $A_1, A_2, \ldots, A_n$ là các sự kiện liên kết với cùng một phép thử thì

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) + \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n)$$

Chẳng hạn khi $n = 3$, ta có

$$P(A_1 + A_2 + A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1A_2) - P(A_2A_3) - P(A_1A_3) + P(A_1A_2A_3)$$

Ví dụ 2.3.25 Một lớp có 100 sinh viên, trong đó có 40 sinh viên giỏi ngoại ngữ, 30 sinh viên giỏi tin hoc, 20 sinh viên giỏi cả ngoại ngữ và tin học. Chọn ngẫu nhiên một sinh viên. Tính xác suất để sinh viên chọn ra học giỏi ít nhất một môn là ngoại ngữ hoặc tin học?

Giải
Gọi A là sự kiện sinh viên được chọn ra giỏi ngoại ngữ hoặc tin học; B là sự kiện sinh viên chọn ra giỏi ngoại ngữ; C là sự kiện sinh viên chọn ra giỏi tin học. Ta có $A = B + C$. Vì B, C không xung khắc nên áp dụng công thức cộng mở rộng, ta có

$$P(A) = P(B) + P(C) - P(BC) = \frac{10}{1000} + \frac{30}{100} - \frac{20}{100} = \frac{1}{2}$$

2.3.2. Xác suất có điều kiện. Định lý nhân xác suất

Cho A, B là hai sự kiện liên kết với cùng một phép thử. Khi đó, ký hiệu A/B là sự kiện A xảy ra khi biết sự kiện B đã xảy ra.

Định nghĩa 2.3.7 (Xác suất có điều kiện) Cho A, B là hai sự kiện liên kết với cùng một phép thử và $P(B) > 0$. Xác suất có điều kiện của sự kiện A khi biết sự kiện B đã xảy ra, ký hiệu $P(A/B)$, được xác định như sau

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

Ví dụ 2.3.26 Một hộp kín có 2 bi xanh và 1 bi đỏ. Lấy ngẫu nhiên lần lượt 2 viên bi không hoàn lại. Tìm xác suất để bi lấy ra lần thứ 2 là bi đỏ, biết rằng bi lấy ra lần thứ nhất là bi xanh?

Giải
Ta ký hiệu hai viên bi xanh là $1, 2$ và bi đỏ là 3. Khi đó mỗi kết quả đồng khả năng (i, j) với $i \neq j, i, j = 1, 2, 3$ nên không gian sơ cấp là

$$\Omega = \{(1,2);\ (1,3);\ (2,1);\ (2,3);\ (3,1);\ (3,2)\}$$

Gọi A là sự kiện bi lấy lần 1 là bi xanh; Gọi B là sự kiện bi lấy lần thứ 2 là bi đỏ. Ta có

$$A = \{(1,2);\ (1;\ 3);\ (2,1);\ (2,3)\}; \qquad B = \{(1,3);\ (2,3)\}; \qquad A \cap B = \{(1,3);\ (2;\ 3)\}$$

Vậy xác suất để lần 2 lấy được bi đỏ khi biết lần thứ nhất lấy được bi xanh là $P(B/A) = \frac{P(AB)}{P(A)} = \frac{2}{6} : \frac{4}{6} = \frac{1}{2}$.

Ví dụ 2.3.27 Có 6 người(gồm 2 nam và 4 nữ) nộp đơn xin việc vào một công ty. Giả sử rằng công ty chỉ tuyển 2 người và khả năng tuyển mỗi người là như nhau.

a) Tính xác suất để có đúng 2 nữ được chọn?

b) Giả sử có ít nhất một nữ được chọn. Tính xác suất để 2 nữ được chọn?

c) Trong 4 nữ có một người tên Huệ. Tính xác suất để Huệ được chọn khi biết có ít nhất một nữ được chọn?

Giải

a) Gọi là là sự kiện 2 nữ được chọn. Ta có

$$P(A) = \frac{C_4^2}{C_6^2} = \frac{2}{5}$$

b) Gọi B là sự kiện có ít nhất một nữ được chọn. Khi đó, sự kiện 2 nữ được chọn khi biết có ít nhất một nữ được chọn là A/B. Vì sự kiện A xảy ra thì sự kiện B xảy ra, nghĩa là $A \subset B \Rightarrow A \cap B = A$.

Ta có: $P(\overline{B}) = \frac{C_2^2}{C_6^2} = \frac{1}{15}$ suy ra $P(B) = 1 - \frac{1}{15} = \frac{14}{15}$.

Vậy xác suất để 2 nữ được chọn khi biết có ít nhất một nữ được chọn là

$$P(A/B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} = \frac{2}{5} : \frac{14}{15} = \frac{3}{7}$$

c) Gọi C là sự kiện Huệ được chọn. Khi đó sự kiện Huệ được chọn khi biết ít nhất một nữ được chọn là C/B. Vì $C \subset B$ nên $C \cap B = C$. Do đó

$$P(C/B) = \frac{P(CB)}{P(B)} = \frac{P(C)}{P(B)} = \frac{C_4^1}{C_6^2} : \frac{14}{15} = \frac{2}{7}$$

Các tính chất:

- $P(\emptyset) = P(\emptyset/B) = 0; \qquad P(\Omega) = P(\Omega/B) = 1.$
- $P((A \cup C)/B) = P(A/B) + P(C/B) - P(AC/B)$

  Đặc biệt: Nếu $AC = \emptyset$ thì $P((A \cup C)/B) = P(A/B) + P(C/B)$
- $P(\overline{A}/B) = 1 - P(A/B)$

Từ định nghĩa xác suất có điều kiện, ta suy ra được định lý sau(Vì sao?):

Định lý 2.3.3 (Định lý nhân xác suất) *Giả sử A, B là hai sự kiện liên kết với cùng một phép thử và $P(A) > 0$. Khi đó*

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

Ta có một công thức tương tự khi $P(B) > 0$ là: $P(AB) = P(B).P(A/B)$

Một cách tổng quát, định lý nhân được phát biểu như sau:

Giả sử n sự kiện $A_1, A_2, \ldots, A_n$ liên kết với cùng một phép thử và $P(A_1A_2 \ldots A_{n-1}) > 0$. Khi đó, ta có

$$P(A_1A_2 \ldots A_n) = P(A_1).P(A_2/A_1).P(A_3/A_1A_2) \ldots P(A_n/A_1A_2 \ldots A_{n-1})$$

Ví dụ 2.3.28 Một hộp kín có 2 chính phẩm và 3 phế phẩm. Lấy ngẫu nhiên lần lượt 2 sản phẩm. Tính xác suất để

a) Hai sản phẩm lấy ra đều là chính phẩm?

b) Hai sản phẩm lấy ra có ít nhất một chính phẩm?

Giải

a) Gọi A_i là sự kiện lấy được chính phẩm lần thứ $i (i = 1, 2)$; A là sự kiện lấy được hai chính phẩm. Ta có $A = A_1A_2$ do đó

$$P(A) = P(A_1).P(A_2/A_1) = \frac{2}{5}.\frac{1}{4} = \frac{2}{20}$$

b) Gọi B là sự kiện lấy được ít nhất một chính phẩm. Ta có $\overline{B} = \overline{A_1 A_2}$ do đó

$$P(\overline{B}) = P(\overline{A_1}).P(\overline{A_2}/\overline{A_1}) = \frac{3}{5}\frac{2}{4} = \frac{6}{10}$$

Vậy $P(B) = 1 - \frac{6}{10} = \frac{7}{10}$.

Ví dụ 2.3.29 Một thủ kho có một chùm chìa khóa gồm 8 chìa, trong đó chỉ có 3 chìa mở được kho. Thủ kho lấy ngẫu nhiên tường chìa một cho đến khi mở được kho thì dừng lại. Tính xác suất để:
a) Đến lần thứ 2 thì mở được kho?
b) Mở được kho không quá 3 lần?

Giải
a) Gọi A_i là sự kiện mở được kho lần thứ $i(i = \overline{1, 8})$. ; A là sự kiện đến lần thứ hai thi mở được khóa. Ta có $A = \overline{A_1}A_2$ và đo đó

$$P(A) = P(\overline{A_1})P(A_2/\overline{A_1}) = \frac{C_5^1}{C_8^1}.\frac{C_5^1.C_3^1}{C_7^1} = \frac{5}{8}.\frac{3}{7} = \frac{15}{56}$$

b) Gọi B là sự kiện mở được kho không quá 3 lần. Khi đó $\overline{B}$ là sự kiện mở được kho ít nhất 4 lần. Ta có $\overline{B} = \overline{A_1 A_2 A_3}$ nên

$$P(\overline{B}) = P(\overline{A_1}).P(\overline{A_2}/\overline{A_1}).P(\overline{A_3}/\overline{A_1 A_2}) = \frac{5}{8}.\frac{4}{7}.\frac{3}{6} = \frac{5}{28}$$

Chú ý rằng ta cũng tính đươc $P(B)$ từ công thức $B = A_1 \cup \overline{A_1}A_2 \cup \overline{A_1 A_2}A_3$ và các sự kiện $A_1, \overline{A_1}A_2$,và $\overline{A_1 A_2}A_3$ đôi một xung khắc.

2.3.3. Tính độc lập của các sự kiện

Định nghĩa 2.3.8 Cho hai sự kiện A, B liên kết với cùng một phép thử. Hai sự kiện A, B gọi là độc lập nếu $P(AB) = P(A).P(B)$

Nhận xét:

- Nếu $P(A) = 0$ thì A, B độc lập với mọi sự kiện B trong cùng một phép thử. Vì $AB \subset A \Rightarrow P(AB) \leqslant P(A) = 0$ nên $P(AB) = P(A).P(B) = 0$
- Nếu A, B là hai sự kiện trong cùng một phép thử sao cho $P(A) > 0, P(B) > 0$ thì A, B độc lập khi và chỉ khi A, B không xung khắc.
- Nếu A, B là hai sự kiện độc lập và $P(A) > 0$ thì $P(B/A) = P(B)$. Điều này có nghĩa là sự kiện A xảy ra không đem lại một thông tin nào cho biết sự kiện B có xảy ra hay không. Tương tự nếu $P(B) > 0$.

Định lý 2.3.4 *Cho A, B là hai sự kiện liên kết với cùng một phép thử và $\overline{A}, \overline{B}$ là hai sự kiện đối lập của A, B. Khi đó các mệnh đề sau là tương đương*

- *A, B độc lập;*
- *$A, \overline{B}$ độc lập;*
- *$\overline{A}, B$ độc lập;*

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

Một cách tổng quát, ta có định nghĩa sau đây:

Định nghĩa 2.3.9 Cho n sự kiện $A_1, A_2, \ldots, A_n$ liên kết với cùng một phép thử.
- Hệ n sự kiện $A_1, A_2, \ldots, A_n$ gọi là độc lập với nhau từng đôi một nếu $P(A_i \cap A_j) = P(A_i).P(A_j), \forall i, j = \overline{1, n}, i \neq j$.
- Hệ n sự kiện $A_1, A_2, \ldots, A_n$ gọi là độc lập toàn bộ nếu với bất kỳ k sự kiện $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ trong n sự kiện đó đều thỏa mãn

$$P(A_{i_1}A_{i_2}\ldots A_{i_k}) = P(A_{i_1})P(A_{i_2})\ldots P(A_{i_k})$$

với $\{i_1, i_2 \ldots, i_k\} \subset \{1, 2, \ldots, n\}, 2 \leqslant k \leqslant n$
Đặc biệt khi $k = n$ ta có $P(A_1A_2\ldots A_n) = P(A_1)P(A_2)\ldots P(A_n)$.

Chú ý rằng tính độc lập toàn bộ thì suy ra độc lập từng đôi nhưng điều ngược lại nói chung không đúng. Để thấy điều này ta xét ví dụ sau.

Ví dụ 2.3.30 Một hộp có 4 quả cầu gồm 1 cầu xanh, 1 cầu đỏ, 1 cầu trắng và 1 cầu gồm 3 màu trên. Lấy ngẫu nhiên một quả cầu. Gọi A, B, C là sự kiện lấy ra được quả cầu xanh, đỏ, trắng. Xét tính độc lập của hệ 3 sự kiện $\{A, B, C\}$.

Ta có $P(A) = P(B) = P(C) = \frac{1}{2}$ và

$$\begin{aligned}
P(AB) &= \frac{1}{4} = \frac{1}{2}.\frac{1}{2} = P(A).P(B)\\
P(AC) &= \frac{1}{4} = \frac{1}{2}.\frac{1}{2} = P(A).P(C)\\
P(BC) &= \frac{1}{4} = \frac{1}{2}.\frac{1}{2} = P(B).P(C)\\
P(ABC) &= \frac{1}{4} \neq P(A).P(B).P(C) = \frac{1}{8}
\end{aligned}$$

Vậy 3 sự kiện A, B, C độc lập từng đôi nhưng không độc lập toàn bộ.

Ví dụ 2.3.31 Một nhà máy có 3 phân xưởng hoạt động độc lập. Xác suất ngừng hoạt động của phân xưởng thứ nhất, thứ hai và thứ ba trong khoảng thời gian T tương ứng là $0,1;\ 0,2;\ 0,3$. Tìm xác suất để trong khoảng thời gian T:
a) Cả 3 phân xưởng đều ngừng hoạt động?
b) Có ít nhất một phân xưởng ngừng hoạt động?
c) Có đúng một phân xưởng ngừng hoạt động?

Giải
a) Gọi A_i là sự kiện phân xưởng i ngường hoạt động trong khoảng thời gian T($i = 1, 2, 3$). Theo giả thiết A_1, A_2, A_3 độc lập toàn bộ và

$$P(A_1) = 0,1; \qquad P(A_2) = 0,2; \qquad P(A_3) = 0,3$$

Gọi A là sự kiện cả 3 phân xưởng ngừng hoạt động trong khoảng thời gian T. Ta có $A = A_1A_2A_3$ và $P(A) = P(A_1)P(A_2)P(A_3) = 0,1.0,2.0,3 = 0,006$.
b) Gọi B là sự kiện có ít nhất một phân xưởng ngừng hoạt động trong khoảng thời gian T. Ta có $\overline{B} = \overline{A_1A_2A_3}$. Do $\overline{A_1}, \overline{A_2}, \overline{A_3}$ độc lập toàn bộ nên

$$P(\overline{B}) = P(> \overline{A_1})P(\overline{A_2})P(\overline{A_3}) = 0,9.0,8.0,7 = 0,504$$

Vậy $P(B) = 1 - P(\overline{B}) = 1 - 0,504 = 0,496$.

Chú ý ta có thể giải câu này từ biểu thức $B = A_1 \cup A_2 \cup A_3$ và áp dụng công thức cộng tổng quát.

c) Gọi C là sự kiện có đúng một phân xưởng ngừng hoạt động trong khoảng thời gian T. Ta có $C = A_1\overline{A_2A_3} \cup \overline{A_1}A_2\overline{A_3} \cup \overline{A_1A_2}A_3$ và 3 sự kiện $A_1\overline{A_2A_3}$, $\overline{A_1}A_2\overline{A_3}$, $\overline{A_1A_2}A_3$ đôi một xung khắc nên

$$P(C) = P(A_1\overline{A_2A_3}) + P(\overline{A_1}A_2\overline{A_3}) + P(\overline{A_1A_2}A_3)$$

Mặt khác, ta cũng có hệ các sự kiện $\{A_1, \overline{A_2}, \overline{A_3}\}$; $\{\overline{A_1}, A_2, \overline{A_3}\}$; $\{\overline{A_1}, \overline{A_2}, A_3\}$ độc lập toàn bộ. Do đó:

$$P(C) = P(A_1)P(\overline{A_2})P(\overline{A_3}) + P(\overline{A_1})P(A_2)P(\overline{A_3}) + P(\overline{A_1})P(\overline{A_2})P(A_3)$$

$$= 0,1.0,8.0,7 + 0,9.0,2.0,7 + 0,9.0,8.0,3 = 0,398$$

2.3.4. Công thức xác suất toàn phần và định lý Bayes

Công thức xác suất toàn phần

Định nghĩa 2.3.10 (Hệ sự kiện đầy đủ) Cho n sự kiện $A_1, A_2, \ldots, A_n$ liên kết với cùng một phép thử. Hệ n sự kiện $A_1, A_2, \ldots, A_n$ được gọi là hệ sự kiện đầy đủ nếu

i) Hệ n sự kiện đã cho đôi một xung khắc, tức là $A_i \cap A_j = \emptyset, \forall i, j (i \neq j)$;

ii) Hợp tất cả n sự kiện là sự kiện tất yếu, tức là $\cup_{i=1}^{n} A_i = \Omega$

Ví dụ tập các sự kiện sơ cấp của một phép thử là một hệ sự kiện đầy đủ.

Định lý 2.3.5 *Giả sử các sự kiện $A_1, A_2, \ldots, A_n$ liên kết với cùng một phép thử tạo thành một hệ đầy đủ các sự kiện sao cho $p(A_i) > 0, \forall i = \overline{1, n}$. Khi đó với mọi sự kiện A ta có*

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

Đẳng thức trên được gọi là **công thức xác suất toàn phần.**

Ví dụ 2.3.32 Có hai hộp giống nhau, hộp I có 6 bi đỏ và 4 bi xanh, hộp II có 8 bi đỏ và 4 bi xanh. Lấy ngẫu nhiên một hộp rồi từ đó lấy ra 2 viên bi. Tìm xác suất để hai bi lấy ra đều là bi đỏ?

Giải

Gọi $A_i (i = 1, 2)$ là sự kiện hộp thứ i được chọn; Gọi A là sự kiện hai bi lấy ra là bi đỏ.

Ta có, hai sự kiện A_1, A_2 tạo thành một hệ đầy đủ các sự kiện và $P(A_1) = P(A_2) = \frac{1}{2}$. Áp dụng công thức xác suất toàn phần, ta có

$$P(A) = P(A_1)P(A/A_1) + P(A_2)P(A/A_2) = \frac{1}{2}\frac{C_6^2}{C_{10}^2} + \frac{1}{2}\frac{C_8^2}{C_{12}^2} = \frac{25}{66}$$

Ví dụ 2.3.33 Một cửa hàng bán bóng đèn, trong đó có 20% do nhà máy thứ nhất sản xuất, 46% do nhà máy thứ 2 sản xuất, 34% do nhà máy thứ 3 sản xuất. Biết rằng tỉ lệ bóng đèn bị hỏng của nhà máy thứ nhất, thứ hai, thứ ba lần lượt là: 3%; 1%; 2%. Một người mua ngẫu nhiên một bóng đèn. Tính xác suất để bóng đèn người đó mua bị hỏng?

Gọi A_i là sự kiện bóng đèn được sản xuất ở nhà máy thứ $i(i = 1, 2, 3)$. Gọi A là sự kiện người mua được bóng đèn hỏng.

Ta có các sự kiên A_1, A_2, A_3 tạo thành hệ đầy đủ các sự kiện và

$$P(A_1) = 0,2; \quad P(A_2) = 0,46; \quad P(A_3) = 0,34$$

Áp dụng công thức xác suất toàn phần, ta có

$$P(A) = P(A_1).P(A/A_1) + P(A_2).P(A/A_2) + P(A_3).P(A/A_3)$$

$$= 0,2.0,03 + 0,46.0,01 + 0,34.0,02 = 0,174$$

Ví dụ 2.3.34 Một hộp có 10 quả bóng tennis, trong đó có 7 quả mới và 3 quả cũ. Lần một lấy ra 2 quả để thi đấu, sau đó bỏ trở lại. Sau đó, lần 2 lấy ra 2 quả để thi đấu. Tính xác suất để hai quả lấy ra lần thứ 2 là quả bóng mới?

Giải

Gọi A_i là sự kiện 2 bi lấy ra lần 1 có i bi mới. Ta có A_0, A_1, A_2 tạo thành hệ sự kiện đầy đủ và

$$P(A_0) = \frac{C_3^2}{C_{10}^2} = \frac{1}{15}; \quad P(A_1) = \frac{C_7^1 C_3^1}{C_{10}^2} = \frac{7}{15}; \quad P(A_2) = \frac{C_7^2}{C_{10}^2} = \frac{7}{15}$$

Gọi A là sự kiện 2 quả cầu lấy ra lần hai là quả cầu mới. Áp dụng công thức xác suất toàn phần, ta có

$$P(A) = P(A_0).P(A/A_0) + P(A_1).P(A/A_1) + P(A_2).P(A/A_2)$$

$$= \frac{1}{15}\frac{C_7^2}{C_{10}^2} + \frac{7}{15}\frac{C_6^2}{C_{10}^2} + \frac{7}{15}\frac{C_5^2}{C_{10}^2} = \frac{1}{15}.\frac{7}{15} + \frac{7}{15}\frac{1}{3} + \frac{7}{15}\frac{2}{9} = \frac{196}{675}$$

Định lý Bayes

Định lý 2.3.6 *Xét một phép thử. Giả sử $A_1, A_2, \ldots, A_n$ là hệ đầy đủ các sự kiện và $P(A_i) > 0.\forall i = \overline{1,n}$ và A là một sự kiện bất kỳ, $P(A) > 0$. Khi đó*

$$P(A_i/A) = \frac{P(A_i)P(A/A_i)}{\sum_{i=1}^{n} P(A_i)P(A/A_i)}$$

Ví dụ 2.3.35 Một nhà máy sản xuất thép tấm gồm 2 phân xưởng sản xuất. Phân xưởng 1 và phân xưởng 2 sản xuất với lượng sản phẩm là 60% và 40%. Biết tỉ lệ phế phẩm của phân xưởng 1 và 2 tương ứng là 3% và 4%. Lấy ngẫu nhiên một tấm thép của nhà máy thì thấy tấm thép là một phế phẩm. Tìm xác suất để tấm thép đó do phân xưởng thứ nhất sản xuất?

Giải

Gọi A_i là sự kiện tấm thép lấy ra do phân xưởng thứ i xản xuất$(i = 1, 2)$. $A - 1, A_2$ tạo thành hệ đầy đủ các sự kiện và

$$P(A_1) = 0,6; \qquad P(A_2) = 0,4$$

Gọi A là sự kiện tấm thép lấy ra là phế phẩm. Áp dụng công thức xác suất toàn phần, ta có

$$P(A) = P(A_1)P(A/A_1) + P(A_2)P(A/A_2) = 0,6.0,03 + 0,4.0,04 = 0,034$$

Áp dụng công thức Bayes, ta có xác suất để phế phẩm lấy ra do phân xưởng 1 sản xuất là

$$P(A/A_1) = \frac{P(A_1)P(A/A_1)}{P(A_1)P(A/A_1) + P(A_2)P(A/A_2)} = \frac{0,6.0,03}{0,034} = \frac{9}{17}$$

Ví dụ 2.3.36 Có hai lô hàng: lô I có 50 sản phẩm, trong đó có 20 sản phẩm xấu; lô II có 40 sản phẩm, trong đó có 15 sản phẩm xấu. Lấy ngẫu nhiên một lô và từ đó lấy ngâu nhiên một sản phẩm.

a) Tính xác suất để sản phẩm lấy ra là sản phẩm tốt?

b) Biết sản phẩm lấy ra là tốt. Tính xác suất để sản phẩm đó thuộc lô II?

Giải

a) Gọi A_1, A_2 là sự kiện sản phẩm lấy ra o lô I, II. Ta có A_1, A_2 tạo thành hệ sự kiện đầy đủ và

$$P(A_1) = P(A_2) = \frac{1}{2}$$

Gọi A là biến cố sản phẩm lấy ra là sản phẩm tốt. Áp dụng công thức xác suất toàn phần, ta có

$$P(A) = P(A_1)P(A/A_1) + P(A_2)P(A/A_2) = \frac{1}{2}\frac{30}{50} + \frac{1}{2}\frac{25}{40} = \frac{49}{80}$$

b) Áp dụng công thức Bayes, ta có xác suất để sản phẩn tốt lấy ở lô II là

$$P(A_2/A) = \frac{P(A_2)P(A/A_2)}{P(A_1)P(A/A_1) + P(A_2)P(A/A_2)} = \frac{\frac{1}{2}\frac{25}{40}}{\frac{49}{80}} = \frac{25}{49}$$

2.3.5. Dãy phép thử độc lập và công thức Bernoulli

Dãy phép thử độc lập - Dãy phép thử Bernoulli Xét một phép thử ε. Thực hiện phép thử n lần và gọi ε_i là phép thử thực hiện lần thứ i.

Định nghĩa 2.3.11 Các phép thử $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ được gọi là độc lập nếu xác suất xảy ra của các sự kiện liên kết với phép thử ε_i nào đó không phụ thuộc vào kết quả của các phép thử khác.

Như vậy, nếu A_i là sự kiện liên kết với phép thử $\varepsilon_i (i = \overline{1, n})$ thì các sự kiện $A_1, A_2, \ldots, A_n$ là độc lập toàn bộ.

Định nghĩa 2.3.12 Cho dãy n phép thử độc lập. Trong mỗi phép thử, ta xét sự kiện A và $\overline{A}$. Giả sử xác suất để sự kiện A xảy ra trong mỗi phép thử là không đổi và bằng $p (0 < p < 1)$ và xác suất để xảy ra biến cố $\overline{A} = 1 - p$. Khi đó n phép thử độc lập trên được gọi là n phép thử Bernoulli. Ký hiệu $B(n;\ p)$.

Định lý 2.3.7 (Định lý Bernoulli) *Thực hiện n phép thử độc lập. Trong mỗi phép thử sự kiện A xảy ra với xác suất không đổi $P(A) = p(0 < p < 1)$. Khi đó, xác suất để sự kiện A xảy ra đúng k lần trong n phép thử đó là*

$$P_n(k) = C_n^k p^k (1-p)^{n-k} \qquad (k = \overline{0, n})$$

Hệ quả 2.3.2 *Với những giả thiết như trong định lý Bernoulli, xác suất để trong n phép thử sự kiện A xảy ra ít nhất k_1 lần và nhiều nhất k_2 lần là*

$$P_n(k_1 \leqslant k \leqslant k_2) = \sum_{i=k_1}^{k_2} C_n^k p^k (1-p)^{n-k}$$

Số lần có khả năng xảy ra nhiều nhất

Định nghĩa 2.3.13 Cho n phép thử Bernoulli. Trong mỗi phép thử, xác xuất để sự kiện A xảy ra $P(A) = p$ và $P(\overline{A} = 1 - p$. Số m gọi là số lần xảy ra sự kiện A nhiều nhất nếu

$$P_n(m) \geqslant P_n(k), \forall k = \overline{0, n}$$

hay $P_n(m) = max\{P_n(0), P_n(1), \ldots, P_n(n)\}$

Định lý 2.3.8 *Cho n phép thử Bernoulli. Trong mỗi phép thử, xác xuất để sự kiện A xảy ra $P(A) = p$ và $P(\overline{A} = 1 - p = q$. Gọi m là số lần sự kiện A xảy ra nhiều nhất, ta có*

$$np - q \leqslant m \leqslant np + q$$

Ví dụ 2.3.37 Có 10 sinh viên thi môn xác suất. Khả năng thi đạt của các sinh viên đều như nhau và bằng 70%.

a) Tìm xác suất để có 8 sinh viên thi đạt?

b) Tìm xác suất để có ít nhất 1 sinh viên thi trượt?

c) Tìm xác suất để có ít nhất 8 sinh viên thi không đạt?

d) Tìm số sinh viên có khả năng thi đạt nhiều nhất trong 10 sinh viên?

Giải

Bài toán tương ứng với một dãy phép thử Bernoulli với $n = 8, p = 0,7$. Áp dụng các định lý trên để giải bài toán.

Chương 3

BIẾN NGẪU NHIÊN

3.1. BIẾN NGẪU NHIÊN

3.1.1. Biến ngẫu nhiên

Khái niệm

Biến ngẫu nhiên là một đại lượng có giá trị thực biến đổi phụ thuộc vào kết quả của phép thử ngẫu nhiên. Ký hiệu biến ngẫu nhiên là $X, Y, Z, \ldots$. Ta có định nghĩa chính xác của biến ngẫu nhiên như sau:

Định nghĩa 3.1.14 Biến ngẫu nhiên là một ánh xạ từ tập Ω các kết quả của một phép thử vào tập các số thực $\mathcal{R}$.

Biến ngẫu nhiên có miền giá trị hữu hạn hoặc đếm được gọi là biến ngẫu nhiên rời rạc.

Biến ngẫu nhiên có miền giá trị là một khoảng(hoặc đoạn) gọi là biến ngẫu nhiên liên tục.

biến ngẫu nhiên còn được gọi là đại lượng ngẫu nhiên.

Định nghĩa 3.1.15 (Biến ngẫu nhiên độc lập) Cho X, Y là hai biến ngẫu nhiên liên kết với một phép thử. X, Y gọi là độc lập nhau nếu $\forall x_1, x_2, y_1, y_2 \in \mathcal{R}$:

$$P((x_1 \leqslant X < x_2).(y_1 \leqslant Y < y_2)) = P(x_1 \leqslant X < x_2).P(y_1 \leqslant Y < y_2)$$

Ví dụ 3.1.38 Gieo một con xúc xắc. Gọi X là số chấm xuất hiện của con xúc xắc thì X là một biến ngẫu nhiên nhận các giá trị có thể là $1, 2, 3, 4, 5, 6$.

X là biến ngẫu nhiên rời rạc.

Ví dụ 3.1.39 Xét phép thử là việc đo thời gian sống(tính bằng giờ) của một con transitor. Gọi Y là thời gian sống của một con transitor thì Y là một biến ngẫu nhiên có miền giá trị là $[0;\ +\infty)$.

Y là biến ngẫu nhiên liên tục.

3.1.2. Hàm phân phối xác suất của biến ngẫu nhiên

Định nghĩa

Cho X là biến ngẫu nhiên liên kết với phép thử T có không gian sơ cấp là Ω. Hàm số ký hiệu và xác định như sau

$$F(x) = P(X < x) \text{ Với } (X < x) = \{\omega in\Omega : X(\omega) < x\}$$

gọi là hàm phân phối xác suất của biến ngẫu nhiên X.

Hàm phân phối xác suất phản ánh mức độ tập trung xác suất về bên trái của điểm x. Từ các tính chất của xác suất ta suy ra các tính chất sau của hàm phân phối xác suất.

Tính chất 3
Giả sử $F(x)$ là hàm phân phối xác suất của biến ngẫu nhiên X, ta có:

- $\forall x \in \mathcal{R} : 0 \leqslant F(x) \leqslant 1$.
- $\lim\limits_{x \to +\infty} F(x) = 1; \qquad \lim\limits_{x \to -\infty} F(x) = 0$.
- $\forall x_1, x_2 \in \mathcal{R} : x_1 < x_2 \Rightarrow F(x_1) \leqslant F(x_2)$.
- $F(x)$ liên tục bên trái tại mọi điểm, tức là $\lim\limits_{x \to x_0^-} F(x) = F(x_0)$.

Hệ quả 3.1.3
- $\forall \alpha, \beta \in \mathcal{R} : P(\alpha \leqslant X < \beta) = F(\beta) - F(\alpha)$
- *Nếu X là biến ngẫu nhiên liên tục thì $F(x)$ liên tục tại mọi điểm trên $\mathcal{R}$ và $P(X = \alpha) = 0, \forall \alpha \in \mathcal{R}$*

Bảng phân phối xác suất và hàm mật độ

Bảng phân phối xác suất: Bảng phân phối xác suất dùng để thiết lập luật phân phối xác suất của biến ngẫu nhiên rời rạc, nó gồm 2 hàng: hàng thứ nhất liệt kê các giá trị có thể $x_1, x_2, \ldots, x_n$ của biến ngẫu nhiên rời rạc X và hàng thứ 2 liệt kê các xác suất tương ứng $p_1, p_2, \ldots, p_n$ của các giá trị có thể đó.

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

Nếu các giá trị của biến ngẫu nhiên X gồm hữu hạn số $x_1, x_2, \ldots, x_n$ thì các sự kiện $X = x_1, X = x_2, \ldots, X = x_n$ lập thành một hệ đầy đủ các sự kiện. Do đó, ta có

- $\sum_{i=1}^{n} p_i = 1$
- $F(x) = \sum_{x_i < x} p_i, \forall x \in \mathcal{R}$.

Ví dụ 3.1.40 Gieo một con xúc xắc đồng chất. Gọi X là số chấm xuất hiện trên mặt của con xúc xắc thì X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

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

Khi đó $P(2, 5) = p_1 + p_2 = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

Hàm mật độ

Định nghĩa 3.1.16 Cho biến ngẫu nhiên liên tục X có hàm phân phối xác suất là $F(x)$. Hàm số $f(x)$ gọi là hàm mật độ của biến ngẫu nhiên liên tục X nếu $f(x) \geqslant 0, \forall x \in \mathcal{R}$, khả tích trên $\mathcal{R}$ và $F(x) = \int\limits_{-\infty}^{x} f(t)dt$.

Tính chất 4
- $\int\limits_{-\infty}^{+\infty} f(t)dt = 1$

- nếu $F(x)$ khả vi tại x_0 thì $F'(x_0) = f(x_0)$. Nếu $F(x)$ khả vi trong khoảng(a, b) thì trong khoảng (a, b) ta có $F'(x) = f(x)$.

- $\forall \alpha, \beta \in \mathcal{R}, \alpha < \beta : P(\alpha \leqslant X < \beta) = \int\limits_{\alpha}^{\beta} f(x)dx.$

Ở đây cần nhấn mạnh rằng: hàm phân phối xác suất $F(x)$ được xác định dựa vào bảng phân phối xác suất nếu X là biến ngẫu nhiên rời rạc và được xác định thông qua hàm mật độ nếu X là biến ngẫu nhiên liên tục.

Ví dụ 3.1.41 Cho biến ngẫu nhiên rời rạc X có bảng phân phối xác suất

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

Tìm hàm phân phối xác suất $F(x)$ của X và vẽ độ thị của nó.

Giải
Nếu $x \leqslant 0$ thì $F(x) = 0$
Nếu $0 < x \leqslant 1$ thì $F(x) = p_0 = \frac{1}{4}$
Nếu $1 < x \leqslant 2$ thì $F(x) = p_0 + p_1 = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$
Nếu $x > 2$ thì $F(x) = p_0 + p_1 + p_2 = \frac{1}{4} + \frac{1}{2} + \frac{1}{4} = 1$
Vậy hàm phân phối xác suất của X là

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

Ví dụ 3.1.42 Biến ngẫu nhiên X có hàm phân phối xác suất như sau

$$F(x) = \begin{cases} 0 & \text{nếu } x \leqslant -1 \\ \frac{3}{4}x + \frac{3}{4} & \text{nếu } -1 < x \leqslant \frac{1}{3} \\ 1 & \text{nếu } x > \frac{1}{3} \end{cases}$$

Tìm xác suất để X nhận giá trị trong khoảng $[0, \frac{1}{3})$

Giải
Theo tính chất của hàm phân phối xác suất, ta có

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

Ví dụ 3.1.43 Hàm phân phối xác suất của biến ngẫu nhiên liên tục X có dạng

$$F(x) = \begin{cases} 0 & \text{nếu } x \leqslant 0 \\ ax^2 & \text{nếu } 0 < x \leqslant 1 \\ 1 & \text{nếu } x > 1 \end{cases}$$

a) Tìm hệ số a?
b) Tìm hàm mật độ xác suất $f(x)$?
c) Tìm xác suất để $X \in (0,25;\ 0,75)$?

Giải

a) Vì hàm phân phối $F(x)$ liên tục tại $x = 1$ tức là $\lim\limits_{x \to 1+0} F(x) = F(1) \Rightarrow a = 1.$

b) Theo định nghĩa của hàm mật độ xác suất, ta có

$$f(x) = F'(x) = \begin{cases} 0 & \text{nếu } x \leqslant 0 \\ 2x & \text{nếu } 0 < x \leqslant 1 \\ 0 & \text{nếu } x > 1 \end{cases}$$

c) $P(0,25 \leqslant X < 0,75) = F(0,75) - F(0,25) = (0,75)^2 - (0,25)^2 = 0,5$

3.2. CÁC THAM SỐ CỦA BIẾN NGẪU NHIÊN

3.2.1. Kỳ vọng toán

Định nghĩa 3.2.17 Giả sử X là một biến ngẫu nhiên. Ta gọi kỳ vọng toán của biến ngẫu nhiên X là một số, ký hiệu $E(X)$, và được xác định như sau:

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

$$\begin{array}{c|cccc} X & x_1 & x_2 & \dots & x_n \\ \hline P & p_1 & p_2 & \dots & p_n \end{array} \text{ hoặc } \begin{array}{c|ccccc} X & x_1 & x_2 & \dots & x_n & \dots \\ \hline P & p_1 & p_2 & \dots & p_n & \dots \end{array}$$

  với $p_i = P(X = x_i)(i = \overline{1,n})$

  thì

$$E(X) = \sum_{i=1}^{n} x_i p_i \text{ hoặc } E(X) = \sum_{i=1}^{\infty} x_i p_i$$

- Nếu X là biến ngẫu nhiên liên tục với hàm mật độ $f(x)$ thì

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

  (với điều kiện tích phân suy rộng ở vế phải hội tụ tuyệt đối)

Ý nghĩa của kỳ vọng: Kỳ vọng của $E(X)$ đặc trương cho giá trị trung bình của biến ngẫu nhiên X.

Ví dụ 3.2.44 Giả sử X là biến ngẫu nhiên có bảng phân phối xác suất:

X	-1	0	1	2
P	$0,1$	$0,2$	$0,2$	$0,5$

Ta có $E(X) = -1.0,1 + 0.0,2 + 1.0,2 + 2.0,5 = 1,1$

Ví dụ 3.2.45 Cho X là biến ngẫu nhiên liên tục có hàm mật độ

$$f(x) = \begin{cases} ax^2 & \text{với } 0 \leqslant x \leqslant 1 \\ 0 & \text{với } x \notin (0;\ 1) \end{cases}$$

Tìm hàm mật độ và tính $E(X)$.

Giải

Theo tính chất của hàm mật độ, ta có $\int\limits_{-\infty}^{+\infty} f(x)dx = 1 \Rightarrow a = 3$

Vậy hàm mật độ

$$f(x) = \begin{cases} 3x^2 & \text{với } 0 \leqslant x \leqslant 1 \\ 0 & \text{với } x \notin (0;\ 1) \end{cases}$$

Từ đó, ta có:

$$E(X) = \int\limits_{-\infty}^{+\infty} xf(x)dx = \int\limits_{-\infty}^{0} xf(x)dx + \int\limits_{0} 1xf(x)dx + \int\limits_{1}^{+\infty} xf(x)dx$$

$$= \int_0^1 x.3x^2 dx = \frac{3}{4}$$

Ví dụ 3.2.46 Theo thống kê được biết tỉ lệ chết của một người ở tuổi trên 30 là $\frac{1}{1000}$. Một công ty bảo hiểm bán bảo hiểm nhân mạng cho người ở độ tuổi trên 30 với số tiền là 100000 đồng. Công ty muốn lãi trung bình khi bán một bảo hiểm là như thế là 70000. Hỏi số tiền mà công ty bảo hiểm phải trả nếu người mua bảo hiểm đó chết là bao nhiêu?

Giải

Gọi a là số tiền mà công ty bảo hiểm phải trả nếu người mua bảo hiểm chết trong năm đó. Gọi X là số tiền lãi khi bán một bảo hiểm thì X là biến ngẫu nhiên có thể nhận hai giá trị $100000-a$, 100000 và có bảng phân phối là

X	$100000-a$	100000
P	$\frac{1}{1000}$	$\frac{999}{1000}$

Ta có $E(X) = (100000-a)\frac{1}{1000} + 100000.\frac{999}{1000} = 100000 - \frac{a}{1000}$

Để tiền lãi trung bình là 70000 khi bán một bảo hiểm thì

$$E(X) = 100000 - \frac{a}{1000} = 70000 \Rightarrow a = 30.000.000$$

Tính chất 5

- $E(c) = 0 (c = const)$
- $E(cX) = cE(X)$
- $E(X \pm Y) = E(X) \pm E(Y)$
- Nếu X, Y là hai biến ngẫu nhiên độc lập thì $E(XY) = E(X)E(Y)$

3.2.2. Phương sai:

Để đo mức độ phân tán của biến ngẫu nhiên X quanh gia trị kỳ vọng $E(X)$, người ta đưa ra khái niệm phương sai như sau.

Định nghĩa 3.2.18 Giả sử X là một biến ngẫu nhiên có kỳ vọng $E(X) = a$. Nếu biến ngẫu nhiên $(X-a)^2$ có kỳ vọng thì giá trị kỳ vọng $E[(X-a)^2]$ được gọi là phương sai của biến ngẫu nhiên X. Ký hiệu $D(X)$.

Ta có: $D(X) = E[(X-a)^2]$

Từ định nghĩa, ta suy ra công thức tính phương sai như sau:

a) Nếu X là biến ngẫu nhiên có bảng phân phối xác suất

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

hoặc

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

$$D(X) = \sum_{i=1}^{n} (x_i - a)^2.p_i \text{ hoặc } D(X) = \sum_{i=1}^{\infty} (x_i - a)^2.p_i$$

b) Nếu X là biến ngẫu nhiên liên tục thì

$$D(X) = \int_{-\infty}^{+\infty} (x-a)^2 f(x) dx$$

Tính chất 6
Từ định nghĩa của phương sai, ta suy ra được các tính chất sau:

- $D(X) = E(X^2) - E^2(X)$
- $D(c) = 0 (c = const)$
- $D(cX) = c^2 D(X)$
- Nếu X, Y là hai biến ngẫu nhiên độc lập thì $D(X \pm Y) = D(X) \pm D(Y)$

Ví dụ 3.2.47 Cho X là biến ngẫu nhiên rời rạc có bảng phân phối xác suất

X	-1	0	1	2
P	$0,1$	$0,2$	$0,3$	$0,4$

Tính $E(X), D(X)$

Giải
Ta có

$$E(X) = -1.0,1 + 0.0,2 + 1.0,3 + 2.0,4 = 1$$

$$D(X) = (-1-1)^2.0,1 + (0-1)^2.0,2 + (1-1)^{\cdot}0,3 + (2-1)^2.0,4 = 1$$

Ta cũng có thể tính được $D(X)$ từ công thức $D(X) = E(X^2) - E^2(X)$ trong đó

$$E(X^2) = (-1)^2.0,1 + 0^2.0,2 + 1^2.0,3 + 2^2.0,4 = 2$$

Vậy $E(X) = 1$ và $D(X) = 1$.

3.2.3. Độ lệch chuẩn

Định nghĩa 3.2.19 Giả sử biến ngẫu nhiên có phương sai $D(X)$. Khí đó, biến $\delta(X) = \sqrt{D(X)}$ gọi là độ lệch chuẩn của X.

Độ lệch chuẩn là một biến đặc trưng cho tính ổn định của biến ngẫu nhiên X. Trong lĩnh vực đầu tư, độ lệch chuẩn đặc trưng cho mức độ rủi ro.

3.3. MỘT SỐ BIẾN NGẪU NHIÊN LIÊN TỤC QUAN TRỌNG

3.3.1. Biến chuẩn $N(a, \delta)$

Định nghĩa 3.3.20 Biến ngẫu nhiên liên tục X gọi là chuẩn $N(a, \delta)$ nếu hàm mật độ của nó có dạng

$$f(x) = \frac{1}{\delta\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\delta^2}}, \forall x \in \mathcal{R};\ a, \delta:\ \text{tham số };\ \delta > 0$$

Tính chất 7 (Các tham số đặc trưng)
Cho X là chuẩn $N(a, \delta)$, ta có

$$E(X) = Med(X) = Mod(X) = a; \qquad D(X) = \delta^2;\ \delta(X) = \delta$$

Cách xác định xác suất của một sự kiện liên kết với biến chuẩn

Định nghĩa 3.3.21 (Hàm Laplace) Hàm laplace là hàm số

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$$

Định lý 3.3.9 *Nếu $F(x)$ là hàm phân phối xác suất của biến chuẩn $N(a, \delta)$ thì*

$$F(x) = \Phi(\frac{x-a}{\delta})$$

Định lý 3.3.10 *Nếu X là biến chuẩn $N(a, \delta), \forall \alpha, \beta \in \mathcal{R}, \alpha < \beta$ thì*

$$P(\alpha \leqslant X < \beta) = \Phi(\frac{\beta - a}{\delta}) - \Phi(\frac{\alpha - a}{\delta})$$

Hệ quả 3.3.4 *Nếu X là biến chuẩn $N(a, \delta), \forall \alpha \in \mathcal{R}, \alpha > 0$ thì*

$$P(\mid X - a \mid < \alpha) = 2\Phi(\frac{\alpha}{\delta})$$

Các định lý về biến chuẩn

Định lý 3.3.11 *Nếu X là biến chuẩn $N(a, \delta)$ thì biến ngẫu nhiên $cX, X \pm c$(với $c = const$) là biến chuẩn với tham $N(ca, \mid c \mid \delta), N(a \pm c, \delta)$.*

Định lý 3.3.12 *Nếu X_i là các biến chuẩn $N(a_i, \delta_i), i = \overline{1,n}$ và các $X_i, i = \overline{1,n}$ độc lập toàn bộ thì biến ngẫu nhiên $X = \sum_{i=1}^n X_i$ là biến chuẩn $N(a, \delta)$ với*

$$a = a_1 + a_2 + \ldots + a_n; \qquad \delta^2 = \delta_1^2 + \delta_2^2 + \ldots + \delta_n^2$$

.

Hệ quả 3.3.5 *Nếu X_i là các biến chuẩn $N(a, \delta), \forall i = \overline{1,n}$ và các $X_i, i = \overline{1,n}$ độc lập toàn bộ thì biến ngẫu nhiên $X = \frac{1}{n}\sum_{i=1}^n X_i$ là biến chuẩn $N(a, \frac{\delta}{\sqrt{n}})$*

Định lý 3.3.13 (Đinh lý Lindebeg- Levi) *Nếu $X_i, i = \overline{1,n}$ là n biến ngẫu nhiên độc lập toàn bộ, cùng phân phối với kỳ vọng $E(X_i) = a$, phương sai $D(X_i) = \delta^2, i = \overline{1,n}$ thì biến ngẫu nhiên $X = \sum_{i=1}^n X_i$ và biến $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$ xấp xỉ biến chuẩn $N(na, \sqrt{n}\delta)$ và $N(a, \frac{\delta}{\sqrt{n}})$.*

Tức là với n khá lớn, ta có

$$P(X < x) \approx \frac{1}{\sqrt{n}\delta\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-na)^2}{2n\delta^2}} dt$$

$$P(\overline{X} < x) \approx \frac{\sqrt{n}}{\delta\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-a)^2 n}{2\delta^2}} dt$$

Hệ quả 3.3.6 (Định lý Moivre - Laplace) *Nếu $X_i, i = \overline{1,n}$ là n biến ngẫu nhiên đơn giản $A(a)$ độc lập toàn bộ có kỳ vọng $E(X_i) = a, D(X_i) = a(1-a), i = \overline{1,n}$ thì biến ngẫu nhiên $X = \sum_{i=1}^n X_i$ là biến nhị thức $B(n,a)$ xấp xỉ biến chuẩn $N(na, \sqrt{na(1-a)})$*

Tức là với n khá lớn, ta có

$$P(B(n,a) < x) \approx \frac{1}{\sqrt{na(1-a).2\pi}} \int_{-\infty}^x e^{-\frac{(t-na)^2}{2na(1-a)}} dt$$

hay

$$P(\alpha \leqslant B(n,a) < \beta) \approx \Phi(\frac{\alpha - na}{na(1-a)}) - \Phi(\frac{\beta - na}{na(1-a)})$$

với $\Phi(x)$ là hàm Laplace.

3.3.2. Biến khi bình phương χ_n^2

Định nghĩa 3.3.22 Cho $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên chuẩn $N(0,1)$ độc lập với nhau. Khi đó biến ngẫu nhiên $X = \sum_{i=1}^{n} X_i^2$ được gọi là biến khi bình phương với n bậc tự do. Ký hiệu K_n hoặc χ_n^2.

Định lý 3.3.14 *Hàm mật độ $f(x)$ của phân phối khi bình phương χ_n^2 với n bậc tự do là*

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

Tính chất 8 (Các tham số đặc trưng)
Cho X là biến khi bình phươngχ_n^2 với n bậc tự do, ta có

$$E(X) = n; \qquad D(X) = 2n$$

Các định lý về biến χ_n^2

Định lý 3.3.15 *Nếu X, Y là các biến χ_n^2, χ_m^2 thì $X + Y$ là biến χ_{n+m}^2.*

Định lý 3.3.16 *Nếu X là các biến χ_n^2 thì biến ngẫu nhiên $Z = \frac{\chi_n^2 - n}{\sqrt{2n}}$ xấp xỉ biến chuẩn $N(0,1)$ khi n khá lớn $(n > 30)$.*

3.3.3. Biến Student T_n

Định nghĩa 3.3.23 Cho X là biến ngẫu nhiên chuẩn $N(0,1)$, Y là biến χ_n^2 với n bậc tự do và X, Y độc lập. Khi đó, biến ngẫu nhiên $\frac{X}{\sqrt{\frac{Y}{n}}}$ được gọi là biến Student với n bậc tự do. Ký hiệu $T_n = \frac{X}{\sqrt{\frac{Y}{n}}}$.

Định lý 3.3.17 *Hàm mật độ $f(x)$ của phân phối Student T_n với n bậc tự do là*

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

Tính chất 9 (Các tham số đặc trưng)
Cho X là biến Student T_n với n bậc tự do, ta có

$$E(T_n) = 0; \qquad D(T_n) = \frac{n}{n-2}$$

Định lý 3.3.18 *Biến Student T_n sẽ xấp xỉ biến chuẩn $N(0,1)$ khi n khá lớn $(n > 30)$*

Chương 4

MẪU VÀ CÁC THAM SỐ MẪU

4.1. MẪU VÀ CÁC PHƯƠNG PHÁP XÂY DỰNG MẪU

4.1.1. Tổng thể và mẫu

Định nghĩa 4.1.24 Tập hợp toàn bộ các đối tượng cần nghiên cứu, khảo sát "đặc tính" nào đó của chúng gọi là tổng thể(hay tập hợp tổng quát hay tập sinh). Ký hiệu tập tổng thể là Ω.

Số phần tử(lực lượng) của Ω gọi là kích thước của tổng thể Ω.

Định nghĩa 4.1.25 Từ tổng thể, ta chọn ngẫu nhiên(theo một cách chọn đã quy định trước) n phần tử(đối tượng), tập n phần tử được chọn gọi là một mẫu. Khi đó, n gọi là kích thước mẫu

4.1.2. Các phương pháp xây dựng mẫu

Mẫu lặp

Lấy mẫu có lặp là lấy mẫu mà phần tử lấy ra, sau khi đã ghi giá trị đặc trưng, được trả trở lại tổng thể trộn đều rồi lấy tiếp phần tử khác.(phân phối nhị thức).

Mẫu không lặp

Lấy mẫu không lặp là lấy mẫu mà phần tử lấy ra, sau khi đã ghi giá trị đặc trưng, không trả trở lại tổng thể mà lấy tiếp phần tử khác.(phân phối siêu bội).

Ta biết rằng, phân phối siêu bội hội tụ về phân phối nhị thức nên khi số phần tử của tổng thể là N rất lớn so với kích thước mẫu $n(N > 100n)$ thì việc lấy mẫu không lặp lại xem như mẫu có lặp. Do đó, trong lý thuyết, ta thường nghiên cứu mẫu lặp.

Xây dựng mẫu theo lối điển hình

Ví dụ 4.1.48 Để ước lượng chiều cao trung bình của học sinh lớp 4 tại địa phương A có 20000 học sinh lớp 4. Trong đó, ở thành phố 7000, ở nông thôn 8000 và ở miền núi 5000 học sinh. Lấy mẫu 2000 học sinh như sau: lấy 700 học sinh ở thành phố, 800 học sinh ở nông thôn và 500 học sinh ở miền núi. Khi đó mẫu được chọn như trên được xây dựng theo lối điển hình.

Xây dựng mẫu theo lối máy móc

Ví dụ 4.1.49 Để kiểm tra một đoạn đường AB dài $3000m$. Bắt đầu từ A cứ cách $30m$ ta lấy một mẫu. Khi đó, ta được một mẫu có kích thước $n = 100$ xây dựng theo lối máy móc.

4.2. Các phương pháp trình bày số liệu

4.2.1. Mẫu ngẫu nhiên và mẫu thực nghiệm

Ta chọn ngẫu nhiên một phần tử từ tập Ω. Khi đó Ω được xem như là không gian các sự kiện sơ cấp. Gọi X là biến ngẫu nhiên biểu thị đặc trưng nghiên cứu trên tập Ω(X liên kết với phép thử lấy ra một phần tử). Ký hiệu ϵ là phép thử lấy ra một phần tử.

Lặp lại phép thử ϵ n lần. Gọi X_i là giá trị đặc trưng của phần tử được lấy ra lần thứ $i(i = \overline{1,n}$. Khi đó các biến $X_1, X_2, \ldots, X_n$ độc lập có cùng quy luật phân phối với X, n biến ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ gọi là mẫu ngẫu nhiên của X.

Sau khi lấy mẫu, ta có $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. Bộ n số $(x_1, x_2, \ldots, x_n)$ được gọi là mẫu cụ thể(mẫu thực nghiệm) của X.

Định nghĩa 4.2.26 Ta gọi mẫu ngẫu nhiên kích thức n của biến ngẫu nhiên X là một bộ n thứ tự $(X_1, X_2, \ldots, X_n)$, trong đó $X_1, X_2, \ldots, X_n$ là n biến ngẫu nhiên độc lập có cùng phân phối xác suất với X.

Sau khi đã lấy mẫu, ta có $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. Bộ n số $(x_1, x_2, \ldots, x_n)$ được gọi là mẫu cụ thể(mẫu thực nghiệm) của X.

4.2.2. Các phương pháp trình bày mẫu

Trình bày một mẫu có ít giá trị khác nhau

Giả sử khi lấy mẫu kích thức n của biến ngẫu nhiên X có mẫu cụ thể với số liệu ban đầu $(x_1, x_2, \ldots, x_n)$ nhưng trong đó chỉ có k giá trị khác nhau: $a_1 < a_2 < \ldots < a_k$

Gọi n_i là số lần $a_i(i = \overline{1,n}$ có trong mẫu thực nghiệm. n_i gọi là tần số.

Gọi $f_i = \frac{n_i}{n}$ là tần suất của giá trị a_i trong mẫu thực nghiệm.

Khi đó, ta có bảng thống kê(Bảng phân phối tần số không chia lớp) sau:

a_i	a_1	a_2	$\ldots$	a_k
n_i	n_1	n_2	$\ldots$	n_k

Ví dụ 4.2.50 Ta lấy mẫu kích thước $n = 20$, ta có 1, 3, 2, 1, 5, 3, 4, 1, 4, 3, 2, 5, 4, 3, 12, 1, 4, 3, 3

Ta có bảng thống kê

a_i	1	2	3	4	5
n_i	5	3	6	4	2

Trình bày một mẫu có nhiều giá trị khác nhau

Trong trường hợp lấy mẫu kích thước n có nhiều giá trị khác nhau hoặc do ý nghĩa thực tế mà ta chia mẫu thành nhiều lớp.

Không có quy tắc chia lớp. Tuy nhiên, theo một số nhà thống kê để nghị chia lớp như sau:

1) Xác định số lượng lớp k

$$\begin{cases} 1 + log_2 n \leqslant k \leqslant 5lgn \\ 6 \leqslant k \leqslant 20 \end{cases}$$

2) Bề rộng của lớp

$$b = \frac{a_{max} - a_{min}}{k}$$

3) Tần số n_i của lớp $a_{i-1} - a_i$ là số lần giá trị của mẫu mà $a_{i-1} \leqslant x < a_i$

$f_i = \frac{n_i}{n}$ là tần suất của lớp $a_{i-1} - a_i$

4) Giá trị chính giữa(trung tâm) của lớp $a_{i-1} - a_i$ là: $a_i^* = \frac{a_{i-1}+a_i}{2}$
Ta có bảng thống kê(Bảng phân phối tần số chia lớp) như sau:

Lớp$[a_i, a_i)$	$a_0 - a_1$	$a_1 - a_2$	$\ldots$	$a_{k-1} - a_k$
n_i	n_1	n_2	$\ldots$	n_k

Chú ý, nếu trong các bảng phân phối tần số thực nghiệm trên ta thay tần số n_i bỡi tần suất tương ứng f_i ta được bảng gọi là bảng phân phối tần suất(chia lớp hoặc không chia lớp) thực nghiệm.

Hàm phân phối thực nghiệm

Định nghĩa 4.2.27 Cho X là một biến ngẫu nhiên và lấy mẫu kích thức n của X. Hàm phân phối thực nghiệm ứng với mẫu được chọn, ký hiệu $F_n(x)$, và được xác định như sau:
+ Nếu mẫu thực nghiệm cho theo bảng không chia lớp (4.2.2.) thì

$$F_n(x) = \sum_{a_i<x} \frac{n_i}{n} = \begin{cases} 0 & \text{Nếu } x \leqslant a_1 \\ \frac{n_1}{n} & \text{Nếu } a_1 < x \leqslant a_1 \\ \frac{n_1+n_2}{n} & \text{Nếu } a_2 < x \leqslant a_3 \\ \ldots\ldots & \\ \frac{\sum_{i=1}^{k-1} n_i}{n} & \text{Nếu } a_{k-1} < x \leqslant a_k \\ 1 & \text{Nếu } x > a_k \end{cases}$$

+ Nếu mẫu thực nghiệm cho theo bảng chia lớp (**??**) thì

$$F_n(x) = \sum_{a_{i-1}<x} \frac{n_i}{n} = \begin{cases} 0 & \text{Nếu } x \leqslant a_0 \\ \frac{n_1}{n} & \text{Nếu } a_0 < x \leqslant a_1 \\ \frac{n_1+n_2}{n} & \text{Nếu } a_1 < x \leqslant a_2 \\ \ldots\ldots & \\ \frac{\sum_{i=1}^{k-1} n_i}{n} & \text{Nếu } a_{k-2} < x \leqslant a_{k-1} \\ 1 & \text{Nếu } x > a_{k-1} \end{cases}$$

Định lý 4.2.19 *Giả sử $F(x)$ là hàm phân phối xác suất của X và $F_n(x)$ là hàm phân phối thực nghiệm của X. Khi đó, với n khá lớn $F_n(x) \approx F(x)$.*

Ví dụ 4.2.51 Tìm hàm phân phối thực nghiệm của X biết

a)

a_i	1	3	7
n_i	2	5	3

; b)

Lớp$[a_i, a_i)$	$0-4$	$4-8$	$8-12$
n_i	1	5	3

Giải
a) Ta có

$$F_{10}(x) = \frac{1}{10} \sum_{n_i<x} n_i = \begin{cases} 0 & \text{Nếu } x \leqslant 1 \\ \frac{2}{10} & \text{Nếu } 1 < x \leqslant 3 \\ \frac{7}{10} & \text{Nếu } 3 < x \leqslant 5 \\ 1 & \text{Nếu } x > 5 \end{cases}$$

b) Ta có

$$F_9(x) = \begin{cases} 0 & \text{Nếu } x \leqslant 0 \\ \frac{1}{9} & \text{Nếu } 0 < x \leqslant 4 \\ \frac{2}{3} & \text{Nếu } 4 < x \leqslant 8 \\ 1 & \text{Nếu } x > 8 \end{cases}$$

4.3. CÁC THAM SỐ ĐẶC TRƯNG CỦA MẪU

Giả sử $(X_1, X_2, \ldots, X_n)$ là mẫu ngẫu nhiên của X và sau khi lấy mẫu ta có mẫu thực nghiệm $(x_1, x_2, \ldots, x_n)$

4.3.1. Các tham số của mẫu ngẫu nhiên

1. Biến ngẫu nhiên $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ được gọi là **trung bình của mẫu ngẫu nhiên**
2. Biến ngẫu nhiên $\delta_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ được gọi là **phương sai của mẫu ngẫu nhiên**

   $\delta_{n-1}^2 = \frac{n}{n-1}\delta_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ được gọi là **phương sai điều chỉnh của mẫu ngẫu nhiên**

   $\delta_n = \sqrt{\delta_n^2}$: Độ lệch chuẩn của mẫu ngẫu nhiên.

   $\delta_{n-1} = \sqrt{\delta_{n-1}^2}$: Độ lệch chuẩn điều chỉnh của mẫu ngẫu nhiên.

4.3.2. Các tham số của mẫu thực nghiệm

1. **Số trung bình của mẫu thực nghiệm:**

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

2. **Số phương sai của mẫu thực nghiệm:**

$$\delta_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

3. **Số phương sai điều chỉnh mẫu thực nghiệm:**

$$\delta_{n-1}^2 = \frac{n}{n-1}\delta_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

   $\delta_n = \sqrt{\delta_n^2}$: Độ lệch chuẩn của mẫu thực nghiệm.

   $\delta_{n-1} = \sqrt{\delta_{n-1}^2}$: Độ lệch chuẩn điều chỉnh của mẫu thực nghiệm.

Từ các công thức trên, ta suy ra công thức tính đối với mẫu thực nghiệm có bảng phân phối không chia lớp và chia lớp như sau:

+ Nếu mẫu thực nghiệm có bảng phân phối tần số không chia lớp dạng

$$\begin{array}{c|cccc} a_i & a_1 & a_2 & \ldots & a_k \\ \hline n_i & n_1 & n_2 & \ldots & n_k \end{array} (\sum_{i=1}^{k} n_i = n)$$

thì

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i a_i$$

$$\delta_n^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(a_i - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i a_i^2 - \overline{x}^2$$

+ Nếu mẫu thực nghiệm có bảng phân phối tần số chia lớp

Lớp$[a_i, a_i)$	$a_0 - a_1$	$a_1 - a_2$	$\dots$	$a_{k-1} - a_k$
n_i	n_1	n_2	$\dots$	n_k

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

Đặt $a_i^* = \frac{a_{i-1}+a_i}{2}$, ta có bảng

a_i^*	a_1^*	a_2^*	$\dots$	a_k^*
n_i	n_1	n_2	$\dots$	n_k

Khi đó

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i a_i^*$$

$$\delta_n^2 = \frac{1}{n}\sum_{i=1}^{k} n_i(a_i^* - \overline{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_i {a^*}_i^2 - \overline{x}^2$$

Ví dụ 4.3.52 Tính $\overline{x}, \delta_n^2$ của mẫu trong các trường hợp sau:

$a)$

a_i	1	3	5
n_i	3	5	2

; $b)$

$[a_i, a_i)$	$0-2$	$2-4$	$4-6$	$6-8$	$8-10$	$10-12$
n_i	5	10	10	5	10	20

Giải

a) Lập bảng tính

a_i	n_i	$a_i.n_i$	$n_i a_i^2$
1	3	3	3
3	5	15	45
5	2	10	50
$\sum$	$n = 10$	28	98

Số trung bình mẫu là $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i a_i = \frac{1}{10}.28 = 2,8$

Số phương sai mẫu $\delta_n^2 = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^2 - \overline{x}^2 = \frac{1}{10}98 - (2,8)^2 = 1,96$

b) Đặt $x_i^* = \frac{x_{i-1}+x_i}{2}$, ta có

x_i^*	1	3	5	7	9	11
n_i	5	10	10	5	10	20

Lập bảng tính

x_i^*	n_i	$x_i^*.n_i$	$n_i {x_i^*}^2$
1	5	5	5
3	10	30	90
5	10	50	250
7	5	35	245
9	10	90	810
11	20	220	2420
$\sum$	$n = 60$	430	3820

Số trung bình mẫu là $\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i x_i^* = \frac{1}{60}.430 = \frac{43}{6}$

Số phương sai mẫu $\delta_n^2 = \frac{1}{n}\sum_{i=1}^{k} n_i {x^*}_i^2 - \overline{x}^2 = \frac{1}{60}3820 - (\frac{43}{6})^2 = 12,31$

Công thức tính toán Khi tính toán các tham số đặc trương của mẫu thực nghiệm để tránh việc tính toán các số có giá trị lớn phức tạp, người ta thường sử dụng các tính chất sau

$$\forall x_0 \in \mathcal{R}, \forall d \neq 0, \sum_{i=1}^{k} n_i = n \text{ ta có}$$
$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k} n_i a_i = \frac{d}{n}\sum_{i=1}^{k} n_i \frac{a_i - x_0}{d} + x_0$$
$$\delta_n^2 = \frac{1}{n}\sum_{i=1}^{k} n_i (a_i - \overline{x})^2 = \frac{d^2}{n}\sum_{i=1}^{k} n_i (\frac{a_i - x_0}{d}) - (\overline{x} - x_0)^2$$

Thông thường ta chọn x_0 là gia trị tại đó tần số lớn nhất, d là khoảng cách đều(nếu có).

Ví dụ 4.3.53 Tìm $\overline{x}, \delta_n^2, \delta_{n-1}$ với

x_i	$3,94$	$3,97$	$4,00$	$4,03$	$4,06$
n_i	1	7	10	5	2

GiảiTa chọn $x_0 = 4,00 = 4, d = 0,03$. Ta có bảng tính:

x_i	n_i	$\frac{x_i-4}{0{,}03}$	$n_i\frac{x_i-4}{0{,}03}$	$n_i(\frac{x_i-4}{0{,}03})^2$
$3,94$	1	-2	-2	4
$3,97$	7	-1	-7	7
$4,00$	10	0	0	0
$4,03$	5	1	5	5
$4,06$	2	2	4	8
$\sum$	$n = 25$	0	0	24

Số trung bình mẫu là

$$\overline{x} = \frac{d}{n}\sum_{i=1}^{k} n_i \frac{a_i - x_0}{d} + x_0 = \frac{0,03}{25}.0 + 4 = 4$$

Số phương sai mẫu và phương sai điều chỉnh

$$\delta_n^2 = \frac{d^2}{n}\sum_{i=1}^{k} n_i (\frac{a_i - x_0}{d}) - (\overline{x} - x_0)^2 = \frac{0,03^2}{25}.24 - (4-4)^2 = 0,000864$$

$$\delta_{n-1} = \sqrt{\frac{n}{n-1}\delta_n^2} = \sqrt{\frac{25}{24}.0,000864} = 0,03$$

Ví dụ 4.3.54 Tìm $\overline{x}, \delta_n^2, \delta_{n-1}$ với

$[a_{i-1}, a_i)$	$10500 - 10550$	$10550 - 10600$	$1060 - 10650$	$10650 - 10700$
n_i	15	55	20	10

Giải

Đặt $x_i^* = \frac{x_{i-1}+x_i}{2}$, ta có

x_i^*	10525	10575	10625	10675
n_i	15	55	20	10

Ta chọn $x_0 = 10575, d = 50$. Ta có bảng tính:

x_i^*	n_i	$\frac{x_i^*-10575}{50}$	$n_i.\frac{x_i^*-10575}{50}$	$n_i(\frac{x_i^*-10575}{50})^2$
10525	15	-1	-15	15
10575	55	0	0	0
10625	20	1	20	20
10675	10	2	20	40
$\sum$	$n = 100$	2	25	75

Số trung bình mẫu là

$$\overline{x} = \frac{50}{100}.25 + 10575 = 10587,5$$

Số phương sai mẫu và phương sai điều chỉnh

$$\delta_n^2 = \frac{50^2}{100}.75 - (10587,5 - 10575)^2 = 1618,75$$

$$\delta_{n-1} = \sqrt{\frac{n}{n-1}\delta_n^2} = \sqrt{\frac{100}{99}.1618,75} \approx 40,44$$

Chương 5

ƯỚC LƯỢNG THAM SỐ

5.1. PHƯƠNG PHÁP ƯỚC LƯỢNG ĐIỂM - HÀM ƯỚC LƯỢNG

5.1.1. Phương pháp ước lượng điểm

Giả sử X là biến ngẫu nhiên biểu thị đặc trưng nghiên cứu trên tập Ω, có phân phối xác suất đã biết nhưng còn phụ thuộc vào tham số θ chưa biết.

Để ước lượng θ, ta lấy mẫu kích thước n. Khi đó, ta có mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$. Sau khi lấy mẫu, ta được mẫu thực nghiệm $(x_1, x_2, \ldots, x_n)$ với $X_i = x_i(i = \overline{1,n})$.

Ứng với mỗi mẫu thực nghiệm $(x_1, x_2, \ldots, x_n)$ ta có một số $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ dùng để ước lượng cho θ. Phương pháp ước lượng đó gọi là phương pháp ước lượng điểm. Vì số $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ ứng với một điểm trên đường thẳng số.

Ta gọi U là tập hợp các mẫu thực nghiệm $(x_1, x_2, \ldots, x_n)$ hay nói cách khác U là miền giá trị của mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$. Khi đó ta có một hàm

$$\hat{\theta}_n : U \to \mathcal{R}, (x_1, x_2, \ldots, x_n) \mapsto \hat{\theta}_n(x_1, x_2, \ldots, x_n)$$

Biến ngẫu nhiên $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ gọi là hàm ước lượng của tham số θ.

Vấn đề là phải tìm hàm $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ sao cho ước lượng được tốt, tức là ước lượng không mắc sai số hệ thống và hiệu quả.

5.1.2. Các tiêu chuẩn ước lượng tham số đặc trưng của X

Ước lượng không chệch

Định nghĩa 5.1.28 Hàm ước lượng $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ của tham số θ được gọi là ước lượng không chệch của tham số θ nếu $E(\hat{\theta}_n(x_1, x_2, \ldots, x_n)) = \theta$.

Ngược lại nếu $E(\hat{\theta}_n(x_1, x_2, \ldots, x_n)) \neq \theta$ thì ta nói hàm ước lượng$\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ là hàm ước lượng chệch của θ.

Ước lượng bền vững

Định nghĩa 5.1.29 Hàm ước lượng $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ của tham số θ được gọi là ước lượng bền vững của tham số θ nếu

$$\forall \epsilon > 0 \text{ ta có } \lim_{n \to +\infty} P(\mid \hat{\theta}_n - \theta \mid < \epsilon) = 1$$

Định lý 5.1.20 *Nếu hàm ước lượng $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ thỏa mãn*

i) $E(\hat{\theta}_n(x_1, x_2, \ldots, x_n)) = \theta$ *hay* $\lim_{n\to\infty} E(\hat{\theta}_n(x_1, x_2, \ldots, x_n)) = \theta$
ii) $\lim_{n\to\infty} D(\hat{\theta}_n(x_1, x_2, \ldots, x_n)) = 0$ *thì* $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ *là ước lượng bền vững của* θ*.*

Ước lượng hiệu quả

Định nghĩa 5.1.30 Ước lượng không chệch $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ của θ được gọi là ước lượng hiệu quả nếu với mọi ước lượng không chệch $\hat{\theta}'_n(x_1, x_2, \ldots, x_n)$ của θ thì $D(\hat{\theta}_n(x_1, x_2, \ldots, x_n)) \leqslant D(\hat{\theta}'_n(x_1, x_2, \ldots, x_n))$.

Định nghĩa 5.1.31 Hàm ước lượng $\hat{\theta}_n(x_1, x_2, \ldots, x_n)$ của tham số θ được gọi là ước lượng tốt của tham số θ nếu nó là ước lượng không chệch, bền vững và hiệu quả của θ

$$\forall \epsilon > 0 \text{ ta có } \lim_{n\to+\infty} P(|\ \hat{\theta}_n - \theta\ | < \epsilon) = 1$$

5.1.3. Ước lượng của kỳ vọng và phương sai của biến ngẫu nhiên X

Định lý 5.1.21 *Cho* X *là một biến ngẫu nhiên và* $(x_1, x_2, \ldots, x_n)$ *là mẫu ngẫu nhiên của* X*. Khi đó*
a) $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ *là ước lượng không chệch, bền vững và hiệu quả của kỳ vọng* $E(X)$ *của biến ngẫu nhiên* X*.*
b) Phương sai điều chỉnh $\delta^2_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ *của mẫu ngẫu nhiên là ước lượng không chệch, bền vững của phương sai* $D(X)$ *đối với biến ngẫu nhiên* X*.*

Ý nghĩa: - Muốn ước lượng kỳ vọng $E(X)$ ta lấy trung bình mẫu ước lượng cho nó.
- Muốn ước lượng phương sai $D(X)$ ta lấy phương sai mẫu điều chỉnh ước lượng cho nó.

5.2. ƯỚC LƯỢNG KHOẢNG

5.2.1. Nguyên lý xác suất nhỏ và lớn

Nguyên lý xác suất nhỏ

Với một số $\alpha > 0$ khá bé và $P(A) = \alpha$(thông thường $0 < \alpha \leqslant 0,05$) thì trong thực tế ta thừa nhận sự kiện A không xảy ra trong một lần thực hiện phép thử.

Nguyên lý xác suất lớn

Với một số $\alpha > 0$ khá bé và $P(B) = 1 - \alpha$(thông thường $0 < \alpha \leqslant 0,05$) thì trong thực tế ta thừa nhận sự kiện B luôn xảy ra trong một lần thực hiện phép thử.

5.2.2. Khoảng tin cậy và độ tin cậy

Định nghĩa

Giả sử X là biến ngẫu nhiên biểu thị đặc trưng nghiên cứu trên tập Ω, có phân phối xác suất đã biết nhưng còn phụ thuộc vào tham số θ chưa biết. Để ước lượng θ ta lấy mẫu kích thức n. Giả sử $(x_1, x_2, \ldots, x_n)$ là mẫu thực nghiệm ứng với mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ của X. Nếu ta tìm được hai biến ngẫu nhiên $f_1(X_1, X_2, \ldots, X_n)$, $f_2(X_1, X_2, \ldots, X_n)$ sao cho

$$P[f_1(X_1, X_2, \ldots, X_n) < \theta < f_2(X_1, X_2, \ldots, X_n)] = \gamma$$

,trong đó $\gamma = 1 - \alpha$ cho trước và gần bằng 1(thông thường $0 < \alpha \leqslant 0,05$), thì khoảng số thực $(f_1(X_1, X_2, \ldots, X_n), f_2(X_1, X_2, \ldots, X_n))$ hay $f_1(X_1, X_2, \ldots, X_n) < \theta < f_2(X_1, X_2, \ldots, X_n)$ gọi là khoảng tin cậy của tham số θ với độ tin cậy γ.

Trong trường hợp khoảng tin cậy đối xứng có dạng

$$(g(X_1, X_2, \ldots, X_n) - \epsilon, g(X_1, X_2, \ldots, X_n) + \epsilon)$$

Khi đó ϵ được gọi là độ chính xác hay sai số ước lượng.

Cần chú ý rằng, cùng một độ tin cậy γ ta có thể tìm được nhiều khoảng tin cậy khác nhau của tham số θ. Khoảng tin cậy nào có độ dài ngắn nhất thì xem khoảng tin cậy đó là tốt nhất. Trong thực hành, ta chỉ tìm khoảng tin cậy tốt nhất.

Nhận xét

Từ định nghĩa trên, nếu ta chọn $\alpha = 1\% \Rightarrow \gamma = 99\%$ và

$$P[f_1(X_1, X_2, \ldots, X_n) < \theta < f_2(X_1, X_2, \ldots, X_n)] = \gamma$$

Điều này nói lên rằng xác suất để lấy ra một mẫu thực nghiệm $(x_1, x_2, \ldots, x_n)$ mà $f_1(x_1, x_2, \ldots, x_n) < \theta < f_2(x_1, x_2, \ldots, x_n)$ là 99%

Như vậy nếu ta lấy ra một mẫu thực nghiệm $(x_1, x_2, \ldots, x_n)$. Ta có khoảng tin cậy cụ thể

$$f_1(x_1, x_2, \ldots, x_n) < \theta < f_2(x_1, x_2, \ldots, x_n)$$

và theo nguyên lý xác suất lớn, ta luôn thừa nhận tham số θ thỏa mãn

$$f_1(x_1, x_2, \ldots, x_n) < \theta < f_2(x_1, x_2, \ldots, x_n)$$

5.2.3. Khoảng tin cậy cho kỳ vọng $E(X) = \mu$ với $X \backsim N(\mu, \delta)$

X là đại lượng ngẫu nhiên biểu thị đặc trưng nghiên cứu trên tập Ω cơ phân phối chuẩn với kỳ vọng $E(X) = \mu$ chưa biết. Để ước lượng μ, ta lấy mẫu ngẫu nhiên kích thước n. Ta có mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ và mẫu thực nghiệm tương ứng $(x_1, x_2, \ldots, x_n)$. Ta tính số trung bình mẫu $\overline{x}$. Tùy theo độ lệch chuẩn δ đã biết hay chưa mà tính độ lệch chuẩn mẫu điều chỉnh δ^2_{n-1}. Với độ tin cậy $\gamma = 1 - \alpha$, ta có

Khoảng tin cậy đối xứng $E(X) = \mu$ là $(\overline{x} - \epsilon, \overline{x} + \epsilon)$ hay $\overline{x} - \epsilon < \mu < \overline{x} + \epsilon$.

Để ước lượng $E(X) = \mu$ ta lấy số trung bình mẫu $\overline{x}$ ước lượng cho nó. Khi đó

$$P(\overline{x} - \epsilon < \mu < \overline{x} + \epsilon) = \gamma \Leftrightarrow P(\mid \overline{X} - \mu \mid < \epsilon) = \gamma$$

Trường hợp độ lệch chuẩn δ đã biết Khi đó $\epsilon = \dfrac{\delta}{\sqrt{n}} \Phi^{-1}\left(\dfrac{\gamma}{2}\right)$

Trường hợp độ lệch chuẩn δ chưa biết Tính số δ^2_{n-1}. Ta có

$$P(\mid \overline{X} - \mu \mid < \epsilon) = 1 - \alpha \Leftrightarrow P\left(\mid \frac{\overline{X} - \mu}{\delta_{n-1}}\sqrt{n} \mid < \frac{\epsilon\sqrt{n}}{\delta_{n-1}}\right) = \gamma$$

$$\Leftrightarrow \epsilon = \frac{\delta_{n-1}}{\sqrt{n}}.t(n-1;\ 1-\gamma)$$

Vì $\dfrac{\overline{X} - \mu}{\delta_{n-1}}\sqrt{n}$ có phân phối Student với $n - 1$ bậc tự do.

Chú ý rằng, khi $n > 30$: $\dfrac{\overline{X} - \mu}{\delta_{n-1}}\sqrt{n}$ có phân phối xấp xỉ phân phối chuẩn $N(0, 1)$ nên $t(n - 1;\ 1 - \gamma) \approx \Phi^{-1}\left(\dfrac{\gamma}{2}\right)$.

Ví dụ 5.2.55 Để ước lượng độ cứng trung bình của một loại bi bằng thép, người ta lấy ra 25 bi để kiểm tra. Tính được độ cứng trung bình $\overline{x} = 10$, với độ tin cậy 99%. Hãy tìm khoảng tin cậy đối xứng của độ cứng trung bình của viên bi thép đó. Biết rằng độ cứng của viên bi có phân phối chuẩn $N(\mu, \delta^2)$ trong hai trường hợp:

a) Độ lệch chuẩn $\delta = 1$

b) Độ lệch chuẩn chưa biết và từ mẫu thực nghiệm tính được $\delta_{n-1} = 1,2$

Giải

Gọi X là độ cứng của viên bi bằng thép thì X là đại lượng ngẫu nhiên có phân phối chuẩn. Khi đó kỳ vọng $E(X) = \mu$ là độ cứng trung bình của một viên bi.

a) Ta có $n = 25, \overline{x} = 10, \delta = 1$, độ tin cậy $\gamma = 99\%$

Tra bảng phân vị chuẩn, ta có $\Phi^{-1}\left(\dfrac{\gamma}{2}\right) = 2,567$.

Từ đó, ta có $\epsilon = \dfrac{\delta}{\sqrt{n}}\Phi^{-1}\left(\dfrac{\gamma}{2}\right) = \dfrac{1}{\sqrt{25}}2,567 \approx 0,515$

Vậy khoảng tin cậy đối xứng của độ cứng trung bình μ là

$$(\overline{x} - \epsilon, \overline{x} + \epsilon) = (9,485;\ 10,515) \text{ hay } 9,485 < \mu < 10,515$$

b) Ta có $n = 25, \overline{x} = 10, \delta_{n-1} = 1,2$, độ tin cậy $\gamma = 99\%$

Tra bảng phân vị Student với $n-1 = 24$ bậc tự do, mức phân vị $1-\gamma = 0,05$, ta có $t(24;\ 0,05) = 2,797$ nên

$$\epsilon = \frac{\delta_{n-1}}{\sqrt{n}} t(n-1;\ 1-\gamma) = \frac{1,2}{5} 2,797 \approx 0,67$$

Vậy khoảng tin cậy đối xứng của độ cứng trung bình μ là

$$(\overline{x} - \epsilon, \overline{x} + \epsilon) = (9,33;\ 10,67) \text{ hay } 9,33 < \mu < 10,67$$

Ví dụ 5.2.56 *Để ước lượng hao phí xăng của một loại ôtô chạy trên đoạn đường AB, người ta theo dõi 100 chuyến xe và tính được $\overline{x} = 10$ lít, $\delta_{n-1} = 0,5$ lít, với độ tin cậy 95%. Giả sử xằng biến ngẫu nhiên X chỉ lượng xăng hao phí của một loại ôtô trên trên đoạn AB có phân phối chuẩn. Hãy ước lượng khoảng tin cậy đối xứng của của mức hao phí xăng trung bình μ?*

Giải

a) Ta có $n = 100, \overline{x} = 10$, độ tin cậy $\gamma = 95\%$. Tra bảng phân vị chuẩn, ta có $\Phi^{-1}\left(\frac{\gamma}{2}\right) = \Phi^{-1}\left(\frac{95\%}{2}\right) = 1,96$.

Từ đó, ta có $\epsilon = \dfrac{\delta}{\sqrt{n}} \Phi^{-1}\left(\dfrac{\gamma}{2}\right) = \dfrac{0,05}{\sqrt{100}} 1,96 \approx 0,1$

Vậy khoảng tin cậy đối xứng của độ cứng trung bình μ là

$$(\overline{x} - \epsilon, \overline{x} + \epsilon) = (9,9;\ 10,1) \text{ hay } 9,9 < \mu < 10,1$$

5.2.4. Ước lượng tỉ lệ phần trăm hay xác suất

Giả sử A là một sự kiện liên kết với một phép thử. Muốn biết xác suất xảy ra biến cố A hay $p = P(A)$, ta lặp lại phép thử n lần(lấy mẫu có lặp kích thước n).

Gọi X_i là số lần xảy ra sự kiện A ở phép thử thứ $i(i = \overline{1,n})$ thì $X_1, X_2, \ldots, X_n$ là các đại lượng ngẫu nhiên độc lập có cùng phân phối

$$\begin{array}{c|cc} X_i & 0 & 1 \\ \hline P & 1-p & p \end{array} (i = \overline{1,n})$$

Khi đó $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = F_n(A)$

Ta có $E(\overline{X}) = p$

Bài toán ước lượng xác suất chính là bài toán ước lượng kỳ vọng $E(X) = p$.

Khi n khá lớn:$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = F_n(A)$ có phân phối xấp xỉ phân phối chuẩn $N(p, p(1-p))$.

Ước lượng tỉ lệ

Giả sử tổng thể Ω có N phần tử, trong đó có M phần tử mang đặc tính A. Do không điều tra toàn bộ nên ta không biết tỉ lệ $p = \frac{M}{N}$. Phép thử lấy ra một phần tử. Gọi C là sự kiện phần tử lấy ra mang đạc tính A. Khi đó $P(C) = \frac{M}{N} = p$.

Suy ra bài toán ước lượng tỉ lệ chính là bài toán ước lượng xác suất.

Để ước lượng tỉ lệ p, ta lấy mẫu kích thước n, ta tính được tỉ lệ f_n và với độ tin cậy $\gamma = 1-\alpha$, ta có:

Khoảng tin cậy cho tỉ lệ p là $(f_n - \epsilon, f_n + \epsilon)$ ϵ được xác định như sau:

Với $n > 50 : nf_n > 5$ và $n(1-f_n) > 5$. Ta có

$$\epsilon = \frac{\sqrt{f_n(1-f_n)}}{\sqrt{n}}.\Phi^{-1}\left(\frac{\gamma}{2}\right)$$

Ví dụ 5.2.57 Để ước lượng tỉ lệ gạch loại 2 của một nhà máy, người ta lấy ra 10000 viên và thấy có 300 viên gạch loại 2. Với độ tin cậy 99%

a) Hãy ước lượng tỉ lệ gạch loại 2?

b) Cần lấy thêm ít nhất bao nhiêu gạch nữa để tỉ lệ gạch loại 2 toàn bộ so với tỉ lệ mẫu có sai số không vượt quá $0,001$. Giả thiết lấy tỉ lệ mẫu f_n thay cho tỉ lệ mẫu của mẫu cần thêm.

Giải

Gọi p là tỉ lệ gạch loại 2 của nhà máy.

Ta có $n = 10000$, tỉ lệ mẫu $f_n = \frac{300}{10000} = 0,03$

Độ tin cậy $\gamma = 1 - \alpha = 99\%$

a)

$$\epsilon = \frac{\sqrt{f_n(1-f_n)}}{\sqrt{n}}.\Phi^{-1}\left(\frac{\gamma}{2}\right) = \frac{\sqrt{0,03.0,07}}{\sqrt{10000}}.2,576 \approx 0,0044$$

Vậy khoảng ước lượng đối xứng của tỉ lệ là

$$f_n - \epsilon < p < f_n + \epsilon \Leftrightarrow 0,03 - 0,0044 < p < 0,03 + 0,0044$$

hay $0,0256 < p < 0,0344$

b) Gọi N là số gạch cần lấy thêm $n = N + 10000$

Tỉ lệ mẫu $f_n = 0,03$ và độ tin cậy $\gamma = 0,99$, ta có $\Phi^{-1}\left(\frac{\gamma}{2}\right) = 2,576$

Theo giả thiết $\epsilon \leqslant 0,001 \Rightarrow$

$$\epsilon = \frac{\sqrt{f_n(1-f_n)}}{\sqrt{n}}\Phi^{-1}\left(\frac{\gamma}{2}\right) = \frac{\sqrt{0,03.0,07}}{\sqrt{n}}.2,576 \leqslant 0,001 \Rightarrow n \geqslant 193101,18$$

suy ra $n_{min} = 193102$. Số gạch cần lấy thêm ít nhất là $N = 193102 - 10000 = 193102$

Ví dụ 5.2.58 Để ước lượng cá trong hồ, người ta bắt 1000 con làm dấu rồi thả lại. Sau đó lại bắt 900 con thấy có 90 con làm dấu. Hãy ước lượng số lượng cá trong hồ với độ tin cậy 95%

Giải

Gọi N là số cá trong hồ.

Gọi p là tỉ lệ cá làm dấu trong hồ, ta có $p = \frac{1000}{N}$

Ta ước lượng p. Ta có $n = 900$, tỉ lệ mẫu $f_n = \frac{90}{900} = 0,1$

Độ tin cậy $\gamma = 1 - \alpha = 95\%$

Tra bảng phân vị chuẩn, ta có $\Phi^{-1}\left(\frac{\gamma}{2}\right) = 1,96$

$$\epsilon = \frac{\sqrt{f_n(1-f_n)}}{\sqrt{n}}\Phi^{-1}\left(\frac{\gamma}{2}\right) = \frac{\sqrt{0,1.0,9}}{\sqrt{900}}.1,96 \approx 0,0196$$

Suy ra $f_n - \epsilon < p < f_n + \epsilon \Rightarrow 0,0804 < \frac{1000}{N} < 0,1196$

Vậy số cá trong hồ là $8362 \leqslant N \leqslant 12437$

5.2.5. Ước lượng phương sai $D(X)$ với $X \backsim N(\mu, \delta)$

Giả sử X là biến ngẫu nhiên biểu thị đặc trưng nghiên cứu trên tập Ω có phân phối chuẩn $D(X) = \delta^2$ chưa biết. Để ước lượng $D(X)$ ta lấy mẫu kích thước n, ta được mẫu ngẫu nhiên $(X_1, X_2, \ldots, X_n)$ và mẫu thực nghiệm $(x_1, x_2, \ldots, x_n)$.

Ta có $X_i \backsim N(\mu, \delta^2) \Rightarrow \frac{X_i - \mu}{\delta} \backsim N(0,1)$. Do đó:

$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\delta}\right)^2 = \frac{ns_0^2}{\delta^2}$ có phân phối khi bình phương với $(n-1)$ bậc tự do.

Trong đó

$$\delta_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2; \qquad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Ta xét các trường hợp sau:

Với kỳ vọng $E(X) = \mu$ đã biết

Tính $s_n = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$
Tra bảng phân vị khi bình phương với n bậc tự do ứng với các mức phân vị $\frac{\alpha}{2}, 1 - \frac{\alpha}{2}$ ta có $\chi^2_{\frac{\alpha}{2}}(n), \chi^2_{1-\frac{\alpha}{2}}(n)$
Khi đó, khoảng tin cậy của phương sai

$$\frac{ns_n^2}{\chi^2_{1-\frac{\alpha}{2}}(n)} < \delta^2 < \frac{ns_n^2}{\chi^2_{\frac{\alpha}{2}}(n)}$$

Với kỳ vọng $E(X) = \mu$ chưa biết

Tính

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \qquad \delta_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Tra bảng phân vị khi bình phương với $(n-1)$ bậc tự do ứng với các mức phân vị $\frac{\alpha}{2}, 1 - \frac{\alpha}{2}$ ta có $\chi^2_{\frac{\alpha}{2}}(n-1), \chi^2_{1-\frac{\alpha}{2}}(n-1)$
Khi đó, khoảng tin cậy của phương sai $D(X) = \delta^2$

$$\frac{n\delta_n^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \delta^2 < \frac{n\delta_n^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}$$

Chú ý: Khi $n > 30$ ta có

$$\chi^2_{\frac{\alpha}{2}}(n) \approx \frac{1}{2}(\sqrt{2n-1} - U_{1-\frac{\alpha}{2}})^2$$

$$\chi^2_{1-\frac{\alpha}{2}}(n) \approx \frac{1}{2}(\sqrt{2n-1} + U_{1-\frac{\alpha}{2}})^2$$

với $U_{1+\frac{\alpha}{2}}$ là phân vị chuẩn mức $_{1-\frac{\alpha}{2}}$ của $N(0,1)$.

Ví dụ 5.2.59 Tuổi thọ X của một van điện có phân phối chuẩn. Ta lấy ra 10 van điện tính được phương sai mẫu $\delta_1^2 = 0,8$. Với độ tin cậy 95% hãy ước lượng phương sai $D(X)$?

GiảiTa có $n = 10$ và kỳ vọng chưa biết. Độ tin cậy $\gamma = 1 - \alpha = 95\% \Rightarrow \alpha = 0,05$
Tra bảng phân vị khi bình phương với $n - 1 = 10 - 1 = 9$ bậc tự do ứng với các mức phân vị $\frac{\alpha}{2} = 0,025, 1 - \frac{\alpha}{2} = 0,975$ ta có $\chi^2_{0,025}(9) = 2,7, \chi^2_{0,975}(9) = 19$
$\frac{n\delta_n^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} = \frac{10.0,8}{2,7} = 2,96$
$\frac{n\delta_n^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} = \frac{10.0,8}{19} = 0,42$
Vậy khoảng tin cậy của phương sai $D(X) = \delta^2$

$$0,42 < \delta^2 < 2,96$$

PHỤ LỤC

1. Hàm Laplace: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt \qquad (x \geqslant 0)$

Chú ý: Với $x < 0$ thì $\Phi(x) = -\Phi(-x)$

x	0	1	2	3	4	5	6	7	8	9
0.0	0.00000	.00399	.00798	.01197	.01595	.01994	.02392	.02790	.03188	.03586
0.1	.03983	.04380	.04776	.05172	.05567	.05962	.06356	.06749	.07142	.07535
0.2	.07926	.08317	.08706	.09095	.09483	.09871	.10257	.10642	.11026	.11409
0.3	.11791	.12172	.12552	.12930	.13307	.13683	.14058	.14431	.14803	.15173
0.4	.15542	.15910	.16276	.16640	.17003	.17364	.17724	.18082	.18439	.18793
0.5	.19146	.19497	.19847	.20194	.20540	.20884	.21226	.21566	.21904	.22240
0.6	.22575	.22907	.23237	.23565	.23891	.24215	.24537	.24857	.25175	.25490
0.7	.25804	.26115	.26424	.26730	.27035	.27337	.27637	.27935	.28230	.28524
0.8	.28814	.29103	.29389	.29673	.29955	.30234	.30511	.30785	.31057	.31327
0.9	.31594	.31859	.32121	.32381	.32639	.32894	.33147	.33398	.33646	.33891
1.0	.34134	.34375	.34614	.34849	.35083	.35314	.35543	.35769	.35993	.36214
1.1	.36433	.36650	.36864	.37076	.37286	.37493	.37698	.37900	.38100	.38298
1.2	.38493	.38686	.38877	.39065	.39251	.39435	.39617	.39796	.39973	.40147
1.3	.40320	.40490	.40658	.40824	.40988	.41149	.41308	.41466	.41621	.41774
1.4	.41924	.42073	.42220	.42364	.42507	.42647	.42785	.42922	.43056	.43189
1.5	.43319	.43448	.43574	.43699	.43822	.43943	.44062	.44179	.44295	.44408
1.6	.44520	.44630	.44738	.44845	.44950	.45053	.45154	.45254	.45352	.45449
1.7	.45543	.45637	.45728	.45818	.45907	.45994	.46080	.46164	.46246	.46327
1.8	.46407	.46485	.46562	.46638	.46712	.46784	.46856	.46926	.46995	.47062
1.9	.47128	.47193	.47257	.47320	.47381	.47441	.47500	.47558	.47615	.47670
2.0	.47725	.47778	.47831	.47882	.47932	.47982	.48030	.48077	.48124	.48169
2.1	.48214	.48257	.48300	.48341	.48382	.48422	.48461	.48500	.48537	.48574
2.2	.48610	.48645	.48679	.48713	.48745	.48778	.48809	.48840	.48870	.48899
2.3	.48928	.48956	.48983	.49010	.49036	.49061	.49086	.49111	.49134	.49158
2.4	.49180	.49202	.49224	.49245	.49266	.49286	.49305	.49324	.49343	.49361
2.5	.49379	.49396	.49413	.49430	.49446	.49461	.49477	.49492	.49506	.49520
2.6	.49534	.49547	.49560	.49573	.49585	.49598	.49609	.49621	.49632	.49643
2.7	.49653	.49664	.49674	.49683	.49693	.49702	.49711	.49720	.49728	.49736
2.8	.49744	.49752	.49760	.49767	.49774	.49781	.49788	.49795	.49801	.49807
2.9	.49813	.49819	.49825	.49831	.49836	.49841	.49846	.49851	.49856	.49861
3.0	0.49865		3.1	.49903	3.2	.49931	3.3	.49952	3.4	.49966
3.5	.49977		3.6	.49984	3.7	.49989	3.8	.49993	3.9	.49995
4.0	.49997		4.1	.49998	4.2	.49999	4.3	.49999	4.4	.49999
4.5	.4999966									
5.0	.4999997									

4. Bảng phân vị Student: $P(X < t_{n,\alpha}) = \alpha$

	α					
n	0.9	0.95	0.975	0.99	0.995	0.9995
1	3.078	6.314	12.706	31.821	63.656	636.578
2	1.886	2.920	4.303	6.965	9.925	31.600
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.012	4.221
14	1.345	1.761	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.518	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.768
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.689
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.660
∞	1.282	1.645	1.960	2.326	2.576	3.291

Bài tập

Câu 1. Tung 3 con xúc xắc. Tính xác suất để

1. **có duy nhất một mặt chia hết cho 3.**
2. **Tổng số chấm 3 mặt bằng 15.**
3. **Có ít nhất một mặt một chấm xuất hiện.**

Câu 2. Có 3 người lên 6 toa tàu trong đó có hai người tên A và B. Tính xác suất để

1. **mỗi người ngồi một toa khác nhau?**
2. **Hai người A, B ngồi cùng một toa?**
3. **có đúng một người ngồi ở toa thứ 2?**

Câu 3. Một hộp có 5 bi đỏ, 4 bi xanh và 6 bi trắng. Lấy ngẫu nhiên lần lượt ra 2 bi. Tính xác suất các sự kiện sau trong hai trường hợp có hoàn lại và không hoàn lại.

1. **Cả hai bi lấy ra đều màu đỏ.**
2. **Hai bi lấy ra cùng màu.**
3. **Hai bi lấy ra khác màu.**

Câu 4. Một hộp có m bi đỏ và 5 bi xanh. Lần 1 lấy ngẫu nhiên 2 bi (không hoàn lại). Sau đó, lần 2 lấy ngẫu nhiên 3 bi. Tính xác suất

a) Lần 1 lấy được hai bi cùng màu ?

b) Cả hai lần đều lấy được toàn bi xanh ?

c) Lần 2 lấy được hai bi xanh ?

d) Giả sử lần 2 lấy được 2 bi xanh. Tính xác suất để lần 1 lấy được hai bi xanh ?

Câu 5. Chon ngẫu nhiên hai điểm $x, y \in (0,1)$. Tính xác suất để chọn được hai điểm có tổng $x + y \leqslant 1$ và $xy \leqslant \frac{2}{9}$.

Câu 6. Trên đoạn thẳng OM có độ dài $m > 0$, ta lấy ngẫu nhiên hai điểm B và C với $OB = x, OC = y$. Tính xác suất để lấy được hai điểm B, C sao cho $BC < \frac{m}{2}$?

Câu 7. Trong 20 sản phẩm có 5 phế phẩm. Ta bỏ vào 3 cái hộp mỗi hộp 5 sản phẩm. Tính xác suất để

1. **Ở hộp 1 chỉ có một phế phẩm?**
2. **Các hộp đều có phế phẩm?**
3. **Các phế phẩm đều ở hộp 1.**

Câu 8. Đề cương ôn thi có 12 câu lý thuyết và 30 câu bài tập. Một đề thi gồm 2 câu lý thuyết và 4 câu bài tập. Một học sinh chỉ học 6 câu lý thuyết và 10 câu bài tập. Tính xác suất để học sinh đó

1. không làm được 2 câu lý thuyết?

2. chỉ làm được một câu lý thuyết và 2 câu bài tập hoặc làm được 2 câu lý thuyết và một bài tập?

3. Học sinh không phải thi lại? Biết học sinh không thi lại nếu làm được ít nhất một câu lý thuyết và 2 câu bài tập hoặc làm được 2 câu lý thuyết và một bài tập.

Câu 9. Trong thành phố A, tỉ lệ người tốt nghiệp đại học là 0.3 và tỉ lệ người tốt nghiệp phổ thông là 0.35. Trong số người đã tốt nghiệp đại học có 60% người có xe máy; trong số người tốt nghiệp phổ thông có 40% người có xe máy; trong số người chưa tốt nghiệp phổ thông có 10% người có xe máy. Chọn ngẫu nhiên một người dân thành phố A. Tính xác suất để

1. người chọn ra có xe máy?

2. Biết người chọn ra có xe máy. Hỏi khả năng người này thuộc nhóm người nào nhiều nhất?

Câu 10. Có hai thùng hàng, thùng I chứa 4 chính phẩm và 5 phế phẩm; thùng II chứa 6 chính phẩm và 1 phế phẩm. Lấy ngẫu nhiên 3 sản phẩm ở thùng I rồi bỏ vào thùng 2. Sau đó lấy ngẫu nhiên 2 sản phẩm từ thùng II. Tính xác suất để hai sản phẩm lấy ra cuối cùng

1. đều tốt?

2. Đều tốt và đều là của thùng I?

Câu 11. Bắn 3 viên đạn độc lập nhau vào một mục tiêu. Xác suất trúng mục tiêu của từng viên đạn lần lượt là $0.7, 0.8$ và 0.9. Biết rằng nếu chỉ 1 viên đạn trúng hoặc 2 viên trúng thì mục tiêu bị phá hủy với xác suất tương tứng là 0.4 và 0.6; còn nếu trúng cả 3 viên thì mục tiêu bị phá hủy. Tính xác suất để mục tiêu bị phá hủy?

Câu 12. Một người trong túi có hai bao diêm, mỗi bao có 30 que. Mỗi lần hút thuốc người đó lấy ngẫu nhiên một bao và dùng mất một que ở bao đó. Tính xác suất để

1. Mỗi bao còn lại đúng 10 que?

2. Một bao hết diêm còn bao kia còn lại 10 que?

Câu 13. Xác suất bắn trúng mục tiêu của một xạ thủ là không đổi và bằng 0.6. Xạ thủ bắn 3 viên vào mục tiêu. Gọi X là số viên đạn bắn trúng.

1. Lập bảng phân phối xác suất và hàm phân phối xác suất của X?

2. Tính các tham số đặc trưng của X?

3. Tính xác suất để xạ thủ phá hủy được mục tiêu. Biết rằng mục tiêu bị phá hủy nếu bắn trúng ít nhất 2 viên.

Câu 14. **Có hai hộp, hộp I có 4 bi đỏ, 6 bi xanh; hộp II có 3 bi đỏ 5 bi xanh. Lấy ngẫu nhiên 2 bi ở hộp I rồi bỏ vào hộp II. Sau đó từ hộp II lấy ngẫu nhiên 2 bi. Gọi X, Y lần lượt là số bi đỏ lấy ra ở hộp I và hộp II.**

1. Lập bảng phân phối xác suất của X và Y?

2. Lập bảng phân phối xác suất của $X \pm Y, X.Y$?

Câu 15. **Cho biến ngẫu nhiên X có hàm phân phối xác suất**

$$F(x) = \begin{cases} 0, & x \leqslant 0 \\ a.\sin 2x, & x \in (0, \frac{\pi}{4}) \\ 0, & x > \frac{\pi}{4} \end{cases}$$

1. Xác đinh a và hàm mật độ xác suất $f(x)$?

2. Tính các tham số đặc trưng của X?

3. Tính xác suất để trong ba lần thực hiện phép thử về X có 2 lần X nhận giá trị trong khoảng $(0,\frac{\pi}{8})$?

Câu 16. **Cho biến ngẫu nhiên X có hàm mật độ xác suất**

$$f(x) = \begin{cases} ax^2, & x \in (0, 2] \\ 0, & x \notin (0, 2] \end{cases}$$

a) Xác định a và hàm phân phối $F(x)$?

b) Tính $E(X), D(X)$?

c) Tính $P(1 < X < 9)$?

Câu 17. **Độ dài X của chi tiết máy là $N(5cm, 0.81cm)$. Tĩnh xác suất để lấy được một chi tiết máy có chiều dài trong khoảng $(4cm, 7cm)$?**

Câu 18. **Chiều cao Y của nam giới trưởng thành là $N(160cm, \delta cm)$. Biết rằng $P(|X - 160| < 2) = 0.2569$, tìm δ và tính xác suất để lấy ngẫu nhiên 4 nam thì có ít nhất một người có chiều cao trong khoảng 158 cm đến 162 cm?**